Study on Impact of Random Telegraph Noise on Scaled MOSFETs

著者	手賀 直樹
year	2014
その他のタイトル	ランダム・テレグラフ・ノイズの微細MOSFETへの影響に関する研究
学位授与大学	筑波大学 (University of Tsukuba)
学位授与年度	2013
報告番号	12102甲第6820号
URL	http://hdl.handle.net/2241/00123312

Study on Impact of Random Telegraph Noise on Scaled MOSFETs

Naoki Tega

February 2014

Study on Impact of Random Telegraph Noise on Scaled MOSFETs

Naoki Tega

Doctoral Program in Nano-Science and Nano-Technology

Submitted to the Graduate School of

Pure and Applied Sciences

in Partial Fulfillment of the Requirements

for the Degree of Doctor of Philosophy in

Engineering

at the

University of Tsukuba

Content

1. Introduction

1.1 Device scaling and reliability

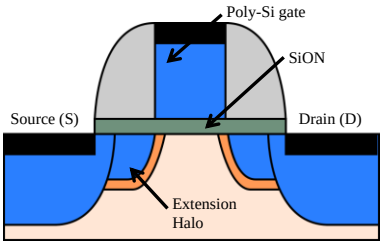
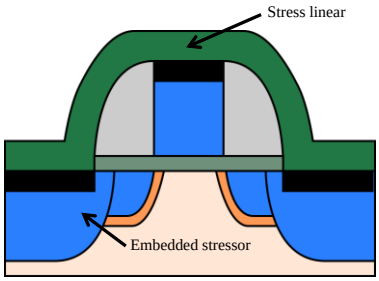
The scaling of metal-oxide-semiconductor field-effect transistor (MOSFET) has played the role of an engine for an expansion of semiconductor industry for forty years [1, 2]. According to the Moore's law, the number of transistors on an integrated–circuits (IC) chip is expected to double every two years [1]. The Moore's law is a prediction and a rule of thumb in the history of computing hardware. Presently, this increase in density of transistors is achieved by constant-field scaling of complementary MOS (CMOS) [2]. In the constant-filed scaling, keeping the electric-field constant makes it possible to reduce power-supply voltage and device dimensions, and as a result, it leads to greater density of MOSFETs on ICs, higher clock speed, and reduction of power consumption.

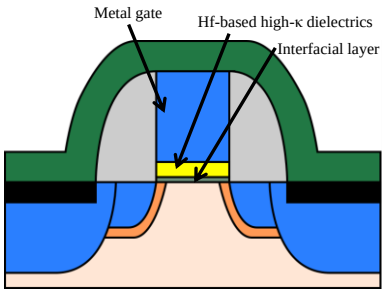
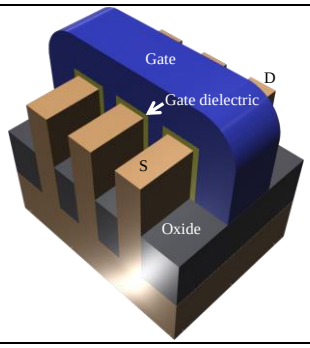
The scaling law greatly contributes to development of the semiconductor industry and, in the broad sense, development of an information society. However, in fact, many researchers and engineers faced various challenges and had to overcome them to achieve the scaling of MOSFET. Table 1.1 summaries a trend of MOSFET structures from 250 nm node, their characteristics, and their challenges. Especially, the author focuses on reliability challenges for each technology. After the 250 nm node, high-concentration extension has been applied to source and drain regions under sidewall spacers, and furthermore, heavy halo implantation has been locally formed close to the source/drain junctions [3-6]. Once a hot-carrier-induced degradation (HCD) was a serious problem and a lightly-doped-drain (LDD) structure was adjusted to improve HCD, however HCD was not a big challenge for the scaled MOSFETs because a voltage drain (Vdd) became lower than 2.5 V in the 250 nm generation and beyond.

In contrast to HCD, a short channel effect (SCE) has arisen, and the extension and the halo structures have been applied to suppress a threshold voltage (V_{th}) roll-off and drain-induced-barrier lowering (DIBL) due to SCE. However, the heavy halo implantation posed a degradation of channel mobility, and then, some effective ways for keeping high drive current, for example, thinner gate dielectrics, were used [5, 7, 8]. In 250 nm generation and beyond, a dual-poly gate technique, in which each dual-poly gate includes an n-type polycrystalline silicon as a gate of nMOSFET and a p-type polycrystalline silicon as a gate of pMOSFET, was also required for reduction of SCE [9, 10].

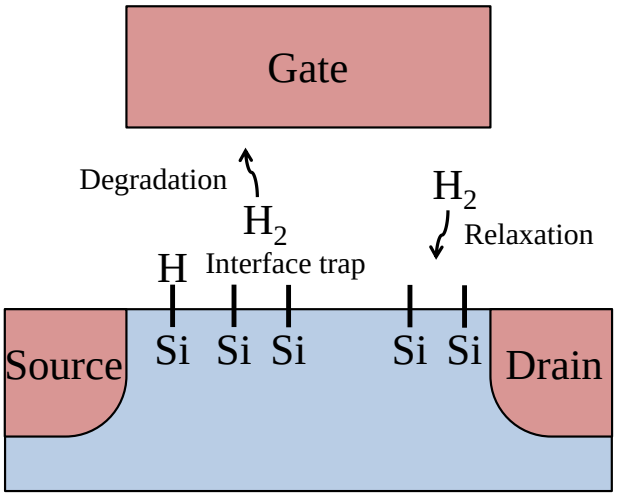
Changes occurred at the gate dielectric with the dual-poly gate technique, and a silicon oxynitride (SiON) came into wide use as gate dielectric to suppress the boron diffusion from the poly gate of pMOSFET from the 180 nm generation [11]. SiON as the gate dielectric brought in a thinner equivalent oxide thickness (EOT) and good time-dependent dielectric breakdown (TDDB) characteristics. On the other hand, SiON raised severe other reliability issue, namely negative bias temperature instability (NBTI) in pMOSFET [12]. Even today, NBTI is one of the

Table 1.1 Trend of MOSFET structure and challege

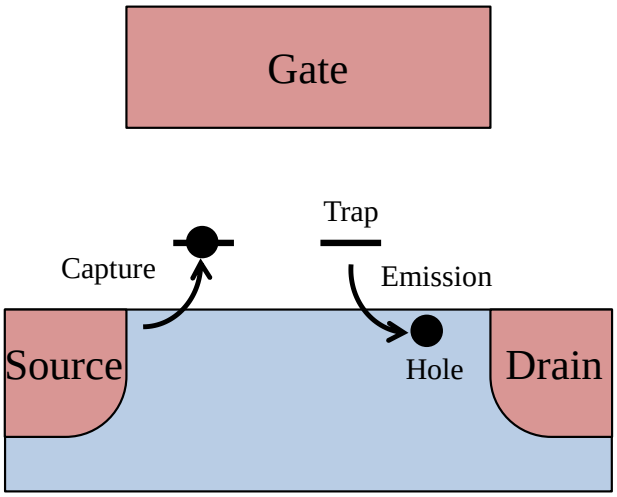
Technology node	250 nm ~	130 nm ~
Device structure		
Characteristics	Source / Drain extension & Halo, Dual-poly gate SiON gate dielectric	Stressors (Stress linear, Embedded SiGe and Si:C) Thin gate dielectric (< 1.5nm)
Challenges	short channel effect (SCE) Degraded mobility Negative bias temperature instability (NBTI)	SCE NBTI Gate leakage current Random dopant fluctuation

Technology node	45 nm ~	22 nm ~
Device structure		
Characteristics	Hf-based high-κ dielectric Metal gate	Multiple gate 3D structure
Challenges	SCE NBTI Variability	NBTI Variability Random telegraph noise

major reliability issues for the scaled pMOSFET. NBTI degradation worsen at high temperature, causing a large negative threshold voltage shift and decrease in drain current (I_d) and transconductance (G_m). Consequently, it causes a decrease in circuit speed and an increase in energy consumption. The mechanism of the NBTI degradation is less certain. Two NBTI



(a) Reaction-diffusion model.



(b) Switching trap model.

Fig. 1.1 Candidate models of NBTI degradation and recovery.

physics have been proposed to understand NBTI degradation so far as shown in Fig. 1.1. One is the reaction-diffusion (R-D) model [13]. The other is the switching trap model [14]. The R-D model is widely accepted as the most likely explanation for the NBTI degradation until quite recently. In the R-D model, the physics of interface trap generation is based on electrochemical reaction. The R-D model states that the NBTI stress induced degradation of pMOSFET is derived by breaking of hydrogen-passivated silicon bonds at the interface and subsequent diffusion of hydrogen as illustrated in Fig. 1.1(a). One of the significant features of the NBTI

behavior is a fast recovery. The fast-recovery component is explained by hydrogen moving back to the interface again in the R-D model. However, recently, the main component in relaxation state is regarded as not the electrochemical reaction stated in the R-D model, as but the hole detrapping phenomenon [14]. On the other hand, T. Grasser introduces the switching trap model more simply to understand the NBTI degradation and recovery. In the switching trap model, the NBTI degradation and recovery are understood by combination of random telegraph noises (RTN) with different time constants as shown in Fig. 1.1(b). RTN is observed as binary fluctuation in V_{th} or I_d in response to the capture and emission of carrier at single trap in the gate oxide. The author focuses on an impact of RTN on scaled MOSFET and intends to introduce RTN in more detail in the next section 1.2.

There are challenges of the 130 nm generation and beyond CMOS technology besides the NBTI degradation. They are gate leakage current and random dopant fluctuation (RDF) [15, 16]. Gate leakage current increasingly posed severe problem to the scaled MOSFET because thickness of the gate dielectric (T_{ox}) became about 1.5 nm. Moreover, the impact of the discrete dopant profiling on the variability of V_{th} in small device became obvious from the 130 nm generation because scaling speed of T_{ox} is not so much fast as gate area, and accordingly, the technology faced a severe problem with V_{th} variability. The two challenges remained a serious threat to the 90 nm and the subsequent 65 nm generations. When new gate stacks structure incorporating Hf-based high-κ dielectric and metal gate (HK / MG) began to be put to practical use in some advanced processers at the 45 nm generation, the adaption of HK / MG stacks led to a significant improvement of the gate leakage current and RDF [17, 18]. On the other hand, some variability problems, for example, line edge roughness (LER), work-function variation comes from the orientation of grains in metal gate, have been remained in scaled MOSFETs to this day [19, 20].

By the way, HK / MG stacks include an interfacial layer (IL) between substrate and Hf-based high-κ dielectric, which usually consists of SiON, because it is expected to prevent or at least minimize an interfacial reaction between the high-κ oxide and the underlying Si [21]. SiON film is likely to be required for the thermal stability for some time in the future. In other words, the above-mentioned NBTI degradation will remain as one of the severe problems until the effect of the SiON film is not needed.

The first processer chips on the 45-nm HK / MG technology were released in 2007. Manufactures worked in research and development of technology for the 45 nm generation and beyond. At around the same time, 2009 International technology roadmap for semiconductors (ITRS), for the first time, mentioned that an attention is needed to be paid to RTN for static random access memory (SRAM) scaling because its acceptable noise margins are becoming narrower due to increasing V_{th} variability including RTN [22]. 2011 ITRS also took up the issue

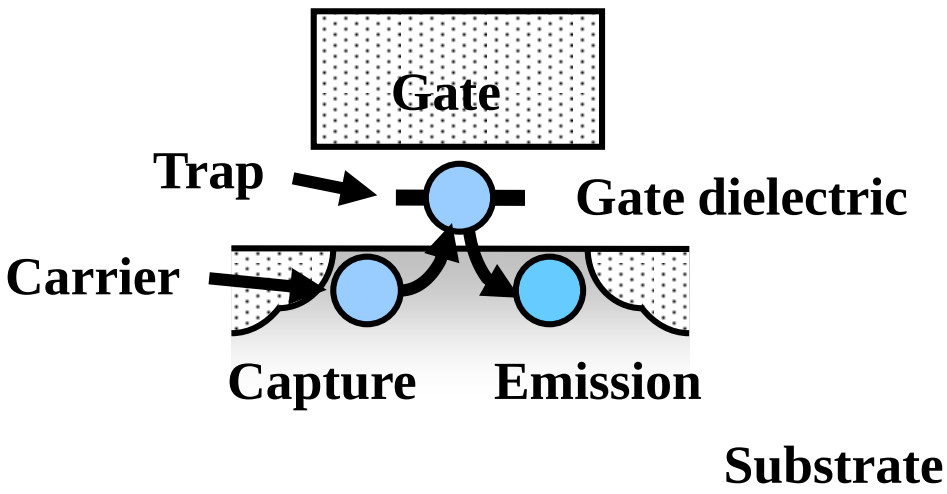


Fig. 1.2 RTN behavior is caused by single trap at gate dielectric.

of RTN for SRAM scaling [23]. The RTN behavior is the most fundamental property of trap at gate dielectric. There are not enough measures yet to improve the RTN problem.

The Manufactures are accelerating research and development of a multi-gate transistor after the practical application of the HK / MG technique to overcome SCE due to the device size decrease. At the moment, 2014, Intel Corporation commercially released the multi-gate transistor called Tri-gate on the 22-nm technology [24]. The other manufactures will be expected to release their multi-gate transistors in the immediate future. However, variability such as LER and RDF, the NBTI degradation, and RTN will be likely to remain as the reliability challenges even in the age of the multi-gate transistor.

1.2 Random telegraph noise (RTN) in MOSFET

The discrete switching in the current in semiconductor has been already known in 1950s and 1960s [25]. The first observation of the single-switching event as RTN is conducted using junction field effect transistor (JFET) in 1978 [26]. The RTN behaviors on semiconductor devices have been reported since early times. However, an observation of RTN behavior on MOSFET and identifying the cause of RTN, namely, interface trap and bulk trap at gate oxide, have not been reported until 1984 [27]. The first report by K. S. Rails et al. indicated that $1/f$ noise on MOSFET is likely to consist of many RTNs from observation. Before that time, A. McWhorter suggested the number fluctuation model using Ge filament in 1957 [28]. This model states that $1/f$ noise is composed of a large number of RTNs.

RTN is caused by single trap at gate dielectric capturing and emitting a carrier as shown in

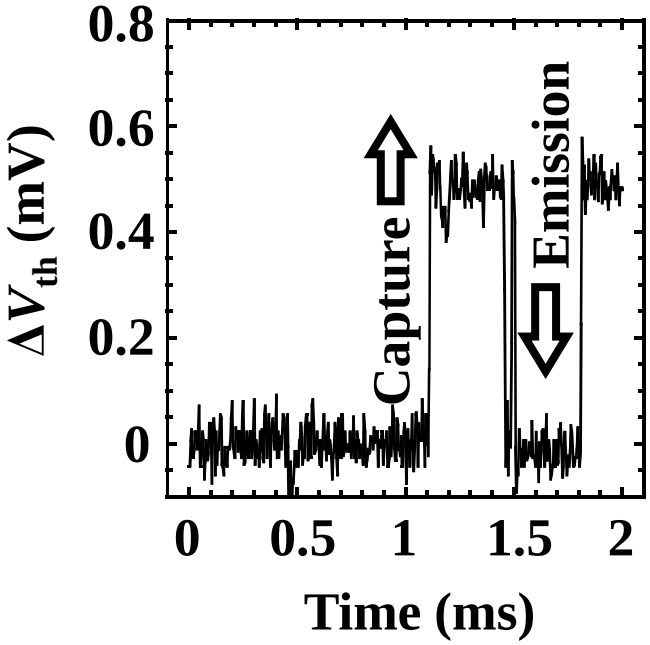


Fig. 1.3 Binary threshold voltage fluctuation due to RTN.

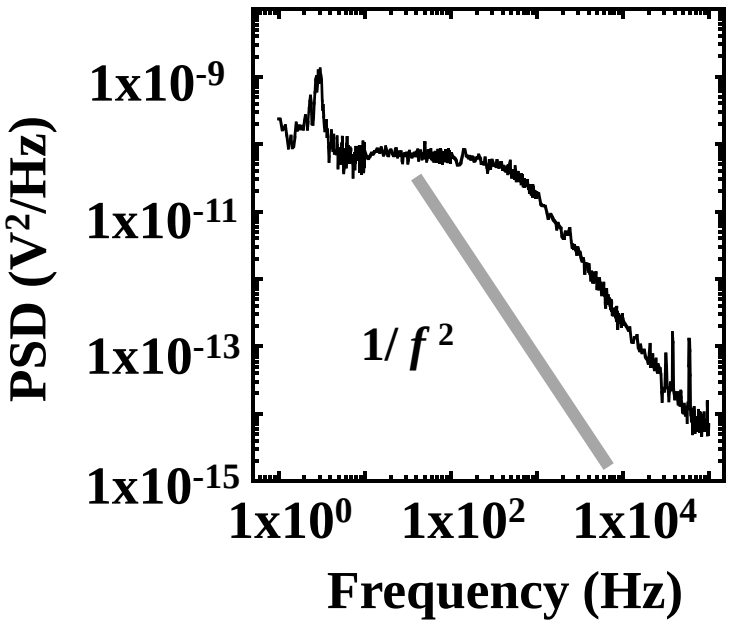


Fig. 1.4 Lorentzian power spectrum density based on RTN waveform in Fig. 1.3.

Fig. 1.2. Figure 1.3 is binary V_{th} variation due to single RTN trap in nMOSFET. The lower states are the duration time to capture (τ_c) and the upper states are the duration time to emission (τ_e).

τ_c and τ_e are known to show exponential distribution. An important property of the exponential distribution is "memorylessness". The name "Random" in Random telegraph noise comes from this memoryless characteristic. The relationship between the exponential function and the Lorentzian function is known as duality in Fourier transform. Therefore, the power spectrum density (PSD) of RTN is shown as below.

$$S(f) = \frac{4(\Delta V)^2}{(\bar{\tau}_c+\bar{\tau}_e)[(1/\bar{\tau}_c+1/\bar{\tau}_e)^2+(2\pi f)^2]} \quad (1.1)$$

f is the frequency, ΔV is the amplitude of RTN, $\bar{\tau}_c$ and $\bar{\tau}_e$ are the average of τ_c and τ_e, respectively. Figure 1.4 is an example of PSD calculated using the data in Fig. 1.3. In general, the plateau part and $1/f^2$ part can be separated from the Lorentzian power spectrum as shown in Fig. 1.3. When $\bar{\tau}_c$ is equal to $\bar{\tau}_e$ and $\bar{\tau}$, Equation (1.1) can be deformed as shown in Eq. (1.2)

$$S(f) = \frac{\Delta V^2}{2} \cdot \frac{\bar{\tau}}{1+\pi^2 f^2 \bar{\tau}^2} \quad (1.2)$$

S. Christensson et al. successfully applied Eq. (1.2) to $1/f$ noise on MOSFET on the assumption that the traps at gate oxide are source of noise, and the trap depth determines the time constant ($\bar{\tau}$) in 1968 [29]. $\bar{\tau}$ is simply expressed as Eq. (1.3) from the trap depth y because $\bar{\tau}$ follows the tunneling probability.

$$\bar{\tau} = \tau_0 \exp(\gamma y) \quad (1.3)$$

τ_0 and γ are a constant. If the RTN traps uniformly spread in gate oxide, and the trap depths distribute from the interface between channel and gate oxide to d, the power spectrum density of $1/f$ noise is shown by the integral from zero to d of Eq. (1.2).

$$\begin{aligned} S(f) &= \frac{\Delta V^2}{2} \int_0^d \frac{\bar{\tau}}{1+\pi^2 f^2 \bar{\tau}^2} \mathrm{d}y \\ &= \frac{\Delta V^2}{2} \int_0^d \frac{\tau_0 \exp(\gamma y)}{1+\pi^2 f^2 [\tau_0 \exp(\gamma y)]^2} \mathrm{d}y \\ &= \frac{\Delta V^2}{2\alpha\pi f} \{\arctan[\tau_0 \pi f \exp(\gamma d)] - \arctan(\tau_0 \pi f)\} \quad (1.4) \end{aligned}$$

S. Christensson et al. noted that τ_0 is of order of 10^{-8} sec, $\exp(\gamma d)$ is nearly equal to 10^{17}. The

range of f is usually from 1 Hz to 100 kHz in the measurement of low-frequency noise. Accordingly, $\arctan[\tau_0 \pi f \exp(\gamma d)]$ is approximately $\pi/2$, and $\arctan(\tau_0 \pi f)$ is about zero. Finally, the power spectrum density of $1/f$ noise is

$$S(f) = \frac{\Delta V^2}{4\gamma f} \quad (1.5).$$

This $1/f$ noise model and the experimental result reported by K. S. Rails et al. help understanding of the low frequency noise in MOSFET. Furthermore, M. J. Kirton and M. J. Uren paved the way for the development of physics of $1/f$ noise and RTN in MOSFET from the Rails' discussion [30]. Until now, many reports' discussions about dependences of RTN on process, device structure, technology node, and etc. are basically based on the Kirton-Uren model. The capture 8 is attempt to advance the RTN physical model to explain a wide distribution of the RTN activation energy and a dependences of RTN time constant on gate voltage more logically and more accurately.

In analog devices, since smallest detectable signal, signal to noise ratio (SNR), and dynamic range are determined by noises, many researchers have earnestly made an investigation about $1/f$ noise as one of the low frequency. In contrast to $1/f$ noise, the attention has not been paid to RTN because the variation of RTN is acceptable level in both analog and digital devices. Researchers were interested in RTN as fundamental characteristics of trap at gate dielectric and, however, placed low priority on RTN as the reliability issue until 2000s.

It was, for the first time, reported at 2006 that the RTN became the reliability issue in high-capacity flash memory [31]. Why did RTN suddenly pose a serious threat to the flash memory? The answers are four points.

1) The RTN impact increases with scaling as indicated in Eq. (1.6).

$$\Delta V_{\mathrm{th}} = -\frac{q}{L_{\mathrm{g}} W_{\mathrm{g}} C_{\mathrm{ox}}} \quad (1.6)$$

ΔV_{th} is the threshold voltage variation, q is the elementary charge, L_g is the gate length, W_{g} is the gate width, and C_{ox} is the gate oxide capacitance of unit area.

2) The flash memory is fabricated by leading-edge technology and its technology continues to drive minimum feature size.

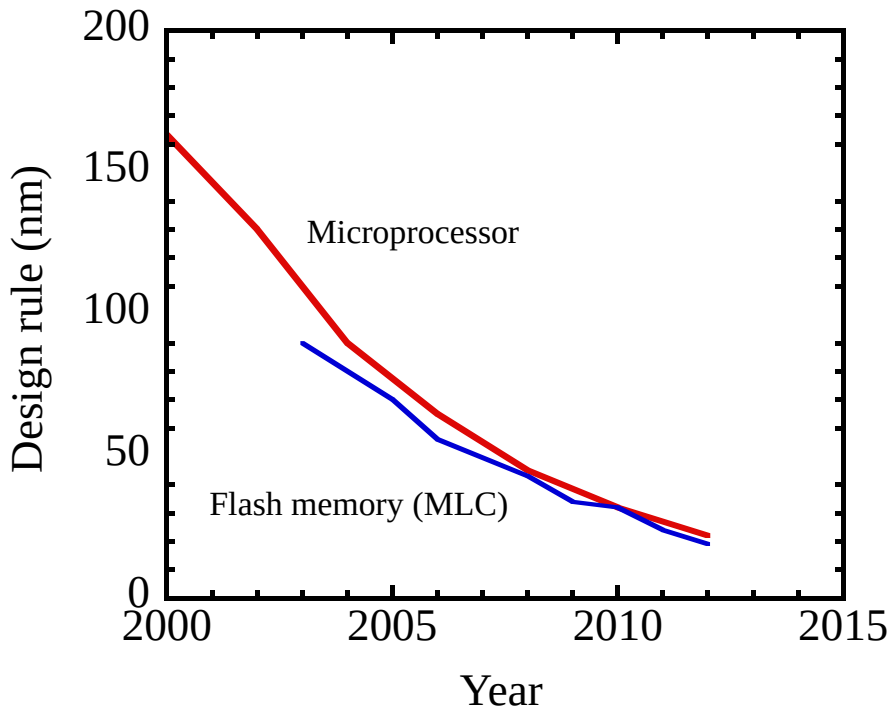


Fig. 1.5 Scaling trends of microprocessor and flash memory. The trend of flash memory with multi-level cell (MLC) technique is based on announcements by Toshiba at International Solid-State Circuits Conference (ISSCC). The trend of microprocessor is based on Intel announcements.

3) High-capacity flash memory adopts the multi-level technique to increase bits per cell, and its V_{th} margin decreases as negative effect.

4) A high field stress on thin oxide, for example, less than 9 ~ 10 nm, increases the current density at low electric field. This problem is known as stress-induced leakage current (SILC) and obstructs decreasing gate-oxide thickness with scaling [32]. Therefore, RTN V_{th} variation in high-capacity flash memory is likely to be large in comparison with the other digital devices with same technology as indicated in Eq.(1.6).

RTN is still one of the reliability issues in high-capacity flash memory.

The author predicted that RTN also has an impact not only on the high-capacity flash memory but also on the scaled logic devices and SRAM in the immediate future because the CMOS technology shows the same scaling trend as the high-capacity flash memory as show in Fig.1.5. The area of the latest 22-nm technology is less than one-tenth of the 90-nm technology in which RTN became obvious in high-capacity flash memory. Therefore, the author needed to plan to accurately study the impact of RTN on scaled CMOS, especially scaled SRAM with

narrow margin, beforehand.

This paper consists of nine chapters. The outline of each chapter is described below.

Chapter 2: The author suggests new statistical method to analyze RTN V_{th} variation. RTN V_{th} variation follows Gumbel distribution, which is one of the extreme value distributions. In this chapter, the author also discusses whether the Gumbel law or the Fréchet law is more suitable for investigation of the impact of RTN on scaled MOS by regression analysis.

Chapter 3: The statistical distribution of RTN V_{th} variation was measured and characterized by use of scaled PDSOI MOSFETs with SiON / polycrystalline Si gate down to 20 nm gate length. Moreover, the dependences of the RTN variation on L_g and W_g are shown.

Chapter 4: The author demonstrates the reduction of RTN in high-κ / metal gate (HK / MG) stacks incorporated in 22 nm generation MOSFETs. Many thousands of such MOSFETs have been fabricated, measured, and analyzed using a statistical technique. Based on a statistical comparison of these MOSFETs, the author finds that high temperature forming gas annealing (HTFGA) can suppress RTN ΔV_{th}. In addition, properly annealed HK FETs have smaller RTN ΔV_{th} than SiON / poly-Si MOSFETs due to fewer traps and to thinner inversion thickness in HK / MG.

Chapter 5: It is clarified that suboxides and interface traps are closely linked to ΔV_{th} due to RTN from an investigation of dependence of ΔV_{th} on silicon-surface orientation: Si(100), (110), and (111). The amount of RTN traps increases with increasing amount of suboxides in the interfacial transition layer. With regard to the total amount of suboxides, the Si(110) surface orientation gives a larger amount than Si(100) and Si(111). Furthermore, the author found that Si(110) has the potential to give fast RTNs with a larger amplitude than Si(111). Accordingly, ΔV_{th} for Si(110) is larger than those of Si(100) and Si(111).

Chapter 6: The impact of RTN on a scaled-down SRAM is shown. To estimate the impact on SRAM, the author statistically analyzed V_{th} variation of n- and pMOSFETs. It is revealed that ΔV_{th} of pMOSFET is larger than that of nMOSFET. This difference can be explained by taking into account both the number- and mobility-fluctuation models of RTN.

Chapter 7: Change in generation of RTNs before and after NBTI stress is demonstrated to reveal NBTI degradation and recovery from statistical perspective. The NBTI stress generates a large number of permanent interface traps and, at the same time, the temporary and the one-time

RTNs. The two types of traps show different features. A re-passivation of interface states is minority in recovery process after the NBTI stress, and in contrast, rapid disappear of the temporary and the one-time RTNs mainly causes the recovery phenomenon. Distinguishing characteristic is that the RTN traps are less likely to become permanent traps in contrast to the interface traps. This two-type trap model simply explains NBTI degradation and recovery in scaled pMOSFET.

Chapter 8: Physical model of RTN is proposed to explain large activation energies for $\bar{\tau}_e/\bar{\tau}_c$, $\bar{\tau}_e$, and $\bar{\tau}_c$ respectively and large gate bias sensitivity of $\bar{\tau}_e/\bar{\tau}_c$. In this chapter, to propose a physical model of RTN in MOSFET, the author newly introduces the coulomb blockade theory, a relationship between a conducting sphere and parallel plates, and an advanced non-radiative muliti-phonon theory (NMP).

Chapter 9: The author summarizes the impact of RTN on technologies for scaling, the relationship between NBTI and RTN, and the proposed RTN physical model. Moreover, the author shows future prospects and remaining problems

1.3 References

[1] G. E. Moore, “Cramming More Components onto Integrated Circuits,” *Electronics*, vol. 38, 19, Apr., 1965.

[2] R. H. Dennard, F. H. Gaensslen, H-N Yu, V. L. Rideout, E. Bassous, and A. R. Leblanc, “Design of Ion-Implanted MOSFET’s with very small physical dimensions,” *IEEE J. Solid-Sate Circuits*, vol. 9, pp.256-268, Oct., 1974.

[3] C. F. Codella, and S. Ogura, “Halo Doping Effects in Submicron DI-LDD Device Design,” in *IEDM Tech. Dig.*, pp. 230–233, 1985.

[4] S. Thompson, P. Packan, T. Ghani, M. Stettler, M. Alavi, I. Post, S. Tyagi, S. Ahmed, S. Yang, and M. Bohr, “Source/Drain Extension Scaling for 0.1 μm and Below Channel Length MOSFETS,” in *VLSI Symp. Tech. Dig.*, pp. 132–133, 1996.

[5] H. Hwang, D-H. Lee, and J. M. Hwang, “Degradation of MOSFETs Drive Current Due to Halo Ion Implantation,” in *IEDM Tech. Dig.*, pp. 567–570, 1996.

[6] M. Bohr, S. S. Ahmed, S. U. Ahmed, M. Bost, T. Ghani, J. Greason, R. Hainsey, C. Jan, P. Packan, S. Sivakumar, S. Thompson, J. Tsai, and S. Yang, “A High Performance 0.25 μm Logic Technology Optimized for 1.8 V Operation,” in *IEDM Tech. Dig.*, pp. 847–850, 1996.

[7] H. S. Momose, M. Ono, T. Yoshitomi, T. Ohguro, S. Nakamura, M. Saito, and H. Iwai, “Tunneling Gate Oxide Approach to Ultra-High Current Drive in Small-Geometry MOSFETs,” in *IEDM Tech. Dig.*, pp. 593–596, 1994.

[8] M. Rodder, S. Hattangady, N. Yu, W. Shiau, P. Nicollian, T. Laaksonen, C. P. Chao, M. Mehrotra, C. Lee, S. Murtaza, and S. Aur, "A 1.2 V 0.1 μm Gate Length CMOS Technology: Design and Process Issues," in *IEDM Tech. Dig.*, pp. 623–626, 1998.

[9] W-H. Chang, B. Davari, M. R. Wordeman, Y. Taur, C. C-H. Hsu, and M. D. Rodriguez, "A High-Performance 0.25-μm CMOS Technology: I-Design and Characterization," *IEEE Trans. Electron Devices*, vol. 39, no. 4, pp. 959–966, Apr., 1992.

[10] B. Davari, W-H. Chang, K. E. Petrillo, C. Y. Wong, D. Moy, Y. Taur, M. R. Wordeman, J. Y-C. Sun, C. C-H. Hsu, and d M. D. Rodriguez, "A High-Performance 0.25-μm CMOS Technology: II-Design and Characterization," *IEEE Trans. Electron Devices*, vol. 39, no. 4, pp. 967–975, Apr., 1992.

[11] D. Wristers, L. K. Han, T. Chen, H. H. Wang, and D. L. Kwong, M. Allen, and J. Fulford, "Degradation of oxynitride gate dielectric reliability due to boron diffusion," *Appl. Phys. Lett.*, vol. 68, no. 15, pp. 2094–2096, Apr., 1996.

[12] J.H. Stathis, and S. Zafar, "The negative bias temperature instability in MOS devices: A review," *Microelectron. Reliab.*, vol. 46, pp. 270–86, 286, Feb. –Apr. 2006.

[13] M. A. Alam, "A critical examination of the mechanics of dynamic NBTI for PMOSFETs," in *IEDM Tech. Dig.*, pp. 14.4.1–14.4.4, 2003.

[14] T. Grasser, "Switching oxide traps as the missing link between negative bias temperature instability and random telegraph noise," in *IEDM Tech. Dig.*, pp. 1–4, 2009.

[15] Y-C. Yeo, T-J. King, and C. Hu, "Direct tunneling leakage current and scalability of alternative gate dielectrics," *Appl. Phys. Lett.*, vol. 81, pp. 2091–2093, Sep., 2002.

[16] A. Asenov, "Random Dopant Induced Threshold Voltage Lowering and Fluctuations in Sub-0.1 m MOSFET's: A 3-D"Atomistic" Simulation Study," *IEEE Trans. Electron Devices*, vol. 45, no. 12, pp. 2505–2513, Dec., 1998.

[17] K. J. Kuhn, "Reducing Variation in Advanced Logic Technologies: Approaches to Process and Design for Manufacturability of Nanoscale CMOS," in *IEDM Tech. Dig.*, pp. 471–474, 2007.

[18] K. Mistry et al., "A 45 nm logic technology with high-k+ metal gate transistors, strained silicon, 9 cu interconnect layers, 193 nm dry patterning, and 100% Pb-free packaging," in *IEDM Tech. Dig.*, pp. 247–250, 2007.

[19] H. Dadgour, V. De, and K. Banerjee, "Statistical Modeling of Metal-Gate Work-Function Variability in Emerging Device Technologies and Implications for Circuit Design," in *Proc. ICCAD*, pp. 270–277, 2008.

[20] A. Asenov, S. Kaya, and A. R. Brown," Intrinsic Parameter Fluctuations in Decananometer MOSFETs Introduced by Gate Line Edge Roughness," *IEEE Trans. Electron Devices*, vol. 50, no. 5, pp. 1254–1260, May, 2003.

[21] G. D.Wilk, R. M.Wallace, and J. M. Anthony, “High-k gate dielectrics: Current status and materials properties considerations,” *J. Appl. Phys.*, vol. 89, pp. 5243–5275, May, 2001.

[22] ITRS 2009 web site: http://www.itrs.net/Links/2009ITRS/Home2009.htm

[23] ITRS 2011 web site: http://www.itrs.net/Links/2011ITRS/Home2011.htm

[24] C. Auth et al., “A 22 nm high performance and low-power CMOS technology featuring fully-depleted tri-gate transistors, self-aligned contacts and high density MIM capacitors,” in *VLSI Symp. Tech. Dig.*, pp. 12–14, 2012.

[25] M.J. Buckingham, *Noise in electronic devices and systems*, Wiley, 1983.

[26] K. Kandiah, and F.B. Whiting, “Low frequency noise in junction field effect transistors,” *J. Appl. Phys.*, vol. 66, pp. 937-948, 1989.

[27] K. S. Rails, W. J. Skocpol, L. D. Jackel, R. E. Howard, L. A. Fetter, R. W. Epworth, and D. M. Tennant, “Discrete Resistance Switching in Submicrometer Silicon Inversion Layers: Individual Interface Traps and Low-Frequency (1/*f*?) Noise,” *Phys Rev. Lett.*, vol. 52, pp. 228–231, 1984.

[28] A. L. McWhorter, “1/f noise and germanium surface properties,” in Semiconductor Surface Physics, R. H. Kingston, Ed. Philadelphia, PA: Univ. Pennsylvania Press, pp. 207–228, 1957.

[29] S. Christensson, I. Lundström, and C. Svensson, “Low frequency noise in MOS transistors-I theory,” *Solid-State Electron*, vol. 11, pp. 797–812, 1968.

[30] M J Kirton and M J Uren, “Noise in Solid-State Microstructures: A New Perspective on Individual Defects, Interface States and Low-Frequency (1/f) Noise,” *Advances in Physics*, vol. 38, p. 367–468, 1989.

[31] H. Kurata, K. Otsuga, A. Kotabe, S. Kajiyama, T. Osabe, Y. Sasago, S. Narumi, K. Tokami, S. Kamohara, and O. Tsuchiya, “The impact of random telegraph signals on the scaling of multilevel Flash memories,” in *VLSI Symp. Circuit Dig.*, pp. 112–113, 2006.

[32] D. Ielmini, A. S. Spinelli and A. L. Lacaita, “Recent developments on Flash memory reliability,” *Micro. Engineering*, vol. 80, pp. 321–328, 2005.

2. Development of statistical method of RTN

2.1 Introduction

To estimate the impact of RTN on the digital devices, especially SRAM, statistical viewpoint is required because, in general, V_{th} variations above 5σ (standard deviation), that is, about 99.9999%, must be considered if the V_{th} variations follow a Gaussian distribution. For example, it was reported that RDF show the Gaussian distribution to at least 5σ [1]. Hence, the researchers make it relatively easy to analyze the V_{th} variations on the basis of the Gaussian distribution because the five-σ value is large as five as the one-σ value. Since V_{th} variations due to RTN, however, show a long-tailed non-Gaussian distribution [2], it is difficult to estimate the ΔV_{th} above the cumulative probability of 99.9999% without new statistical method of RTN. Furthermore, prior to estimate the impact of RTN on technology and scaling, the author introduces a suitable statistical theory for RTN V_{th} variations.

In this chapter, the extremely value distribution, especially the Gumbel distribution, is introduced to the analysis RTN V_{th} variations.

2.2 Device fabrication

A small transistor gate of less than 100 nm can provide clear ΔV_{th} of more than 10 mV, and more than 100 transistors with same size are needed to investigate critical values above 95% cumulative probability. Narrow-channel nMOSFETs with SiON / poly-Si gate stacks, which are designed to clearly show RTN behavior, were prepared. actual gate length and width were 45 nm and 25 nm, respectively, and the number of measured devices was over two hundreds. Figure 2.1 shows the cross-sectional scanning-electron-microscope (SEM) image of a shallow trench isolation (STI) with the actual width of 25 nm. Figure 2.2 shows the cross-sectional SEM image of a MOS structure with the actual gate length of 45 nm. Both the STI and the MOS structures are fabricated by mix-and-match lithography process, which uses electron beam system and KrF stepper. The gate dielectric is fabricated by dry oxidation process and oxidation anneal in nitric oxide. Numbers of nitrogen are 4 atom% near interface between gate dialectic and substrate.

