Physical-Based RTN Modeling of Ring Oscillators in 40-nm SiON and 28-nm HKMG by Bimodal Defect-Centric Behaviors

Azusa Oshima, Takuya Komawaki, Kazutoshi Kobayashi, Ryo Kishida*, Pieter Weckx§†, Ben Kaczer†, Takashi Matsumoto‡, and Hidetoshi Onodera††
∗Kyoto Institute of Technology, § KU Leuven, †imec, ‡University of Tokyo, ††Kyoto University
*{aoshima, kobayasi, rkishida, tkomawaki}@vlsi.es.kit.ac.jp

Abstract

We apply the bimodal trap model of Random Telegraph Noise (RTN) to predict its impact on the circuit level instead of the conventional unimodal trap model. Two different trap distributions represented by the average number of traps N and the average impact to threshold voltage Vth per trap η in gate dielectric make measured RTN distributions of ring oscillators (ROs) to follow the bimodal trap model. It replicates distributions of frequency fluctuation of ROs caused by RTN on 40-nm SION and 28-nm HKMG processes. RO with standard size-transistors can be modeled by scaled parameters from RO with minimum-size transistors.

I. Introduction

With the rapid downscaling of MOSFETs to atomistic levels, dynamic variations such as Random Telegraph Noise (RTN) become dominant as comparable as static process variations [1]. Fig. 1 depicts RTN caused by capture and emission of carriers at gate-oxide traps [2], [3]. Since traps capture and emit alternately with various time constants, $\Delta V_{\rm th}$ fluctuates when supply power is applied to MOSFETs. It has been reported that RTN has a big impact on CMOS image sensors, flash memories and SRAMs [4]–[6]. It is mandatory for circuit designers to accurately predict the impact of RTN to the circuit level.

This paper deals with the circuit-level modeling of RTN by using the bimodal defect-centric model [7]. Section II introduces the unimodal and bimodal defect-centric models. The unimodal model can be applied to the process without higk-k gate dielectric, while the bimodal one can be applied to those with high-k. Section III shows measurement results of RTN-induced frequency fluctuations of ring oscillators fabricated in 40 nm and 28 nm processes. Section IV explains our fitting methodology on the circuit level with the bimodal model. Section V reveals that the measurement results can accurately be replicated with the fitting results. Section VI concludes this paper.

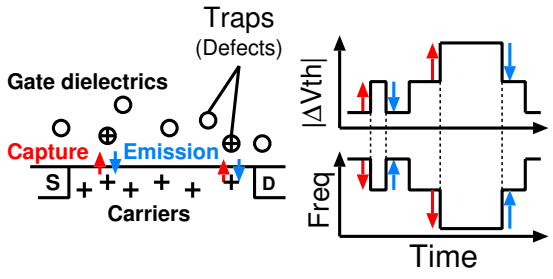


Fig. 1: RTN is caused by capture and emission of carriers at gate oxide traps.

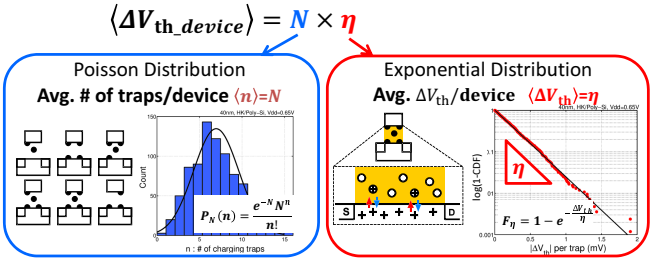
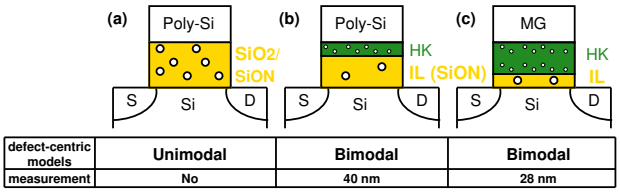


Fig. 2: Predicttion of the total RTN-induced $\Delta V_{\rm th}$ distribution on a chip by defect-centric models.



defect-centric models	Unimodal	Bimodal	Bimodal
measurement	No	40 nm	28 nm

Fig. 3: Dielectric materials and defect-centric models (a) SiO_2/Poly-Si, (b) Ultra thin HK/Poly-Si, (c) HKMG process.

II. RTN Modeling by Unimodal and Bimodal Defect-Centric Distributions

Recently, the defect-centric distribution has been proposed to predict RTN-induced degradations [1], [8]. It is based on the physics of trap behaviors in gate dielectrics. RTN-induced $\Delta V_{\rm th}$ is charaterized by the average number of charged traps N and the average impact to threshold voltage $V_{\rm th}$ per trap η as shown in Fig. 2. The device-to-device threshold variation $\langle \Delta V_{\rm th\_d} \rangle$ is computed as a convolution of exponential distributions of individual charged defects expressed as the Poisson distribution.

We replicate frequencies of Ring Oscillators (ROs) with dynamic RTN-induced threshold voltage ($V_{\rm th}$) variations by the Monte-Carlo method. Fig. 3 shows three defect-centric models according to gate dielectronic materials. The unimodal defect-centric model accurately predict defect distributions in SiO_2 and SiON processes, but it is not applicable in High-K Metal Gate (HKMG) processes [9]. The threshold fluctuation ($\Delta V_{\rm th}$) by RTN in HKMG is characterized by two independant N, η values attributed to traps in HK and IL (Interface Layer) regions [7], respectively. The average numbers of traps in these regions are defined as $N_{\rm HK}$ and $N_{\rm IL}$. The average impacts to $V_{\rm th}$ per trap to RTN are defined as $\eta_{\rm HK}$ for HK and $\eta_{\rm IL}$ for IL, respectively.

978-1-5090-0818-6/16/$31.00 ©2016 IEEE

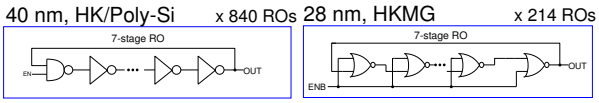


Fig. 4: Test structures to measure frequency fluctuation in 40 nm and 28 nm.

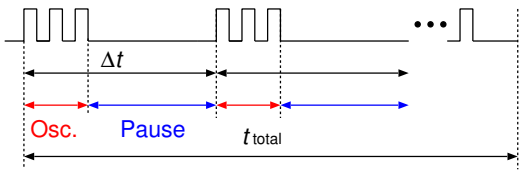


Fig. 5: Timing chart to measure RTN-induced frequency fluctuation.

III. Measuring Impacts of RTN

Test structures are fabricated in 40 and 28 nm processes. They both have HK gate dielectric as shown in Fig. 3 (b) and (c). In 40 nm, the thin HK layer is attached on IL. HK is doped to modify gate work functions to control $V_{\rm th}$ [10]. In 28 nm, the HKMG process in Fig. 3 (c) is used as gate stacks to increase permittivity.

Fig. 4 shows two RO structures to measure frequency fluctuation in 40 and 28 nm. RO consists of 6 inverters and one NAND gate in 40 nm, while RO in 28nm is composed of 7 NOR gates. We measure frequency fluctuations of 840 ROs in 40 nm and 216 ROs in 28 nm. Note that all measurement data are taken from a single chip to eliminate chip-to-chip process variations. Fig. 5 shows the timing chart to measure frequency fluctuation in a single RO. It periodically repeats oscillation and pause. Embedded counters are used to measure number of oscillations periodically during each oscillation period. Measured numbers of oscillations (frequency) dynamically fluctuate when RTN changes $V_{\rm th}$. Table I shows measurement conditions of 40 nm and 28 nm. The duration time of measurement are 20 sec. and 2.7 sec. in 40 nm and 28 nm respectively. It is because of the limitation of measurement facilities.

Fig. 6 (a), (b) shows measured frequency fluctuations of 40 nm and 28 nm at 0.65 V. $F_{\rm max}$ and $\Delta F = F_{\rm max} - F_{\rm min}$ are defined as the maximum frequency and the maximum frequency fluctuation respectively. $\Delta F/F_{\rm max}$ represents the impact of RTN to circuit operations. RTN fluctuates oscillation frequencies by 8.61% in 40 nm and 2.64% in 28 nm respectively. Fig. 7 shows distributions of measured $\Delta F/F_{\rm max}$ (triangles) of ROs in 40 nm and 28 nm. At $+6\sigma$ points, oscillation frequencies of ROs in 40 nm fluctuates 50%.