2.3 Experimental

As shown in Fig. 2.3, the RTN measurement system consists of a fast-Fourier-transform (FFT) analyzer (Ono Sokki CF-5210) for observing a time series of ΔI_d, a current to voltage amplifier with gain of 10^5 (This amplifier is powered by a battery to suppress power-supply noise), a semiconductor parameter analyzer (Agilent 4156C) for measuring V_{th}, a switching matrix (Agilent B2200A fA leakage switch mainframe), a prober (Cascade S300), and a control PC. This measurement system has been set up at Central Research Laboratory, Hitachi. RTN

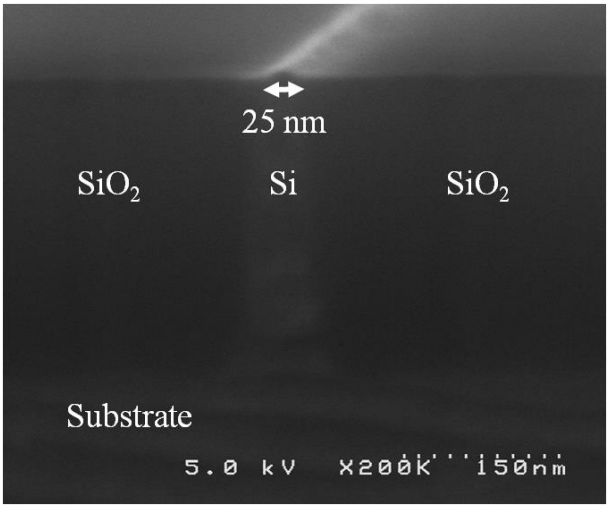


Fig. 2.1. Cross-sectional SEM image of STI with actual gate width of 25 nm.

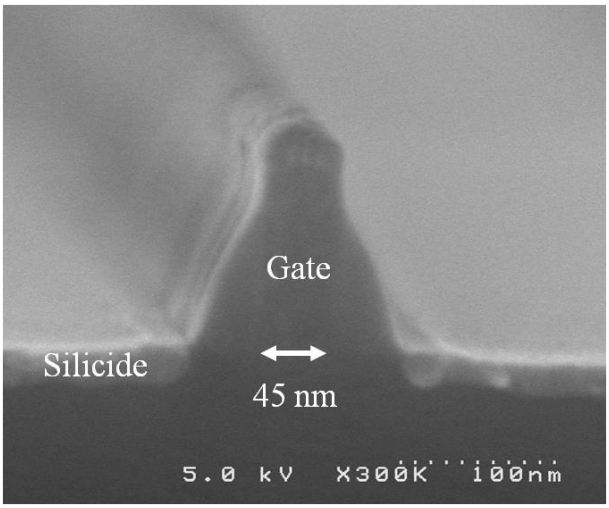


Fig. 2.2. Cross-sectional SEM image of MOS structure with actual gate length of 45 nm.

was measured under the following conditions: gate voltage (V_g) of 1.0 V and drain voltage (V_d) of 50 mV in the linear region, measurement frequency from 1 Hz to 100 kHz and measurement temperature of 25 ºC. For statistical analysis, this measurement was automatically repeated 244 times (i.e., for 244 different devices). With regard to data analysis, a hidden Markov model is used for extracting ΔI_d and decomposing overlapped RTNs [3]. The detailed hidden Markov model is introduced in chapter 7. The overlapped RTNs can be decomposed into up to 5 single RTNs and residue. I_d variation is defined as total amplitude of extracted single RTNs. ΔI_d is transformed into ΔV_{th} by using transconductance. In addition to Hitachi's MOSFETs, I prepared IBM's MOSFETs. The IBM's MOSFETs are introduced in the next chapter.

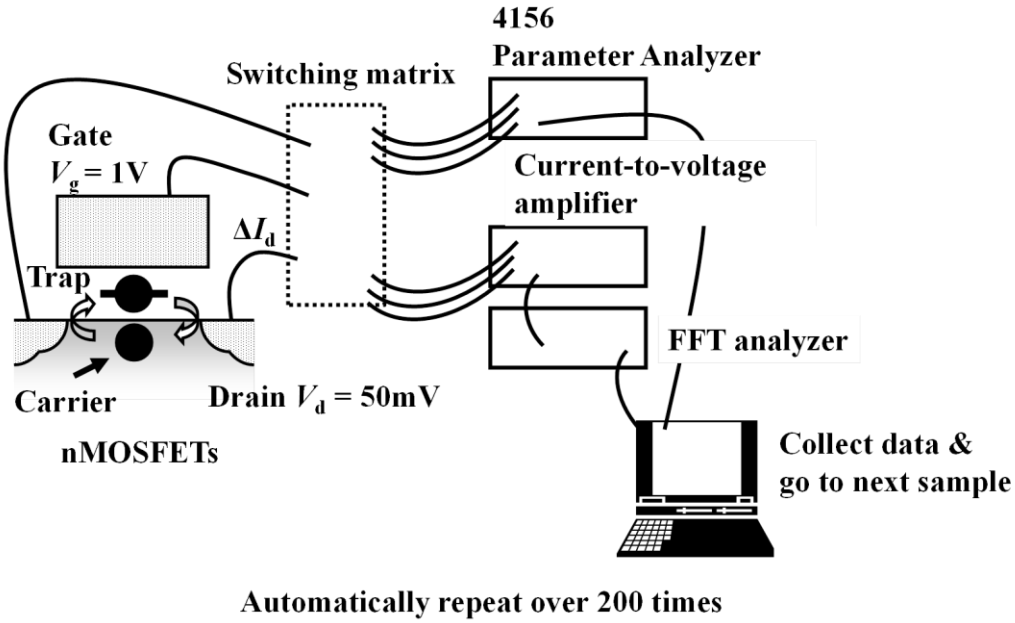


Fig. 2.3. Automatic measurement system for observation of RTN ΔV_{th}.

2.4 Statistical analysis RTN V_{th} variations

Examples of RTN waveforms are shown in Figs. 2.4 to 2.7. The typical binary fluctuation due to single RTN trap at gate dielectric is clearly shown in Fig. 2.4. The arrows indicate clear states. The amplitude depends on a location of RTN trap and discreteness of impurities in channel [4, 5]. According to these reports, RTN V_{th} variation maximizes in the case of the trap location at the center of channel and the edge of shallow trench isolation (STI). Moreover, ΔV_{th} becomes bigger if channel width becomes narrower due to the discreteness of impurities. That is, the binary V_{th} variations due to single RTN trap are not constant and are distributed. Figure 2.5 shows that complex RTN with 4 states is caused by two RTN traps, and each RTN has different V_{th} variations. The author defined the difference between the upper state and the lower state as RTN V_{th} variation to estimate the impact of RTN accurately. Multiple levels states because of more than 3 RTN traps appear in the time series as indicated Fig. 2.6. The RTN fluctuations with the different amplitudes and the different time constants are overlapped in this figure. Figure 2.7 demonstrates that a large-amplitude RTN and a small-amplitude RTN are overlapped. V_{th} variations due to RTNs are often complicated by multiple factors including the RTN trap positions and the discreteness of impurities around the RTN trap.

From these data, a useful statistical theory to meet the multiple factors should be introduced to compare the effectiveness of process for suppression of RTN V_{th} variation. First of all, the RTN V_{th} variations are plotted on a Gaussian distribution as demonstrated on Figs 2.8 and 2.9. The devices, which are used in the demonstration of Fig. 2.8, were fabricated in Hitachi

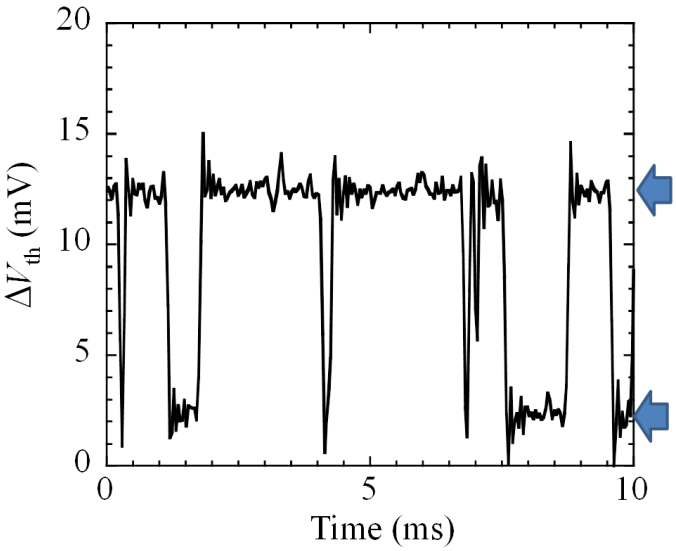


Fig. 2.4. Binary RTN due to single trap.

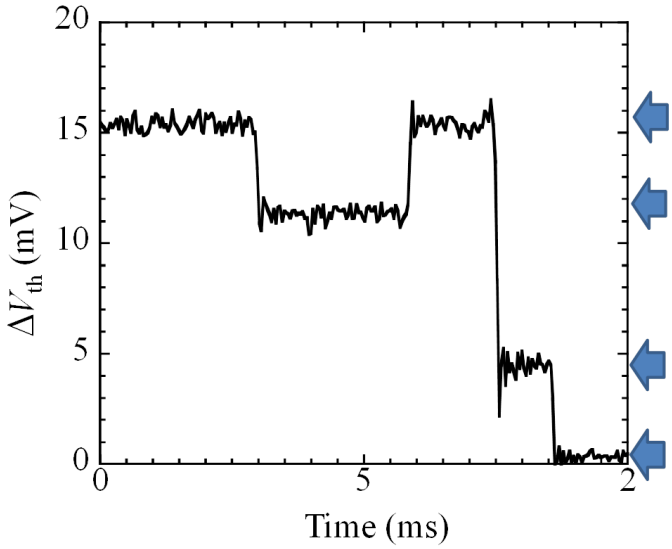


Fig. 2.5. Complex RTN with four states due to two RTN traps.

laboratory, and on the other hand, Ones, which are used in Fig. 2.9, were fabricated and measured in IBM laboratory. Both data shows a long-tailed non-Gaussian distribution and my conclusion is that the curious distribution is universal. The long tail on Gaussian plot is a quite characteristic of RTN distribution. The events on the tail are rare. These large RTN V_{th} variations cannot be explained by Gaussian distribution. For example, the large RTN V_{th} variations as demonstrated in Figs. 2.6 and 2.7 are rare events above 90 % cumulative probability. The causes of the large V_{th} variations look like complex RTN with multiple states and narrower channel due to the discreetness of impurities. Determining a suitable statistical

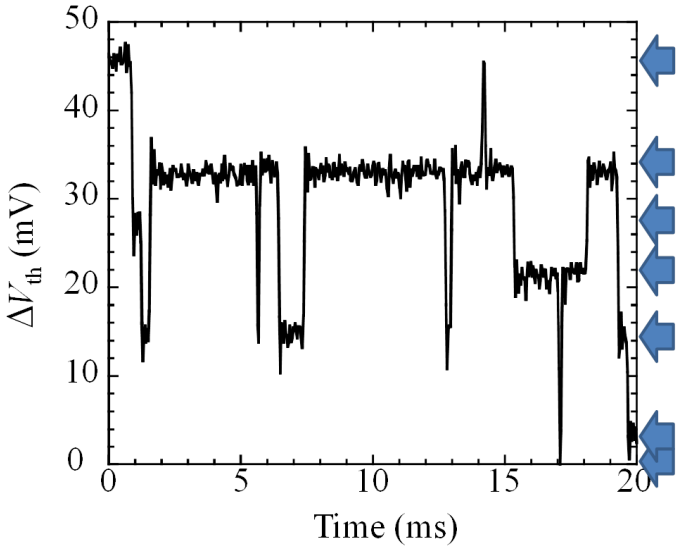


Fig. 2.6. Complex RTN with multiple states due to more than three RTN traps

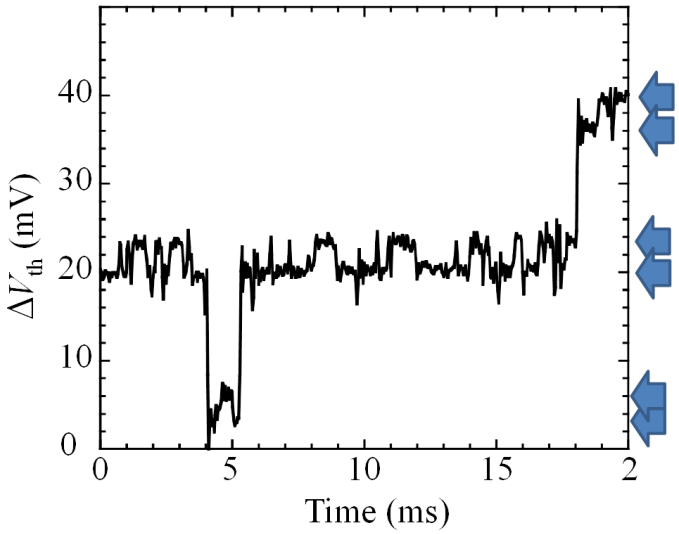


Fig. 2.7. Complex RTN consisted of both small- and large-amplitude RTNs.

distribution is required for a comparison of processes and an estimation of a critical value above high cumulative probability of 99.9999%.

My goal in this chapter is what I suggest a suitable statistical distribution which can show how often extremely large RTN V_{th} variations appear. One of the possible statistic theories for showing extremely rare and large RTN V_{th} variations is an extreme value theory. Extreme value distributions are the limiting distributions for the minimum or the maximum of a very large collection of random observations in each sample. Extreme value theory is useful theory of

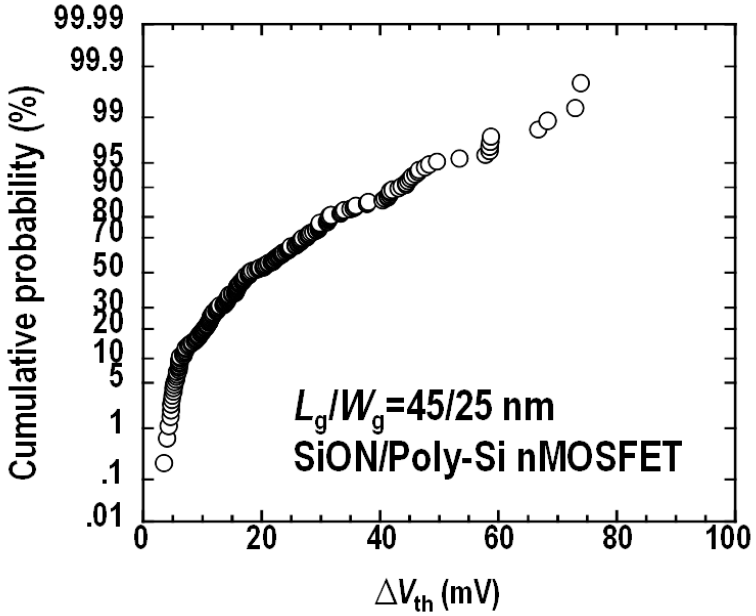


Fig. 2.8. RTN V_{th} variations on Gaussian plot using SiON / Poly-Si nMOFET with L_g / W_g of 45 / 25 nm. These devices were fabricated at Central Research Laboratory, Hitachi.

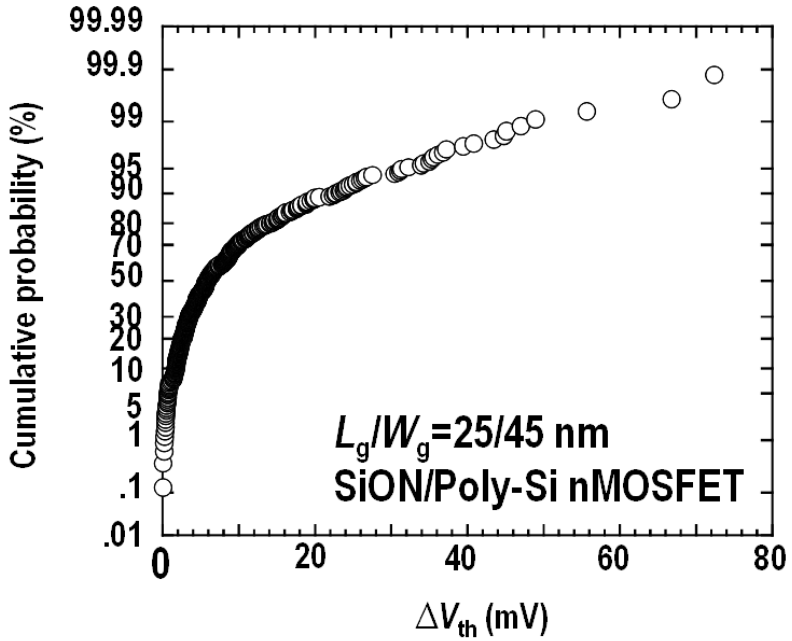


Fig. 2.9. RTN V_{th} variations on Gaussian plot using SiON / Poly-Si nMOFET with L_g / W_g of 25 / 45 nm. These devices were fabricated at T. J. Watson Research Center, IBM.

modeling rare event with extremely large or small values. This theory is also effective for assessing risk, namely rare events with extremely large or small values.

The extreme value distributions are categorized into three types: specifically, the Gumbel, the Fréchet, and the Weibull distributions, whose cumulative distribution functions are described below.

The Gumbel distribution: Type I extreme value distribution

$$F(x) = \exp\left\{-\exp\left[-\left(\frac{x-\mu_1}{\sigma}\right)\right]\right\},\ -\infty < x < \infty \quad (2.1)$$

$$f(x) = \frac{\partial F(x)}{\partial x} = \frac{1}{\sigma}\exp\left[-\left(\frac{x-\mu_1}{\sigma}\right)\right]\exp\left\{-\exp\left[-\left(\frac{x-\mu_1}{\sigma}\right)\right]\right\} \quad (2.2)$$

The Fréchet distribution: Type II extreme value distribution

$$F(x) = \exp\left[-\left(\frac{x-\mu_1}{\sigma}\right)^{-k_1}\right],\ \mu \le x \quad (2.3)$$

$$f(x) = \frac{\partial F(x)}{\partial x} = \frac{k_1}{\sigma}\left(\frac{x-\mu_1}{\sigma}\right)^{-k_1-1}\exp\left[-\left(\frac{x-\mu_1}{\sigma}\right)^{-k_1}\right] \quad (2.4)$$

The Weibull distribution: Type III extreme value distribution

$$F(x) = \exp\left\{-\left[\frac{-(x-\mu_1)}{\sigma}\right]^{k_1}\right\},\ \mu \ge x \quad (2.5)$$

$$f(x) = \frac{\partial F(x)}{\partial x} = \frac{k_1}{\sigma}\left[\frac{-(x-\mu_1)}{\sigma}\right]^{k_1-1}\exp\left\{-\left[\frac{-(x-\mu_1)}{\sigma}\right]^{k_1}\right\} \quad (2.6)$$

$F(x)$ is the cumulative distribution function. $f(x)$ is the frequency function. k_1 is the shape parameter. μ_1 is the location parameter. σ is the scaling parameter.

The Weibull distribution is used to show minimum values; that is, lifetimes and degradations of objects based on the weakest link theory. In reliability of MOSFET, the Weibull distribution is used to investigate the reliability of gate dielectric and, from the result of the statistical analysis, the percolation model was proposed [7]. The Gumbel distribution is used in the field of science and hydrology for extreme event. For example, in hydrology, the Gumbel distribution is used to analyze variations with rare events as monthly and annual maximum daily rainfall and river, [8] and also to describe droughts [9]. The Fréchet distribution is used for applications of

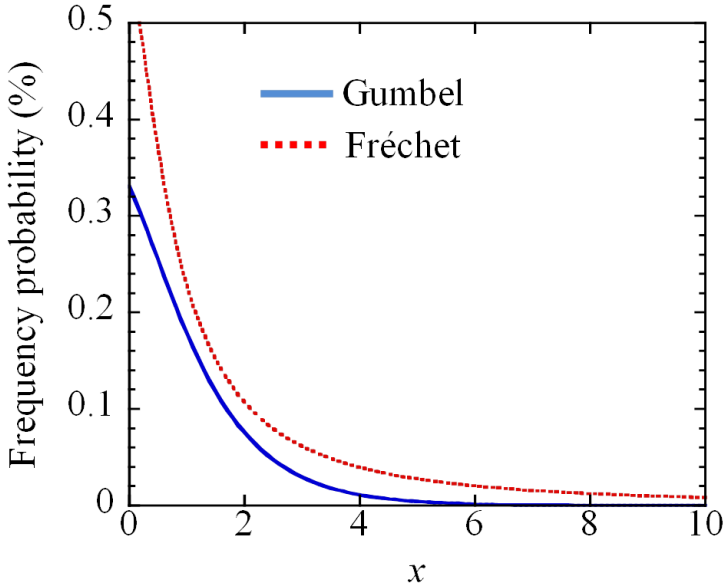


Fig. 2.10. Comparison of tail between Gumbel distribution and Fréchet distribution.

insurance and finance, which involve taking heavy risks because tail of the Fréchet distribution is heavier than that of the Gumbel distribution [10].

The Weibull distribution is the limiting distribution for the smallest observation in each sample. Hence, the Gumbel distribution or the Fréchet distribution is likely to show the impact of RTN V_{th} variations. Figure 2.10 demonstrates the comparison of tails between the Gumbel distribution and the Fréchet distribution. The shape, scaling and location parameters are 1, 1, and –0.5, respectively. The tailed component of the Fréchet distribution is definitely heavier than that of Gumbel distribution, and the Fréchet distribution is suitable to show higher risk. A regression analysis determines whether the Gumbel distribution or the Fréchet distribution is more likely to show RTN V_{th} variations. Equations (2.1) and (2.2) are deformed as described below

Linearity of the Gumbel plot

$$-\ln\left(-\ln\left(F(x)\right)\right) = \frac{x-\mu_1}{\sigma} \quad (2.7)$$

Linearity of the Fréchet plot

$$-\ln\left(-\ln\left(F(x)\right)\right) = k_1 \ln(x-\mu_1) - k_1 \ln\sigma \quad (2.8)$$

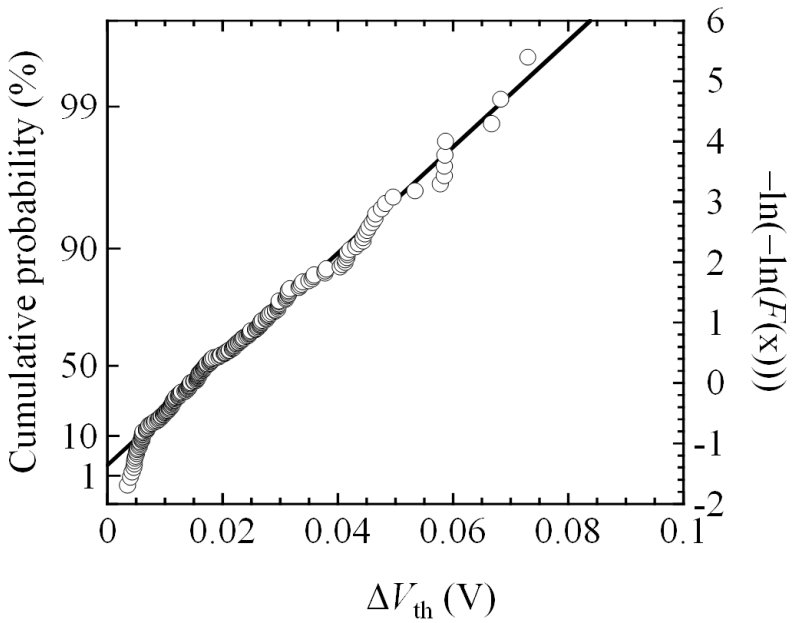


Fig. 2.11. RTN V_{th} variations on Gumbel plot.

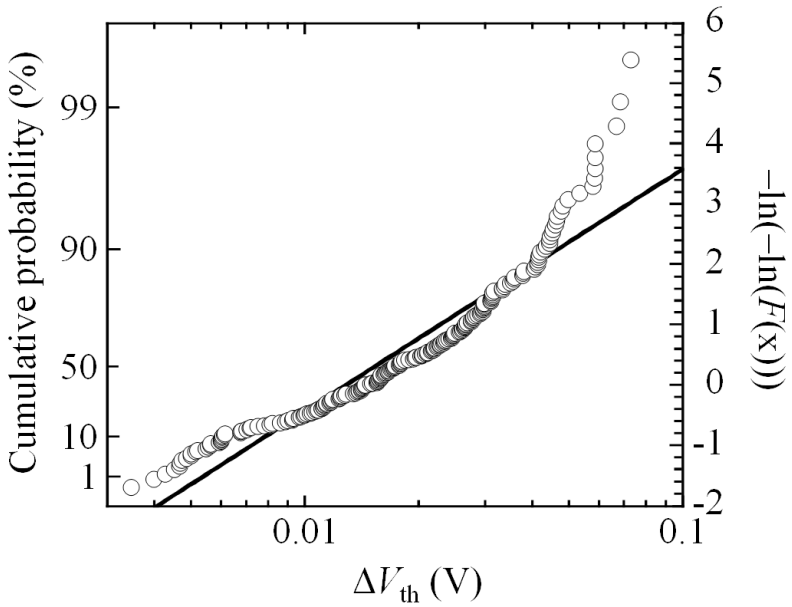


Fig. 2.12. RTN V_{th} variations on Fréchet plot.

Equation (2.7) shows a linearity of the Gumbel distribution. And equation (2.8) shows a linearity of the Fréchet distribution. These equations are used to compute the regression analysis. Figure 2.11 shows RTN V_{th} variations on the Gumbel plot. On the other hand, figure 2.12 shows RTN V_{th} variations on the Fréchet plot. A multiple R squared of the Gumbel plot is 0.98, and on

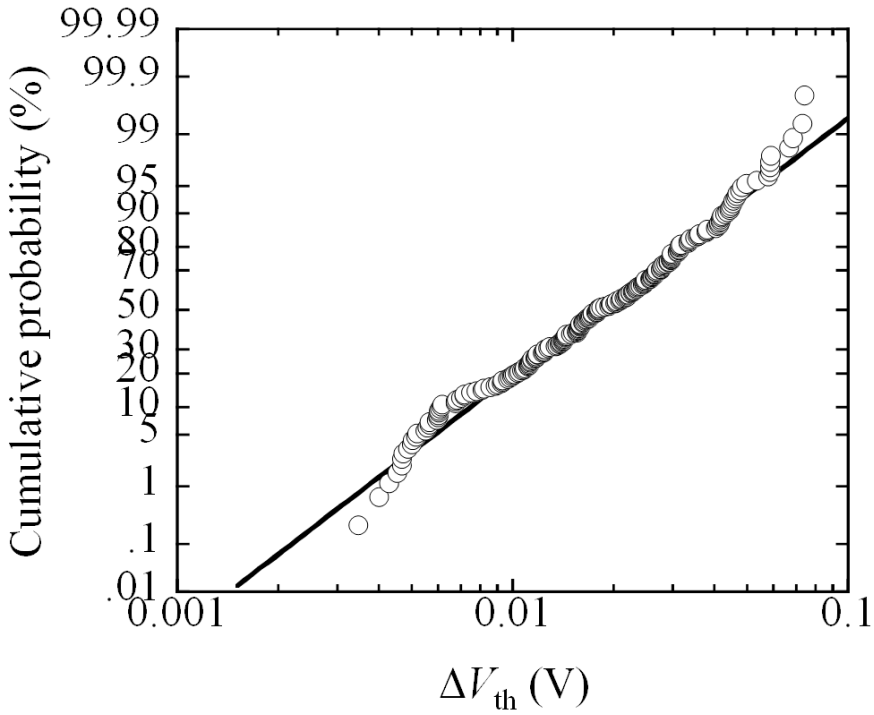


Fig. 2.13. RTN V_{th} variations on log normal plot.

the other hand, that of the Fréchet plot is 0.90; that is, the Gumbel plot more fit than the Fréchet plot. Furthermore, the Fréchet plot above 90% cumulative probability is out of the regression line. Consequently, the Fréchet plot over-estimates the impact of RTN V_{th} variations, and the Gumbel plot is more suitable to predict and judge RTN V_{th} variations.

It was reported that not only the Gumbel distribution but also log-normal distribution matches RTN V_{th} variations as shown in Fig. 2.13 [2, 4, 11]. The reason is clear. The cumulative probability distributions of the extreme value theory have properties which are called the domain of attraction. The domain of attraction is the specific distribution which approach asymptotically to the extreme value distribution. The log-normal distribution is one of the well-known domains of attraction of the Gumbel distribution and has moderately heavy tail to show RTN V_{th} variations.

2.5 Conclusions

The extreme value theory is useful for showing RTN V_{th} variations. Especially, the Gumbel best meets to show the impact of RTN V_{th} variations. Moreover, the log-normal distribution also matches the tail components of RTN V_{th} variations because the log-normal distribution is the domain of attraction of the Gumbel distribution. The Gumbel distribution helps us estimate the impact of RTN V_{th} variations above 99.9999 on SRAM operation. From the dependence of the

slop and the intercept of the Gumbel distribution, we can identify effective processes for improvement of RTN V_{th} variations.

After this chapter, the Gumbel distribution and the log-normal distribution are used to discuss the impact of RTN on scaling, and moreover, the effect of processes on RTN V_{th} variations.

2.6 References

[1] T.Tsunomura, A.Nishida, F.Yano, A.T.Putra, K.Takeuchi, S.Inaba, S.Kamohara1, K.Terada, T.Mama, T.Hiramoto, T.Mogami, "Analysis of Extra VT Variability Sources in NMOS Using Takeuchi Plot, " in *VLSI Symp. Tech. Dig.*, pp. 110–111, 2009.

[2] H. Kurata, K. Otsuga, A. Kotabe, S. Kajiyama, T. Osabe, Y. Sasago, S. Narumi, K. Tokami, S. Kamohara, and O. Tsuchiya, "The impact of random telegraph signals on the scaling of multilevel Flash memories," in *VLSI Circuit Dig.*, pp. 112–113, 2006.

[3] D. J. Frank, "Random Telegraph Noise – Measurement, Analysis, and Consequences," in *Tutorial IRPS*, 2012.

[4] K. Sonoda, K. Ishikawa, T. Eimori, and O. Tsuchiya, "Discrete Dopant Effects on Statistical Variation of Random Telegraph Signal Magnitude, " *IEEE Trans. Electron Devices*, vol. 54, no. 8, pp. 1918–1925, Aug, 2007.

[5] A. Ghetti, C. M. Compagnoni, F. Biancardi, A. L. Lacaita, S. Beltrami, L. Chiavarone, A. S. Spinelli, and A. Visconti, "Scaling trends for random telegraph noise in deca-nanometer Flash memories," in *IEDM Tech. Dig.*, pp. 1–4, 2008.

[6] A. Ghetti, C. M. Compagnoni, A. S. Spinelli, and A. Visconti, "Comprehensive Analysis of Random Telegraph Noise Instability and Its Scaling in Deca–Nanometer Flash Memories," *IEEE Trans. Electron Devices*, vol. 56, no. 8, pp. 1746–1752, Aug, 2009.

[7] J. H. Stathis, "Percolation models for gate oxide breakdown," *J. Appl. Phys.*, vol. 86, pp. 5757–5766, 1999.

[8] S. Yue, "The Gumbel logistic model for representing a multivariate storm event," *Adv. Water Resour*. vol. 24, pp. 179–185, Nov., 2001.

[9] E. J. Gumbel, "STATISTICAL FORECAST OF DROUGHTS," *International Association of Scientific Hydrology. Bulletin*, vol. 8, no. 1, pp.5–23, 1963.

[10] R. Cont, "Empirical properties of asset returns: stylized facts and statistical issues," *QUANTITATIVE FINANCE*, vol. 1 pp.223–236, 2001.

[11] K. Takeuchi, T. Nagumo, S. Yoko gawa, K. Imai and Y. Hayashi, "Single-Charge-Based Modeling of Transistor Characteristics Fluctuations Based on Statistical Measurement of RTN Amplitude," in *VLSI Symp. Tech. Dig.*, pp. 54–55, 2009.

3. Increasing threshold voltage variation due to RTN in MOSFETs

3.1 Introduction

According to the 2009 and 2011 international technology roadmap for semiconductors (ITRS), maintaining adequate noise margin in SRAM is one of the difficult challenges for the 22 nm technology generation [1, 2]. In this context, most work has focused on RDF in very small MOSFETs and its impact on SRAM margin, and various countermeasures have been proposed [3-6] because V_{th} variations due to RDF is inversely proportional to the square root of gate area [7].

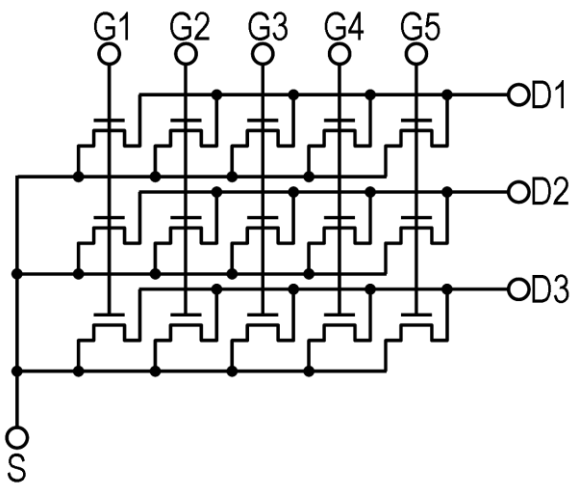
In this chapter, the author, however, focuses on the statistical distribution of RTN and its impact on MOSFETs scaled down to the dimensions of the 22 nm technology node because RTN V_{th} variations is proportional to the gate area as indicated in Eq. (1.6), and moreover, the anomalously large V_{th} variations can be observed. The anomalously large V_{th} variations are rare events, and the enhancement is expected to be caused by complex RTNs and the percolation path due to discreteness of impurities in channel. Therefore, RTN is likely to have a larger impact on scaled MOSFETs than RDF. The author quantifies the impact of RTN V_{th} variation in scaled MOSFETs and shows that RTN is expected to have the impact SRAM design by using the statistical method, which was introduced from the extreme value theory in chapter 2.

3.2 Device fabrication

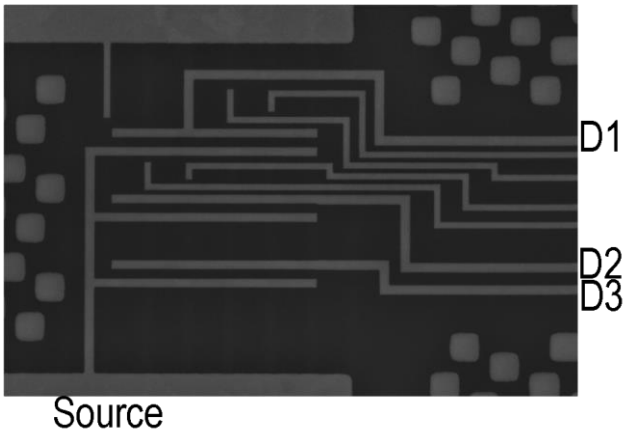
nMOSFETs used in chapters 3 and 4 were fabricated at T. J. Watson Research Center, IBM to research the impact of RTN on scaling in more detail. Mixed e-beam / optical processing has been used to provide devices with L_g from 20 to 90 nm and W_g from 25 to 180 nm. Test array structures were designed for easy measurement of large numbers of devices (27K/die) to enable statistical analysis of RTN V_{th} variation as shown in Fig. 3.1. nMOSFETs were fabricated using standard poly-Si gate and SiON gate dielectric processes and a single shallow source / drain implant as indicated in Fig. 3.2.

3.3 Experimental

This section provides detailed explanation about measurement system for RTN characterization at T. J. Watson Research Center, IBM. Figure 3.3 shows a schematic diagram of measurement system for RTN characterization, and table 3.1 shows this system specifications. When setting up the system for measuring RTN in I_d, there are two important points. First point is a sampling speed and stability against environment. The author selected Agilent 3458A as a digital multi meter (DMM), which provides both speed and accuracy. This DMM has a reading rate of 100,000 readings / sec for maximal test throughput and achieves the highest levels of precision with up to 8.5 digits of measurement resolution and 0.1 ppm transfer accuracy. Second



(a) Circuit schematic of array for statistical analysis of RTN.



(b) SEM top view.

Fig. 3.1. Terminals of array: A matrix has 15 same-size devices.

point is an ability of *IV* converter. Since the DMM reads only voltage, the *IV* converter is needed for reading I_d. Moreover, the *IV* converter should have a high-resolution current-to-voltage conversion. The author selected Stanford Research Systems SR 570 as the *IV* converter. Its features are 1pA / V maximum gain, 1 MHz maximum bandwidth, low noise and low drift. The combination of Agilent 3458A and SR 570 can provide wide frequency range from 1 Hz to 50 kHz and meets required specifications as shown in table 3.1. When measuring RTN in I_d, the DMM applies a drain voltage to MOSFETs. On the other hand, gate voltage is applied by the Keithley semiconductor parametric tester 4200-SCS.

Each chip contains over 27,000 FETs. A probe card with 25 pin is needed for measuring many FETs with a same word line and data line. A switching matrix is needed for automatic switching. Keithley 707-A was selected as the switching matrix. All instruments, including a probe station

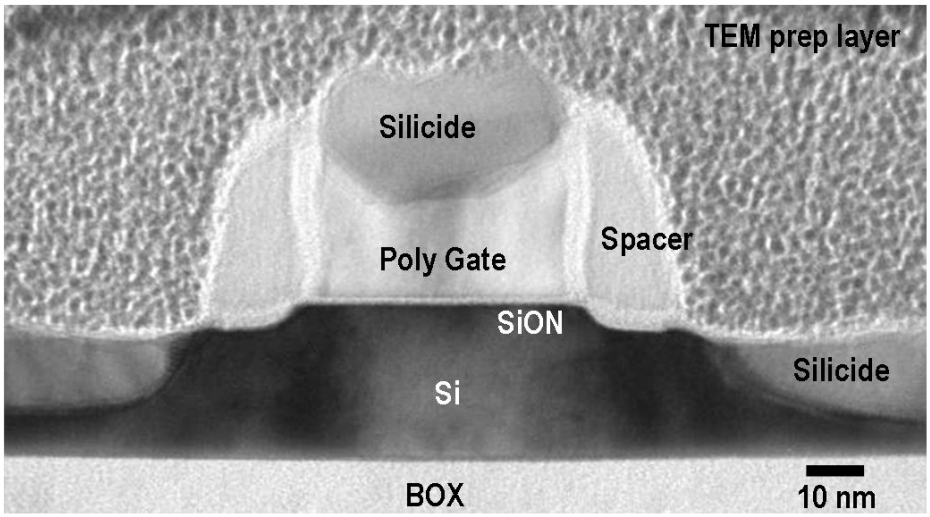


Fig. 3.2. TEM image of device after silicide: The device with L_g = 45 nm was fabricated by basic techniques such as poly gate and SiON gate dielectric.

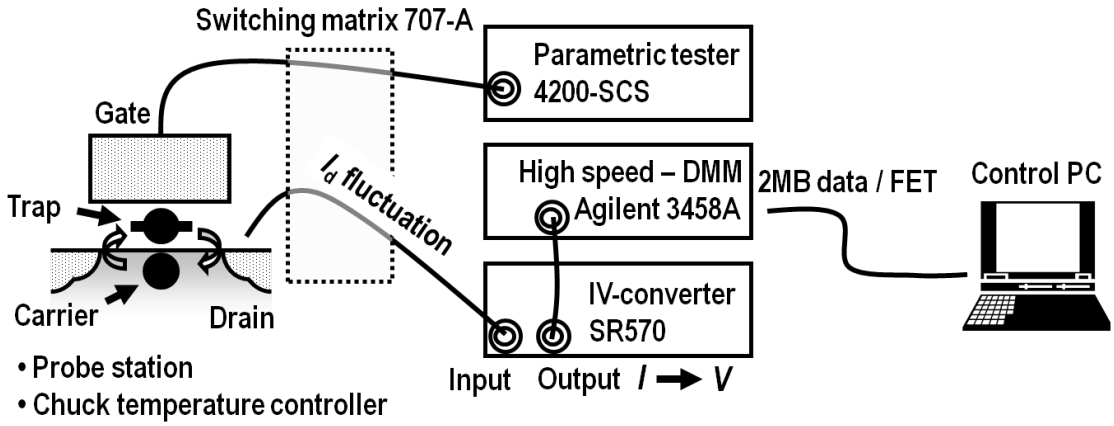


Fig. 3.3. System for high-speed measurement of RTN.

Table 3.1. System specifications.

Items	Required specs	Performances
Measuring speed	20sec / MOSFET	About 20sec / MOSFET
Frequency range	1 ~ 50 kHz	1 ~ 50 kHz
System noise level	Low noise level < 1 mV	< 1 mV

and a check temperature controller, are under computer control for automatic measurement as show in Fig. 3.3. As the result of making the system, it achieves high-speed measurement, around 20 sec / device, and can provide both high speed and resolution. The author checked a

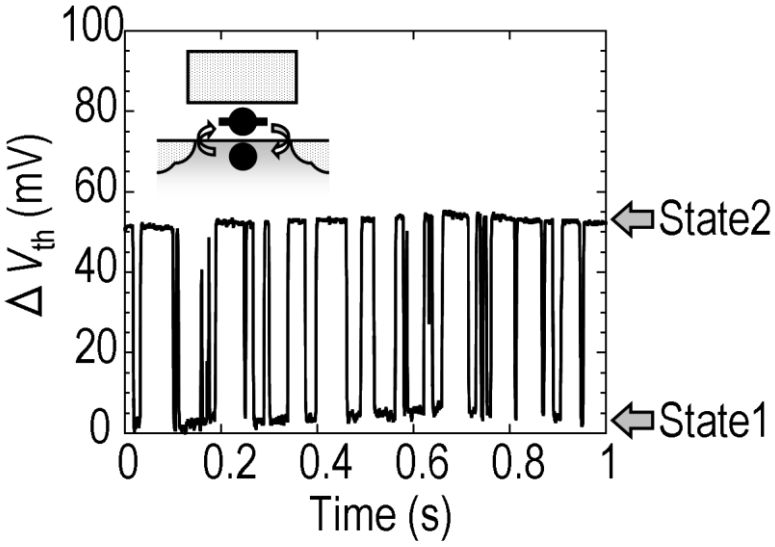


Fig. 3.4. Single RTN in the small nMOSFET with $L_g / W_g = 30/45$ nm: A binary fluctuation is caused by trapping and detrapping of carrier at a single trap in near-interface gate oxide.

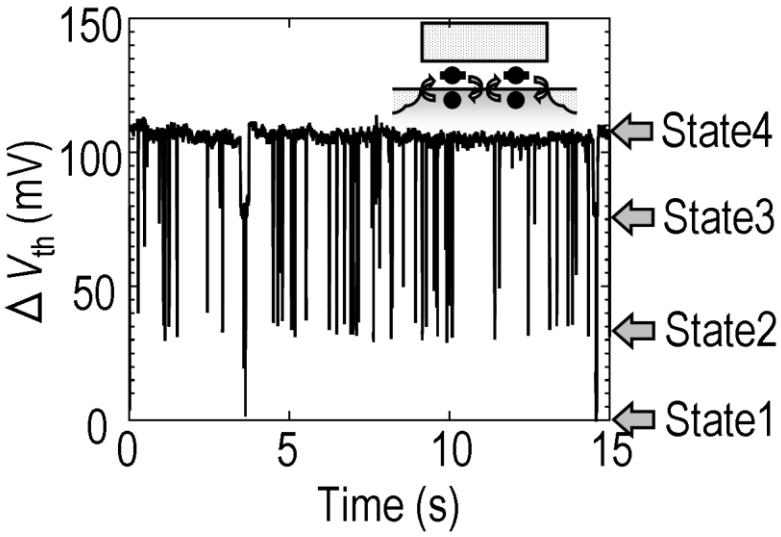


Fig. 3.5. Complex RTN in the small nMOSFET with $L_g / W_g = 30/45$ nm: 4-state fluctuation is caused by 2 traps in near-interface gate oxide.

noise floor level which comes from this system. As a result, it is less than 1 mV. The value is enough to measure RTN because V_{th} variability of more than 1 mV has the impact on the reliability of MOSFETs. The system estimates a suitable sensitivity of SR 570 and a transconductance of measurement point before measuring RTN in I_d. ΔI_d is defined as the peak-to-peak value and is transformed into ΔV_{th} by G_m.

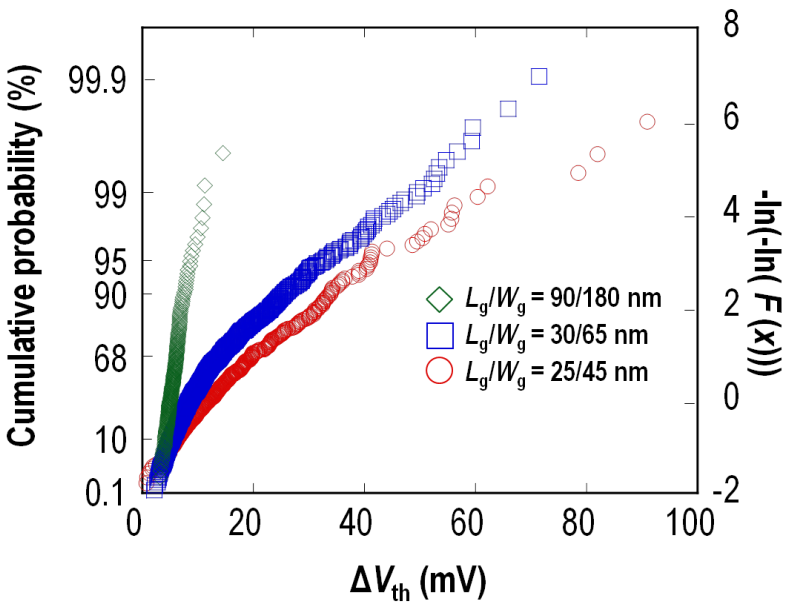


Fig. 3.6. RTN V_{th} variations on Gumbel distribution: The device sizes (L_g/W_g) are 25 / 45, 30 / 65, and 90 / 180 nm, respectively.

3.4 Statistical analysis on V_{th} variation in 22 nm MOSFETs

Figure 3.4 shows a typical binary fluctuation due to trapping and detrapping of a carrier at a single trap in near-interface gate oxide. Because of the small device with $L_g / W_g = 30/45$ nm, the amplitude of this single RTN reaches around 50 mV. According to equation (1.6), a predictable V_{th} variation is less than 10 mV. However, the actual V_{th} variation is more than five time the predictable value. The likeliest model of the enhancement is the percolation path model; that is, RTN V_{th} variation could be enhanced because the effective channel becomes narrower due to the discreteness of impurities.

Moreover, even though the average number of traps is less than one per 20 nm MOSFET, complex RTN with multiple simultaneous trapping events was also observed because traps are discretely distributed as shown in Fig. 3.5. Figures 3.4 and 3.5 are strong evidences that both single RTN and complex RTN have an impact on scaling.

Figure 3.6 shows RTN V_{th} variations of approximately 1,000 MOSFETs per device size on the Gumbel distribution. The devices with L_g / W_g = 25 / 45 nm (25 nm MOSFET), 30 / 65 nm (30 nm MOSFET), and 90 / 180 nm (90 nm MOSFET) are used for 22, 32, 90 nm technology node, respectively. The 90 nm MOSFET has relatively small V_{th} variations, around 10 mV at 95% point (95% point is equivalent to 2σ of Gaussian distribution). This level seems to have little effect on SRAM operation. In contrast, both the 30 nm and 25 nm MOSFETs have large V_{th} variations as a heavy tail above 68% point (68% point is equivalent to 1σ of Gaussian

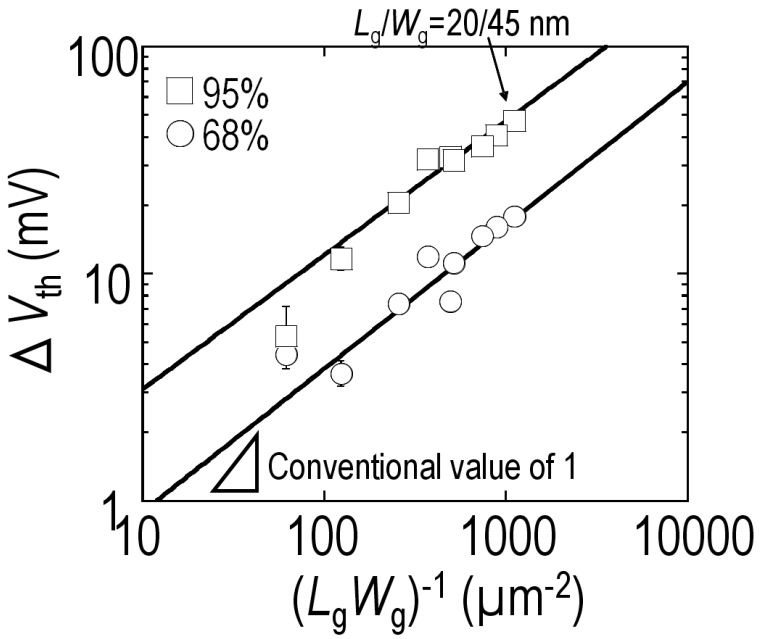


Fig. 3.7. Device-size dependences of ΔV_{th}: The power law exponent is less than 1.

distribution). In fact, the 30 nm MOSFET's values reach around 50 mV at 95% point. In addition, the 25 nm MOSFET shows the heaviest tail of the three devices and its V_{th} variations reach more than 70 mV.

Figure 3.7 indicates the device-size dependence of RTN V_{th} variation at 68% and 95% points. Regarding the power law exponent, RTN V_{th} variation is thought to be inversely proportional to device size and the conventional power law exponent is 1. The power law exponent of 0.6, however, is less than the conventional value of 1 as shown in Fig. 3.7. This exponent is important because it determines the scaling of RTN. The reason for the small power law exponent of 0.6 compared to 1 is incompletely understood. However, according to the percolation path model, it was reported that the influence of gate-width scaling on RTN V_{th} variation is relatively larger than that of gate-length scaling [8]. In fact, my data shows the same tendency as explained below. I may think that the difference in sensitivities for W_g, L_g, and so on causes my result.

The power law exponent of RTN V_{th} variation is larger than that of RDF. The power law exponent of RDF is 0.5 as shown in the Pelgrom plot [9, 10]. Therefore, RTN has the potential to show larger impact on V_{th} margin in logic and SRAM than RDF. In the next chapter, the author demonstrates a comparison between RTN and RDF.

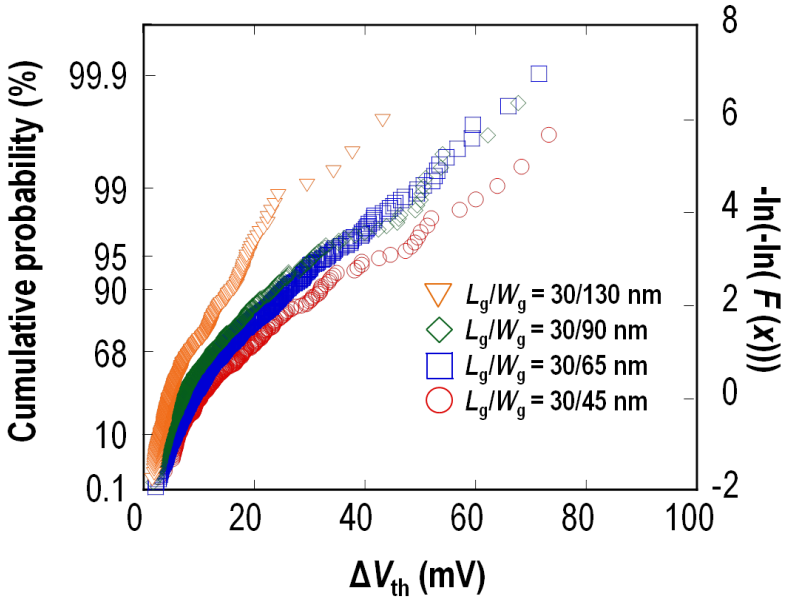


Fig. 3.8. RTN V_{th} variations on Gumbel distribution for constant $L_g = 30$ nm.

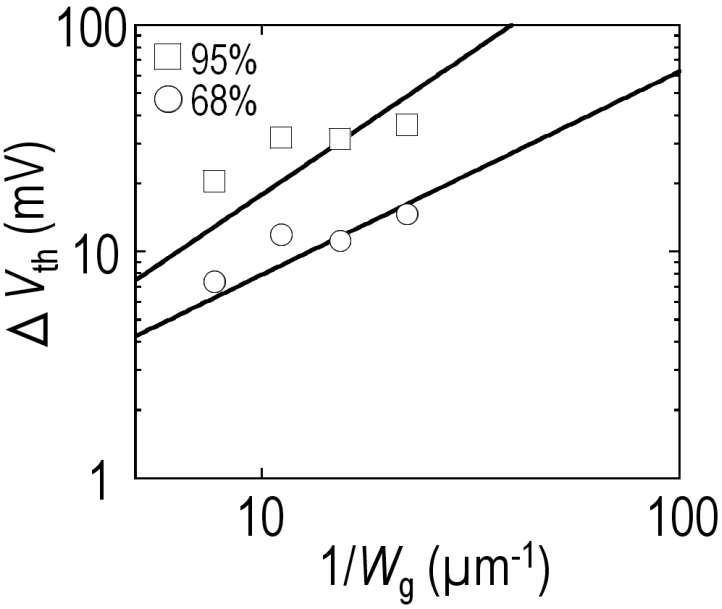


Fig. 3.9. W_g dependences of ΔV_{th}: The percolation path depends on W_g rather than L_g.

3.5 Gate-width and gate-length dependences of RTN V_{th} variation

As mentioned above, the two mechanisms are proposed to explain the extremely large V_{th} variations, which are rare even; that is, the simultaneous overlapped multiple RTNs, and the

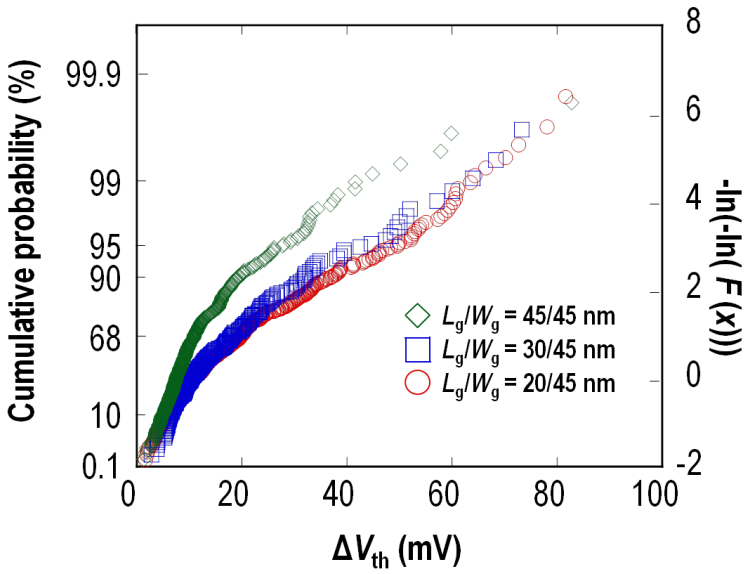


Fig. 3.10. RTN V_{th} variations on Gumbel distribution for constant $W_g = 45$ nm.

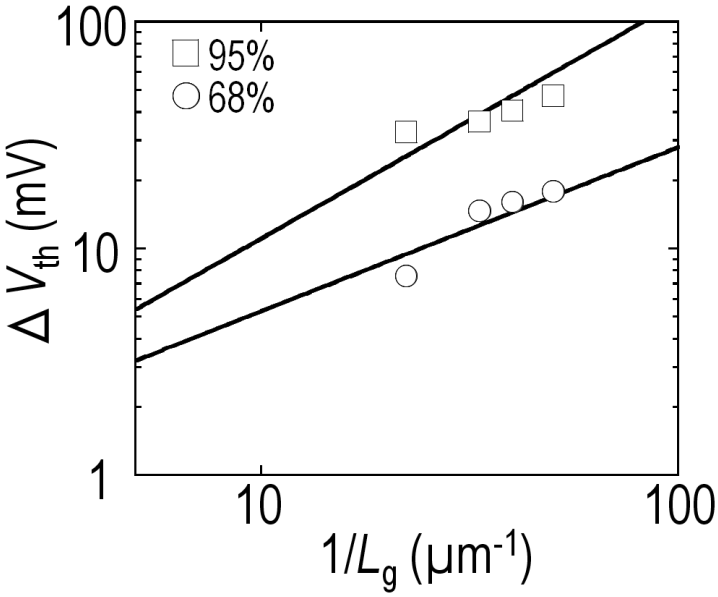


Fig. 3.11. L_g dependences of ΔV_{th}: L_g dependences are smaller than W_g dependences.

interaction of traps with percolation paths. It was reported for the first time by A. Asenov *et al.* that one of the enhancement phenomena is the effect of discreteness of channel dopants [11]. Moreover, it was simulated by A. Ghetti *et al.* in detail that source-drain path is affected by

random dopant, and RTN variation depends on trap location in gate dielectric [8].

Complex RTN has been already confirmed as shown in Fig. 3.5. On the other hand, the gate-width and gate-length dependences are essential for confirming the percolation path model. As noted before, it was reported that the effect of gate-width scaling on RTN V_{th} variation is relatively larger than that of gate-length scaling [8]. However, this result came from the simulation and there is not strong evidence such as measurement data to support the percolation path theory. This work can support this theory by measurement data. First, figures 3.8 and 3.9 show the gate-width dependence of RTN V_{th} variation. W_gs are from 45 to 130 nm and L_g is a constant of 30 nm. Then, figures 3.10 and 3.11 show the gate-length dependence of RTN V_{th} variation. L_gs are from 20 to 45 nm and W_g is a constant of 45 nm. It is difficult to quantitatively discuss these dependences. However, data is likely to allow the comparison of the dependent difference. In fact, the gate-width dependence relatively large compared to the gate-length dependence as shown in Figs. 3.9 and 3.11. This result is consistent with the percolation path model. Thus, this work successfully shows that both complex RTN and percolation paths enhance V_{th} variation from experimental data.

3.6 Conclusions

The author demonstrates that RTN V_{th} variations increase with scaling from large quantities of data. The RTN V_{th} variation is inversely proportional to device seize. This trend shows the power law. The enhanced mechanisms of RTN V_{th} variation are the simultaneously overlapped RTNs and the percolation of drain current due to the discreteness of impurities in channel.

3.7 References

[1] ITRS 2009 web site: http://www.itrs.net/Links/2009ITRS/Home2009.htm

[2] ITRS 2011 web site: http://www.itrs.net/Links/2011ITRS/Home2011.htm

[3] D. Burnett, K. Erington, C. Subramanian, and K. Baker, “Implications of fundamental threshold voltage variations for high-density SRAM and logic circuits,” in *VLSI Symp. Tech. Dig.*, pp. 15–16, 1994.

[4] B. Cheng, S. Roy, and A. Asenov, “The impact of random doping effects on CMOS SRAM cell,” *Proc. ESSCIRC*, pp. 219–222, 2004.

[5] A. Asenov, S. Roy, R. A. Brown, G. Roy, C. Alexander, C. Riddet, C. Millar, B. Cheng, A. Martinez, N. Seoane, D. Reid, M. F. Bukhori, X. Wang, and U. Kovac, “Advanced simulation of statistical variability and reliability in nano CMOS transistors,” in *IEDM Tech. Dig.*, pp. 15–17, 2008.

[6] T.Tsunomura, A.Nishida, F.Yano, A.T.Putra, K.Takeuchi, S.Inaba, S.Kamohara1, K.Terada, T.Mama, T.Hiramoto, T.Mogami, “Analysis of Extra VT Variability Sources in NMOS Using Takeuchi Plot, ” in *VLSI Symp. Tech. Dig.*, pp. 110–111, 2009.

[7] M. J.M. Pelgrom, H. P. Tuinhout, and M. Vertregt, “Transistor matching in analog CMOS applications,” in *IEDM Tech. Dig*., pp. 915–918, 1998.

[8] A. Ghetti, C. Monzio Compagnoni, F. Biancard, A. L. Lacaita, S. Beltrami, L. Chiavarone, A.S. Spinelli, and A. Visconti, “Scaling trends for random telegraph noise in deca-nanometer Flash memories,” in *IEDM Tech. Dig*., pp. 835–838., 2008.

[9] M. J. M. Pelgrom, A. C. J. Duinmaijer, and A. P. G. Welbers, “Matching properties of MOS transistors,” *IEEE J. Solid-State Circuits*, vol. 24, no. 5, pp. 1433–1440, Oct., 1989.

[10] M. J.M. Pelgrom, H. P. Tuinhout, and M. Vertregt, “Transistor matching in analog CMOS applications,” in *IEDM Tech. Dig*., pp. 915–918, 1998.

[11] A. Asenov, A. R. Brown, J. H. Davies, S. Kaya, and G. Slavcheva, “Simulation of Intrinsic Parameter Fluctuations in Decananometer and Nanometer-scale MOSFETs,” *IEEE Trans. Electron Devices*, vol. 50, no. 9, pp. 1837–1852, Sept., 2003.

4. Impact of RTN on high-κ / metal-gate stacks and future device scaling

4.1 Introduction

Variation due to RTN is becoming a new threat to the future devices because V_{th} variation of RTN rapidly rises with scaling, just as variation due to random dopant fluctuation (RDF) as discussed in chapter 3 [1]. The distinctive features of RTN are the size dependence and statistical distribution of V_{th} variation. The size dependence of RTN is relatively stronger than RDF. Moreover RTN V_{th} variation shows the Gumbel distribution and log-normal distribution with extremely large V_{th} variations, which are rare event, whereas RDF has a Gaussian distribution. Therefore RTN ΔV_{th} could exceed that of RDF at high cumulative probability, in especially advanced devices. RTN is one of the newly envisioned challenges to SRAM stability at the next generations such as 15 nm and beyond.

High-κ / metal-gate (HK / MG) stacks will be incorporated in these generations as standard technique because of their more optimal power and performance compared to SiON / Poly-Si gate stacks [2, 3]. Study on the reliability of HK / MG stacks has made progress in recent years. The HK / MG stacks have a great advantage not only for gate leakage current, but also for variability such RDF because the HK / MG stacks can provide thinner EOT than the conventional SiON / Poly-Si gate stacks [4]. This effect is expected to reduce RTN V_{th} variations as well as RDF. In contrast, HK / MG stacks have a serious problem associated with an interfacial layer (IL) between the high-κ dielectric and the substrate. IL is expected to prevent or at least minimize an interfacial reaction between the high-κ oxide and the underlying Si [5]. However, since IL often consists of SiON film, the many researchers express concern about its potential to reduce the reliability, especially its potential to degrade BTI characteristics [6]. On the other hand, a relationship between RTN and HK / MG stacks has been unknown because the thin EOT of HK / MG stacks is likely to contribute to suppressing the impact of RTN as well as RDF, and in contrast, IL including SiON film has a potential to degrade RTN V_{th} variations as well as the BTI characteristics.

The author evaluates the RTN impact on HK / MG MOSFETs and future device scaling from a statistical viewpoint and demonstrates an effect of high temperature forming anneal (HTFGA) on an improvement of HK / MG stacks.

4.2 Device fabrication

In this work, mixed e-beam / optical processing has been used to fabricate small devices, nMOSFETs, with gate length from 20 to 90 nm and width from 25 to 180 nm. And these small devices incorporate HfO_2-based HK / MG stacks. For a comparison, devices with SiO_2 and SiON / Poly-Si gate stacks were also fabricated. The HK MOSFETs have smaller T_{inv} and larger transconductance than the pure-SiO_2 and SiON MOSFETs, as shown in Table 4-1. To evaluate

Table 4.1 Features of nMOSFETs.

Gate dielectric	SiO_2	High-κ
Gate	Poly Si	Metal Gate
Split N conditions	With or W/O	
Split HTFGA		With or W/O
T_{inv} (nm)	1.83(SiON)	1.49
G_m	Small	Large

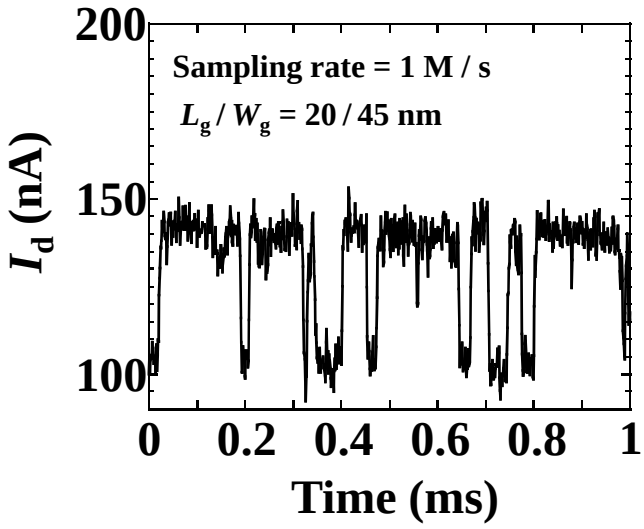


Figure 4.1. Typical RTN dependence on time.

the effect of hydrogen passivation of interface traps [7], HTFGA was performed on some HK MOSFETs. The annealing temperature of HTFGA is 475 °C. It was reported by K. Onishi et al., that HTFGA can reduce the interface trap density of HK MOSFET [8, 9]. HTFGA is, therefore, expected to passivate RTN traps and suppress RTN ΔV_{th}. On the other hand, since RTN shows the extreme value distribution such as the Gumbel distribution, statistical analysis of many MOSFETs is vital. Test array structures were designed for easy measurement of large numbers of devices (27000 / die) to enable statistical analysis of RTN ΔV_{th} variation the same way as SiON MOSFETs in chapter 3.

4.3 Measurement and analysis methods

A fast measurement unit (Agilent 1530A) which enables a wide band width of up to 1 M/s was used to measure the RTN signals. Figure 4.1 shows the typical RTN dependence on time obtained by high sampling rate 1 M/s. 10^5 sampling points were used for the measurement. A sampling interval was 1 μs. Drain current fluctuation was measured under 50 mV of drain voltage. This RTN behavior has two obvious states due to trapping and detrapping of a carrier at a single

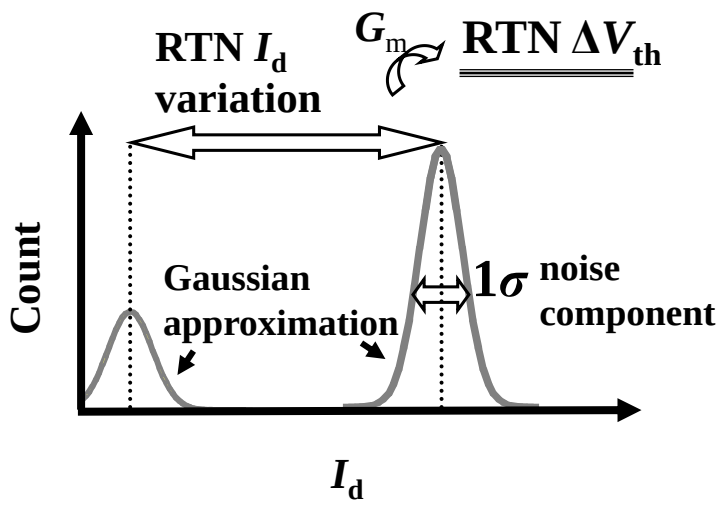


Figure 4.2. Separation between RTN and other noise components.

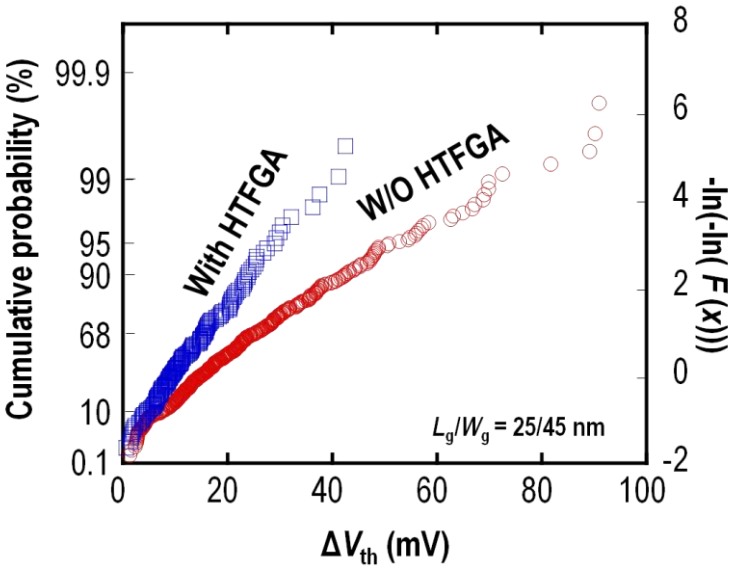
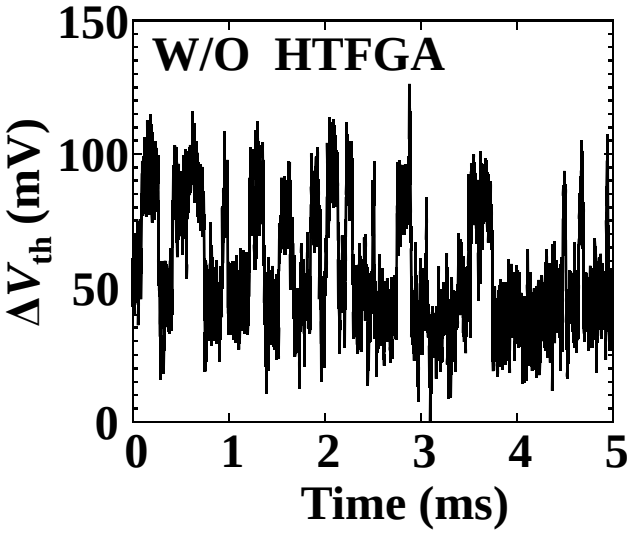
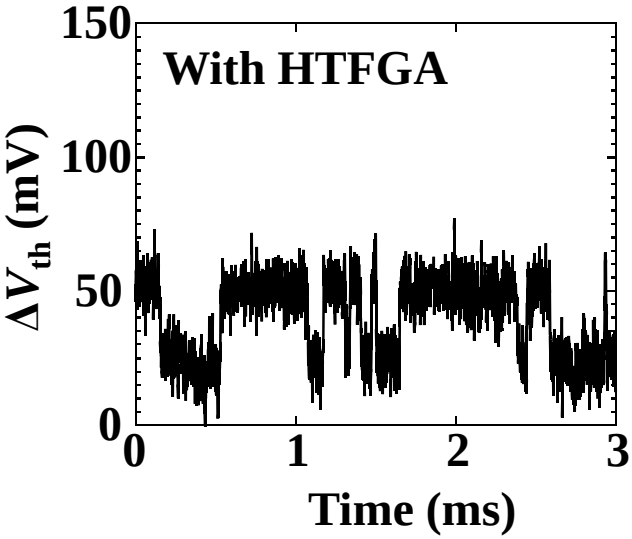


Figure 4.3. Comparison of RTN in 25 nm MOSFETs with or without HTFGA.

trap in gate dielectric. Figure 4.2 illustrates the histogram separation method. RTN I_d variation is defined as the peak-to-peak value in this figure. Other noise components are removed from the RTN component by this extraction. RTN variation is transformed from ΔI_d to the input-referred RTN voltage noise (ΔV_{th}) using G_m, in addition to the separation.



(a) With HTFGA.



(b) Without HTFGA.

Figure 4.4. RTNs at 95% level in 25 nm MOSFET.

.

4.4 Inhibitory effect of HK / MG techniques on RTN

As mention above, HTFGA is expected to passivate RTN traps and suppress RTN ΔV_{th}. Figure 4.3 demonstrates the effect of HTFGA on the distributions of RTN ΔV_{th} in 25 nm MOSFETs. As expected, RTN is well suppressed by HTFGA. HTFGA can reduce ΔV_{th} at the cumulative probability of 95% from 40 mV to 20 mV. For instance, Figure 4.4 shows the samples with and without HTFGA at the 95% level. The RTN magnitude at the 95% level for

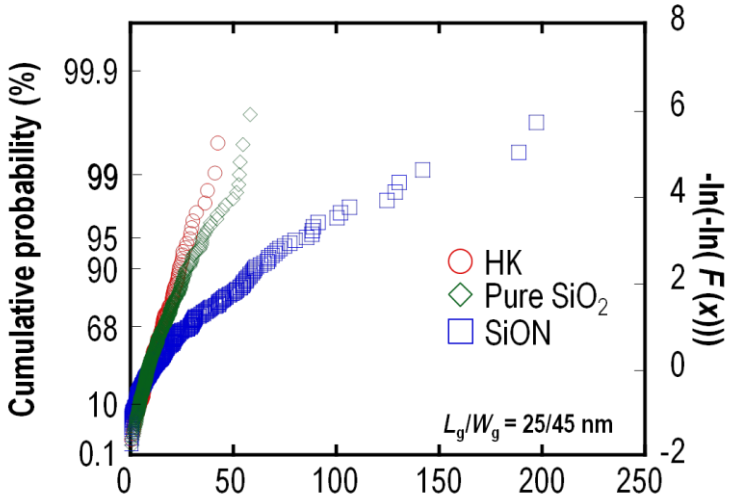


Figure 4.5. Comparison of RTN between HK with HTFGA, pure-SiO_2, and SiON FETs.

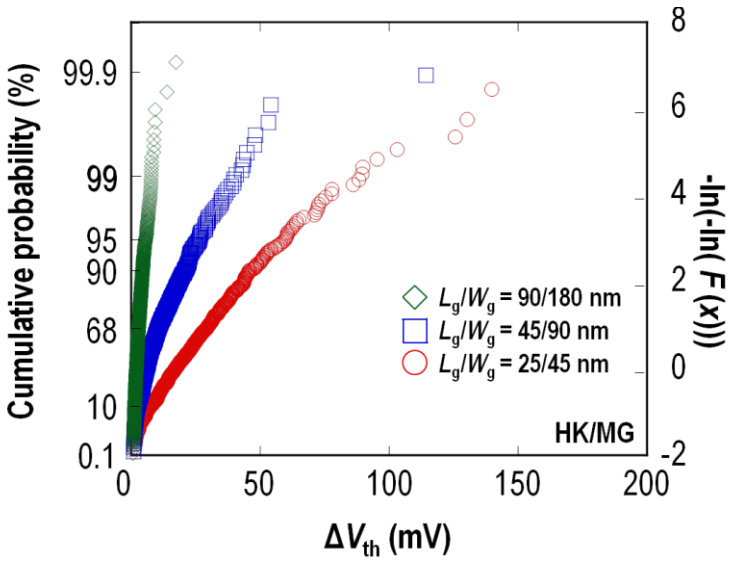
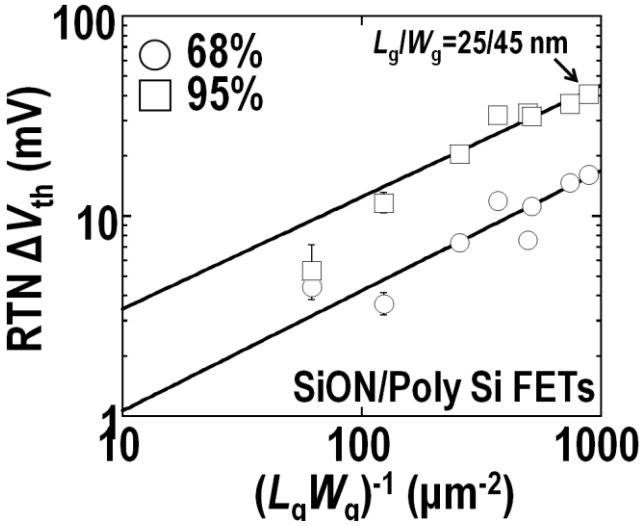
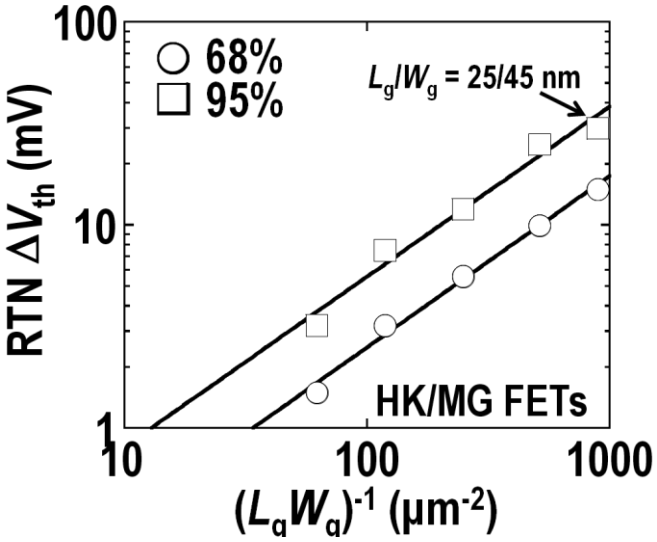


Figure 4.6. Device-size dependence of RTN ΔV_{th} distributions.

these two cases differs by ~2 times because hydrogen is capable of passivating traps. As I will discuss again in the after chapters, the RTN V_{th} variations is affected significantly by the processes which can reduce the interface trap density. Consequently, I am sure that the RTN trap is located very near the interface between gate dielectric and substrate. I will prove my hypotheses in chapter 8. By the way, R .J. Carter et al. reported the reason that the HTFGA is useful for passivating interface traps in the case of HfO_2-based HK dielectric [10]. They revealed



(a) SiON/poly-Si MOSFETs.



(b) HK/MG MOSFETs.

Figure 4.7. Device-size dependences of RTN ΔV_{th}.

that HTFGA is not necessary for the hydrogen to diffuse through the conventional gate dielectric, SiON, but through HfO_2-based material. The best annealing temperature is 520 ºC. The interface trap density (D_{it}) controlled by HTFGA at 520 ºC is one tenth that uncontrolled by FGA at 420 ºC. Thus, I succeed in controlling the RTN traps near interface by HTGA at 475 °C in the same way as D_{it} and believe that HTFGA will be able to be optimized much further.