One reason why 28 nm has smaller $\Delta F/F_{\rm max}$ is the transistor size. Transistors in the 40 nm ROs are minimized to measure RTN-induced frequency fluctuation distinctly, while the 28 nm ROs are designed with standard-size transistors. In 40 nm, ROs with standard size inverters have less than 2.0% frequency fluctuation at $+3\sigma$ as later shown in Fig. 11

Table I: Measurement conditions.

Process	40 nm	28 nm
Δt	2.2 ms	0.174 ms
# of measurements	9,024	15,000
$t_{\rm total}$	20 s	2.7 s
# of ROs	840	216
$V_{\rm dd}$	0.65 V	0.65 V

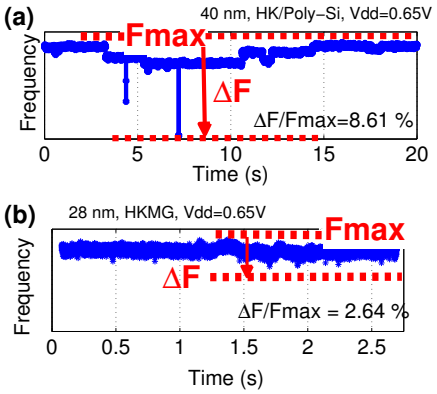
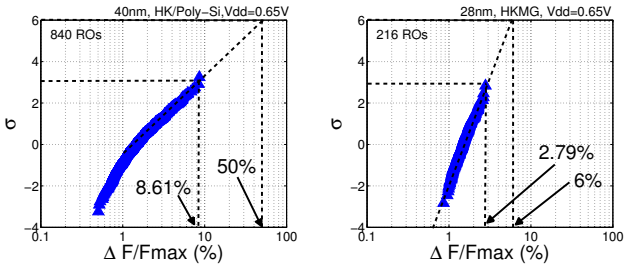


Fig. 6: Measurement results of RTN-induced frequency fluctuation of single RO. (a) 40nm HK/Poly-Si, (b) 28 nm HKMG.



(a) 40 nm HK/Poly-Si (b) 28 nm HKMG

Fig. 7: Distributions of $\Delta F/F_{\rm max}$ of ROs on 40 nm (a) and 28 nm(b)

IV. Fitting Methodology Using Bimodal Model

Fig. 8 shows our simulation flow to obtain the distributions of $\Delta F/F_{\rm max}$ of ROs by using these four parameters: average number of traps and average $\Delta V_{\rm th}$ per traps in interface and high-k layers, $N_{\rm IL}, N_{\rm HK}, \eta_{\rm IL}$,and $\eta_{\rm HK}$.

In the simulation flow, we assume that the RO with the maximum frequency $F_{\rm max}$ has no RTN fluctuation. The RO with the minimum frequency $F_{\rm min}$ has the maximum RTN fluctuation when all possible defects are trapped to decrease oscillation frequency. The value of $\Delta F/F_{\rm max}$ is computed from Eq. 1.

$$\Delta F/F_{\rm max} = \frac{F_{\rm max} - F_{\rm min}}{F_{\rm max}} \quad (1)$$

Initial parameters are heuristically optimized to fit the $\Delta F/F_{\rm max}$ distribution from circuit simulations with the measurement data.

V. Fitting with the Bimodal Model

Fig. 9 (a) shows distributions of $\Delta F/F_{\rm max}$ by the defect-centric distributions and the measurement results in 40 nm HK/Poly-Si. While the unimodal model fails to follow the distributions at the tail, the bimodal (HK+IL) model follow them from head to tail. The bimodal defect-centric model clearly replicates the CDF distributions by the combination