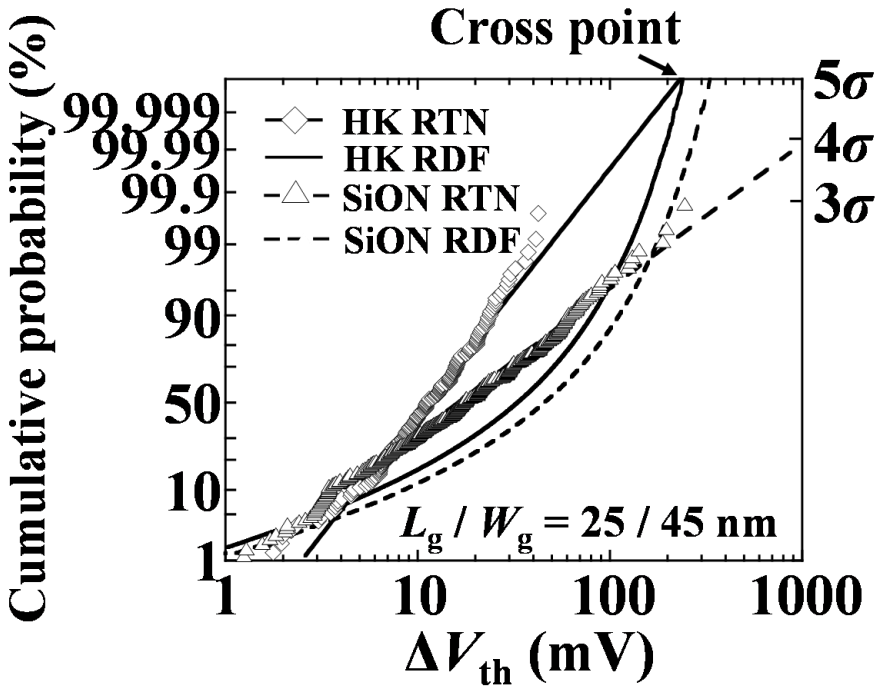


Figure 4.8. Comparison between RTN and RDF impacts on log-normal distribution.
* RDF data are based on the reference 11.

Figure 4.5 compares RTN ΔV_{th} between HTFGA HK, pure-SiO_2 and SiON MOSFETs. RTN ΔV_{th} of HK and pure-SiO_2 MOSFETs are quite comparable. On the other hand, both the median value and the variation of HK MOSFETs are obviously smaller than those of SiON MOSFETs. The main reason is that trap density in our optimized HK MOSFETs becomes lower because of HTFGA. A secondary reason is that the T_{inv} of HK MOSFETs is 25% smaller than that of SiON MOSFETs. Optimizing HK/MG stacks is useful for not only the RDF variation but also the RTN variation in the advanced MOSFETs.

Figure 4.6 indicates the cumulative distribution of RTN ΔV_{th} of HK MOSFETs with strong device-size dependence. All measured devices were fabricated on same wafer, so process variation is suppressed as much as possible to evaluate only the impact of scaling on RTN. As mentioned above, the RTN V_{th} variations at high cumulative probability are of greatest importance. The MOSFET with L_g and W_g = 90/180 nm has small ΔV_{th} of approximately 4 mV at 95% point. ΔV_{th} of the MOSFET with 45/90 nm is more than 20 mV. And the smallest device with 25/45 nm has approximately 40 mV of ΔV_{th}. Thus, the variation increases with device scaling.

The comparison of the device-size dependence of ΔV_{th} between SiON and HK MOSFETs is seen in Fig. 4.7. RTN ΔV_{th} steeply increases with device size in both HK MOSFETs and SiON MOSFETs. However, the RTN V_{th} variations of HK MOSFETs are smaller than these of SiON

MOSFETs anywhere in the generations because of the effect of HTFGA and the thinner EOT of HK MOSFET than that of SiON MOSFET.

The statistical distribution of RTN V_{th} variation shows the Gumbel or the log-normal based on the extreme value theory, while the statistical distribution of RDF has been shown to be Gaussian out to at least 5σ [11]. This work compares RTN with RDF in HK MOSFET by arranging data of reference 11 as demonstrated in Fig. 4.8. The data are plotted on the log-normal distribution. 22 nm generation RTN V_{th} variations exceed RDF V_{th} variations at the 3σ level in the SiON MOSFET, while the cross point can be extended to more than 5σ level in the case of the HK MOSFET. The author, therefore, concludes that the influence of RTN is less than that of RDF in the 22 nm generation. Considering the size dependence of RTN as shown in Fig. 4.8, however, RTN may pose a difficult challenge for the 15 nm generation.

4.5 Conclusions

The RTN V_{th} variation in HK MOSFET can be suppressed by suitable annealing, such as HTFGA, and by thin T_{inv}. As a consequence, properly annealed HK MOSFETs can have smaller RTN variation than SiON MOSFETs. The RTN ΔV_{th} dependence on scaling was demonstrated using both SiON MOSFETs and HK MOSFETs. RTN impact may, however, become severe in 15 nm generation and beyond because even though HK appears to offer lower RTN, the dimensions will be so small that RTN will become large.

4.6 References

[1] K. Takeuchi, T. Nagumo, S. Yokogawa, K. Imai, and Y. Hayashi, “Single-Charge-Based Modeling of Transistor Characteristics Fluctuations Based on Statistical Measurement of RTN Amplitude,” in *VLSI Tech. Dig.*, pp.54–55, 2009.

[2] C. H. Diaz et al., “32nm Gate-First High-k/Metal-Gate Technology for High Performance Low Power Applications” in *IEDM Tech. Dig.*, pp. 629–632, 2008.

[3] C.-H. Jan, M. Agostinelli, H. Deshpande, M. A. El-Tanani, W. Hafez, U. Jalan, L. Janbay, M. Kang, H. Lakdawala, J. Lin, Y-L Lu, S. Mudanai, J. Park, A. Rahman, J. Rizk, W.-K. Shin, K. Soumyanath, H. Tashiro, C. Tsai, P. VanDerVoorn, J.-Y. Yeh, and P. Bai, “RF CMOS Technology Scaling in High-k/Metal Gate Era for RF SoC (System-on-Chip) Applications,” in *IEDM Tech. Dig.*, pp. 604–607, 2010.

[4] K. Mistry et al., “A 45nm Logic Technology with High-k+Metal Gate Transistors, Strained Silicon, 9 Cu Interconnect Layers, 193nm Dry Patterning, and 100% Pb-free Packaging,” in *IEDM Tech. Dig.*, pp.247–250, 2007.

[5] G. D.Wilk, R. M.Wallace, and J. M. Anthony, “High-k gate dielectrics: Current status and materials properties considerations,” *J. Appl. Phys.*, vol. 89, pp. 5243–5275, May, 2001.

[6] C. Choi, C.-S. Kang, C. Y. Kang, S. J. Rhee, M. S. Akbar, S. A. Krishnan, M. Zhang, and J. C. Lee, “Positive Bias Temperature Instability Effects of Hf-Based nMOSFETs With Various Nitrogen and Silicon Profiles,” *IEEE Electron Devic Lett.*, vol. 26, no. 1, pp. 32–34, Jan., 2005.

[7] H. Aono, E. Murakami, K. Shiga, F. Fujita, S. Yamamoto, M. Ogasawara, Y. Yamaguchi, K. Yanagisawa, and K. Kubota, “A STUDYOF SRAM NBTI BY OTF MEASUREMENT,” in *IRPS Proc.*, pp. 67–71, 2008.

[8] K. Onishi, C. S. Kang, R. Choi, H.-J. Cho, S. Gopalan, “Effects of High-Temperature Forming Gas Anneal on Hf02 MOSFET Performance,” in *VLSI Tech. Dig.*, pp.22–23., 2002.

[9] K. Onishi, C. S. Kang, R. Choi, H.-J. Cho, S. Gopalan, R. E. Nieh, S. A. Krishnan, and J. C. Lee, “Improvement of Surface Carrier Mobility of HfO2 MOSFETs by High-Temperature Forming Gas Annealing,” *IEEE Trans. Electron Devices*, vol. 50, no. 2, pp. 384–390, Feb., 2003.

[10] R. J. Carter, E. Cartier, A. Keber, L. Pantisano, T. Schram, S. De. Gendt, and M. Heyns, “Passivation and interface state density of SiO2 / HfO2 –based/polycrystalline-Si gate stacks,” *App. Phys. Lett.*, vol. 83, no. 3, pp. 533–535, Jul., 2003.

[11] T. Tsunomura, A. Nishida, F. Yano, A. T. Putra, K. Takeuchi, S. Inaba, S. Kamohara, K. Terada, T. Hiramoto, T. Mogami, “Analyses of 5σ Vth Fluctuation in 65nm-MOSFETs Using Takeuchi Plot,” in *VLSI Tech.*, pp.156–157, 2008.

5. Investigation of RTN using MOSFETs fabricated on Si(100), (110), and (111)

5.1 Introduction

In this chapter, the author has a perspective on the impact of RTN on future devices and focuses how a difference in silicon-surface orientation affects ΔV_{th}. The reason for emphasizing the importance of silicon-surface orientation is that three-dimensional (3D) transistors such as finFETs and tri-gate transistors are just about to take the place of the planar devices of the 22 and 15-nm generations [1−3]. Other silicon-surface orientations in addition to (100) will be used as the interface of channel if the 3D transistors become practical [4]. Furthermore, Si(110)/ <110> is expected to increase channel mobility in pMOSFETs [5]. On the other hand, interface trap density (N_{it}) at the Si / SiO_2 interface also depends on silicon-surface orientation [6]. Moreover, silicon-surface orientation also affects the amount of suboxide, which creates traps in the interfacial transition layer [7]. To maintain progress in silicon technology, it is, therefore, essential to confirm whether or not RTN ΔV_{th} is affected by silicon-surface orientation as well as the gate-stack techniques.

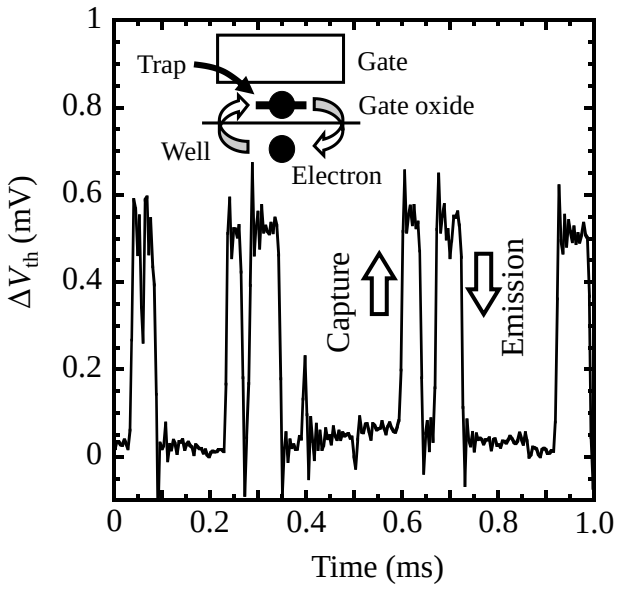
In this chapter, the dependence of RTN V_{th} variation on silicon-surface orientation is demonstrated by statistical analysis. Samples, nMOSFETs, were fabricated on Si(100), (110), and (111) substrates. They were designed for statistical analysis of RTN behaviors. The amounts of suboxides and interface trap densities are shown to depend on silicon-surface orientation. The effect of silicon-surface orientation on RTN ΔV_{th} distributions is thoroughly discussed.

5.2 Samples Preparation and Experimental Methods

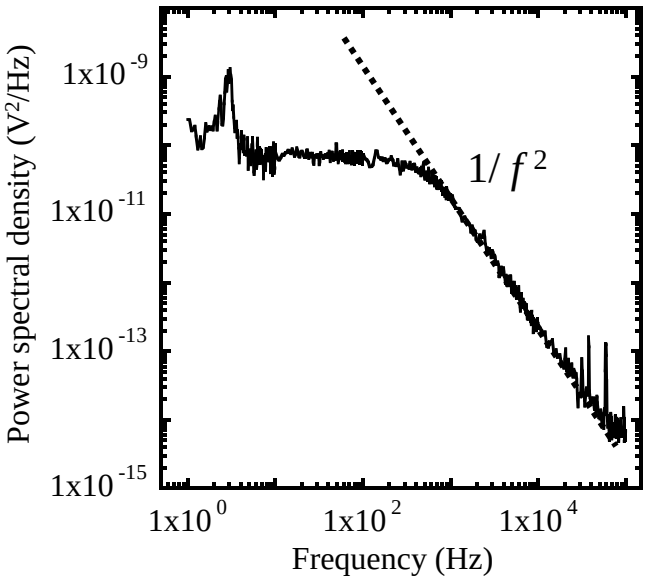
nMOSFETs were used in this chapter to study a relationship between RTN V_{th} variation and silicon-surface orientation. For statistical analysis of ΔV_{th}, many nMOSFETs with the same device size are needed, so more than 200 nMOSFETs were fabricated and used for analyzing uncommon RTN event with large ΔV_{th}. Silicon (100), (110), and (111) substrates were used to compare the effect of silicon-surface orientation on RTN V_{th} variation. L_{eff} and W_{eff} of the nMOSFETs are 0.13 μm and 0.22 μm, respectively. The gate dielectric was pure SiO_2 film with a thickness of about 3 nm. Hydrogen annealing was conducted for all samples.

The amount of suboxide in the gate oxide was analyzed with X-ray photoelectron spectroscopy (XPS). The peak positions of Si^{1+}, Si^{2+}, and Si^{3+} observed at 1.0, 1.7, and 2.6 eV, respectively [7]. The percentage of the total amount of suboxides was estimated from the ratio of the peak intensities in the spectra.

Interface trap density was measured by the charge-pumping technique, which can evaluate the interface states of a small MOSFET with thin gate dielectric. Agilent 4156C with pulse generators provided a stepping-pulse base voltage with fixed amplitude. Both lead and trail slopes of the pulses were 1 MV/s. The base voltage started at −2.0 V and stopped at 2.0 V.



(a) Time series of V_{th} fluctuation due to RTN.



(b) Power spectral density of RTN.

Fig. 5.1. Typical fluctuation due to RTN: single RTN trap at gate dielectric capturing and emitting a carrier.

The RTN measurement system was introduced in chapter 2. Typical result of the binary V_{th} fluctuations is shown in Fig. 5.1(a). The power spectral density (PSD) determined by FFT is shown in Fig. 5.1(b). When V_{th} randomly changes between the upper and lower states as shown in Fig. 5.1(a), the PSD has a Lorentzian shape with $1/f^2$ roll-off as shown in Fig. 5.1(b).

5.3 Dependence of Si/SiO_2 Interface State on Si Surface Orientation

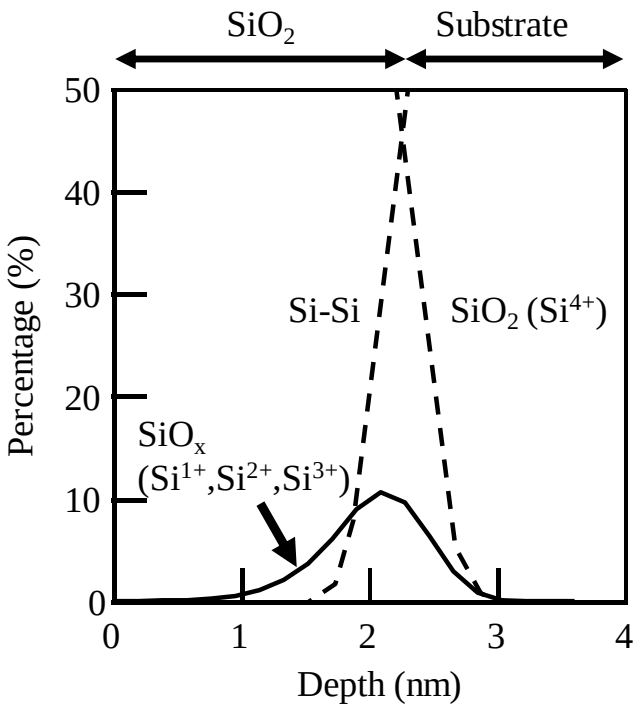
As RTN V_{th} variations are due to the interface state between the silicon substrate and the gate dielectric, the author, first, investigated the effect of silicon-surface orientation on the interface state density using XPS and charge pumping. In XPS, silicon atoms in intermediate oxidation states, so-called "suboxide", are observed in the interfacial transition layer. Some of the suboxides break and become fixed charges or interface traps since the suboxide bonding is energetically costly [8, 9]. Specifically, it has been suggested that the suboxide bonds, $(Si_{4-x}O_x)$::::Si where x means the number of chemical bonds between the right side silicon and oxygen; x = 0, 1, 2, 3, or 4, are ruptured to give dangling-bond configurations given by, for example, $Si_{3-y}O_y \equiv Si\bullet$ where y = 0, 1, 2, or 3 [9]. Of them, (y = 0) $Si_3 \equiv Si\bullet$ is known as a P_b center or a P_{b0} center. $Si_3 \equiv Si\bullet$ is trivalent silicon and has a dangling-bond defect. The P_b and P_{b0} centers are thought to be electrically similar and to be a majority of the interface traps. As remarked above, the suboxides in the intermediate oxide layer are closely linked to the dangling bonds as interface trap. Park et al. reported that both the amount of the suboxide and the interface trap density show similar dependence on gate oxidation processes [10]. Therefore, in the investigation of the connection between the silicon-surface orientations and interface states, it is expected that the amounts of the suboxides, which are estimated by XPS results, are coincident with N_{it} measured by charge pumping.

Depth dependence of amount of suboxides is shown in Figs. 5.2(a) to 5.2(c). The solid curves represent SiO_x (i.e., Si^{1+}, Si^{2+}, and Si^{3+}) percentage to sum of Si-related spectral intensities. The dashed curves increasing and decreasing with depth represent a Si-Si bond and Si^{4+}, respectively, the latter corresponding to stoichiometric SiO_2.

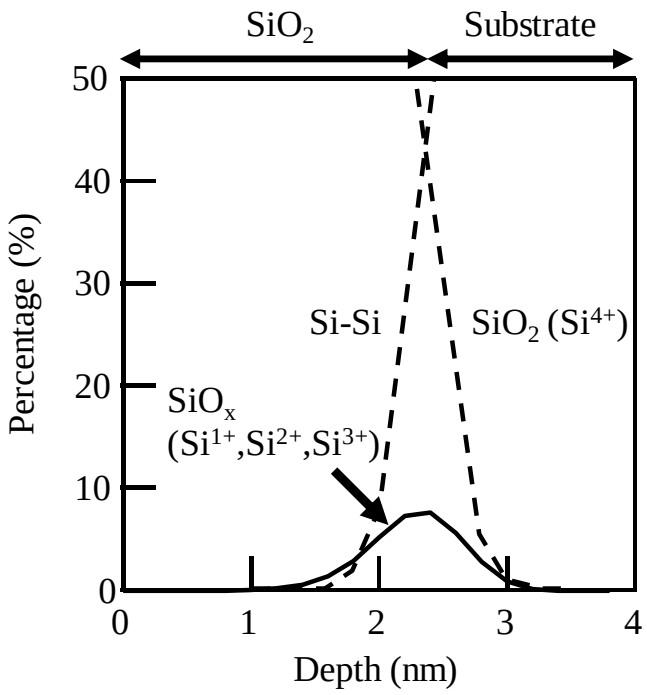
The peak amounts of the suboxides are 18, 11, and 8% for Si(110), Si(100), and Si(111), respectively. The peak for Si(110) is clearly higher than that for Si(111) or Si(100), whereas Si(100) shows slightly higher intensity than Si(111). Peak positions of all suboxides are within 1 nm from the interface. When the peak amounts of suboxides are converted to mean surface densities of suboxides, they are 3.0×10^{14}, 1.8×10^{14}, and 1.3×10^{14} cm^{-2} for Si(110), Si(100), and Si(111), respectively. These results are better than the report by Grunthaner et al., $(5.1 \sim 6.5)\times10^{14}$ cm^{-2} [7].

The trap densities at the Si/SiO_2 interface measured by the charge-pumping method were shown as a function of silicon-surface orientation in Fig. 5.3. N_{it} of Si(110), 2.4×10^{12} cm^{-2}, is the largest of all the three orientations, showing good agreement with the XPS results mentioned above. Since the N_{it} value is about one hundredth of the surface density of suboxides, there is likely to be the substantial amount of the potentially-unstable suboxides giving the dangling-bond defect.

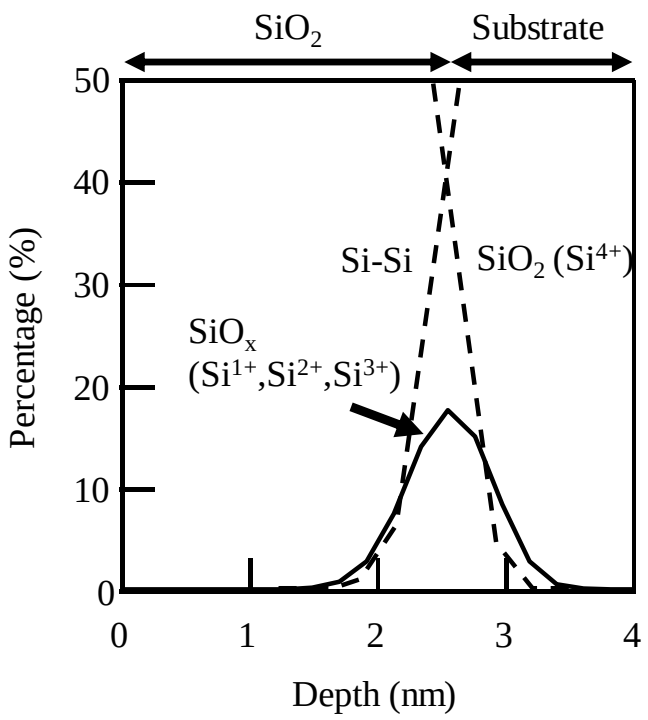
However, N_{it} of Si(111) is larger than that of Si(100), which is a reverse order of the amount of suboxides as shown in Fig. 5.3. Difference in the atomic configuration of the interface defect



(a) Si(100).



(b) Si(111).



(c) Si(110).

Fig. 5.2. Depth profiles of amount of suboxides.

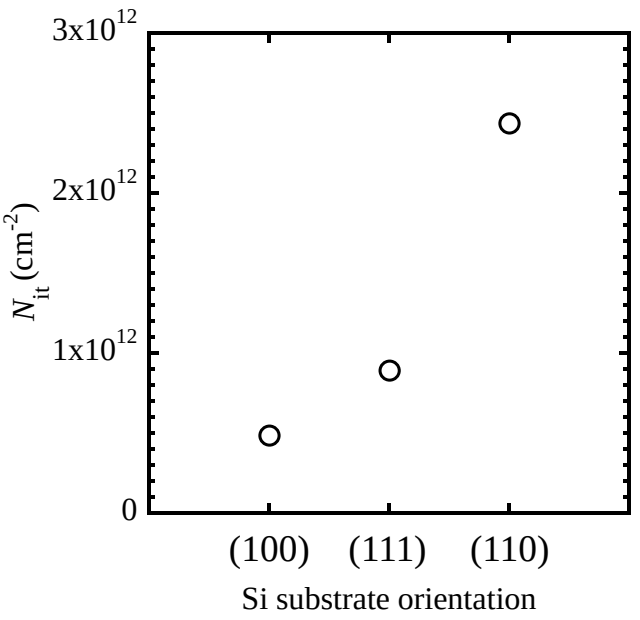


Fig. 5.3. Dependence of interface trap density on silicon-substrate orientation.

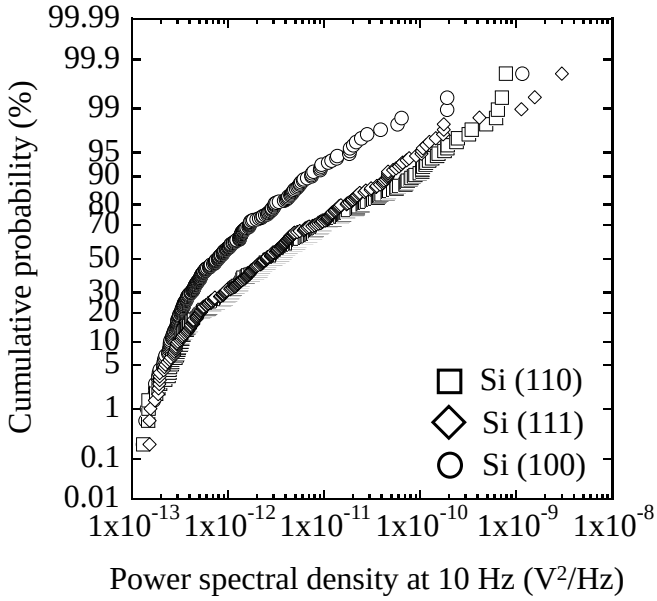


Fig. 5-4. Distributions of intensity of PSD at 10 Hz.

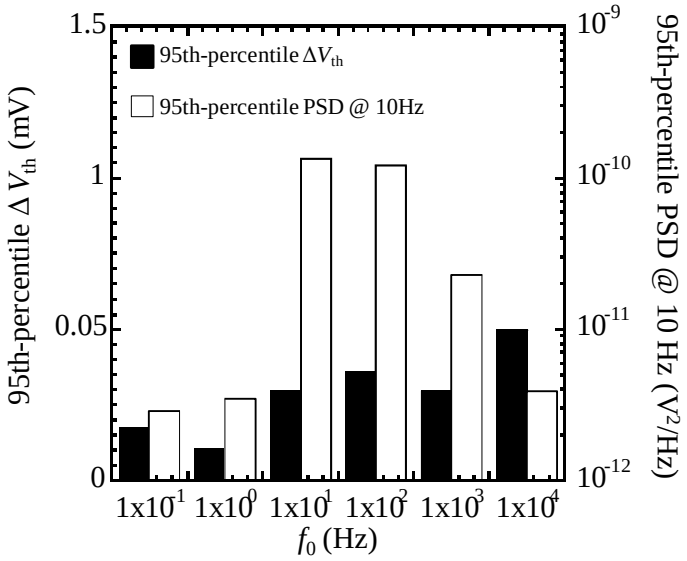
corresponds to the P_b center or the P_{b0} center. The P_{b0} center is generated at a silicon protrusion into the interfacial transition layer of the Si(100) and Si(110) substrates, whereas the P_b center only exists at Si(111) interface. In addition, The P_b and P_{b0} centers are generated by rupture of the Si^{1+} suboxide bond. From a crystallographic approach by Grunthaner et al., the surface density of Si^{1+} of Si(111) is larger than that of Si(100) [7, 10]. The surface densities of Si^{1+} on an ideal Si/SiO_2 interface are about 7.8×10^{14} cm^{-2}, and near zero (0.1×10^{14} ~ 0.5×10^{14} cm^{-2} in their real data) for Si(111) and Si(100). In contrast, the surface density of Si^{2+} of Si(100) is larger than that of Si(111). The author thinks that even though the total amount of the suboxides of Si(100) is larger than that of Si(111), N_{it} of Si(111) is worse than that of Si(100) because the interface of Si(111) gives a larger amount of Si^{1+} with potentiality from which the P_b center arises as interface trap.

The author clarified the close connection between the interface trap density and the surface density of suboxides from the investigation of the dependence of Si/SiO_2 interface state on silicon-surface orientation.

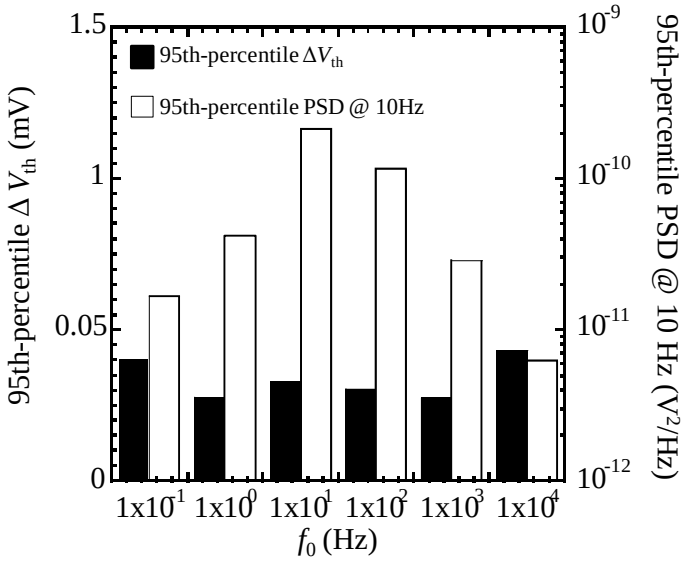
5.4 Relationship between Si Orientation and RTN

The difference between the statistical distributions of RTN V_{th} variation in the cases of Si(100), (110), and (111) substrates were investigated. Statistical analysis becomes more important in scaled-down devices since device-to-device variation increases as an average number of traps in a device decreases.

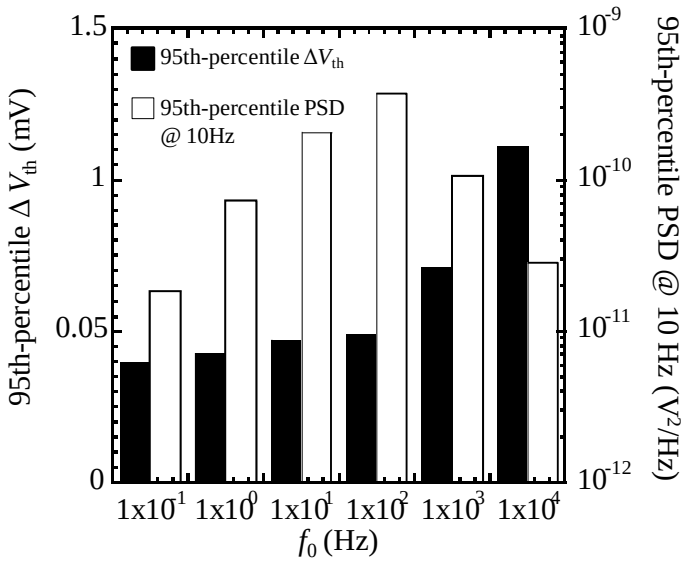
In the previous studies of a relationship between 1/*f* noise and silicon surface orientation, not many devices are used since 1/*f* noise is easily observed in a device with large area. Moreover, the noise intensities were usually discussed at a specific frequency, for example, 10 Hz or 1 kHz, as its frequency dependence is always 1/*f* [11-13]. On the other hand, a large number of samples are needed to understand the impact of RTN because ΔV_{th} follows the Gumbel distribution in



(a) Si(100).



(b) Si(111).



(c) Si(110).

Fig. 5.5. Histograms of 95th-percentile ΔV_{th} and 95th-percentile PSD at 10 Hz.

which anomalously large ΔV_{th}'s are observed as an uncommon RTN event as indicated in chapter 2. Device-to-device variation in frequency dependence of PSD of RTN addresses another issue in comparing power spectral densities between samples.

Figure 5.4 shows the distributions of PSD at 10 Hz in the case of Si(100), (110), and (111) substrates. The intensities distribute broadly from 1×10^{-13} to 1×10^{-8} V^2/Hz. The data show that the Si(111) and Si(110) substrates give higher intensities than Si(100) substrate. However, there is not a significant difference between the intensities for Si(111) and Si(110) substrates. The author found that this confusing result is attributable to the fact that the fixed-frequency intensity depends not only on ΔV_{th} but also on time constants of RTN. The intensity of power spectral density (S) of RTN is given by

$$S = \frac{4(\Delta V_{th})^2 \left[\frac{\overline{\tau_c}^2 \overline{\tau_e}^2}{(\overline{\tau_c}+\overline{\tau_e})^3} \right]}{1+\left(f/f_0\right)^2}, \quad (5.1)$$

where f_0 is a corner frequency given by

$$f_0 = \frac{1}{2\pi} \frac{\overline{\tau_c}+\overline{\tau_e}}{\overline{\tau_c}\cdot\overline{\tau_e}}. \quad (5.2)$$

As indicated by Eq. (5.1), S is determined not only by ΔV_{th} but also by $\overline{\tau_c}$ and $\overline{\tau_e}$. It is therefore difficult to compare the impact of RTN correctly using the intensity at the specific frequency such as 10 Hz, when V_{th} variations are dominated by RTN.

The author, therefore, reevaluates the conventional fixed-frequency comparison as follows. Figures 5.5 show dependences of ΔV_{th} at a cumulative distribution function of 95% on the corner frequencies. The primary Y-axis (left side) means the 95th-percentile V_{th} variation due to the uncommon RTN whose f_0 is within a given frequency range. Complex RTNs are divided into single RTNs in this estimation. The secondary Y-axis (right side) means the 95th-percentile value which is obtained from the cumulative probability of PSD at $f = 10$ Hz as indicated in Eq. (5.1). Each PSD is not an estimate from the drain-current waveform, but a calculated value by Eq. (5.1) using extracted RTN time constants. The 95th-percentile PSD reaches a peak around $f_0 = 10$ Hz and decreases sharply with an increase of f_0 from the peak because of the strong dependence of PSD on frequency as indicated in Eq. (5.2). The peaks of the 95th-percentile PSD at $f_0 = 10$ Hz are 2.1×10^{-10}, 2.1×10^{-10}, and 1.3×10^{-10} V^2/Hz for Si(110), Si(111), and Si(100), respectively. The value of Si(111) is nearly equal to that of Si(110). The result clearly reveals that it leads to significant underestimation of impacts of RTN with higher corner frequencies if the conventional fixed-frequency comparison for $1/f$ noise is applied to RTN.

Furthermore, it is noteworthy that the 95th-percentile ΔV_{th} of Si(110) increases with corner frequency as shown in Fig. 5.5(c). There is a gradual change in ΔV_{th} from 0.1 to 100 Hz in the Figure. ΔV_{th} significantly increases with corner frequency from 100 Hz. ΔV_{th}'s are 0.49, 0.71, and 1.1 mV at 100 Hz, 1 kHz, and 10 kHz, respectively. In general, interface traps closer to the silicon conduction band minima (E_c) show shorter time constants [14]. If the fast interface trap shows RTN behavior with $\overline{\tau_c}$ and $\overline{\tau_e}$ of 10^{-5} s, its corner frequency is 3×10^4 Hz. As noted in section 5.3, the P_b and P_{b0} centers are the main origin of the interface trap. It has been reported by Edwards [15], and Jupina and Lenahan [16] that the energy levels of the P_b and P_{b0} centers within the silicon band gap correspond to a change in back-bond angles to the dangling-bond silicon atom at the silicon interface. The trap level increases as the silicon atom behind the dangling-bond silicon atom moved closer to its nearest neighbor atoms.

The author supposes that the electrical property of the P_{b0} center is different from that of the P_b center in the case of RTN. The P_{b0} centers near E_c have the potential to give the fast RTNs with the larger V_{th} variation than the P_b centers near E_c. The Si(110) substrate not only gives the largest amount of suboxides and the largest interface trap density, but also has the potential to give the fast RTNs with the large ΔV_{th}. The P_{b0} center is generated at the silicon protrusion into the interfacial transition layer of the Si(100) and Si(110) substrates. Moreover, the P_{b0} center

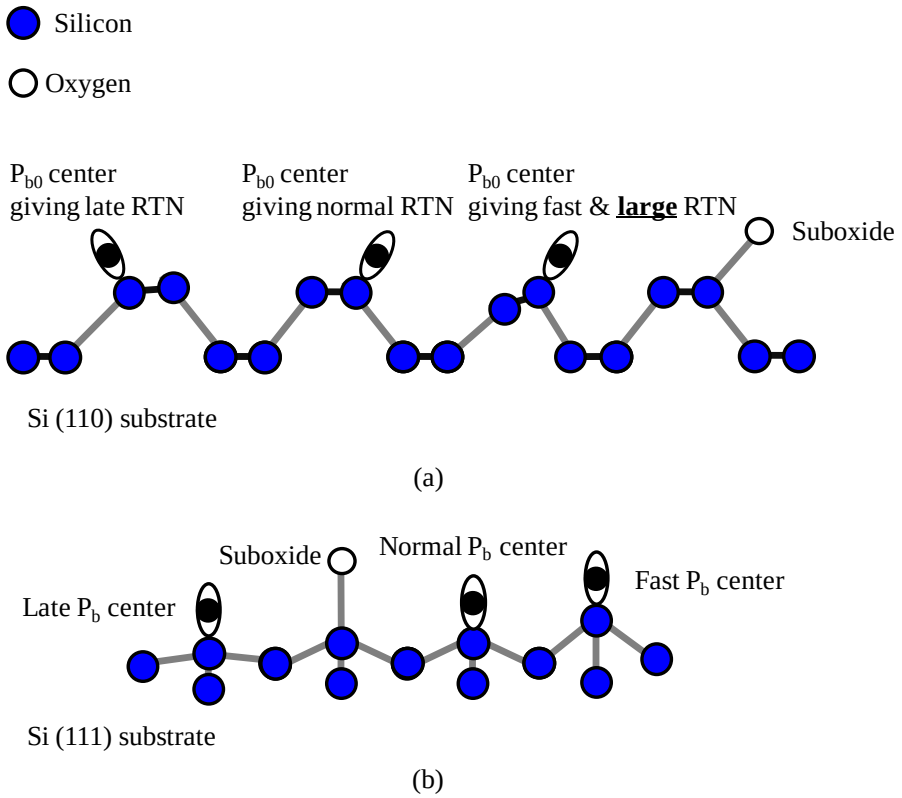


Fig. 5.6. Structural image of RTN traps and suboxide on Si(110) and Si(111) substrates: fast P_{b0} center on Si(110) giving large ΔV_{th}, on the other hand, all P_b centers on Si(111) giving similar impact on RTN ΔV_{th}. (a) Si(110). (b) Si(111).

gives the fast RTN with large ΔV_{th} if its configuration sharpens as shown in Fig. 5.6(a). Accordingly, as well as the Si(110) substrate, ΔV_{th} of Si(100) modestly increases with f_0 in Fig. 5.5(a). On the other hand, regardless of whether the P_b center's configurations sharpen or not, they make the similar impact on the RTN ΔV_{th} as shown in Figs. 5.5(b) and 5.6(b). The author, therefore, thinks that the RTN V_{th} variations strongly depend on a local atomic structure, namely, there is a high possibility of the larger ΔV_{th} due to the uncommon RTN if the local atomic structure around RTN trap is more protuberant. In this chapter, interface trap is assumed to be one of the causes of RTN. Probably, both interface trap and oxide trap near interface as border trap cause RTN behavior. Significantly, there is likely to be no distinction between interface trap and oxide trap as border trap.

RTN V_{th} variations with use of a statistical distribution based on the extreme-value theory is most suitable to enable correct evaluation of the influence of silicon-surface orientation on the RTN V_{th} variations in scaled MOSFETs. Figure 5.7 is the distributions of ΔV_{th} in accordance with a Gumbel function of extreme-value theory. This evaluation method is not affected by the

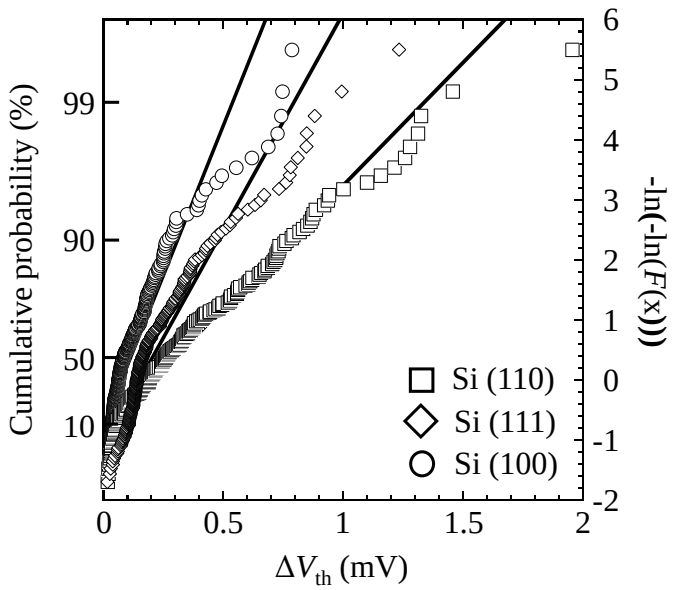


Fig. 5.7. Gumbel distributions of RTN ΔV_{th}.