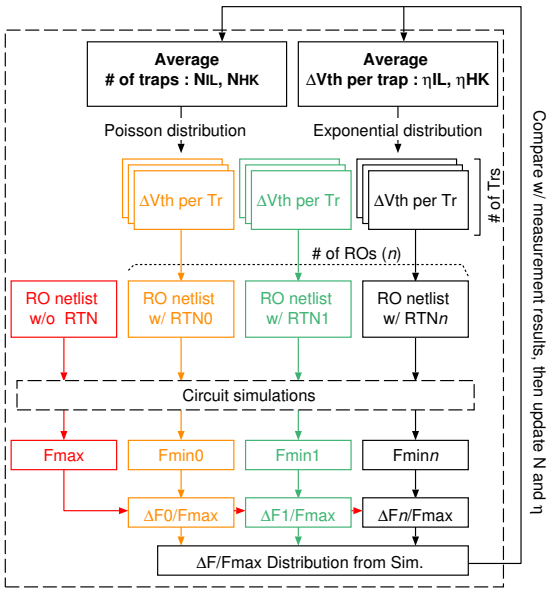


Fig. 8: Simulation flow to obtain $N_{\mathrm{IL}}, N_{\mathrm{HK}}, \eta_{\mathrm{IL}}, \eta_{\mathrm{HK}}$.

Table II: Parameter values to obtain $\Delta F/F_{\max}$ distributions of ROs with minimum and standard size transistors as shown in Fig. 11.

	Size ratio	N_{HK}	η_{HK}	N_{IL}	η_{IL}
Min. size	0.21	4.5	0.20	0.15	2.7
Std. size	1	21.4	0.042	0.71	0.57
Min/Std		0.21	1/0.21	0.21	1/0.21

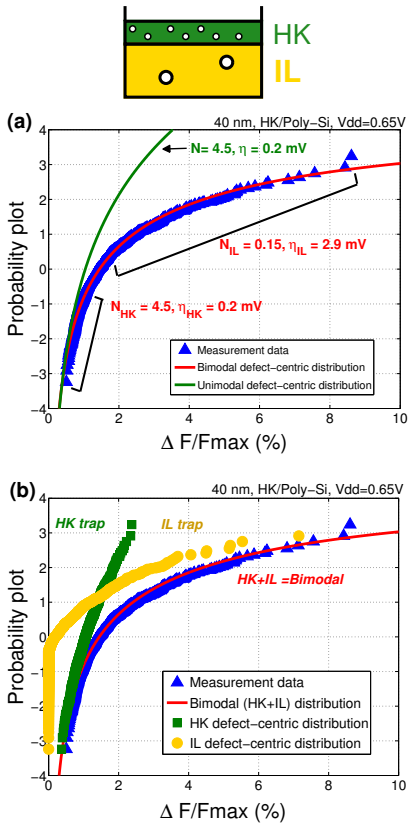


Fig. 9: 40nm HK/Poly-Si gate stack results. (a) Bimodal distribution is better to describe the measured $\Delta F/F_{\max}$ than unimodal distribution. (b) Bimodal distribution is a combination of HK/IL trap.

of HK and IL traps in Fig. 9 (b). Fig. 10 (a), (b) shows that the bimodal defect-centric model well describes $\Delta F/F_{\max}$ distributions than the unimodal defect-centric model in 28 nm HKMG. The measured $\Delta F/F_{\max}$ distributions cannot be modeled by the unimodal distribution.

As shown in Figs. 9 (b) and 10 (b), defects in High-K dielectric fluctuates frequency widely but the impact to frequency fluctuations is smaller than traps in interface layers. On the other hand, the number of traps in interface layers is much less than those in HK, but the impacts to frequency fluctuation are significant.

When the transistor width is doubled, the number of traps per transistor N are doubled while the average ΔV_{th} per traps η becomes half. From these assumption, the fitting parameters are obtained from the distribution of ROs with the minimum-size transistors. The measured $\Delta F/F_{\max}$ distribution of ROs with the standard-size inverters is almost identical with the simulated distribution with the parameters scaled from the minimum-size inverters. Fig. 11 (a) shows the fitting results of $\Delta F/F_{\max}$ distributions of ROs with two different transistor widths. Fig. 11 (b) The ratio of transistor widths is 1:0.21. Table II show the parameters to obtain the distributions.

As shown in Fig. 12, η_{IL} decreases when reverse bias at the back gate is applied. On the other hand, when forward bias is applied, it increases. N_{HK}, N_{IL}, η_{HK} are constant.

VI. Conclusion

We accurately replicate RTN-induced frequency fluctuation distribution of ROs on both ultra thin and general HK processes, using a bimodal defect-centric distribution with the number of traps N_{HK}, N_{IL} and the impact to V_{th} per trap η_{HK}, η_{IL}. The parameters scaled from ROs with minimum-size transistor can replicate RTN-induced frequency distributions of ROs with standard-size transistor.