Table 5.1. Summary of statistical analysis for each orientation.

	Average (mV)	Median (mV)	Variance (mV^2)
Si(100)	0.13	0.13	0.016
Si(110)	0.35	0.30	0.098
Si(111)	0.23	0.20	0.032

wide distributions of $\overline{\tau_c}$ and $\overline{\tau_e}$. An average, a median, and a variance of the Gumbel function are expressed as

$$average = \mu_1 + \sigma \times \gamma_{EM}, \quad (5.3)$$

$$median = \mu_1 - \sigma \times \ln[\ln(2)], \quad (5.4)$$

$$variance = \frac{\pi^2}{6} \times \sigma^2, \quad (5.5)$$

where γ_{EM} is the Euler-Mascheroni constant.

Table 5.1 summarizes the average, median, and variance values, which are estimated from the RTN ΔV_{th} distributions. The average, median, and variance values for the Si(110) substrate are larger than those for Si(100) and(111) substrates. The clear difference between characteristics of RTN for Si(110) and Si(111) substrates can be found not in the distribution of the intensity of PSD, but in the distribution of ΔV_{th} because the distribution of ΔV_{th} sufficiently reflects not only the influence of late RTN traps but also the influence of fast RTN traps, which are deeply linked to suboxides.

5.5 Conclusions

The (110) silicon-surface gives the largest total amount of suboxides of all three orientations, namely, Si(100), Si(110), and Si(111). Furthermore, the interface trap density (N_{it}) in the case of Si(110) substrate is the largest because interface trap density increases with increased amount of suboxides. On the other hand, N_{it} of Si(111) is larger than that of Si(100), which is a reverse order of the amount of suboxides. The possible reason is that the interface of Si(111) gives a larger amount of Si^{1+} from which the P_b center arises as interface trap more readily than the interface of Si(100), even though the total amount of suboxides of Si(100) is larger than that of Si(111).

It was also demonstrated statistically that RTN V_{th} variation of the Si(110) substrate is larger than those of the Si(100) and Si(111) substrates because Si(110) gives the largest total amount of suboxide which is deeply linked to the interface trap and contributes greatly to creating the fast RTN traps with large ΔV_{th}. The difference between ΔV_{th} for Si(110) and Si(111) substrates could not be found in the distribution of intensity of PSD. In contrast, the distribution of ΔV_{th} clearly showed that RTN V_{th} variation of Si(110) substrate is larger than those of Si(111) substrate. The reason is that unlike the distribution of intensity of PSD, the distribution of ΔV_{th} accurately reflects the influence of the fast RTN traps. From the result, the author thinks that the P_{b0} center near E_c has the potential to give the larger RTN V_{th} variation than the P_b center near E_c.

The author concludes that the Si(110) as the vertical plane of a 3D channel gives the largest amount of suboxides and the largest interface trap density. It follows that RTN traps on a Si(110) surface pose a considerable threat to 3D-device reliability.

5.6 References

[1] D. Hisamoto, W.-C. C. Lee, J. Kedziershi, H. Takeuchi, K. Asano, C. Kuo, E. Anderson, T.-J. King, J. Bokor, and C. Hu, “FinFET-a self-aligned double-gate MOSFET scalable to 20 nm,” *IEEE Trans. Electron Devices*, vol.47, no. 12, pp. 2320–2325, Dec., 2000.

[2] B. S. Doyle, S. Datta, M. Doczy, S. Hareland, B. Jin, J. Kavalieros, T. Linton, A. Murthy, R. Rios, and R. Chau, "High performance fully-depleted tri-gate CMOS transistors," *IEEE Electron Device Lett*., vol.24, no. 4, pp. 263–265, Apr., 2003.

[3] M. Guillorn, J. Chan, A. Bryant, N. Fuller, O. Dokumaci, X. Wang, J. Newbury, K. Babich, J. Ott, B. Haran, R. Yu, C. Lavoie, D. Klaus, Y. Zhang, E. Sikorski, W. Graham, B. To, M. Lofaro, J. Tornello, D. Koli, B. Yang, A. Pyzyna, D. Neumeyer, M.Khater, A. Yagishita, H. Kawasaki, and W. Haensch, "FinFET performance advantage at 22nm: An AC perspective," *VLSI Tech. Dig*., pp. 12–13, 2008.

[4] J. Kavalieros, B. Doyle, S. Datta, G. Dewey, M. Doczy, B. Jin, D. Lionberger, M. Metz, W. Rachmady, M. Radosavljevic, U. Shah, N. Zelick, and R. Chau, "Tri-Gate Transistor Architecture with High-k Gate Dielectrics, Metal Gates and Strain Engineering," in *VLSI Tech. Dig.*, pp. 50–51, 2006.

[5] H. Irie, K. Kita, K. Kyuno, and A. Toriumi, "In-plane mobility anisotropy and universality under uni-axial strains in nand p-MOS inversion layers on (100), [110], and (111) Si," in *IEDM Tech. Dig.*, p. 225–228, 2004.

[6] M. H. White, and J. R. Cricchi, "Characterization of thin-oxide MNOS memory transistors," *IEEE Trans. Electron Devices*, vol.19, no. 12, pp. 1280–1288, Dec., 1972.

[7] P. J. Grunthaner, M. H. Hecht, F. J. Grunthaner, and N.M. Johnoson, "The localization and crystallographic dependence of Si suboxide species at the SiO_2/Si interface," *J. Appl. Phys.* vol.61, no.2, pp. 629–638, Sep., 1987.

[8] S. T. Pantelides, S. N. Rashkeev, R. Buczko, D. M. Fleetwood, and R. D. Schrimpf, "Reactions of hydrogen with Si-SiO_2 interfaces," *IEEE Trans. Nucl. Sci*. vol.47 no.6, pp. 2262–2268, Dec., 2000.

[9] C.-T. San, *Fundamentals of Solid-State Electronics: Solution Manual*, World Scientific, Singapore, p. 110, 1996.

[10] Y-B. Park, X. Li, and S-W. Rhee, "Characterization of the Si/SiO2 interface formed by remote plasma enhanced chemical vapor deposition from SiH4/N2O with or without chlorine addition," *J. Vac. Sci. Technol*. B, vol.14, pp. 2260–2266, Jul., 1996.

[11] T. Sato, Y. Takeishi, and H. Hara, "Effects of Crystallographic Orientation on Mobility, Surface State Density, and Noise in *p*-Type Inversion Layers on Oxidized Silicon Surfaces," *Jpn. J. Appl. Phys.*, vol. 8, pp. 588–598, May, 1969.

[12] H. S. Momose, T. Ohguro, S. Nakamura, Y. Toyoshima, H. Ishiuchi, and H. Iwai, "Ultrathin gate oxide CMOS on (111) surface-oriented Si substrate," *IEEE Trans. Electron Devices*, vol. 49, no. 9, pp. 1597–1605, Sep., 2002.

[13] H. S. Momose, T. Ohguro, K. Kojima, S. Nakamura, and Y. Toyosuma, “1.5-nm gate oxide CMOS on [110] surface-oriented Si substrate,” *IEEE Trans. Electron Devices*, vol. 50, no. 4, pp. 1001–1008, Apr., 2003.

[14] E. H. Nicollian, and J. R. Brews, “MOS (Metal Oxide Semiconductor) Physics and Technology,” Wiley, New York, 2003, p.296.

[15] A. H. Edwards, “Theory of the P_b center at the <111> Si/SiO_2 interface,” *Phys. Rev. B*, vol. 36, pp. 9638–9648, Dec., 1987.

[16] M. A. Jupina, and P. M. Lenahan, “Spin dependent recombination: a ^{29}Si hyperfine study of radiation-induced P_b centers at the Si/SiO_2 interface,” *IEEE Trans. Nucl. Sci*. vol. 37, no. 6, pp.1650–1657, 1990.

6. Impact of RTN V_{th} variation on CMOS

6.1 Introduction

From chapters 2 to 5, the author demonstrated the impact of RTN on scaled nMOSFETs and the gate stacks techniques. However, to estimate the impact of RTN on logic devices and SRAM, discussion not only about RTN in nMOSFET, but also about RTN in pMOSFET is essential. In recent years, the reliability of scaled pMOSFET has been paid attention because the NBTI degradation in pMOSFETs is one of the big challenges for scaling [1-3]. Especially, the SRAM margin of pMOSFET is becoming narrower with scaling than that of nMOSFET [4-6]. The investigation of RTN behaviors in pMOSFETs is significant as well as that of impacts of RTN on scaling and the gate stacks techniques.

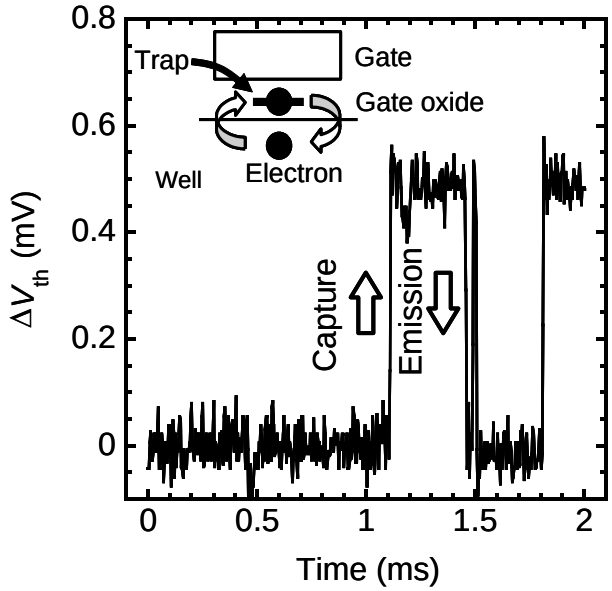
In chapters 6 and 7, RTN in pMOSFET is discussed. In this chapter, the author statistically analyzes RTN V_{th} variations in both n- and pMOSFETs to estimate the impact of RTN on the scaled-down SRAM. It is found that the difference in the distribution of ΔV_{th} of both n- and pMOSFETs can be explained by considering the unified number- and mobility-fluctuation models. Moreover, a SRAM margin enclosed by read/write V_{th} curves with or without RTN is simulated and the impact of RTN in comparison with the erratic phenomena on SRAM operation is discussed.

6.2 Experimental setup for noise measurement

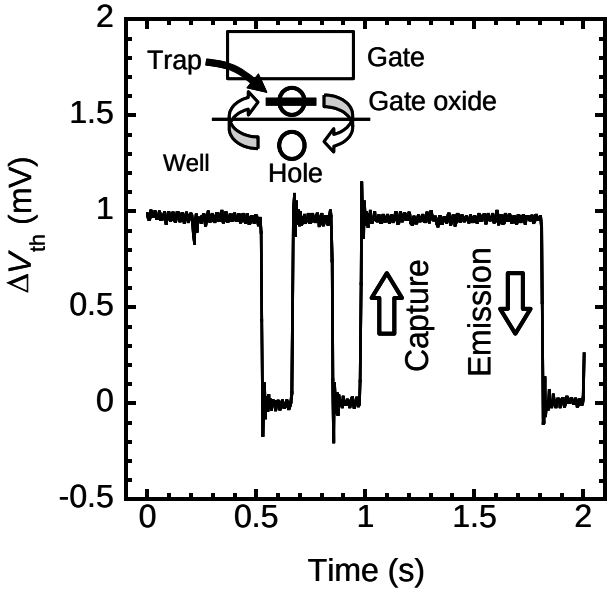
To statistically analyze the impact of the V_{th} variations due to RTN, the author measured electronic noises in many n- and pMOSFETs with same device size. The author, thus, used more than 200 n- and p-MOSFETs with the L_{eff} of 0.13 and W_{eff} of 0.22 μm, respectively. The gate oxide is pure SiO_2. When measured device is nMOSFET, the author measured the time dependences of drain current fluctuation (ΔI_d) of all samples under gate voltage (V_g) of 1 or 1.5 V and drain voltage (V_d) of 50 mV in the ohmic region by using the system introduced in chapter 2. On the other hand, when the device is pMOSFET, the author measured ΔI_d under gate voltage of −1 or −1.5 V and drain voltage (V_d) of −50 mV. RTN I_d variation is defined as the peak-to-peak value as demonstrated in chapter 4. ΔI_d is transformed into ΔV_{th} by using transconductance.

6.3 Observation of RTN in n- and pMOSFETs

Typical results of noise measurement using n- and pMOSFETs are shown in Figs. 6.1 (a) and (b), respectively. Binary fluctuations of V_{th} due to capture and emission of carrier at single trap in the gate oxide are clearly recognized, indicating that RTN is caused by trapped hole as well as electron. Moreover, complex RTN is sometimes observed both for n- and pMOSFETs as shown in Figs. 6.2 (a) and (b). Therefore, similar RTN behaviors are more likely to occur in both n- and



(a) nMOSFET.

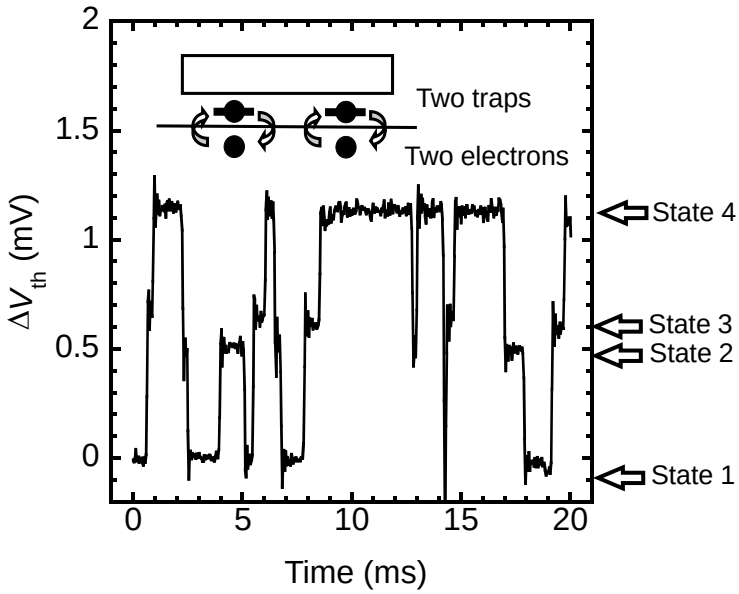


(b) pMOSFET.

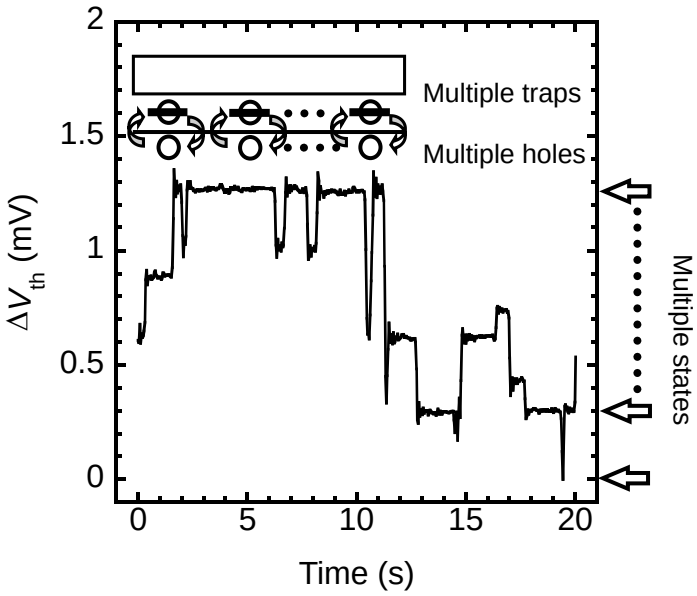
Fig. 6.1. Time dependence of V_{th} fluctuation due to RTN.

pMOSFETs. In this chapter, the author applies the Gumbel distribution to compare the RTN V_{th} variations in pMOSFETs with those in nMOSFETs.

As shown in Fig. 6.3, the Gumbel plot of ΔV_{th} of pMOSFET is on a straight line as well as that of nMOSFET. And an interesting point is that ΔV_{th} of pMOSFET is obviously larger than that of nMOSFET. In the observation of $1/f$ noise, it has been reported that the strength of $1/f$



(a) Complex RTN caused by two traps of nMOSFET.



(b) Complex RTN caused by multiple traps of pMOSFET.

Fig. 6.2. Complex RTN.

noise in pMOSFET is larger than that in nMOSFET [4]. However, the comparison of the RTN V_{th} variations in n- and pMOSFETs is, for the first time, reported.

Two physical models have been proposed for the electronic noises such as $1/f$ noise in MOFFET as shown in Figs. 6.4. A possible cause of ΔV_{th} is capture/emission of charge to a

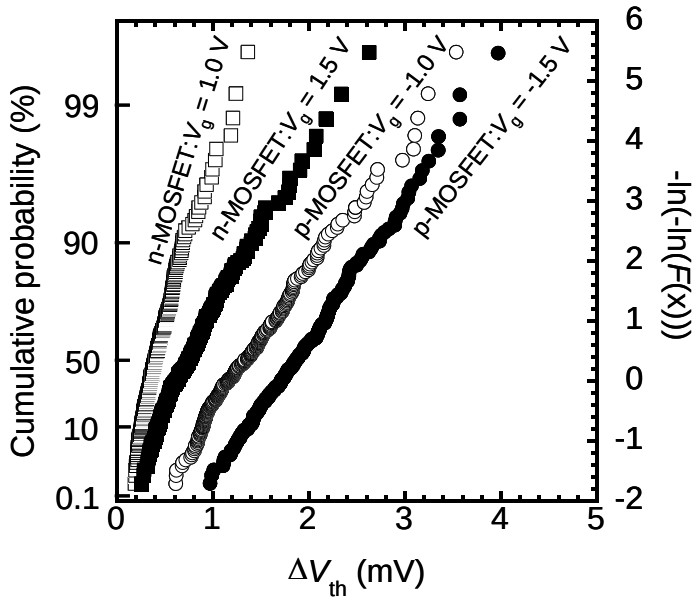


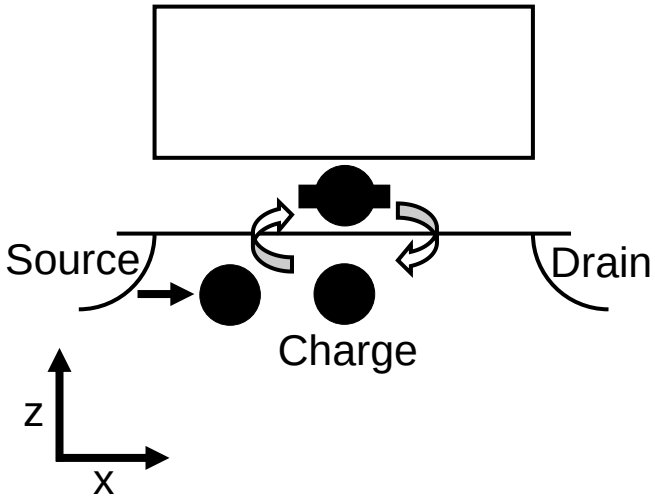
Fig. 6.3. Cumulative distributions of V_{th} fluctuation.

defect at gate dielectric (number-fluctuation model). Another model is a mobility-fluctuation model, which is interpreted in terms of fluctuations in the free path length of the carriers [7, 8]. Moreover, a unified model of the number- and mobility-fluctuation models has been also suggested for analysis of the 1/*f* noise. It was reported by T. Boutchacha *et al*. that a mobility scattering factor (α), which is estimated from the mobility-fluctuation model, is different between n- and pMOSFET. The scattering factor of p-MOSFET (8.8×10^{-15} Vs) is about one order bigger than that of the n-MOSFET (4.5×10^{-16} Vs) [7]. The strength of 1/*f* noise in pMOSFET is larger than that in nMOSFET.

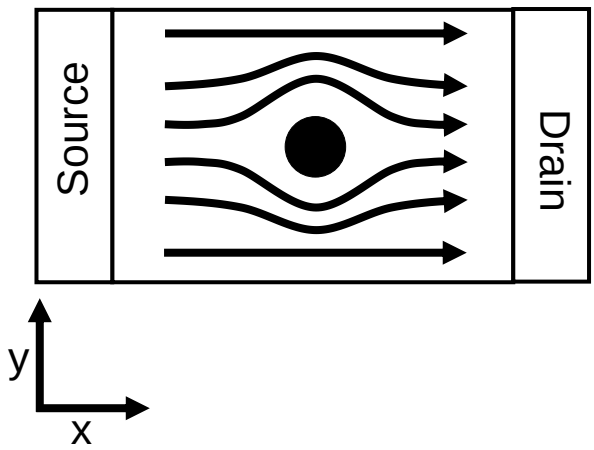
However, to discuss the difference in V_{th} fluctuations due to RTN between the n- and pMOSFETs, the unified model and the difference in the scattering factor have never been considered. The author derives the function of unified model in the case of RTN behavior to investigate the difference in RTN variations between n- and pMOSFETs. G. Ghibardo *et al*. assumed that a charge fluctuation derived from the number-fluctuation model is corresponded to a flat band voltage fluctuation (ΔV_{fb}) and the author introduces their concept [8]. Assuming that the all traps causing RTN are located at Si/SiO_2 interface, ΔV_{fb} is expressed as

$$\Delta V_{fb} = \frac{qn}{L_{eff}W_{eff}C_{ox}} = \frac{qN_t}{C_{ox}} \quad (6.1)$$

where n is the number of traps and N_t is the trap density per unit area. On the other hand,



(a) Number-fluctuation model: ΔV_{th} is caused by changing number of free carriers in inversion layer.



(b) Mobility-fluctuation model: the mobility fluctuates based on coulomb potential of the trapped charges in the gate oxide.

Fig. 6.4. Number- and mobility-fluctuation models proposed as RTN fluctuation model.

when I_d is differentiated partially with respect to V_{fb} and the effective mobility (μ_{eff}), ΔI_d is written as

$$\Delta I_d = \Delta V_{fb} \frac{\partial I_d}{\partial V_{fb}} + \Delta \mu_{eff} \frac{\partial I_d}{\partial \mu_{eff}} \quad (6.2)$$

Matthiessen's rule for the mobility is

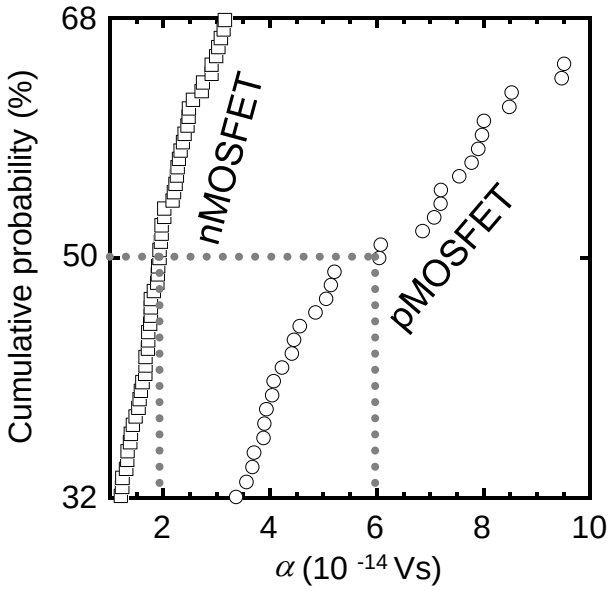


Fig. 6.5. Estimation of mobility scattering factor.

$$\frac{1}{\mu_{\mathrm{eff}}} = \frac{1}{\mu_{\mathrm{eff0}}} + \frac{1}{\mu_{\mathrm{ox}}} = \frac{1}{\mu_{\mathrm{eff0}}} + \alpha N_{\mathrm{t}} \quad (6.3)$$

where μ_{eff0} is the primary mobility and μ_{ox} is the mobility limited by oxide charge scattering. In addition, I_{d} in the ohmic region is expressed as

$$I_{\mathrm{d}} = \mu_{\mathrm{eff}} \frac{W_{\mathrm{eff}}}{L_{\mathrm{eff}}} Q_{\mathrm{i}} V_{\mathrm{d}} \quad (6.4)$$

where Q_{i} is the inversion charge per unit area. Therefore, from Eq. (6.2) to Eq. (6.4), ΔI_{d} can be written as

$$\Delta I_{\mathrm{d}} = g_{\mathrm{m}} \Delta V_{\mathrm{fb}} \pm \alpha q I_{\mathrm{d}} \mu_{\mathrm{eff}} N_{\mathrm{t}} \quad (6.5)$$

The ± sign in Eq. (6.5) depend on the charge state of the scattering centre. The – sign holds in the case of nMOSFET, while the + sign holds in the case of pMOSFET [9]. In this work, ΔI_{d} is defined as the difference between the maximum and minimum I_{d} and ΔI_{d} was transformed into ΔV_{th} by g_{m}. Therefore, ΔV_{th} can be eventually expressed as

$$\Delta V_{\mathrm{th}} = \left(1 \pm \alpha \mu_{\mathrm{eff}} C_{\mathrm{ox}} \frac{I_{\mathrm{d}}}{g_{\mathrm{m}}}\right) \Delta V_{\mathrm{fb}} \quad (6.6)$$

Equation (6.6) means that ΔV_{fb} derived from the number fluctuation is enhanced by the effect of the mobility fluctuation. The author checked validity of application of the mobility-fluctuation model to RTN by using Eq. (6.6).

The scattering factor can be estimated by measuring ΔV_{th} by using different gate voltages (V_{g} = 1.0 and 1.5 V) as shown in Fig. 6.3. Figure 6.5 shows cumulative distributions of the scattering factor of the n- and p-MOSFET. As shown in Fig. 6.5, the scattering factor of the nMOSFET is several times bigger than that of pMOSFET at median, i.e., about 2×10^{-15} Vs and about 6×10^{-15} Vs for n- and pMOSFET, respectively. This result implies that the flat band voltage fluctuation of pMOSFET is more enhanced by the effect of the mobility-fluctuation than that of nMOSFET.

Why is the scattering factor of pMOSFET larger than that of nMOSFET? C. T. Sah and T. H. Ning reported that the mobility limited by surface oxide charges is shown as below [10].

$$\mu_{\mathrm{ox}} = \frac{16 \varepsilon_{\mathrm{av}} \hbar k T}{\pi m^* q^3 N_{\mathrm{it}}} \quad (6.7)$$

$$\varepsilon_{\mathrm{av}} = \frac{\varepsilon_{\mathrm{Si}} + \varepsilon_{\mathrm{SiO2}}}{2}$$

where m^* is the carrier effective mass, $\hbar$ is the Planck constant, k is the Boltzmann factor, T is the temperature, $\varepsilon_{\mathrm{av}}$ is the average dielectric constant of Si and SiO_2. From equations (6.3) and (6.7), the scattering factor is expressed by

$$\alpha = \frac{\pi m^* q^3}{16 \varepsilon_{\mathrm{av}} \hbar k T}. \quad (6.8)$$

Thus, the scattering factor is proportional to the effective mass [11]. The electron effective mass is $0.19m_0$. The hole effective mass is $0.54m_0$. Accordingly, the scattering factor of pMOSFET is expected about three times larger than that of nMOSFET. This result is generally consistent with the median data as shown in Fig. 6.5.

6.4 Impact of RTN on SRAM operation

The author estimated the impact of RTN on the operation of SRAM on the basis of the RTN

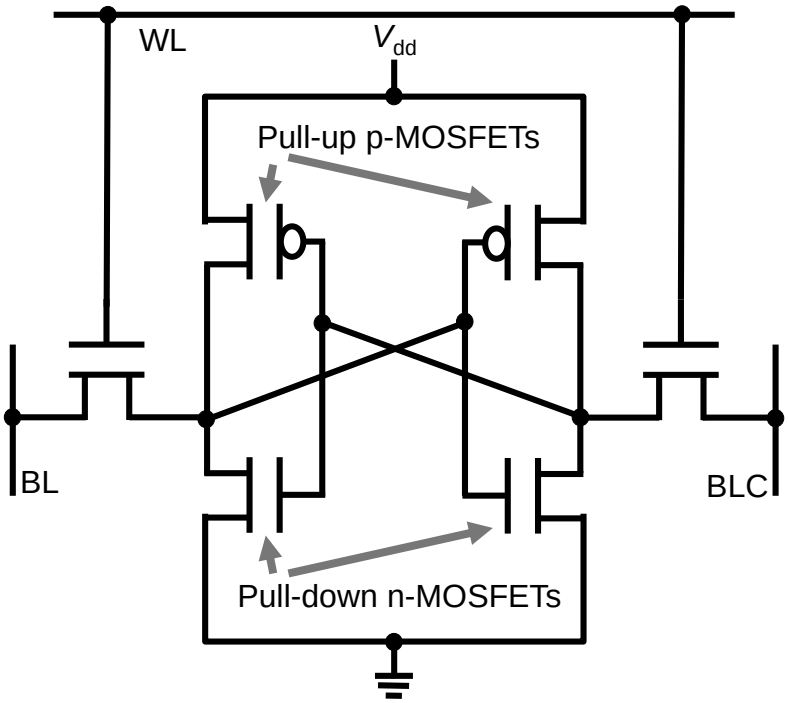


Fig. 6.6. Six transistors of SRAM bit cell.

data. In general, the SRAM read stability decreases when the threshold voltage of the nMOSFET is smaller and the absolute value of the threshold voltage of pMOSFET is larger than the typical V_{th} condition. Since the write operation is opposite the read one, the write stability is guaranteed in this V_{th} region. U. Tsukamoto *et al.* simulated the margin of SRAM to estimate the impact of RDF on SRAM operation [12]. The author applied this method to the analysis of the V_{th} fluctuations due to RTN as well as RDF.

If RTN occurs in a 6-transistor SRAM cell, RTNs affect pull-down nMOSFETs and pull-up pMOSFETs as shown in Fig. 6.6. Assuming that the 6-transistor SRAM cell is fabricated using the 65 nm node, this impact can be visually expressed using a graph that has the V_{th} variation of the pull-down nMOSFET as horizontal and that of the pull-down pMOSFET as vertical axes, respectively as shown in Fig. 6.7. This graph shows that RTN affects read / write stabilities because read / write operations are realized by the valance of MOSFET's performance [13]. According to Tsukamoto's report, the upper line (read V_{th} curve) of this graph indicates a read boundary determined by restriction, namely the DC margin becomes zero. The lower line (write V_{th} curve) of this graph indicates a write boundary. The region enclosed by this read / write V_{th} curves indicates the SRAM operation margin. In this work, when the V_{th} curves were estimated, the author examined the impact of RTN as well as that of RDF on the SRAM operation margin. The dark region is the V_{th} window. If the 6-transistor SRAM cells consist of ideal and typical MOSFETs, their ideal value is located at the center of this dark region.

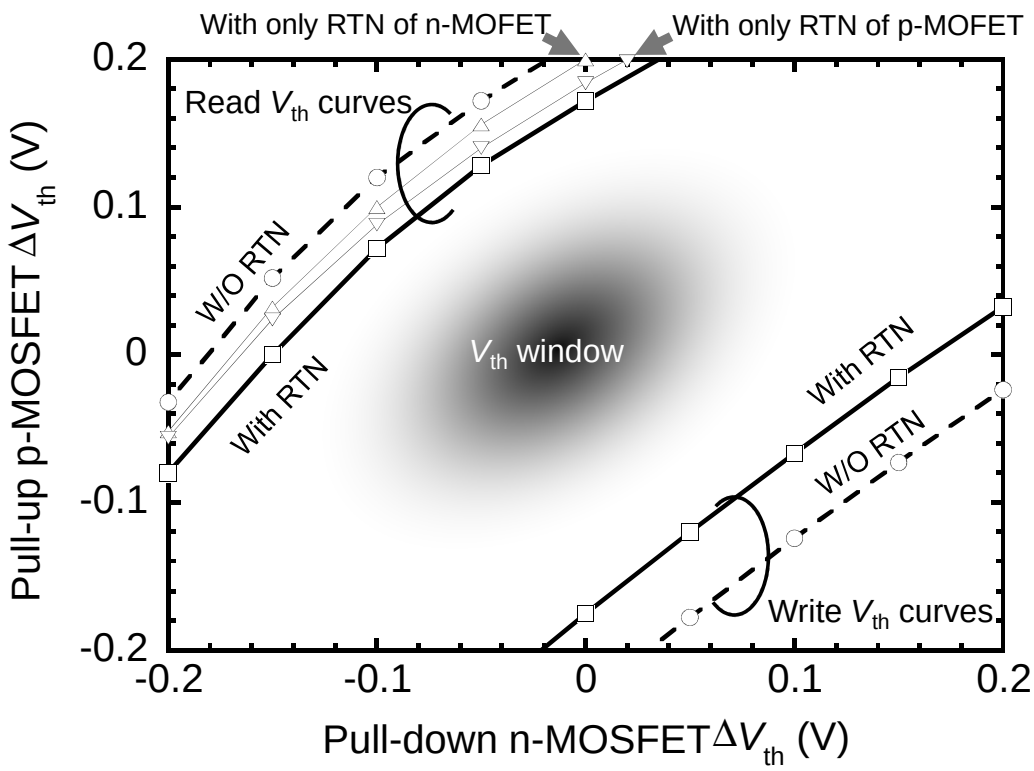


Fig. 6.7. SRAM margin enclosed by read/ write V_{th} curves at 65 nm generation.

Regarding the read/write V_{th} curves with RTN, the author assumed that the V_{th} variation of nMOSFET is 10 mV, while the V_{th} variation of the pMOSFET is 30 mV at three standard deviations (3σ). The V_{th} margin becomes small by RTNs at the 65 nm node. The impact of the V_{th} variation due to RTN in pMOSFET on the read operation is larger than that of the n-MOSFET. It is noteworthy that, even at the 65 nm node, the V_{th} margin comes close to the V_{th} window of the SRAM by considering RTN in addition to RDF. As for pMOSFET, it is also affected by NBTI [14]. In conclusion, the researchers should pay more attention to the operation margin of SRAM, especially the pull-up pMOSFET.

6.5 Conclusion

The author found that ΔV_{th} of pMOSFET is larger than that of nMOSFET from the statistical viewpoint. The reason for the difference in ΔV_{th} is that the effect of mobility fluctuation on V_{th} variation of pMOSFET is bigger than that of nMOSFET.

Based on the data of V_{th} fluctuations, assuming that RTN occurs in a 6-transistor SRAM cell fabricated at the 65 nm generation, the impact of ΔV_{th} due to RTN is simulated and can be visually expressed by a graph showing both V_{th} margin with RTN and V_{th} window. As a result, the author found that the V_{th} margin comes close to the V_{th} window of SRAM by considering the effect of RTN on ΔV_{th}, even at the 65 nm generation.

6.6 References

[1] G. Chen, M. F. Li, C. H. Ang, J. Z. Zheng, and D. L. Kwong, “Dynamic NBTI of p-MOS Transistors and Its Impact on MOSFET Scaling,” *IEEE Electron Device Lett.*, vol. 23, no. 23, pp.734–736, Dec., 2002.

[2] A. T. Krishnan, V. Reddy, S. Chakravarthi, J. Rodriguez, S. John, S. Krishnan, “NBTI Impact on Transistor & Circuit: Models, Mechanisms & Scaling Effects,” in *IEDM Tech. Dig.*, pp. 349–352, 2003.

[3] G. Ghibaudo, O. Roux dit Buisson, C. N. Duc, F. Balestra, and J. Brini, “Improved Analysis of Low Frequency Noise in Field-Effect MOS transistors,” *Journal of physca status solidi* (a). vol. 124, pp. 571–581, Apr., 1991.

[4] S. V. Kumar, C. H. Kim, and S. S. Sapatnekar, “Impact of NBTI on SRAM Read Stability and Design for Reliability,” P*roc. ISQED*, pp. 212–218, 2006.

[5] K. Kang, H. Kufluoglu, K. Roy, and M. A. Alam, “Impact of Negative-Bias Temperature Instability in Nanoscale SRAM Array: Modeling and Analysis,” *IEEE Trans. Comput. Aided Des. Integer Circuits Syst.*, vol. 26, no. 10, pp. 1770–1781 Oct. 2007.

[6] V. Huard, C. Parthasarathy, C. Guerin, T. Valentin, E. Pion, M. Mammasse, N. Planes, and L. Camus, “NBTI Degradation: From Transistor to SRAM Arrays,” *Proc. IRPS*, pp. 286–300, 2008.

[7] L. K. J. Vandamme, “CORRELATION BETWEEN MOST 1/*f* NOISE AND CCD TRANSFER INEFFICIENCY,” *Solid-State Elecronics* vol. 28, no. 10, pp. 1049–1056, Jan., 1985.

[8] L. K. J. Vandamme, X. Li, and D. Rigaud, “1/*f* Noise in MOS Devices, Mobility or Number Fluctuations?,” *IEEE Trans. Electron Devices*, vol. 41, no. 11, pp. 1936–1945, Nov., 1994.

[9] E. Simoen, and C. Claeys, “Random Telegraph Signal: a local probe for single point defect studies solid-state devices,” *Journal of Materials Science and Engineering*, B91-92, pp.136–143, Apr. 2002.

[10] C. T. Sah, T. H. Ning, and L. L. Tschopp, “The scattering of electrons by surface oxide charges and by lattice vibrations at the silicon-silicon dioxide interface,” *Surf. Sci*., 32, pp. 561–575, Apr. 1972.

[11] K. K. Hung, P. K. Ko, C. Hu, and Y. C. Cheng, “A unified model for the Flicker noise in metal-oxide-semiconductor field-effect transistor,” *IEEE Trans. Electron Devices*, vol. 37, no. 3, pp. 654–665, Mar., 1990.

[12] U. Tsukamoto, K. Nii, S. Imaoka, U. Oda, S. Ohbayashi, T. Yoshizawa, H. Makino, K. Ishibashi, and H. Shinohara, “Worst-Case Analysis to Obtain Stable Read/Write DC Margin

of High Density 6T-SRAM-Array with Local Vth Variability," in *Proc. ICCAD*, pp. 398–405, 2005.

[13] M. Yamaoka, N. Maeda, Y. Shinozaki, Y. Shimazaki, K. Nii, S. Shimada, K. Yanagisawa, and T. Kawahara, " Low-Power Embedded SRAM Modules with Expanded Margin for Writing" in *ISSCC Tech.l Dig.*, 2005, pp. 480–482.

[14] M. E. Weybright, "SRAM Cell Stability: Part I. SRAM Basics," in *Proc. IRPS*, Tutorial, 2007.

7. Statistical analysis of relationship between NBTI and RTN in small pMOSFETs

7.1 Introduction

The previous chapter discusses the comparison of RTN V_{th} variation between n- and pMOSFET. This chapter focuses on RTN behaviors before and after an application of stress voltage because the author is interested in a relationship between RTN and NBTI. The physical mechanism of NBTI remains unknown even nearly half a century after the first report [1]. As the author noted in chapter 1, the most popularized model is a reaction-diffusion (R-D) model, which states that the degradation of pMOSFET is driven by breaking of hydrogen-passivated silicon bonds at the interface and subsequent diffusion of hydrogen as illustrated in Fig. 1.1(a) [2, 3]. Recently, a switching-trap model has been introduced to explain NBTI degradation and recovery in scaled pMOSFETs. In the switching trap model, NBTI features are understood by a combination of RTNs with different time constants as indicated in Fig. 1.1(b) [4-6]. Furthermore, V. Huard et al. and S. Mahapatra et al. attempted to explain the NBTI characteristics in terms of both interface trap generation and hole trapping / detrapping [7, 8]. Nevertheless, up to the present, complicated NBTI behaviors in scaled pMOSFETs have never been statistically analyzed to connect to a considerable research on NBTI-stress-induced traps in large-size devices [9-12].

Furthermore, RTN is regarded as the most basic kind of trap behavior and is caused by a single trap at the gate dielectric capturing and emitting a carrier. RTN fluctuations are further complicated and changeful because each RTN shows different amplitudes and different time constants, and moreover, a number of RTNs are overlapped as shown Fig. 2.7. Therefore, several selected RTN data are not enough to universally discuss the relationship between RTN and NBTI. In our conclusion, a study on RTN behaviors before and after the application of NBTI stress in scaled pMOSFETs requires a statistical viewpoint.

From the statistical perspective, the author investigates the change of the RTN V_{th} variations before and after the NBTI stress by using small narrow-channel pMOSFETs, whose gate length (L_g) and width (W_g) are 45 and 25 nm, respectively. It is demonstrated that the distribution of the RTN V_{th} variation is drastically changed before and shortly after NBTI stress, and subsequently, fast relaxation of the RTN V_{th} variation is shown in the statistical distribution. Moreover, the interface trap density (N_{it}) and V_{th} shift due to the NBTI stress are observed by using large-size transistors with both L_g and W_g of 100 μm, in which standard trap behaviors can be observed, to compare with degradation and the recovery of the RTN V_{th} variation.

7.2 Sample preparation and experimental methods

A small transistor gate of less than 100 nm can provide clear threshold-voltage variation (ΔV_{th}) of more than 10 mV, and over 100 transistors are needed to investigate critical values

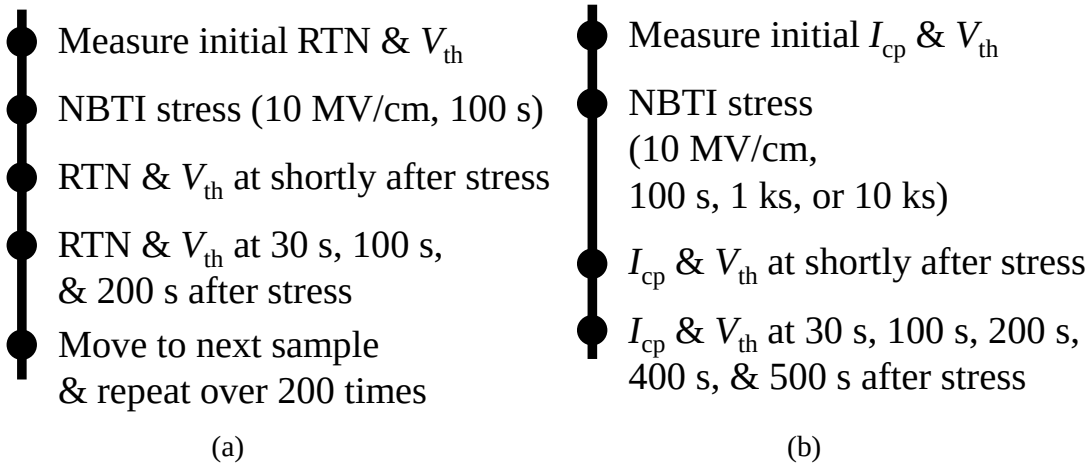
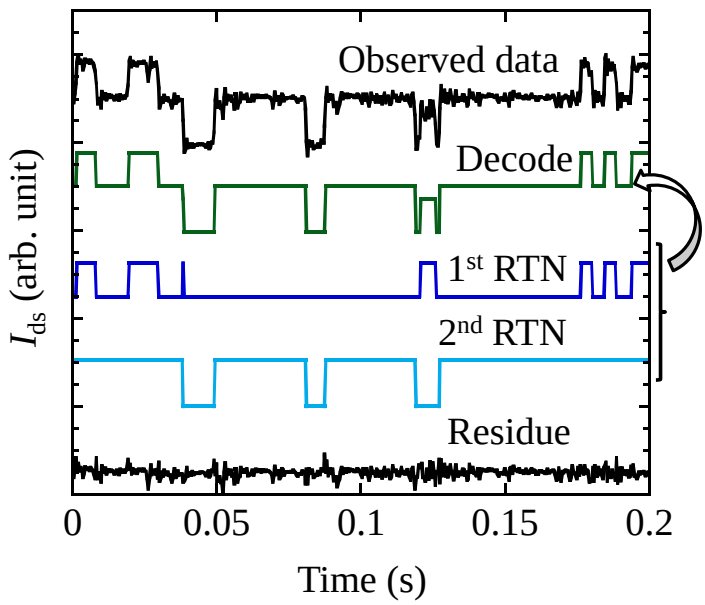


Fig. 7.1 Flows of measurement. (a): Statistical RTN measurements before and after NBTI stress. (b): I_{cp} and V_{th} are measured before and after NBTI stress.

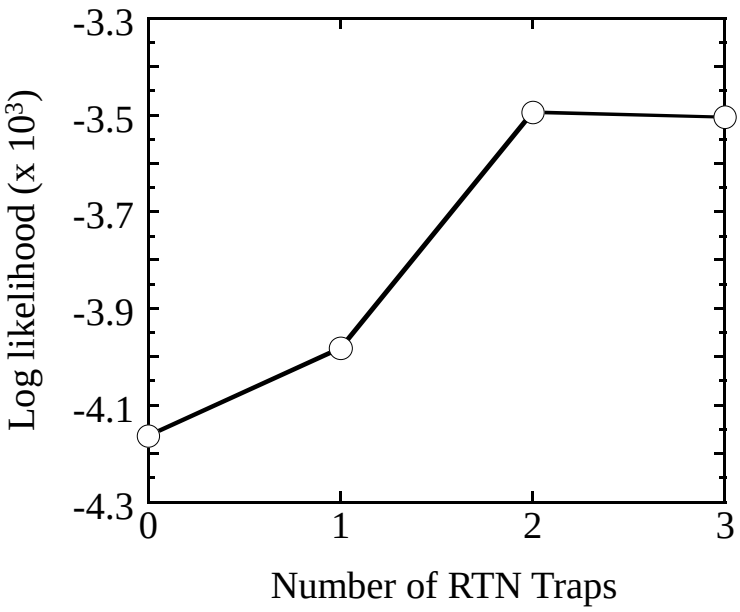
above 95% cumulative probability. Narrow-channel pMOSFETs with SiON / poly-Si gate stacks were designed to clearly show RTN. The gate dielectric was formed by dry oxidation, nitric oxide annealing, and forming gas annealing. The actual gate length and width were 45 nm and 25 nm, respectively. The number of measured devices was more than 200.

The new RTN measurement system is similar to the system in Fig. 2.3. It can apply NBTI stress in addition to the RTN measurement. RTN was measured under the following conditions: V_g of −1.0 V and V_d of −50 mV, and sampling time of 200 ms (because switching between RTN-measurement mode and NBTI-stress mode needs a few hundred milliseconds; hence, the system is suitable for tracing the recovery within several hundred milliseconds). As shown in Fig. 7.1(a), RTNs are monitored before and after the NBTI stress. This flow is automatically repeated over two hundred times (i.e., for over two hundred different devices). The conditions of the stress are shown in Fig. 7.1(a). The measurement temperature is 125 °C. With regard to data analysis, a hidden Markov model is used for extracting ΔI_d and decomposing overlapped RTNs [13]. The hidden Markov model means that the system is assumed to follow a Markov process with unobserved (hidden) internal states. In the observation of RTN behavior, I_d is equivalent to the output of the system, and the hidden internal state is equivalent to the charged state of the RTN trap. The transition of the hidden internal state (i.e., unobserved charged state) is based solely on the present conditions of system, and its past is independent (Markov process). Moreover, a Viterbi algorithm is used to find the unknown parameters, namely, RTN time constants and RTN amplitudes [14]. The overlapped RTNs can be decomposed into up to 5 single RTNs and residue. ΔI_d is transformed into ΔV_{th} by using G_m.

Figure 7.2 is an example of analysis of overlapped RTNs by the hidden Markov model. The data are from the initial state before the application of the NBTI stress. As shown in Fig. 7.2(a),



(a)



(b)

Fig. 7.2 Decomposition of multi-state RTN by using hidden Markov model. (a) Complex RTN states. The top data are the observed recovery process. The second top data are decoded using extracted 1st and 2nd single RTNs. (b) Estimation of optimum number of RTN traps.

the top data are from observation. The second top data are decoded using extracted 1st and 2nd single RTNs. The observed signal can be decomposed in two RTNs and residue, and the decoded signal closely reproduce the original one. The number of traps is estimated by a Baum–Welch algorithm. Figure 7.2(b) shows the result of estimation of the number of traps. The optimum number of traps is two because its log-likelihood value is the highest. The calculation of the number of traps stops if the log-likelihood value shows a decreasing trend.

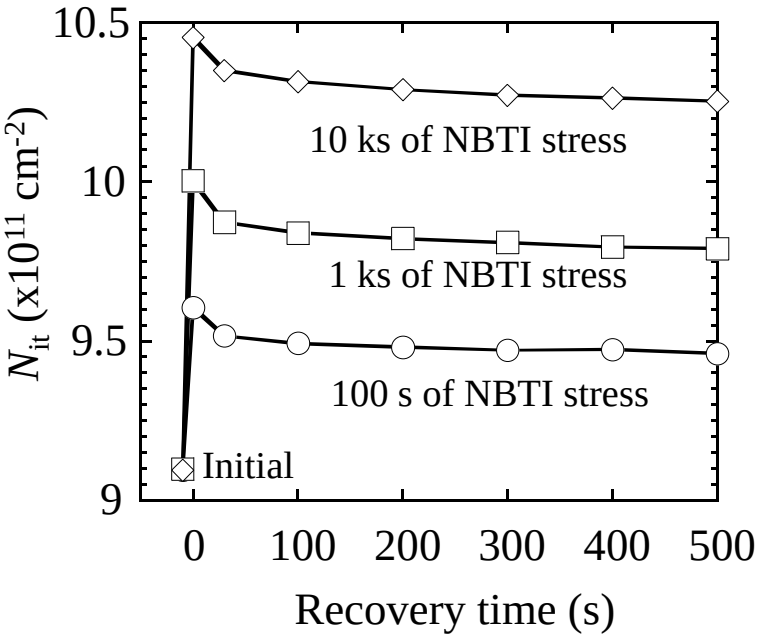
In regard to the measurement of interface trap density (N_{it}), as shown in Fig. 7.1(b), a similar measurement flow to that used for RTN was used. The charge pumping current (I_{cp}) was measured in the shortest time possible. The charge-pumping frequency was 100 kHz, the base voltage was +1.5 V, and the pulse peak voltage was −1.0 V. To make it possible to observe standard trap behaviors, the size of the transistor used was large; both the gate length and width were 100 μm.

7.3 V_{th} shift and change of N_{it} before and after NBTI stress

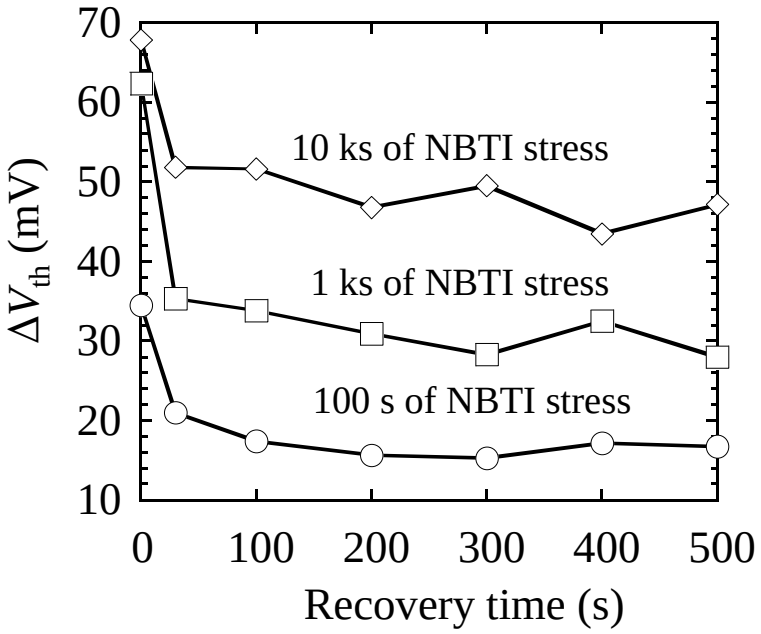
It was reported that the magnitude of the recovery of N_{it} under positive bias is much smaller than that of ΔV_{th} [15, 16] because both the interface traps and bulk traps such as oxide vacancies equally increase with increasing stress voltage and stress time, and in contrast, in the relaxation process, re-passivation of interface states is a negligible change, and hole de-trapping dominates. Prior to starting the RTN measurement, as shown in Fig. 7.3, N_{it} and ΔV_{th} were also measured. N_{it} and ΔV_{th} increase with increasing NBTI stress time. The recovery of N_{it} is clearly less than that of ΔV_{th}. For example, the recovery of N_{it} after 30 s under the stress time of 10 ks is only about 7.4%. In contrast, the recovery of ΔV_{th} reaches about 25% in the same time. The similarity is that most of the recovery occurs within 30 s. From these results, the author also assumes that hole de-trapping is likely to be dominants in the recovery process. However, it is still not known exactly how each trap relates to the NBTI degradation and the NBTI relaxation.

7.4 NBTI-stress-induced RTN

Next, the author attempted to observe RTN behaviors in the recovery process by the measurement method, as shown in Fig. 7.1(a). Examples of the recovery process are shown in Figs. 7.4(a) and (b). The RTN data were obtained shortly after the NBTI stress application. In these figures, since I_d decreases with increasing recovery time, it can be concluded that the recovery component appears in the measurement as intended. The observed signal in Fig. 7.4(a) can be decomposed to five single RTNs and residue. The temporary RTNs, for example, the 5th RTN with the largest amplitude, disappear halfway through the recovery. The disappearance of the 5th RTN mainly causes the recovery process. When it comes to the other RTNs, the 1st to 3rd RTNs with small amplitude remain, and 4th RTN disappears. It is, however, not clear how long

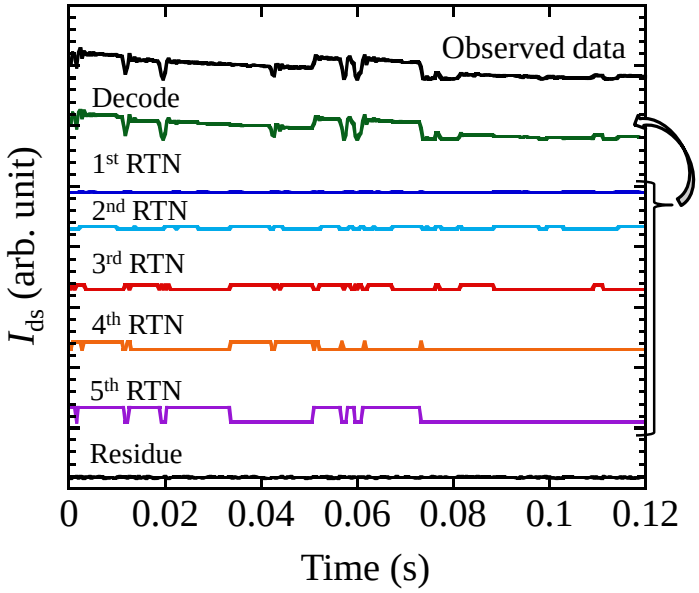


(a)

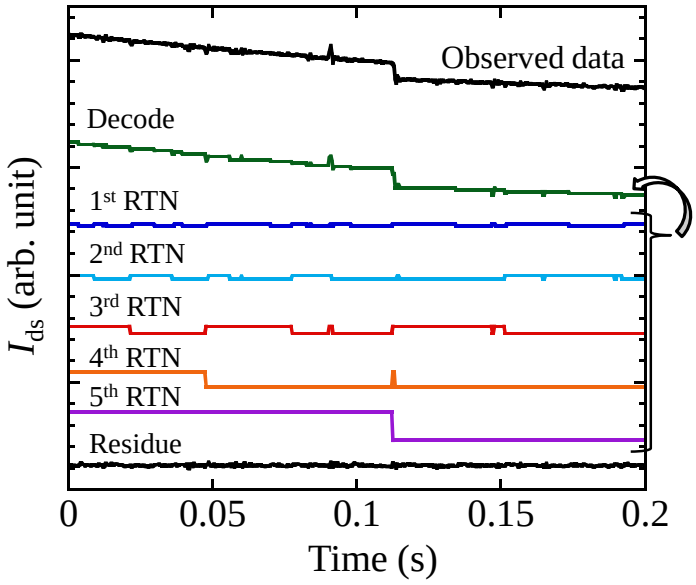


(b)

Fig. 7.3 Influence of NBTI stress. Circles, boxes, and diamonds show the recovery processes after 100 s, 1 ks, and 10 ks of the NBTI stress. (a): Influence of NBTI stress on N_{it}. (b): NBTI-stress-induced V_{th} shift from initial state.



(a)



(b)

Fig 7.4 Decomposition of recovery components obtained by measurement of RTN shortly after NBTI stress application. The top data are the observed recovery process. The second top data are decoded using extracted 1st – 5th single RTNs.

(a): RTNs exist in the recovery process. The 5th RTN disappears halfway through recovery.

(b): One-time RTN (5th RTN) boosts the recovery process.

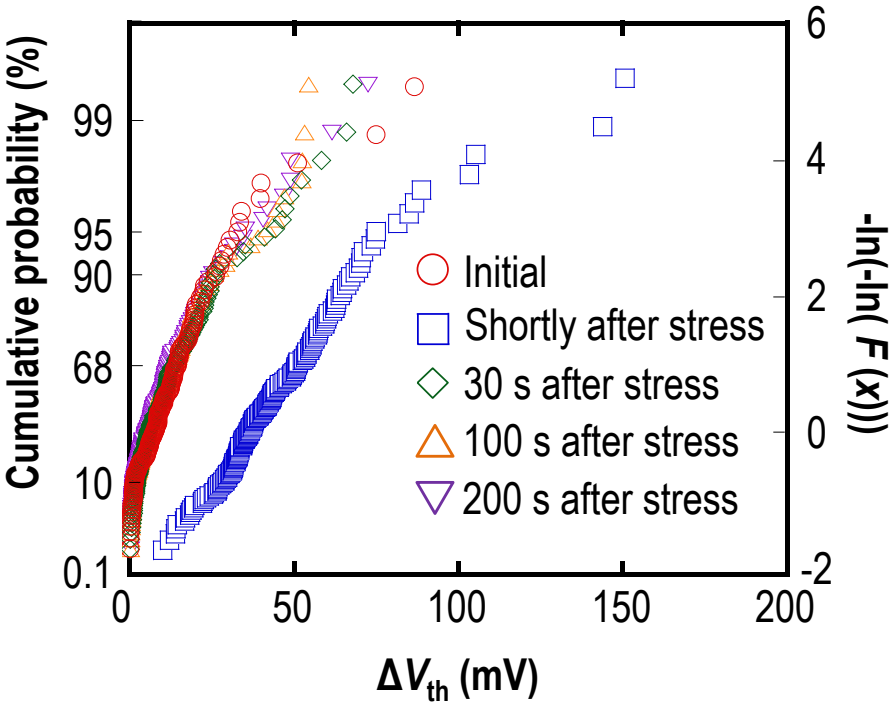


Fig. 7.5 Cumulative distribution of RTN V_{th} variation.

the 1st to 3rd RTNs remain in an active state. As the result of Fig. 7.4(a), the author concludes that NBTI stress causes the temporary RTNs, and sequentially, the disappearance of the temporary RTNs cause the recovery process.

Furthermore, a one-time RTN, the 5th RTN, boosting the recovery is found in Fig. 7.4(b) in addition to the temporary RTNs. Reportedly, the time to jump of each one-time RTN follows an exponential distribution in response to the charge-discharge property of a trap [4, 17]. This means that each one-time RTN has the potential to show the time to jump in an extremely wide range from microseconds to seconds under the same conditions. In the R-D model, the recovery time is determined by whether the reaction or the diffusion is dominant [18]. If the extremely wide distribution of recovery time was explained by the R-D model, it would be recognized that each interface trap has varieties of rate constants of reaction or diffusion coefficients under the same conditions. Therefore, it is reasonable to assume that the one-time RTN is a type of charge-discharge property of traps.

In the recovery process shortly after NBTI stress, the author observed the waveforms with a mixture of one-time RTNs and temporary RTNs. From the result of the statistical analysis, the percentage of devices in which one-time RTNs are observed simultaneously with temporary RTNs or permanent RTNs is about 81%. In contrast, the devices with only one-time RTNs cannot be observed. Therefore, the recovery process results from a complex combination of the

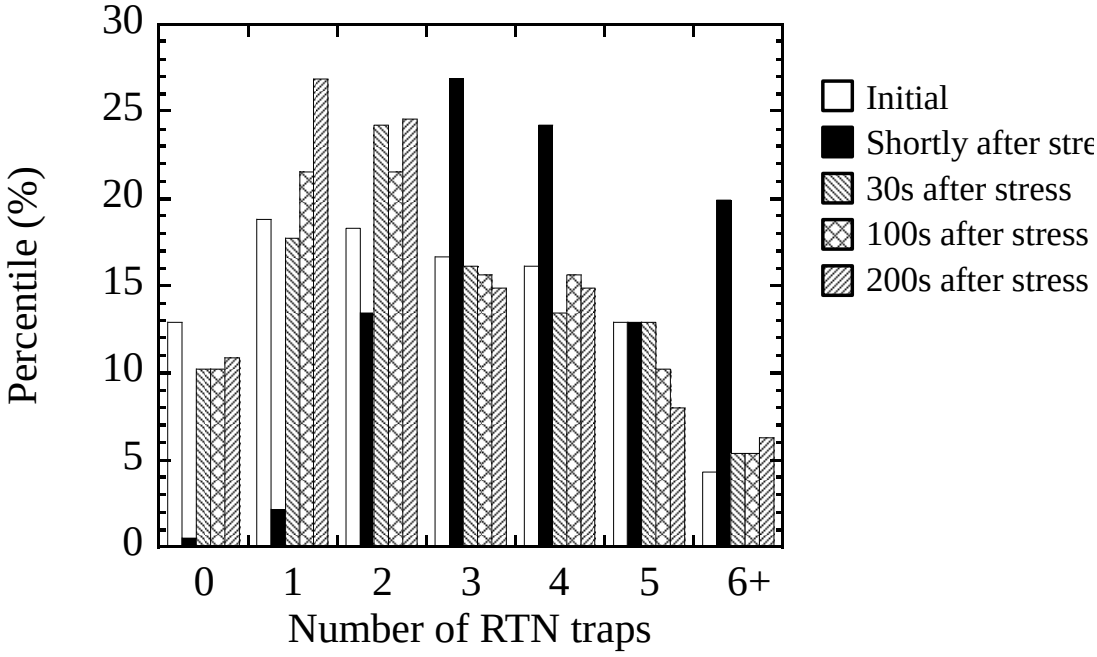


Fig. 7.6 Histogram of number of RTN traps.

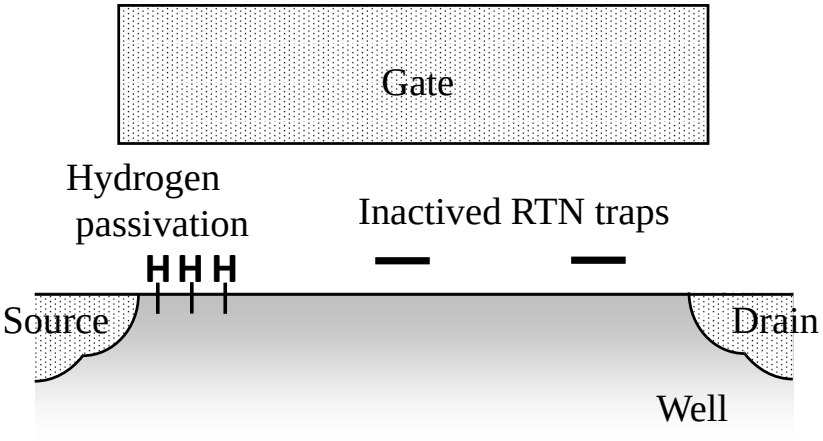
temporary RTNs and the one-time RTNs.

From these data, the recovery process is more likely to be caused by disappearance of the temporary RTN and the one-time RTN. Since RTN V_{th} variation is thought to show the extreme value distribution in the case of scaled MOSFETs as indicated in chapter 2, a statistical viewpoint is required for investigating the change of the RTN V_{th} variations before and after the NBTI stress.

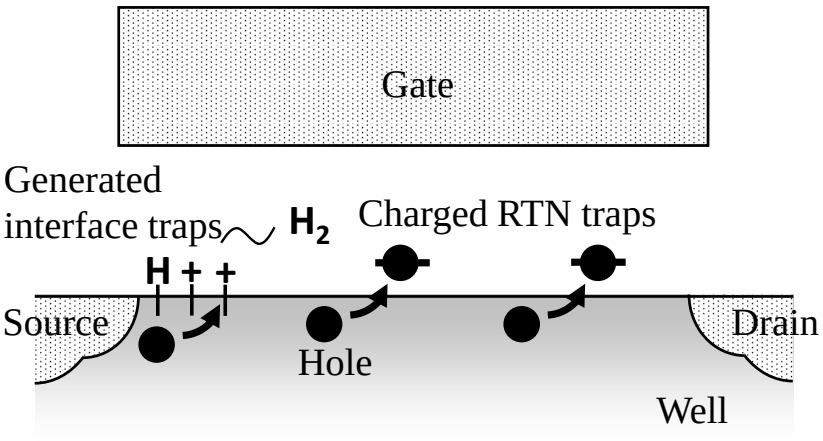
A cumulative distribution of the V_{th} variation due to RTN on the Gumbel plot is shown in Fig. 7.5. In this figure, the V_{th} variation is defined as the total amplitude of extracted single RTNs. The median value (40 mV) of the plots shortly after stress is four times larger than that of the initial plots (10 mV). The data at 100 and 200 s after NBTI stress show similar distribution to the data at 30 s after NBTI stress. The reason that the distribution of ΔV_{th} shortly after stress shows a large shift is that the temporary RTN and one-time RTN probably create the recovery process. Interestingly, the 30-s-after-stress plot shows a similar distribution to the initial plot, and then, both the 100-s-after-stress plot and the 200-s-after-stress plot show almost the same distribution as the 30-s-after-stress plot because most temporary RTNs and the one-time RTNs disappear within 30 s in the same manner as the recovery component as shown in Fig. 7.3(b).

The number of extracted single RTN traps per device is summarized in a histogram in Fig. 7.6. For example, the two devices used for observing five RTNs, as shown in Figs. 7.4(a) and (b) are

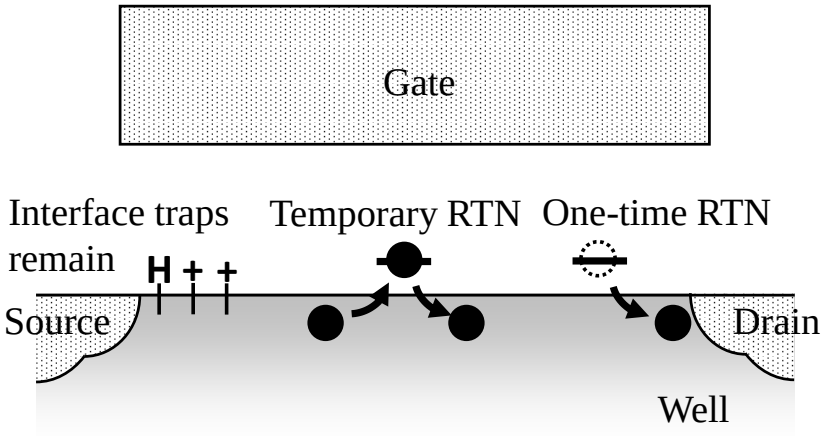
counted as the data shortly after stress at five RTN traps. The devices in which more than 6 RTNs are observed are counted as the data at +6 RTN traps, and in the calculation of the average number of RTN traps, they are counted as 6 RTN traps. The histograms roughly follow a Poisson distribution. The average number of RTN traps is initially 2.5. The average values shortly after NBTI stress (i.e., 30, 100, and 200 s) are 3.9, 2.7, 2.6, and 2.5, respectively. The shortly-after-stress histogram shows that its average is the largest because many temporary RTNs and one-time RTNs, which continue to exist in the recovery process, are counted. Moreover, the average number of traps approaches that of the initial state over time. However, it cannot return fully to that of the initial state; that is, the figure shows that the number of devices



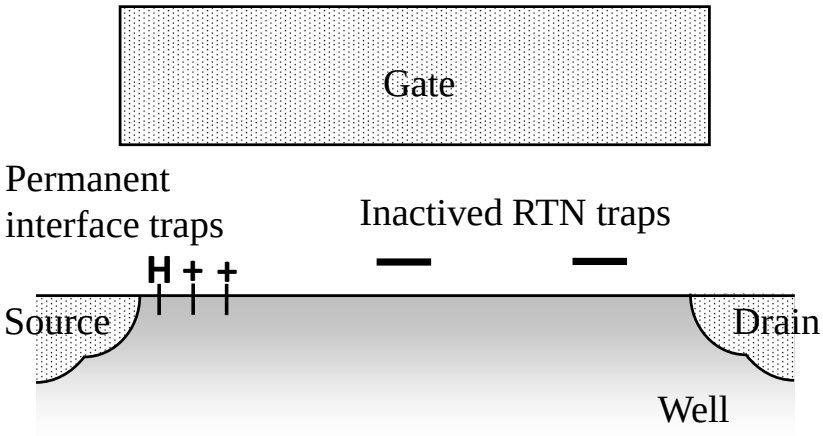
(a) Before application of NBTI stress.



(b) During NBTI-stress period.



(c) Shortly after application of NBTI stress.



(d) 30 s after application of NBTI stress.

Fig. 7.7 Two-type trap model.

without RTN traps after the NBTI stress is not completely equal to that of the initial state.

On the basis of the above results, the author can establish an advanced model to complement both the R-D model and the hole trapping / detrapping model as illustrated in Fig. 7.7. Before the application of NBTI stress, silicon bonds at the interface are passivated by hydrogen and native RTN traps remain inactive as illustrated in Fig. 7.7(a). During the steady NBTI-stress period, the interface traps are generated by breaking of hydrogen-passivated silicon bonds at the interface, and at the same time, the RTN traps become activated and are charged as shown in Fig. 7.7(b). Accordingly, the threshold voltage shifts drastically. Shortly after NBTI stress, the interface traps remain, and the activated RTN traps cause the temporary RTNs or the one-time RTNs, as shown in Fig. 7.7(c). Subsequently, the temporary RTNs and the one-time RTNs

rapidly disappear. At 30 s after NBTI stress, the interface traps cannot sufficiently recover, and most of them become permanent, as shown in Fig. 7.7(d). In contrast, almost all the temporary RTNs and the one-time RTNs disappear, and again, most of them become inactivated. As a result, only the threshold-voltage shift due to the temporary RTNs and the one-time RTNs recovers.

The combination of two types of trap with different characteristics, namely, the interface trap and RTN trap, cause curious NBTI degradation and recovery in pMOSFETs.

7.5 Conclusions

Statistical analysis of RTN behaviors before and after the NBTI stress reveals that the RTN traps and the normal interface traps cause curious NBTI degradation and recovery. Many permanent interface traps are generated by the NBTI stress, and at the same time, temporary and one-time RTNs are also generated. In the recovery process, the re-passivation of the interface states is the minor cause of the recovery, and in contrast, the rapid disappearance of the temporary RTN and the one-time RTN is the main cause of the recovery.

From the statistical analysis of RTN V_{th} variation before and after the application of NBTI stress, the distribution of RTN V_{th} variation shortly after NBTI stress shows a large shift in comparison with the initial distribution. Subsequently, 30-s-after-stress data shows a similar distribution to the initial data because almost all temporary RTNs and one-time RTNs disappear within 30 s.

From our statistical analysis, the author proposed an advanced model, which explains the NBTI degradation and recovery by the two types of traps with different characteristics, namely, the interface trap and RTN trap. It can complement both the R-D model and the hole trapping / detrapping model and will be useful for the understanding and improvement of the NBTI degradation in the near future.

7.6 References

[1] B. E. Deal, M. Sklar, A. S. Grove, and E. H. Snow, “Characteristics of the Surface-State Charge (Qss) of Thermally Oxidized Silicon,” *J. Electrochem. Soc.*, vol. 114, no. 3, Mar., pp. 266–274, 1967.

[2] M. A. Alam, and S. Mahapatra “A comprehensive model of PMOS NBTI degradation,” *Microelectron. Reliab.*, vol. 45 no. 1, Jan., pp. 71–81, 2005.

[3] C. Ma, H. J. Mattausch, M. Miyake, K. Matsuzawa, T. Iizuka, S. Yamaguchi, T. Hoshida, A. Kinoshita, T. Arakawa, J. He, and M. Miura-Mattausch “Unified Reaction--Diffusion Model for Accurate Prediction of Negative Bias Temperature Instability Effect,” *Jpn. J. Appl. Phys*. vol.51, no.2, (2012) 02BC07, 2012.

[4] T. Grasser, H. Reisinger, P.-J. Wagner, F. Schanovsky, W. Goes, and B. Kaczer “The Time Dependent Defect Spectroscopy (TDDS) for the Characterization of the Bias Temperature Instability,” in *Proc. IRPS*, pp. 16–25, 2010.

[5] T. Grasser, “Stochastic charge trapping in oxides: From random telegraph noise to bias temperature instabilities,” *Microelectron. Reliab.*, vol. 52, pp. 39–70, Oct., 2011.

[6] T. Grasser, B. Kaczer, W. Goes, H. Reisinger, T. Aichinger, P. Hehenberger, P.-J. Wagner, F. Schanovsky, J. Franco, M. T. Luque, and M. Nelhiebel, “The paradigm shift in understanding the bias temperature instability: from reaction-diffusion to switching oxide traps,” *IEEE Trans. Electron Devices*, vol. 58, no. 11, pp. 3652–3666, Nov., 2011.

[7] V. Huard, C. Parthasarathy, N. Rallet, C. Guerin, M. Mammasel, D. Barge, and C. Ouvrard, “New characterization and modeling approach for NBTI degradation from transistor to product level,” in *IEDM Tech. Dig.*, pp. 797–800, 2007.

[8] S. Mahapatra, N. Goel, S. Desai, S. Gupta, B. Jose, S. Mukhopadhyay, K. Joshi, A. Jain, A. E. Islam, and M. A. Alam “A Comparative Study of Different Physics-Based NBTI Models,” *IEEE Trans. Electron Devices*, vol. 60, no. 3, pp. 901–916 Mar. 2013.

[9] K. Kushida-Abdelghafar, K. Watanabe, J. Ushio, and E. Murakami “Effect of nitrogen at SiO2 /Si interface on reliability issues—negative-bias-temperature instability and Fowler–Nordheim-stress degradation,” *Appl. Phys. Lett.*, vol. 81, no. 23, pp. 4362–4364, Dec., 2002.

[10] S. Fujieda, Y. Miura, M. Saitoh, E. Hasegawa, S. Koyama, and K. Ando “Interface defects responsible for negative-bias temperature instability in plasma-nitrided SiON/Si(100) systems,” *Appl. Phys. Lett.*, vol. 82, no. 21, pp. 3677–3679, Mar., 2003.

[11] J. P. Campbell, P. M. Lenahan, A. T. Krishnan, and S. Krishnan “Observations of NBTI-Induced Atomic-Scale Defects,” *IEEE Trans. Device Mater. Rel.*, vol. 6, no. 2, pp. 117–122, Jun., 2006.

[12] S. Mahapatra, D. Saha, D. Varghese, and P. B. Kumar “On the generation and recovery of interface traps in MOSFETs subjected to NBTI, FN, and HCI stress,” *IEEE Trans. Electron Devices*, vol. 53, no. 7, pp. 1583–1592, Jul., 2006.

[13] D. J. Frank “Random Telegraph Noise – Measurement, Analysis, and Consequences,” in *Tutorial IRPS*, 2012.

[14] H. Miki, N. Tega, M. Yamaoka D. J. Frank, A. Bansal, M. Kobayashi, K. Cheng, C. P. D’Emic, Z. Ren, S. Wu, J-B. Yau, Y. Zhu, M. A. Guillorn, D.-G. Park, W. Haensch, E. Leobandung, and K. Torii “Statistical measurement of random telegraph noise and its impact in scaled-down high-κ/metal-gate MOSFETs” in *IEDM Tech. Dig.*, pp. 19.1.1. –19.1.4., 2012.

[15] M. Denais, A. Bravaix, V. Huard, C. Parthasarathy, G. Ribes, F. Perrie, Y. Rey-Tauriac, and N.Revil “On-the-fly characterization of NBTI in ultra-thin gate oxide PMOSFET's,” in *IEDM Tech. Dig*., pp. 109–112, 2004.

[16] J. H. Stathis, and S. Zafar “The negative bias temperature instability in MOS devices: A review,” *Microelectron. Reliab.*, vol. 46, pp. 270–286, Sep., 2006.

[17] H. Miki, M. Yamaoka, N. Tega, Z. Ren, M. Kobayashi, C. P. D’Emic, Y. Zhu, D. J. Frank, M. A. Guillorn, D.-G. Park, W. Haensch, and K. Torii “Understanding short-term BTI behavior through comprehensive observation of gate-voltage dependence of RTN in highly scaled high-κ / metal-gate pFETs,” in *Proc. VLSI Tech. Dig.*, pp. 148–149, 2011.

[18] M. A. Alam “A critical examination of the mechanics of dynamic NBTI for PMOSFETs,” in *IEDM Tech. Dig.*, p. 345–349, 2003.

8. Physical model of RTN in MOSFET

8.1 Introduction

From chapter 2 to 7, the author discussed the impact of RTN on scaled MOSFETs. However, not only the impact of RTN on MOSFET, but also the physics of RTN in MOSFET remains incompletely understood. In this chapter, the author proposes the physical model of RTN which can explain RTN behaviors such as wide distributions of activation energy and a bias sensitivity.