Acknowledgment:

The chips in this study have been fabricated by Renesas Electronics and ST Microelectronics and designed by EDA system supported by VDEC, Synopsys Inc., Cadence Design System and Mentor Graphics.

References

[1] B. Kaczer, J. Franco, P. J. Roussel, G. Groeseneken, T. Chiarella, N. Horiguchi, and T. Grasser, "Extraction of the random component of time-dependent variability using matched pairs," *IEEE Electron Device Letters*, vol. 36, no. 4, pp. 300–302, April 2015.

[2] T. Matsumoto, K. Kobayashi, and H. Onodera, "Impact of random telegraph noise on CMOS logic circuit reliability," in *Custom Integrated Circuit Conference*, Sept. 2014, pp. 14–4.

[3] K. Ito, T. Matsumoto, S. Nishizawa, H. Sunagawa, K. Kobayashi, and H. Onodera, "The impact of RTN on performance fluctuation in CMOS logic circuits," in *Proc. Int. Reliability Phys. Symp.*, Apr. 2011, pp. CR.5.1–CR.5.4.

[4] K. H. Joo, C. R. Moon, S. N. Lee, X. Wang, J. K. Yang, I. S. Yeo, D. Lee, O. Nam, U. I. Chung, J. T. Moon, and B. I. Ryu, "Novel charge trap devices with NCBO trap layers for NVM or image sensor," in *International Electron Devices Meeting*, Dec 2006, pp. 1–4.

[5] H. Kurata, K. Otsuga, A. Kotabe, S. Kajiyama, T. Osabe, Y. Sasago, S. Narumi, K. Tokami, S. Kamohara, and O. Tsuchiya, "Random telegraph signal in flash memory: Its impact on scaling of multilevel flash memory beyond the 90-nm node," *IEEE Journal of Solid-State Circuits*, vol. 42, no. 6, pp. 1362–1369, June 2007.

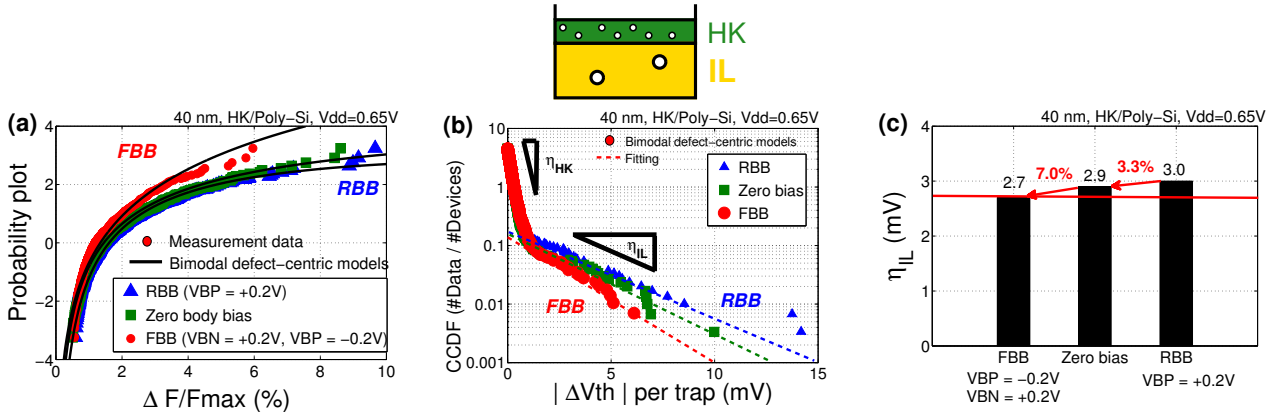


Fig. 12: Back-gate bias dependence of (a) $\Delta F/F_{\max}$ and (b) $\Delta V_{\rm th}$ per trap with 40 nm HK/Poly-Si gate stack. FBB reduces RTN-induced $\Delta F/F_{\max}$, while RBB increases it. This is because the FBB reduces $\eta_{\rm IL}$ and the RBB increases it (c).