Some reports discussed a specific location of the trap causing the RTN behavior at gate dielectric and its level in bandgap [1−5]. Kirton and Uren proposed their model introduced by both thermodynamic approach and the Shockley-Read-Hall statistics to account for RTN behaviors characterized by change in state of drain current. According to the Kirton-Uren model, a ratio of average duration times to capture and emission ($\bar{\tau}_e/\bar{\tau}_c$) is expressed by

$$\frac{\bar{\tau}_e}{\bar{\tau}_c} = \frac{P_c}{P_e} = g^{-1}\exp\left(-\frac{E_T - E_F}{kT}\right) = g^{-1}\exp\left(-\frac{E_a}{kT}\right) \quad (8.1)$$

where P_c is the probability of capture, P_e is the probability of emission, T is temperature, k is the Boltzmann factor, E_T is the trap level in the gate dielectric, E_F is the Fermi level of bulk Si, E_a is the activation energy of RTN, and g is the degeneracy factor. In the event of single RTN, g is equal to one. Figure 8.1 is the energy band diagram to explain the Kirton-Uren model. A device is nMOSFET. The activation energy for $\bar{\tau}_e/\bar{\tau}_c$ is caused by difference between the trap level in the gate dielectric and the Fermi level of bulk Si. Moreover, Kirton and Uren introduced the non-radiative multi-phonon (NMP) model to explain extremely large energies for capturing electron. Recently, T. Grasser calculated specific barrier of NMP model [6]. On the other hand, it is difficult to explain a continuously broadened activation energies in a range of plus or minus hundreds meV.

Some reports considered an effect of gate-voltage modification on the trap level in the band gap. Specific trap depths are estimated using modification of Eq. (8.1) as illustrated in Fig. 8.2. The difference between E_1 and E_2 is equivalent to E_T–E_F. Dependence of $\bar{\tau}_e/\bar{\tau}_c$ on bias is shown below [1−3, 7].

$$\frac{\bar{\tau}_e}{\bar{\tau}_c} = \exp\left(\frac{q\alpha V_{ox} + C_1}{kT}\right) \quad (8.2)$$

$$V_{ox} = V_g - V_{fb} - \psi_s$$

$$\alpha = \frac{z_T}{T_{ox}}$$

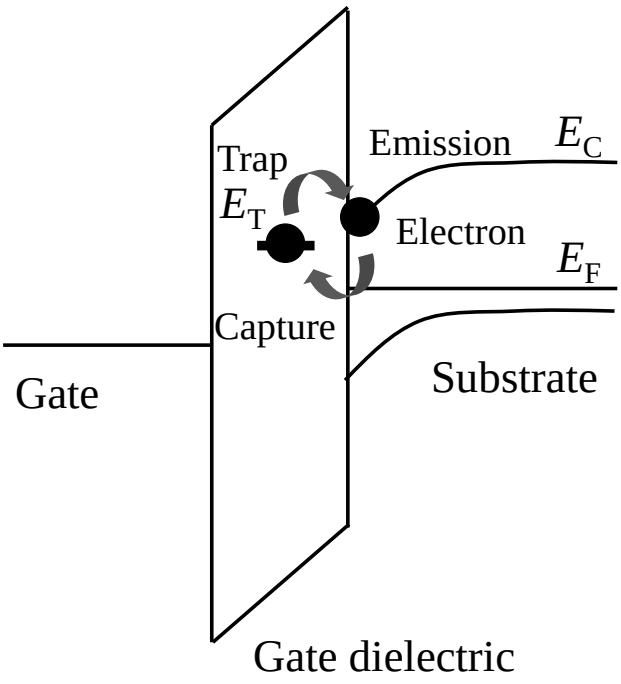


Fig. 8.1. Energy band diagram of nMOSFET taking into account the trap level and the Fermi level to estimate a balance between the probability of capture and the probability of emission.

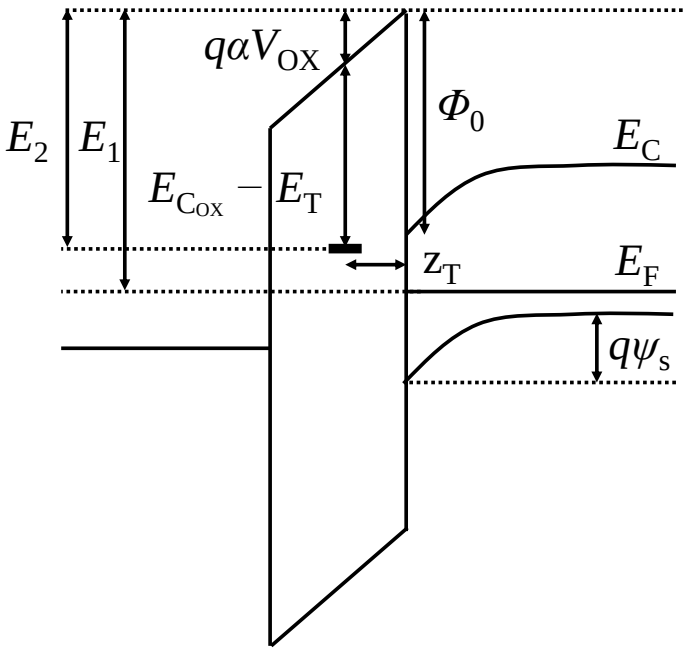


Fig. 8.2. Energy band diagram of nMOSFET showing a relationship between the trap level and the trap depth.

$$C_1 = \left(E_{C_{OX}} - E_T\right) - (E_C - E_F) - \Phi_0 + q\psi_s$$

In these equations, q is the elementary electric charge, α is gate-bias sensitivities, V_{ox} is the oxide voltage drop, V_g is the gate voltage, V_{fb} is the flat band voltage. ψ_s is the surface potential, z_T is the trap depth, T_{ox} is the thickness of the gate dielectric, $E_{C_{OX}}$ is the conduction band edge of gate dielectric, E_C is the conduction band edge of silicon, Φ_0 is the difference between the electron affinities of gate and gate dielectric, and C_1 is a constant. Equation (8.2) is introduced by considering both the band bending because of an electrostatic energy (qV_{ox}) and the geometric position of the trap at the energy band diagram. However, if the relationship between the gate-bias sensitivity and the trap location is fully explained by this equation, it is required to presume some unexpectedly deeper traps from the channel to explain gate-bias sensitivities. Moreover, the author demonstrated that the HTFGA process can passivate RTN traps in chapter 4 and the RTN V_{th} variations strongly depend on the silicon-surface orientations in chapter 5. The results mean that RTN traps are likely to be located near interface between gate dielectric and substrate. It is consistent with my results to identify the RTN traps as deep traps in gate dielectric.

Some researchers took into consideration an electrostatic energy induced by capacitances of trap at gate oxide, specifically, the coulomb-blockade effect to elucidate both the large activation energies and the large gate-bias sensitivities [8−11]. However, the application of this effect to the RTN model has never reached a consensus in the same way that the Kirton-Uren model has. In particular, the coulomb blockade model has not been verified through the use of a wide variety of data.

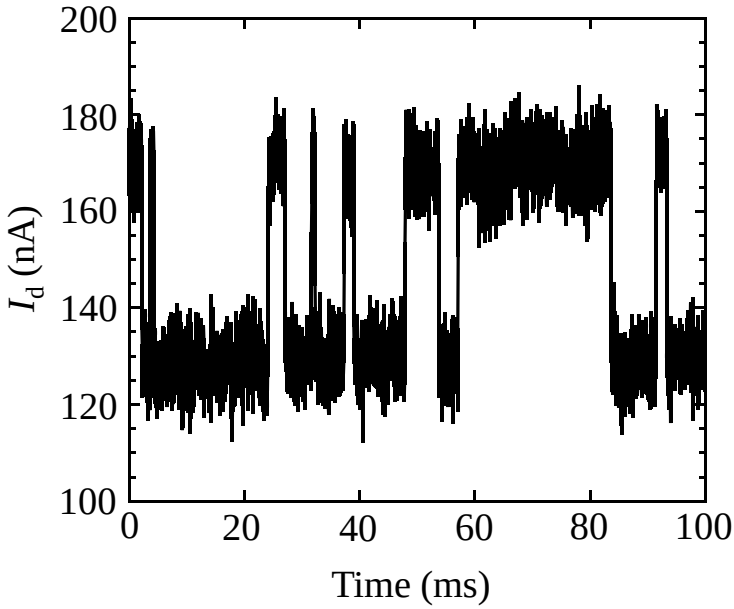
In this chapter, to explain both the strong bias sensitivity and the large activation energies, the author extends the Kirton-Uren model by introducing the single-electron-transfer process which is based on the thermal dynamics. The extended model provides more detailed information concerning RTN defect. Furthermore, the author introduces an extended NMP model as a possible solution to the problem about the larger gate bias sensitivities than expected. The author demonstrates an effectiveness of the extended NMP models by explaining the variety of the RTN behaviors, especially the large bias sensitivities and the wide distributions of the energies.

8.2 Experimental

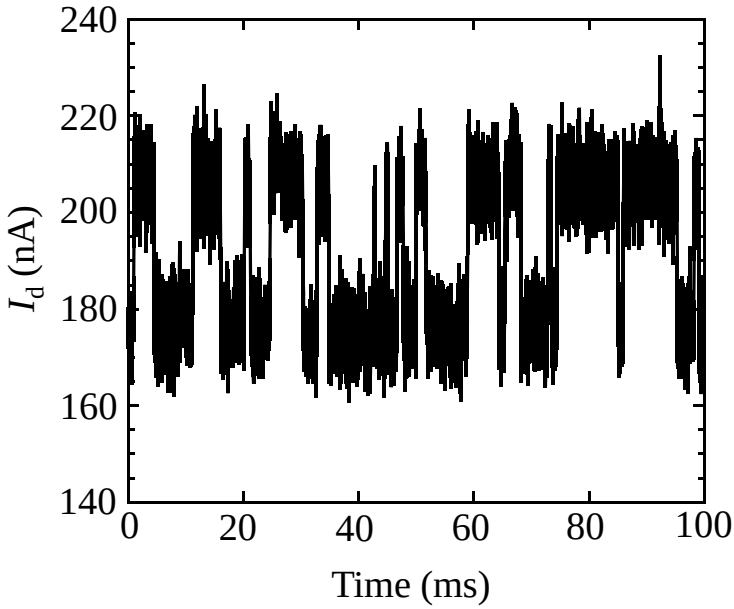
The devices are the same in chapter 4. The device is nMOSFETs with 25-nm gate length and 45-nm width. And these small devices incorporate HfO_2-based high-κ / metal-gate stacks because HK / MG stacks are essential for small devices to obtain sufficient on current. However, this work will not focus on the characteristics of HK / MG device but universally discuss data in

order to make the appropriate RTN physical model.

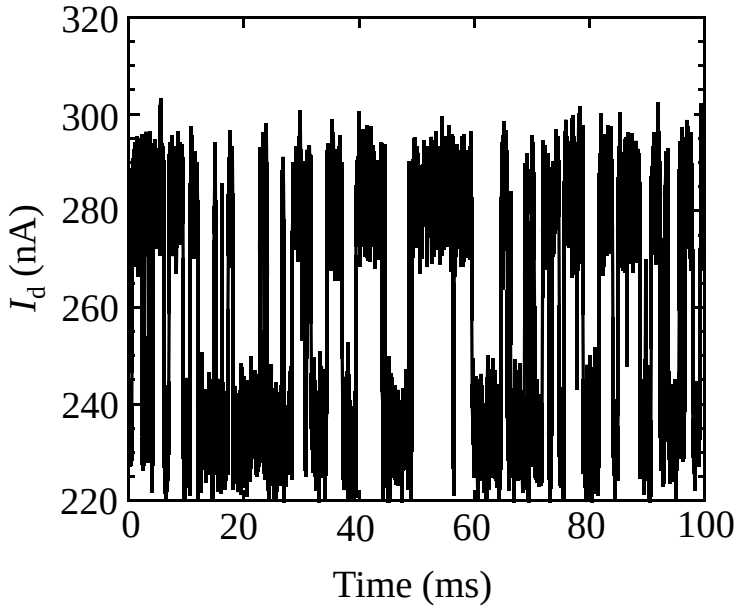
A fast measurement unit (Agilent 1530A) which enables a wide band width of up to 1 MHz was used to measure RTN signals. 10^5 sampling points were used for each time series of I_d. A sampling interval was 1 μs. All data are I_d fluctuations at a drain voltage (V_d) of 50 mV.



(a) Time series of I_d at T = 25 ºC. $\bar{\tau}_e$, $\bar{\tau}_c$, and, $\bar{\tau}_e/\bar{\tau}_c$ are 1.2×10^{-3} s, 9.0×10^{-4} s, and 1.4, respectively.



(b) Time series of I_d at T = 40 ºC. $\bar{\tau}_e$, $\bar{\tau}_c$, and $\bar{\tau}_e/\bar{\tau}_c$ are 4.2×10^{-4} s, 3.8×10^{-4} s, and 1.1, respectively.

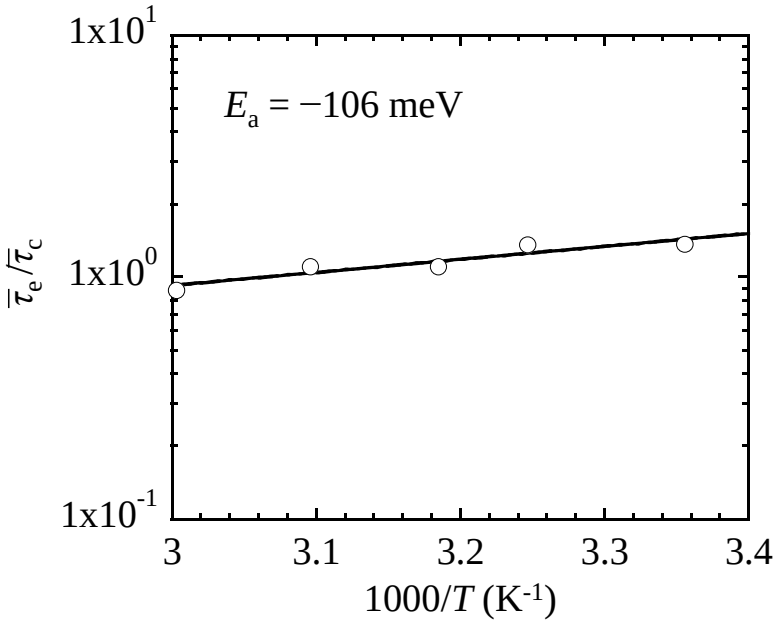


(c) Time series of I_d at T = 60 ºC. $\bar{\tau}_e$, $\bar{\tau}_c$, and $\bar{\tau}_e/\bar{\tau}_c$ are 1.3×10^{-4} s, 1.5×10^{-4} s, and 8.9×10^{-1}, respectively.

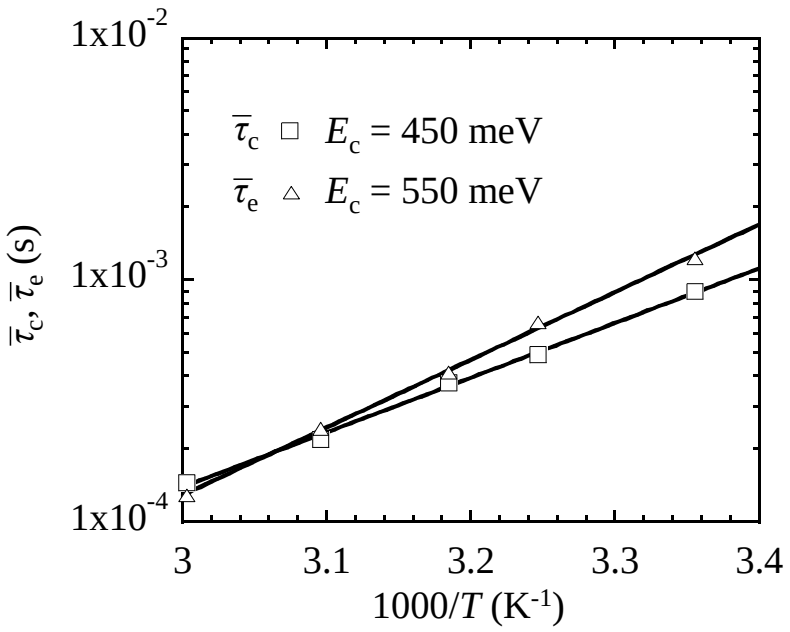
Fig. 8.3. Demonstration of dependence of RTN on measurement temperature. V_d is 50 mV. Sample is nMOSFET with 25-nm gate length and 45-nm gate width. It incorporates HfO_2-based high-κ / metal-gate stacks.

In this work, the author measured gate-voltage dependences of RTN at each temperature. Measurement temperatures were from 15 to 65 ºC at an interval of 5 ºC. Figure 8.3 is a demonstration of the temperature dependence of RTN. Figure 8.3(a) shows time series of I_d at T = 25 ºC. Apparently $\bar{\tau}_c$ is shorter than $\bar{\tau}_e$ in the figure. $\bar{\tau}_e$ is 1.2×10^{-3} s, $\bar{\tau}_c$ is 9.0×10^{-4} s, and $\bar{\tau}_e/\bar{\tau}_c$ is 1.4. These values change with the temperature. Figure 8.3(b) is time series of I_d at T = 40 ºC. $\bar{\tau}_e$ is nearly equal to $\bar{\tau}_c$. $\bar{\tau}_e$, $\bar{\tau}_c$, and $\bar{\tau}_e/\bar{\tau}_c$ are 4.2×10^{-4} s, 3.8×10^{-4} s, and 1.1, respectively. Furthermore, the gap between $\bar{\tau}_e$ and $\bar{\tau}_c$ expands with increase of the temperature. Figure 8.3(c) is time series of I_d at T = 60 ºC. $\bar{\tau}_e$ is 1.3×10^{-4} s, $\bar{\tau}_c$ is 1.5×10^{-4} s, and $\bar{\tau}_e/\bar{\tau}_c$ is 8.9×10^{-1}. Figure 8.4(a) is the Arrhenius plot of $\bar{\tau}_e/\bar{\tau}_c$. The activation energy for $\bar{\tau}_e/\bar{\tau}_c$ is −106 meV. As previously noted, the activation energies are broadened in a range of plus or minus hundreds meV. It is difficult to allege that the larger activation energies than imaged consist only of the difference between the trap level and the Fermi level. Furthermore, both $\bar{\tau}_e$ and $\bar{\tau}_c$ show the extremely large activation energy as show in Fig. 8.4(b). For example, $\bar{\tau}_e$ becomes enormously shorter from 1.2×10^{-3} s to 1.3×10^{-4} s by increase in temperature of 35 ºC as demonstrated in Fig. 8.3.

In addition to the temperature dependence of RTN, the author measured time series of I_d near V_{th} under the constant gate-voltage step of 10 mV to thoroughly investigate the



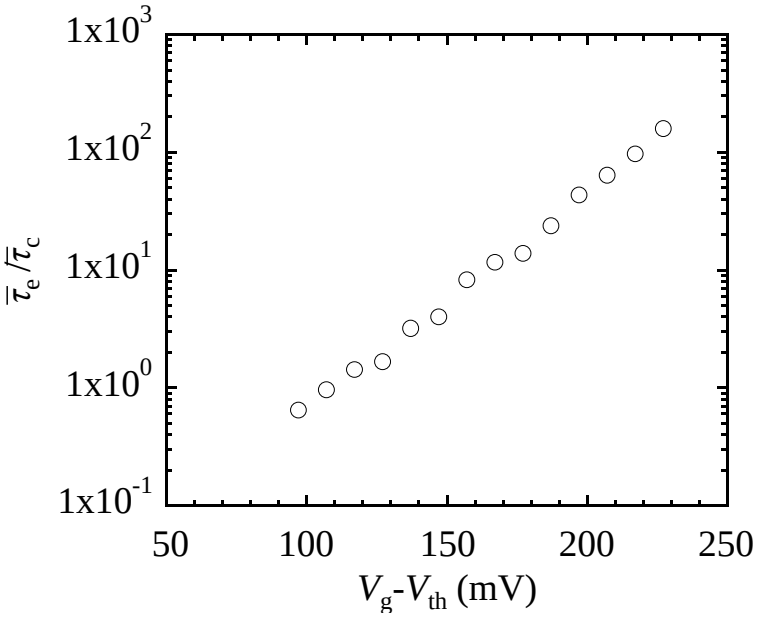
(a)



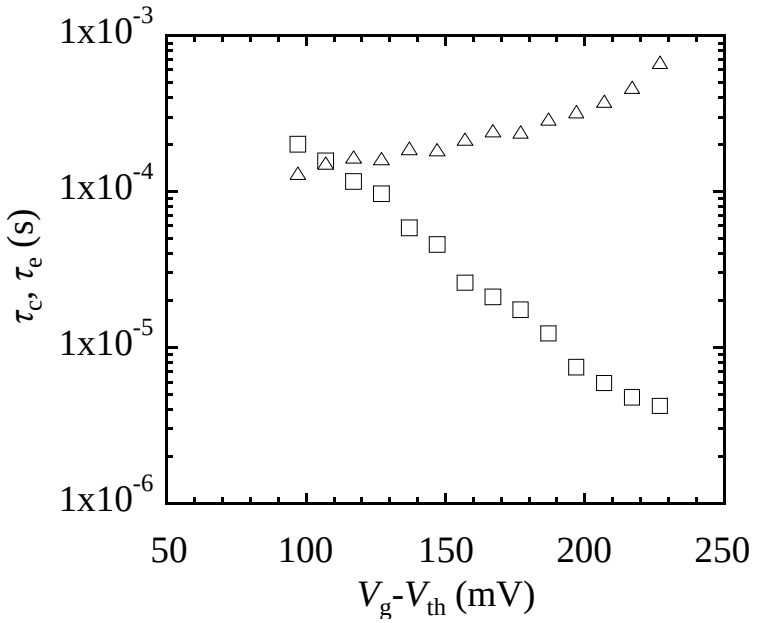
(b)

Fig. 8.4. Dependences of RTN behavior on measurement temperature. (a) Temperature dependence of $\bar{\tau}_e/\bar{\tau}_c$. (b) Temperature dependences of $\bar{\tau}_e$ and $\bar{\tau}_c$.

dependence of RTN on gate voltage. RTN frequently shows the strong V_g dependence. Figure 8.5(a) is a clear V_g dependence of $\bar{\tau}_e/\bar{\tau}_c$ which is based on the time series data as a function of overdrive voltage. The gate-bias sensitivity is extremely large. This large gate-bias sensitivity is caused by both large bias dependences of $\bar{\tau}_c$ and $\bar{\tau}_e$ as shown in Fig. 8.5(b).



(a)



(b)

Fig. 8.5. Dependences of RTN behavior on gate voltage. (a) V_g dependence of $\bar{\tau}_e/\bar{\tau}_c$. (b) V_g dependences of $\bar{\tau}_e$ and $\bar{\tau}_c$.

8.3 Modeling RTN in MOSFET

It was reported by M. J. Kirton and M. J. Uren that equation (8.1) was developed from the thermodynamics. In specific, they used the grand partition function to express the ratio of the capture and emission states of the RTN defect [1]. The author assents their approach. However,

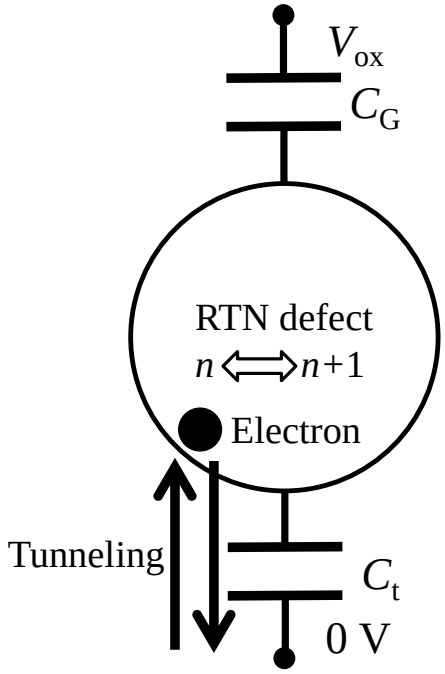


Fig. 8.6. System of RTN defect connecting with environments by two junctions.

before the thermodynamic approach, this work more clarifies the energy change when single RTN defect at gate dielectric captures and emits single electron in order to discuss the RTN physical model in more detail. Specifically, the author assumes a system of the RTN defect connecting with environments by two junctions as illustrated in Fig. 8.6. The first junction is a capacitor between the RTN defect and the gate (C_g). And the second one is a tunnel junction between the RTN defect and the channel (C_t). The other capacitors between the RTN defect and drain, source, or adjoin defects, etc. are ignored simply because they have a little influence in comparison with the gate capacitor. The important assumption is that electron can transfer only through the second junction between the RTN defect and the channel. This concept is based on the single-electron-tunneling theory and the single-electron memory [12–14]. Basically, the assumed capacitance areas are limited to areas surrounding the RTN defect like the capture cross section. Their radiuses do not extend beyond nanometer order as discussed later.

The charging energy at the RTN defect ($Q(n)$) is indicated as Eq. (8.3) when the number of electron is n. And equation (8.4) shows work (W) done by the gate voltage to make the number of electron be from zero to n.

$$Q(n) = \frac{C_g C_t V_{ox}^2 + (nq)^2}{2C_{total}} \quad (8.3)$$

$$C_{total} = C_g + C_t$$

$$W(n) = Q_g V_{ox} = \frac{C_g(C_t V_{ox} + nq)V_{ox}}{C_{total}} \quad (8.4)$$

Q_g is the polarization charge on the gate capacitor. The free energy of the system with n electrons ($F(n)$) is expressed as

$$\begin{aligned} F(n) &= Q(n) - W(n) \\ &= \frac{-C_g C_t V_{ox}{}^2 - 2nqC_g V_{ox} + (nq)^2}{2C_{total}}. \quad (8.5) \end{aligned}$$

The free energy change that is a transition of the number of electron from n to n+1 ($E(V_{ox})$) is shown as Eq. 8.6.

$$\begin{aligned} E(V_{ox}) &= F(n+1) - F(n) \\ &= \frac{(2n+1)q^2 - 2qC_g V_{ox}}{2C_{total}} \\ &= E_0 - \beta q V_{ox} \quad (8.6) \end{aligned}$$

In Eq. (8.6), E_0 is the initial energy when V_{ox} is zero, and β is gate-bias sensitivity.

The author thermodynamically characterizes the RTN behavior in consideration of $E(V_{ox})$ and E_T as references [12, 15]. According to the grand canonical ensemble, a balance between the probability of capture and the probability of emission is properly expressed by

$$\begin{aligned} \frac{\bar{\tau}_e}{\bar{\tau}_c} &= \frac{\gamma_c P_c}{\gamma_e P_e} = \frac{\gamma_c}{\gamma_e}\exp\left[-\frac{E(V_{ox}) + E_T - E_F}{kT}\right] \\ &= \frac{\gamma_c}{\gamma_e}\exp\left(-\frac{E_a}{kT}\right) \quad (8.7) \\ &= \frac{\gamma_c}{\gamma_e}\exp\left(\frac{q\beta V_{ox} - E_{a0}}{kT}\right) \quad (8.8) \end{aligned}$$

where γ_c and γ_e correspond to frequencies of attempt to escape from each state. For example, in the Shockley-Read-Hall theory, γ_c includes a charge density in the inversion layer, the average thermal velocity, and the capture cross section [16, 17]. The spin degeneracy is not considered because it has an insignificant influence on my discussion in this work. E_{a0} is the

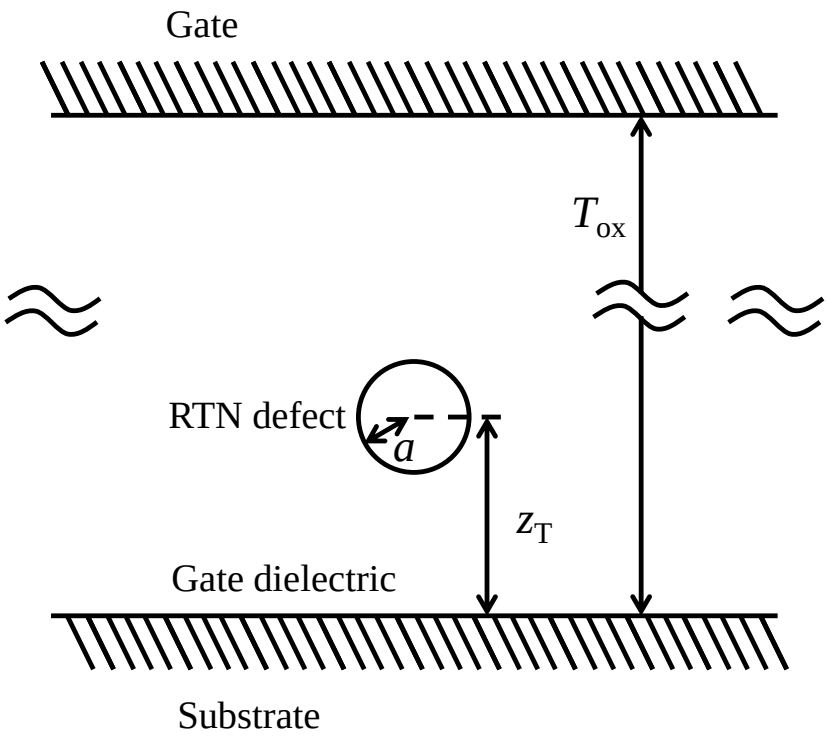


Fig. 8.7 Relationship between a conducting sphere (RTN defect) and borderless parallel plates (substrate and gate).

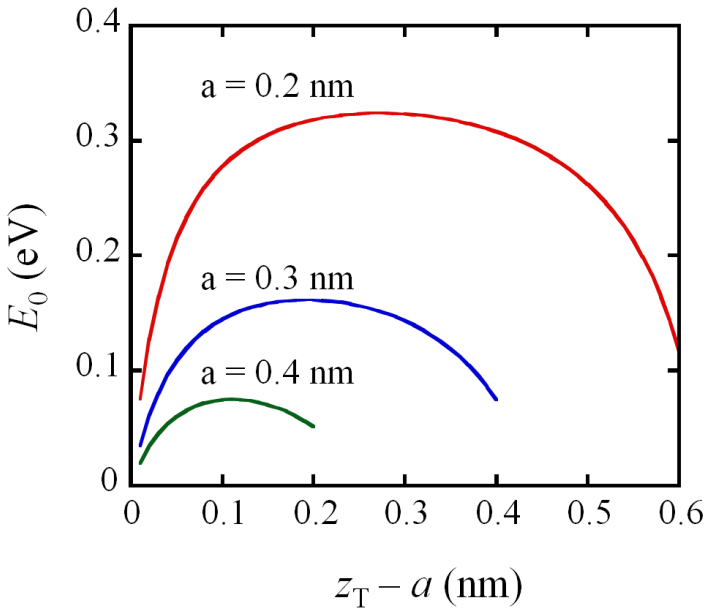
initial activation energy without bias effect.

In Eq. (8.8), β is determined not by a centroid of trapped space charges but by the contribution of the single RTN trap. C_g and C_t, which are required to estimate the gate-bias sensitivity, can be roughly calculated by a relationship between a conducting sphere and parallel plates. The RTN trap is used to resemble the conducting sphere as illustrated in Fig. 8.7. The channel and the gate are used as the parallel boundless plate. a is a radius of the RTN defect. According to the image, C_g and C_t are expressed as below.

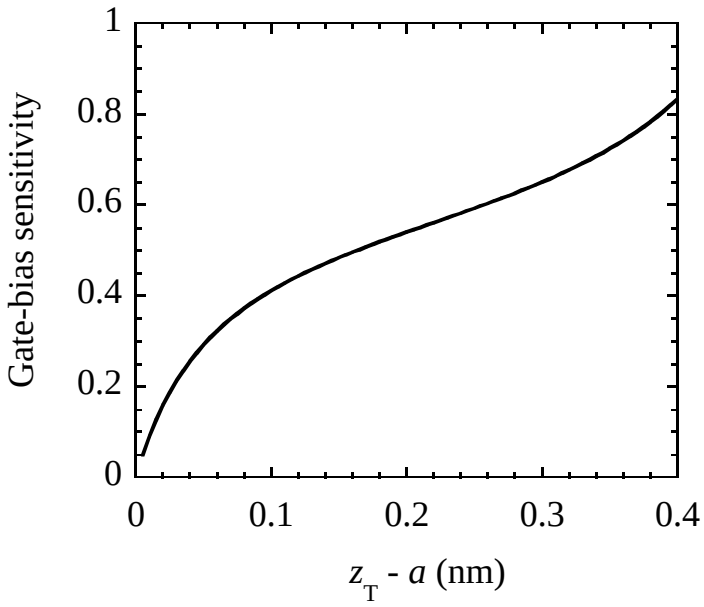
$$C_t = 4\pi\varepsilon_{SiO_2}\varepsilon_0\left(\frac{1}{a} - \frac{1}{2z_T - a}\right)^{-1} \quad (8.9)$$

$$C_g = 4\pi\varepsilon_{SiO_2}\varepsilon_0\left(\frac{1}{a} - \frac{1}{2(T_{ox} - z_T) - a}\right)^{-1} \quad (8.10)$$

ε_{SiO_2} is the relative permittivity of the silicon dioxide. ε_0 is the vacuum permittivity. The gate dielectric simply consists of only SiO_2 in Fig. 8.7. In fact, the gate stacks consist of an interfacial layer, the high-κ dielectric, and the metal gate in scaled devices. As an example,



(a)



(b)

Fig. 8.8. RTN defect characteristics calculated by relationship between a conducting sphere (RTN defect) and parallel plates (gate and substrate). (a) Relationship between initial energy and location of RTN defect. (b) Relationship between gate-bias sensitivity and location of RTN defect under the assumption that a is 0.3 nm.

figure 8.8(a) shows the dependences of the initial energy (E_0) on the radius and the location of the RTN defect. Figure 8.8(b) shows the gate-bias sensitivity (β) under the assumption that a is 0.3 nm. The relative permittivity of the interfacial layer is 3.9, the relative permittivity of the

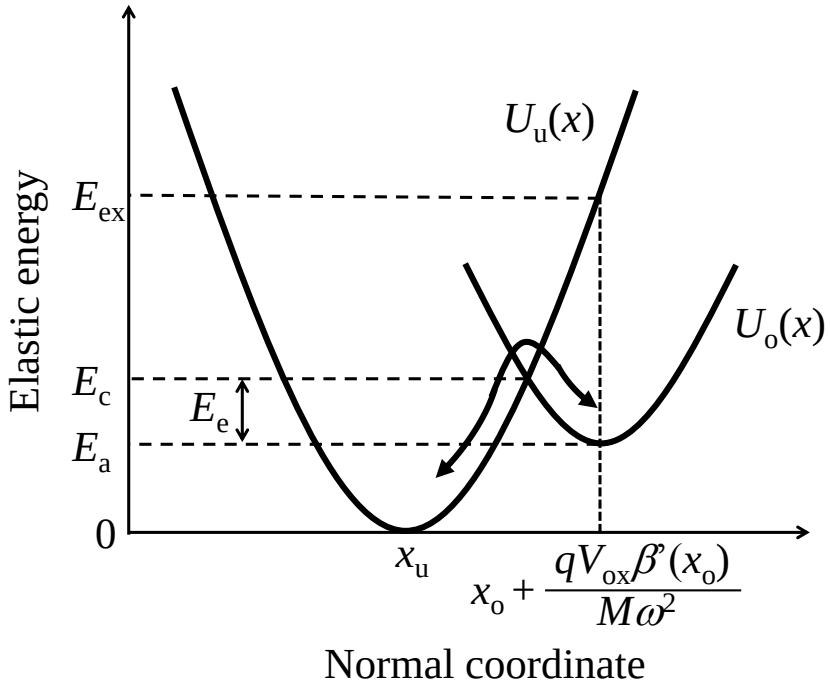


Fig. 8.9. Configuration coordinate diagram showing non-radiative multi-phonon model.

high-κ dielectric is 25, the interfacial-layer thickness is 1 nm, and the high-κ dielectric thickness is 3 nm. Equation (8.10) is changed to Eq. (8.11).

$$C_g = \left\{ \left(4\pi\varepsilon_{HfO_2}\varepsilon_0\right)^{-1} \left(\frac{1}{T_{ox} - z_T} - \frac{1}{T_{ox} - z_T + T_{HfO_2}} \right) + \left(4\pi\varepsilon_{SiO_2}\varepsilon_0\right)^{-1} \left[\frac{1}{a} - \frac{1}{2(T_{ox} + z_T) - a} - \frac{1}{T_{ox} - z_T} + \frac{1}{T_{ox} - z_T + T_{HfO_2}} \right] \right\}^{-1} \quad (8.11)$$

The radius of the RTN defect of 0.3 nm is a suitable value for determining the initial energies as I discuss below. The gate-bias sensitivity reaches 0.8 when the trap is located at 0.4 nm from channel. This value is all but identical with the radius of the capture cross section because this area, 2.8×10^{-15} m^2, is due to typical neutral trap within in the range of 1×10^{-18} to 1×10^{-15} m^2 [18]. From a physical point of view, if the RTN defect is an oxide vacancy in the interfacial layer, the radius nearly corresponds to a distance between neighboring oxides. The RTN defect could directly interfere with neighboring atoms within approximately 0.3 nm.

Next, energies for capturing (E_c) and emitting electron (E_e) are discussed to derive the equations of $\bar{\tau}_c$ and $\bar{\tau}_e$ in addition to $\bar{\tau}_e/\bar{\tau}_c$. According to the NMP theory, The RTN defect has two states. One is an unoccupied state with electron, and the other is an occupancy state with electron. In each state, the atomic equilibrium configuration is different. Moreover, the proposed

model considers the dependence of the gate-bias sensitivity (β) on position because this dependence crucially affects the activation energy in Eq. (8.8). The ideal is equivalent to a harmonic oscillator in an electric field. This is a possible solution to the problem about the larger gate bias sensitivity than expected. Figure 8.9 shows two harmonic oscillator potentials. One is the potential of the unoccupied state ($U_u(x)$). The other is the potential of the occupied state ($U_o(x)$). The barrier due to two harmonic oscillators must be overcome in order that the RTN behavior occurs. x is the reaction coordinate. An equilibrium position of $U_u(x)$ is defined as x_u, and a minimum energy of $U_u(x_u)$ is also zero. $U_u(x)$ and $U_o(x)$ are described below.

$$U_u(x) = \frac{1}{2}M\omega^2(x - x_u)^2 \quad (8.12)$$

$$U_o(x) = \frac{1}{2}M\omega^2(x - x_o)^2 + E_{a0} - q\beta V_{ox} \quad (8.13)$$

M is the effective mass. x_o is the equilibrium position of the occupied state. ω is the angular frequency. A different energy between $U_u(x_u)$ and $U_o(x_o)$ is E_a. Meanwhile the state changes from the unoccupied state to the occupied state, β is not always constant. β changes corresponding to the charge centroid of the RTN defect. As a highly simplified example, if the change in the charge centroid of the RTN defect with the radius of about 3 nm is 0.02 nm in a parallel direction from gate to channel, the change in β could reach up to 0.1 as show in Fig. 8.8(a). This value is considerable large. The author supposes that the potential of the occupied state is affected by the change in β because it can reflect the position shift of the RTN defect even if the shift does not exceed 0.1 nm.

In this work, simply, the variation of β is limited near x_o. A first-order Taylor expansion approximates $\beta(x)$ on the neighborhood of x_o as below.

$$\beta(x) = \beta(x_o) + \beta'(x_o)(x - x_o) \quad (8.14)$$

From Eqs. (8.13) and (8.14), $U_o(x)$ can be described in greater detail.

$$U_o(x) = \frac{1}{2}M\omega^2(x - x'_o)^2 + E_a \quad (8.15)$$

$$x'_o = x_o + \frac{qV_{ox}\beta'(x_o)}{M\omega^2} \quad (8.16)$$

$$E_a = E_{a0} - q\beta(x_0)V_{ox} - \frac{q^2\beta'(x_o)^2 V_{ox}{}^2}{2M\omega^2} \quad (8.17)$$

To be exact, the vertex of the parabola is affected by V_{ox}. The activation energy has a second-order term as shown in Eq. (8.17). The introduction of the second-order term corrects the larger gate bias sensitivities than expected.