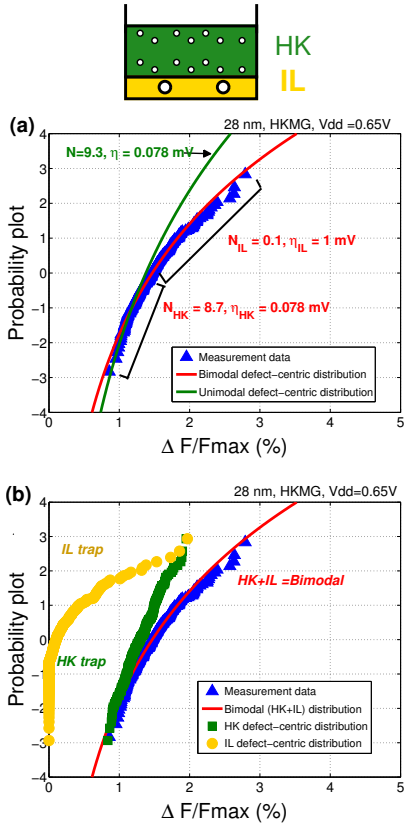
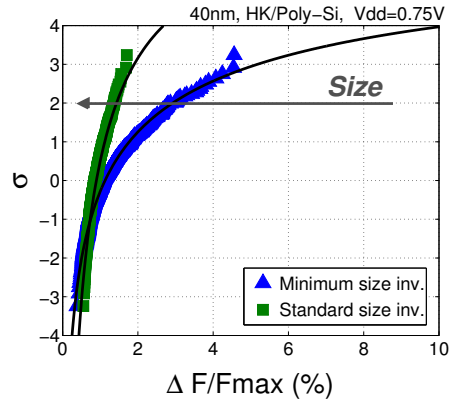
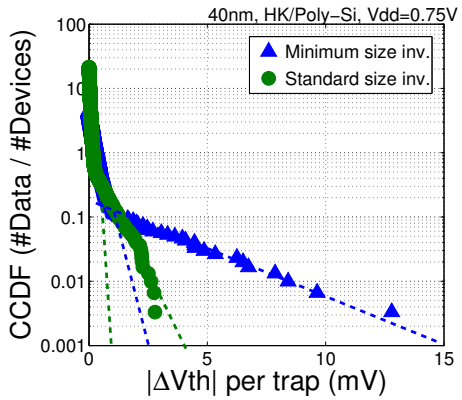


Fig. 10: 28nm HKMG gate stack results. (a) Bimodal distribution replicates the measured $\Delta F/F_{\max}$ (b) HK trap, IL trap and HK+IL traps.



(a) Distributions of freq. fluctuations.



(b) Threshold voltage distribution.

Fig. 11: Dependence of Transistor size for frequency fluctuation and $\Delta V_{\rm th}$ distributions in 40 nm

[6] M. Yamaoka, H. Miki, A. Bansal, S. Wu, D. J. Frank, E. Leobandung, and K. Torii, "Evaluation methodology for random telegraph noise effects in SRAM arrays," in *Electron Devices Meeting (IEDM), 2011 IEEE International*, Dec 2011, pp. 32.2.1–32.2.4.

[7] P. Weckx, B. Kaczer, C. Chen, J. Franco, E. Bury, K. Chanda, J. Watt, P. Roussel, F. Catthoor, and G. Groeseneken, "Characterization of time-dependent variability using 32k transistor arrays in an advanced HK/MG technology," in *IEEE International Reliability Physics Symposium*, April 2015, pp. 3B.1.1–3B.1.6.

[8] M. Toledano-Luque, B. Kaczer, J. Franco, P. Roussel, T. Grasser, T. Hoffmann, and G. Groeseneken, "From mean values to distributions of BTI lifetime of deeply scaled FETs through atomistic understanding of the degradation," in *Symposium on VLSI Technology*, June 2011, pp. 152–153.

[9] A. Oshima, P. Weckx, B. Kaczer, K. Kobayashi, and T. Matsumoto, "Impact of random telegraph noise on ring oscillators evaluated by circuit-level simulations," in *ICICDT*, June 2015.

[10] H. Nakamura, Y. Nakahara, N. Kimizuka, T. Abe, I. Yamamoto, T. Fukase, T. Nakayama, K. Taniguchi, K. Masuzaki, K. Uejima, T. Iwamoto, T. Tatsumi, and K. Imai, "55nm CMOS technology for low standby power/generic applications deploying the combination of gate work function control by HfSiON and stress-induced mobility enhancement," in *Symposium on VLSI Technology*, 2006, pp. 158–159.