In addition to the accurate calculation of E_a, both E_c and E_e can be calculated by taking into account Fig. 8.9.

$$E_c = \frac{(E_{ex} + E_a)^2}{4E_{ex}} \quad (8.18)$$

$$E_e = \frac{(E_{ex} - E_a)^2}{4E_{ex}} \quad (8.19)$$

$$E_{ex} = U_u(x'_o) = \frac{1}{2}M\omega^2(x'_o - x_u)^2 \quad (8.20)$$

E_{ex} is the excitation energy if a radiative transition can occur. Therefore, $\bar{\tau}_c$ and $\bar{\tau}_e$ are expressed as

$$\bar{\tau}_c = \frac{1}{\gamma_c P_t}\exp\left(\frac{E_c}{kT}\right) \quad (8.21)$$

$$\bar{\tau}_e = \frac{1}{\gamma_e P_t}\exp\left(\frac{E_e}{kT}\right) \quad (8.22)$$

where P_t is the tunneling probability. P_t is approximately developed by the Wentzel-Kramers-Brillouin approximation and the trapezoidal tunneling barrier.

$$k(z) = \sqrt{\frac{2m^*\left(E_{C_{ox}} - E_T - q\frac{z_T}{T_{ox}}V_{ox}\right)}{\hbar^2}}$$

$$P_t = \exp\left[-2\int_0^{z_T} k(z)\,\mathrm{d}z\right]\delta(E_c - E_{Tz})$$

$$= \exp\left\{-\frac{4\sqrt{2}T_{ox}m^{*\frac{1}{2}}}{3q\hbar V_{ox}}\left[-\left(E_{C_{ox}} - E_T - q\frac{z_T}{T_{ox}}V_{ox}\right)^{\frac{3}{2}} + \left(E_{C_{ox}} - E_T\right)^{\frac{3}{2}}\right]\right\}\delta(E_c - E_{Tz})$$

In the case that E_c is equal to E_{Tz}

$$P_t = \exp\left\{-\frac{4\sqrt{2m^*}T_{ox}\left(E_{C_{ox}} - E_T\right)^{\frac{3}{2}}}{3q\hbar V_{ox}}\left[1 - \left(1 - \frac{z_T q V_{ox}}{T_{ox}\left(E_{C_{ox}} - E_T\right)}\right)^{\frac{3}{2}}\right]\right\} \quad (8.23)$$

k is the wave vector, m^* is the electron effective mass, $\hbar$ is the Planck constant, and E_{Tz} is the trap energy when applying V_{ox}. Technically, the quantum effect of surface state in MOSFET is perhaps one of the items to be taken into account in calculating Eq. (8.23) [19]. This work, however, does not consider quantized levels for simplicity because the surface quantization has little effect on my results.

To properly explain the RTN behaviors, the author proposes Eqs. (8.17), (8.18), and (8.19) as the energies for ratio for $\bar{\tau}_e/\bar{\tau}_c$, $\bar{\tau}_c$ and $\bar{\tau}_e$. The RTN physical models are extended in consideration of the free energy change which is based on the single-electron tunneling theory and the NMP theory.

8.4 Analysis of electrical properties

In this section, first, the gate-bias sensitivities of $\bar{\tau}_e/\bar{\tau}_c$ are analyzed using Eq. (8.17). Then, the author demonstrates that the activation energies for $\bar{\tau}_e/\bar{\tau}_c$ broaden, and the introduction of the charging effect is required for understanding the RTN behavior. Finally, the energies for capturing and emitting electron are investigated from the aspect of the proposed NMP theory.

8.4.1 Gate-bias sensitivity of $\bar{\tau}_e/\bar{\tau}_c$

Figure 8.10 shows a cumulative probability of the gate-bias sensitivity (β) of $\bar{\tau}_e/\bar{\tau}_c$. The sensitivities were determined by fitting of quadratic curve with the expression of E_a in Eq. (8.17). For example, the sensitivity of $\bar{\tau}_e/\bar{\tau}_c$ is about 0.8 in the event of Fig. 8.5. This value is relatively small when compared to α ($\alpha = 1.1$ in this case) estimated by linear regression using Eq. (8.2). The estimated sensitivities widely spread as shown in Fig. 8.10. The median value of a normal distribution is about 0.6. A relatively large number of RTN defects are located

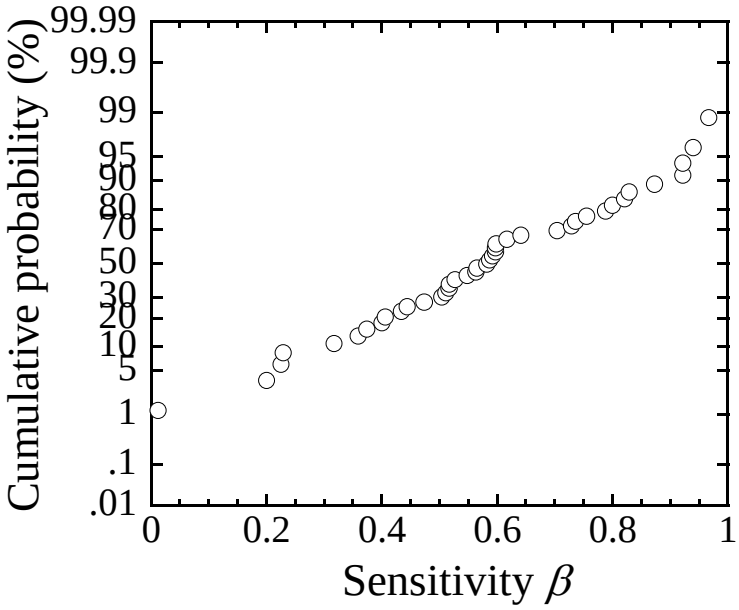


Fig. 8.10. Cumulative probability of gate-bias sensitivities calculated by Eq. (8.16).

around 0.3 nm from the channel to the edge of the defect under the assumption as shown in Fig. 8.8. This distance is equivalent to a side length of a SiO_4 tetrahedron. Hence, Oxide vacancies are likely to cause RTN behavior. The RTN trap is categorized into border trap, which can communicate with the channel [20]. It is, however, possible that the interface traps cause RTN because some RTN behaviors show little dependency on gate-bias.

The McWhorter model, the number fluctuation model, was proposed to understand an origin of $1/f$ noise in the germanium filaments [21]. The model states that carrier fluctuation due to charge trapping in surface causes $1/f$ noise. S. Christensson et al. were the first to apply the McWhorter theory to $1/f$ noise in MOSFETs [22]. They explained that $1/f$ noise is generated by a number of the RTN traps with different time constants, which are mainly determined by difference in the RTN trap depth, and possible trap depth is less than 2 nm. Indeed, the difference in the trap depths is likely to make a change in the RTN time constant as shown in Eqs. (8.21), (8.22), and (8.23). On the other hand, it is theoretically possible that the RTN time constants distributes not only because of a variety of the trap depths but also because of a variety of the RTN energies.

8.4.2 Activation of $\bar{\tau}_e/\bar{\tau}_c$

Figure 8.11 clarifies that the activation energies for $\bar{\tau}_e/\bar{\tau}_c$ broaden. The activation energies were estimated from the Arrhenius plot of $\ln(\bar{\tau}_e/\bar{\tau}_c)$ under the constant V_{th} because the author concerned about the influence of the difference in gate over drive on the activation energies for

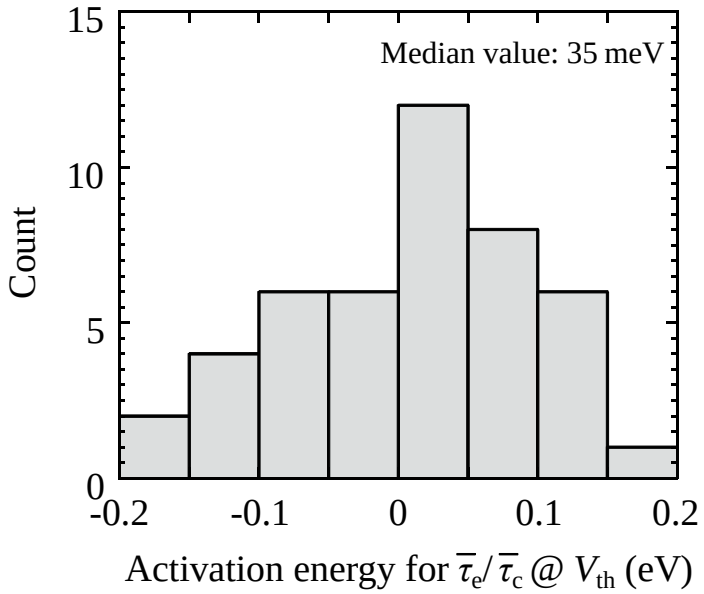


Fig. 8.11. Histogram of activation energies for $\overline{\tau}_e/\overline{\tau}_c$.

$\overline{\tau}_e/\overline{\tau}_c$. The activation energies widely spread from approximately 200 to −200 meV. Even if E_T is nearly equal to E_F, equation (8.7) indicates that the activation energies can become larger because the energies can is determined by the free energy change. $E(V_{ox})$ depends on the capacitances, the number of electrons occupying the defect (n), and the gate bias condition as indicated in Eq. (8.6).

The author cited an example as the large $E(V_{ox})$ when n is zero (neutral trap) as shown in Fig. 8.8(a). The maximum value is about 150 mV if the radius of RTN defect is 0.3 nm. This value is in good agreement with the observed data. In contrast, if the RTN defect has the radius of 0.4 nm, the maximum value is only about 75 mV. If the radius of the RTN defect is 0.2 nm, the maximum value is no fewer than about 320 mV. The author, therefore, supposes that the appropriate radius is about 0.3 nm. Likewise, coulomb-attractive traps ($n = -1$) give the distribution opposite in sign to Fig. 8.8(a). Coulomb-repulsive traps ($n = 1$) give larger activation energies than the neutral traps. The number of the coulomb-repulsive traps is smaller than either the neutral traps or the coulomb-attractive traps because enormously energies were not observed in this work. The total number of the neutral traps and the coulomb-repulsive traps, however, is likely larger than the coulomb-attractive traps because the positive shift (the median value is about 35 meV) of histogram is observed in Fig. 8.11.

8.4.3 Influence of non-radiative multi-phonon emission

If the influence of the non-radiative multi-phonon emission mainly determines the activation

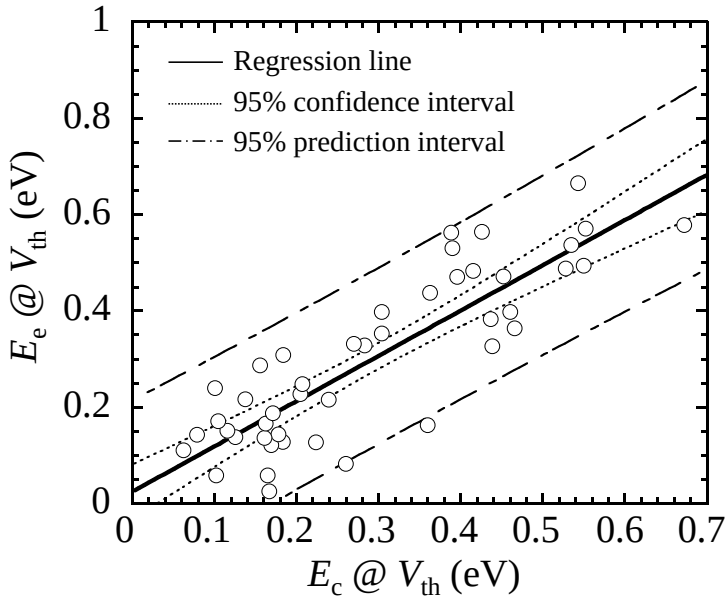


Fig. 8.12. Correlation between E_c and E_e.

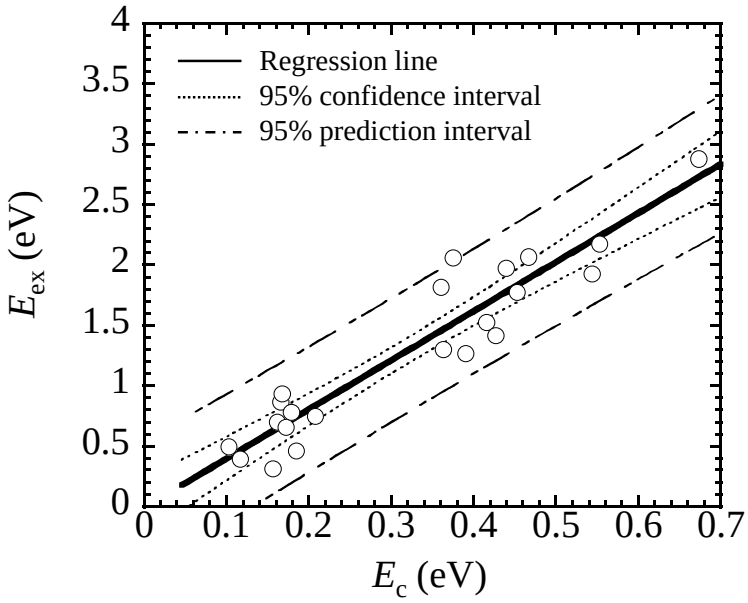


Fig. 8.13 Correlation between E_c and E_{ex}.

energies for capturing and emitting electron, E_c and E_e are sure to show a definite correlation on the basis of Eqs. (8.18) and (8.19). Figure 8.12 clarifiers the correlation between E_c and E_e. The multiple R-squared is 0.74 and the P-value based on the statistical hypothesis testing is less than 0.05. Consequently, the correlation between E_c and E_e are statistically warranted. The activation energies for capture and emission were estimated from the Arrhenius plots of $\ln(\bar{\tau}_c)$ and $\ln(\bar{\tau}_e)$

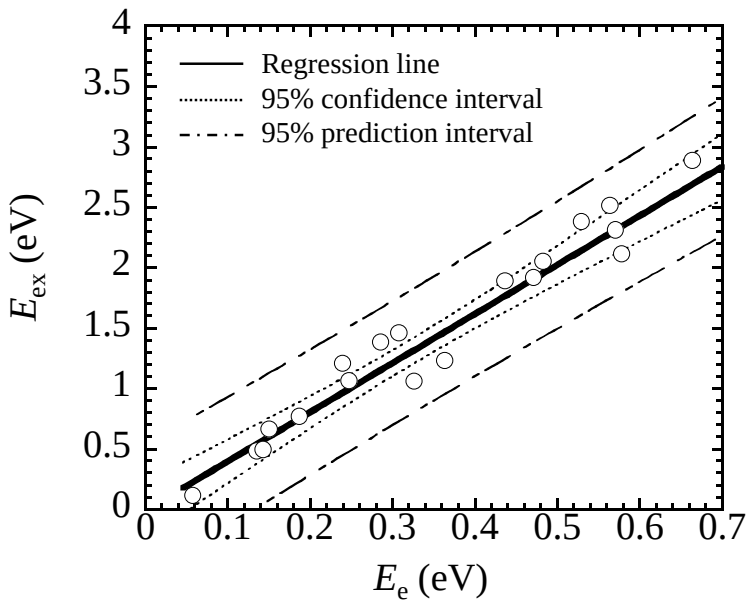


Fig. 8.14. Correlation between E_e and E_{ex}.

under the constant V_{th}. The solid line is a regression line. The dotted lines are a 95% confidence interval. The chain lines are a 95% prediction interval. The activation energies for capture and emission broaden within about 700 meV. The positive numbers are only observed in respect of the energies for capture and emission because they are the barrier caused by the intersection of the two harmonic oscillator potentials as illustrated in Fig. 8.9.

Since the energies for capture and emission widely spread from several tens meV to 700 meV and show the strong correlation, the author can expect that the excitation energies are extremely large enough to determine both the capture and the emission energies. The excitation energies were estimated by the activation energies for $\bar{\tau}_e/\bar{\tau}_c$ and the energies for capture and emission. Figure 8.13 shows the relationship between E_c and E_{ex}. The multiple R-squared is 0.89. P-value is less than 0.05. It means that E_c strongly depends on E_{ex}. Figure 8.14 shows the relationship between E_e and E_{ex}. The multiple R-squared is 0.96. P-value is less than 0.05. E_e also depends heavily on E_{ex} as well as E_c. The excitation energies broaden from several hundreds meV to about 3 eV. Sicne E_{ex} is much larger than E_a, it can be concluded that an electron-phonon coupling of RTN is very strong. Consequently, equations (8.18) and (8.19) are simply expressed as below [23].

$$E_c = \frac{E_{ex}}{4} + \frac{E_a}{2} \quad (8.24)$$

$$E_e = \frac{E_{ex}}{4} - \frac{E_a}{2} \quad (8.25)$$

E_c and E_e are nearly equal to $E_{ex}/4$. E_{ex} evidently determines the energies for capture and emission. From Eqs. (8.24) and (8.25), both E_c and E_e roughly show a quadratic dependence on V_{ox} as well as E_a.

8.5 Conclusion

The author expanded the conventional RTN model in MOSFEF in consideration of the change in the free energy through the single-electron tunneling and the non-radiative multi-phonon emission. Accordingly the author can provide more suitable models for real RTN parameters. Moreover the author demonstrated that my proposed models can explain the RTN behaviors particularly.

According to the proposed models, the activation energies for $\bar{\tau}_e/\bar{\tau}_c$ are determined by not only the difference between the trap level and the Fermi level but also the free energy change (E(V_{ox})) through the single-electron tunneling. This model adequately answers the question why activation energies widely spread from approximately 200 meV to −200 meV. The reason for the larger value than expected is that E(V_{ox}) is large enough to contribute to E_a even if $E_T - E_F$ is relatively small. The proposed models suggest that the gate-bias sensitivity (β) of $\bar{\tau}_e/\bar{\tau}_c$ is determined by the contribution of the capacitors related to the single RTN trap, mainly the gate capacitor (C_g) and the tunnel capacitor (C_t). They can be roughly calculated by the relationship between the conducting sphere and the parallel plates. The calculation provides curious information that the gate-bias sensitivity reaches 0.8 when the trap is located at 0.4 nm from channel in this work. The result means that the even RTN traps close to the channel can respond sensitively to the gate bias. The radius of the RTN defect is estimated to be 0.3 nm. This value is a suitable for determining the initial energies.

The estimation of the energies for capturing (E_c) and emitting electron (E_e) was demonstrated using the proposed non-radiative multi-phonon (NMP). The energy for capturing electron (E_c) shows the definite correlation with energy for emitting electron (E_e). They broaden continuously with about 700 meV. The reason for the strong correlation is that the excitation energy mainly determines both E_c and E_e as described in Eqs. (8.18) and (8.19).

The proposed RTN physical models can explain various types of RTN. Estimating RTN parameters based on my proposed models will become useful for identifying the kind of the defect structure.

8.6 References

[1] M. J. Kirton, and M. J. Uren, “Noise in solid-state microstructures: A new perspective on individual defects, interface states and low-frequency (1/f) noise,” *Adv. Phys.*, vol. 38, no. 4, pp. 367–468, Jul., 1989.

[2] K. K. Hung, P. K. Ko, C. Hu, and Y. C. Cheng, “A unified model for the flicker noise in metal-oxide-semiconductor field-effect transistors,” *IEEE Electron Dev. Lett.*, vol. 11, no. 3, pp. 90–92, Feb., 1990.

[3] Z. Shi, J. -P. Miéville, and M. Dutoit, “Random telegraph signals in deep submicron n-MOSFET’s,” *IEEE Trans. Electron Devices*, vol. 41, no. 7, pp. 1161–1168, Jul., 1994.

[4] D. Kang, J. Kim, D. Lee, B.-G Park, J. D. Lee, and H. Shin, “Extraction of Vertical, Lateral Locations and Energies of Hot-Electrons-Induced Traps through the Random Telegraph Noise,” *Jpn. J. Appl. Phys.*, vol. 48, 04C034, Apr., 2009.

[5] K. Abe, A. Teramoto, S. Sugawa, and T. Ohmi, “Understanding of traps causing random telegraph noise based on experimentally extracted time constants and amplitude,” i*n Proc. IRPS*, pp. 4A.4.1–4A.4.6, 2011.

[6] T. Grasser, “Stochastic charge trapping in oxides: From random telegraph noise to bias temperature instabilities,” *Microelectron. Reliab.*, vol. 52, pp. 39–70, Oct., 2011.

[7] T. Grasser, B. Kaczer, W. Goes, H. Reisinger, T. Aichinger, P. Hehenberger, P.-J. Wagner, F. Schanovsky, J. Franco, M. T. Luque, and M. Nelhiebel, “The paradigm shift in understanding the bias temperature instability: from reaction-diffusion to switching oxide traps,” *IEEE Trans. Electron Devices*, vol. 58, no. 11, pp. 3652–3666, Nov., 2011.

[8] M. Schulz, “Coulomb energy of traps in semiconductor spacecharge regions,” *J. Appl. Phys.*, vol. 74, no. 4, pp. 2649–2657, Apr., 1993.

[9] A. Palma, A. Godoy, J. A. Jiménez-Tejada, J. E. Carceller, and J. López-Villanueva, “Quantum two-demensional calculation of time constants of random telegraph signals in metal-oxide-semiconductor structures,” *Phys. Rev. B*, vol. 56, no. 15, pp. 9565–9574, Oct., 1997.

[10] E. Simoen, and C. Claeys “Random Telegraph Signal: a local probe for single point defect studies in solid-state devices,” *Mater. Sci. Eng. B* vol. 91–92, pp. 136–143, Apr., 2002.

[11] Ming-Pei Lu, and Ming-Jer Chen, “Oxide-trap-enhanced Coulomb energy in a metal-oxide-semiconductor system,” *Phys. Rev. B*, vol. 72, 235417, Dec., 2005.

[12] C. W. J. Beenakker, “Theory of Coulomb-blockade oscillations in the conductance of a quantum dot,” *Phys. Rev. B*, vol. 44, no. 4, pp. 1646–1656, Jul., 1991.

[13] K. Yano, T. Ishii, T. Hashimoto, T. Kobayashi, F. Murai, and K. Seki, "Room-temperature single-electron memory," *IEEE Trans. Electron Devices*, vol. 41, no. 9, pp. 1628–1638, Sep., 1994.

[14] R. Waser, *Nanoelectronics and Information Technology*. Weinheim, Germany: WILEY-VCH, pp. 391–393, 2012.

[15] C. W. J. Beenakker, "Theory of thermopower of a quantum dot," *Phys. Rev. B*, vol. 46, no. 15, pp. 9667–9676, Oct., 1992.

[16] R. N. Hall, "Electron-Hole Recombination in Germanium," *Phys. Rev.*, vol. 87, no. 2, pp. 387–387, May, 1952.

[17] W. Shockley, and W. T. Read, "Statistics of the Recombination of Holes and Electrons," *Phys. Rev.*, vol. 87, no. 5, pp. 835–842, Apr., 1952.

[18] M. Lax, "Cascade Capture of Electrons in Solids," *Phys. Rev*, vol. 119, no. 5, pp. 1502–1523, Sep., 1960.

[19] Z. A. Weinberg, "On tunneling in metal-oxide-silicon structures," *J. Appl. Phys.*, vol. 53, no.7, pp.5052–5056, Jul., 1982.

[20] D. M. Fleetwood, P. S. Winokur, R. A. Reber, Jr., T. L. Meisenheimer, J. R. Schwank, M. R. Shaneyfelt, and L. C. Rieweb, "Effects of oxide traps, interface traps, and "border traps" on metal-oxide-semiconductor devices," *J. Appl. Phys.*, vol. 73, no.10, pp.5058–5074, Jan., 1993.

[21] A. L. McWhorter, "l/*f* noise and germanium surface properties," in *Semiconductor Surface Physics*. Philadelphia: University of Pennsylvania Press, p. 207, 1957.

[22] S. Christensson, I. Lundström, and C. Svensson, "Low frequency noise in MOS transistors–I. theory," *Solid-state Electronics*, vol. 11, pp. 797–812, Sep., 1968.

[23] A. Avellan, D. Schroeder, and W. Krautschneider, "Modeling random telegraph signals in the gate current of metal-oxide-semiconductor field effect transistors after oxide breakdown," *J. Appl. Phys.* vol. 94, pp. 703–708, Apr., 2003.

9. Conclusions

The author mainly showed the impact of RTN on scaled MOSFETs and the physical model of RTN from the statistical view in this doctoral thesis. Specifically, the author established the statistical model of RTN V_{th} variation to discuss the impact of RTN on scaling, the gate-stack techniques, the 3D devices, and the reliability after the application of NBTI stress. Furthermore, the author developed the RTN physical model. As the author noted in this work, the gate-stack techniques can prevent a large increase in RTN V_{th} variation in the scaled MOSFETs. However, not only process level techniques but also circuit level and system level techniques for tolerance of RTN are required to suppress the impact of RTN on future devices as much as possible because there are some point defects brought about in gate dielectric due to favorable thermodynamics of formation. My outlook for the impact of RTN on scaled MOSFETs and proposed physical model of RTN will be certainly useful for development of the circuit level and the system level techniques for tolerance of the RTN impact. The contributions of this work are summarized and the future prospects and remaining problems are shown as below.

9.1 Contributions

i. In chapter 2, the author newly introduced the extreme value theory to investigate and control extremely large RTN V_{th} variations above the high cumulative probability of 99.999%. The Gumbel distribution and the log-normal distribution are suitable for showing RTN V_{th} variations from the regression analysis.

ii. In chapter 3, the author, for the first time, showed the impact of RTN on scaling by using scaled SiON / poly-Si MOSFETs. The RTN V_{th} variation is inversely proportional to device size. This trend shows the power law. RTN V_{th} variations depend on W_g more strongly than L_g because the effective channel width becomes narrower due to the discreteness of impurities in channel. And the anomalously large RTN V_{th} variations as extremely rare event appear in small MOSFET because the RTN V_{th} variations can be enhanced by the overlapped RTNs and the percolation pass due to discreteness of impurities in channel.

iii. In chapter 4, the author demonstrated that the RTN V_{th} variation in HK/MG MOSFET can be suppressed by suitable annealing, such as HTFGA, and by thin T_{inv}. As a consequence, properly annealed HK MOSFETs can have smaller RTN variation than SiON MOSFETs.

iv. In chapter 5, the author showed the relationship between RTN V_{th} variations and silicon-surface orientation. The (110) silicon-surface gives the largest total amount of suboxides of all three orientations, namely, Si(100), Si(110), and Si(111). Furthermore, the interface trap density (N_{it}) in the case of Si(110) substrate is the largest because interface trap density increases with increased amount of suboxides. It was also demonstrated statistically that RTN V_{th} variation of the Si(110) substrate is larger than those of the Si(100) and Si(111) substrates because Si(110) gives the largest total amount of suboxide which is deeply linked to the interface trap and contributes greatly to creating the fast RTN traps with large ΔV_{th}.

v. In chapter 6, the author found that ΔV_{th} of pMOSFET is larger than that of nMOSFET from the statistical viewpoint. This reason is that that the effect of mobility fluctuation on V_{th} variation of pMOSFET is bigger than that of nMOSFET. From the simulation of the impact of RTN on 6-transistor SRAM, the author found that the V_{th} margin comes close to the V_{th} window of the SRAM by taking into account the effect of RTN on ΔV_{th}, even at 65 nm generation.

vi. In chapter 7, the author investigated the relationship between NBTI and RTN. As a result, many permanent interface traps are generated by NBTI stress, and at the same time, the temporary and the one-time RTNs are also generated. In the recovery process, the re-passivation of interface states is the minor cause of recovery, and in contrast, rapid disappearance of the temporary RTN and the one-time RTN is the main cause of recovery.

vii. In chapter 8, the author expanded the conventional RTN model in MOSFEF in consideration of the change in the free energy through the single-electron tunneling and the non-radiative multi-phonon emission. The proposed model explain that the activation energies for $\bar{\tau}_e/\bar{\tau}_c$ are determined by not only the difference between the trap level and the Fermi level but also the free energy change (E(V_{ox})) through the single-electron tunneling. Moreover, my model revealed that the even RTN traps close to the channel can respond sensitively to the gate bias. The radius of the RTN defect is estimated to be 0.3 nm. This value is suitable for determining the initial energies. The estimation of the energies for capturing (E_c) and emitting electron (E_e) was demonstrated by using the advanced non-radiative multi-phonon (NMP). E_c shows the definite correlation with the energy for E_e. Both the energies continuously broaden with about 700 meV. The reason for the strong correlation is that the excitation energy mainly determines both E_c and E_e.

Moreover, I can get new insight into traps at gate dielectric through the overall studies of RTN. First of all, although the average number of traps per device decreases with scaling, the impact of RTN rare events with the extremely large V_{th} variation increases. Consequently, the importance of statistical analysis is increasing in the field of reliability of scaled MOSFETs. Secondly, I conclude that the RTN traps are located near interface. Therefore, HTFGA can reduce the impact of the RTN V_{th} variation, and silicon-surface orientations also affect the RTN V_{th} variation.

9.2 Future prospects and remaining problems

Next, the author shows the future prospects and remaining problems as follows:

i. Mechanisms of enhancement of RTN V_{th} variation are not completely clear. The mechanisms other than the overlapped RTNs and the percolation pass model are possible. Recently, K. P. Cheung reported a new model called “hole in the inversion layer” [1]. However, at the moment, this model has not achieved consensus. The other mechanisms will be clear by using intrinsic-channel MOSFETs in which the percolation passes of drain current do not occur.

ii. The author revealed that the RTN traps are located near interface between channel and gate dielectric. However, a specific 2D position coordinate of the RTN trap with large amplitude is not yet understood. Identifying the RTN trap position intimately links to complete understanding of enhancement of RTN V_{th} variation. Moreover, the type of trap causing RTN, for example, the oxygen vacancy or interface trap, is not yet identified.

iii. The energies for capture and emission reach to approximately 0.8 eV. These values are larger than expected. Furthermore, they widely spread. The author revealed that the results can be explained by using the non-radiative multi-phonon (NMP) model. On the other hand, nobody has ever given a physical image, what determines the large barrier height between the emission and capture states.

iv. The relationship between RTN and $1/f$ noise in MOSFET is still unclear. The hidden Markov model succeeded in extracting each RTN from complex RTNs. If $1/f$ noise arises exclusively from innumerable RTNs, the residue which is obtained by using the hidden Markov model cannot show $1/f$ noise. However, I occasionally observe that the residue shows $1/f$ noise. This fact means that $1/f$ noise does not necessarily consist of overlapped

RTNs, and there is other mechanism causing 1/*f* noise.

v. RTN behavior is the most basic kind of trap behavior. Recently, it is reported that various kind of traps cause RTN behavior. For example, a trap at junction causes RTN in junction leakage current. It is called the variable junction leakage (VJL). Variable retention time (VRT) at DRAM is likely attributed to VJL [2, 3]. Furthermore, a trap at grain boundary of poly-Si channel of 3D flash memories also causes RTN behavior in drain current reportedly [4]. Recently, it was reported that RTN has a potential to affect resistance random access memory (RRAM) operation [5]. Therefore, understanding of the other types of RTN will be required. My proposed statistical analysis method of RTN and my proposed RTN physical model will play increasingly important roles in the future digital devices.

The study of the impact of RTN on digital devices such as CMOS and memory has only just begun. I am sure that this doctoral thesis contributes to the further development of the study on RTN.

9.3 References

[1] K. P. Cheung, and J. P. Campbell, "On the magnitude of Random telegraph noise in ultra-scaled MOSFETs," in *Proc. ICICDT*, pp. 1–4, 2011.

[2] P. J. Restle, J. W. Park, and B. F. Lloydand, "DRAM Variable Retention time," in *IEDM Tech. Dig.*, pp. 807–810, 1992.

[3] Yuki Mori, Kiyonori Ohyu, Kensuke Okonogi, and Renichi Yamada, "The Origin of Variable Retention Time in DRAM" in *IEDM Tech. Dig.*, pp. 1034–1037, 2005.

[4] M.-K. Jeong, S.-M. Joe, C.-S. Seo, K.-R. Han, E. Choi, S.-K. Park, and J.-H. Lee, "Analysis of Random Telegraph Noise and Low Frequency Noise Properties in 3-D Stacked NAND Flash Memory with Tube-Type Poly-Si Channel Structure," in *VLSI Symp. Tech. Dig.*, pp. 55–56, 2012.

[5] D. Ielmini, F. Nardi, and C. Cagli, "Resistance-dependent amplitude of random telegraph-signal noise in resistive switching memories," *Appl. Phys. Lett.* vol. 96, 053503, Feb., 2010.

Acknowledgement

First of all, I would like to express my deep and sincere gratitude to my supervisors, Prof. Keisaku Yamada and Assoc. Prof. Kenji Ohmori at Doctoral Program in Nano-Science and Nano-Technology, Graduate School of Pure and Applied Sciences, University of Tsukuba for their elaborated guidance and considerable encouragement that I compile my theses about RTN in MOSFETs. I am very grateful to Prof. Kikuo Yamabe, Prof. Nobuyuki Sano, Prof. Yuzo Ohno at University of Tsukuba, and Prof. Takanobu Watanabe at Waseda University for invaluable comments on this paper.

My doctoral thesis is based on overall findings of the research at Central Research Laboratory, Hitachi, Ltd. I would like to express my deep gratitude to Dr. Nobuyuki Osakabe General Manager, Central Research Laboratory, Hitachi, Ltd., Dr. Kazuyoshi Torii CTO, Hitachi Europe Ltd., Dr. Shinji Yamada General Manager , Electronics Research Center, Dr. Yasuhiro Shimamoto Department Manager, Energy Conversion Electronics Research Department, Dr. Ryuta Tuchiya, and Dr. Akio Shima for providing me this opportunity as a Ph. D student.

I also appreciate the direct technical help received from Dr. Hiroshi Miki, and Dr. Renichi Yamada. They have given me their continuing supports and encouragement ever since I joined Central Research Laboratory, Hitachi, Ltd. from 2004. And I thank Mr. Toshiyuki Mine, Dr. Takeshi Ishida, Ms. Yuki Mori, and Dr. Masanao Yamaoka for providing many advices and helping my research. For the contributions to the sample fabrication, I would acknowledge the stuffs in S3 clean room.

I wish to extend my appreciation to Dr. Zhibin Ren, Dr. Christoper P. D'Emic, Dr. Yu Zhu, Dr. David J. Frank, Dr. Michael A. Guillorn, Dr. Dae-Gyu Park, and Dr. Wilfriend Haensch at T. J. Watson Research Center, IBM for their kind supports and invaluable suggestions. For the contributions to the sample fabrication, I would acknowledge the personal of T. J. Watson MRL facility. I lived a truly extraordinary life as researcher in USA for two years from 2008.

Finally, I pay my respects to my parents and thank to my wife Yukiko Tega, and my little daughter Honami Tega for supporting my life.

List of publications

Papers

[1] Naoki Tega, Hiroshi Miki, Toshiyuki Mine, and Kazuyoshi Torii, "Investigation of Relationship between Interface State and Random Telegraph Noise Using Metal–Oxide–Semiconductor Field-Effect Transistors Fabricated on Si(100), (110), and (111) Substrates," *J. J. Appl. Phys.*, vol. 51, 114001, Oct., 2012.

[2] Naoki Tega, Hiroshi Miki, Toshiyuki Mine, Kenji Ohmori, and Keisaku Yamada, "Statistical analysis of relationship between negative-bias temperature instability and random telegraph noise in small p-channel metal-oxide-semiconductor field-effect transistors," to be published in *J. J. Appl. Phys*.

[3] Naoki Tega, Hiroshi Miki, and Kazuyoshi Torii, "Comprehensive Physical Model of Random Telegraph Noise in MOSFET," submitted to *IEEE Trans. Device Mater. Reliab*.

[4] Takeshi Ishida, Naoki Tega, Yuki Mori, Hiroshi Miki, Toshiyuki Mine, Hitoshi Kume, Kazuyoshi Torii, Renichi Yamada, and Kenji Shiraishi, "State Transition of a Defect Causing Random-Telegraph-Noise Fluctuation in Stress-Induced Leakage Current of Thin SiO2 Films in a Metal–Oxide–Silicon Structure," *J. J. Appl. Phys.*, vol. 52, 110203, Oct., 2013.

International conferences

[1] Naoki Tega, Hiroshi Miki, Taro Osabe, Akira Kotabe, Kazuo Otsuga, Hideaki Kurata, Shiro Kamohara, Kenji Tokami, Yoshihiro Ikeda, and Renichi Yamada, "Anomalously Large Threshold Voltage Fluctuation by Complex Random Telegraph Signal in Floating Gate Flash Memory," in *IEDM Tech. Dig.*, pp. 1–4, 2006.

[2] Naoki Tega, Hiroshi Miki, Masanao Yamaoka, Hitoshi Kume, Toshiyuki Mine, Takeshi Ishida, Yuki Mori, Renichi Yamada, and Kazuyoshi Torii, "IMPACT OF THRESHOLD VOLTAGE FLUCTUATION DUE TO RANDOM TELEGRAPH NOISE ON SCALED-DOWN SRAM," in *Proc. IRPS*, pp. 541–546, 2008.

[3] N. Tega, H. Miki, F. Pagette, D. J. Frank, A. Ray, M. J. Rooks, W. Haensch, and K. Torii, "Increasing Threshold Voltage Variation due to Random Telegraph Noise in FETs as Gate Lengths Scale to 20 nm," in *VLSI Tech. Dig.*, pp. 50–51, 2009.

[4] Naoki Tega "Study on Variability in Transistor Characteristics due to Random Telegraph Noise," *IEEE/ACM Workshop on Variability Modeling and Characterization CVM Collocated Workshop ICCAD*, 2009.

[5] N. Tega, H. Miki, Z. Ren, C. P. D'Emic, Y. Zhu, D. J. Frank, J. Cai, M. A. Guillorn, D.-G. Park, W. Haensch, and K. Torii, "Reduction of Random Telegraph Noise in

High-κ / Metal-gate Stacks for 22 nm Generation FETs," in *IEDM Tech. Dig.*, pp. 772–775, 2009.

[6] N. Tega, H. Miki, Z. Ren, C. P. D'Emic, Y. Zhu, D. J. Frank, J. Cai, M. A. Guillorn, D.-G. Park, W. Haensch, and K. Torii, "On the need for a new model: Inconsistencies between observations and physical model for random telegraph noise in HKMG MOSFET," *IWDTF*, pp. 153–154, 2011.

[7] Naoki Tega, Hiroshi Miki, Zhibin Ren, Christoper P. D'Emic, Yu Zhu, David J. Frank, Michael A. Guillorn, Dae-Gyu Park, Wilfried Haensch, and Kazuyoshi Torii, "Impact of HK / MG stacks and Future Device Scaling on RTN," in *Proc. IRPS*, pp. 630–635, 2011.

[8] H. Miki, T. Osabe, N. Tega, A. Kotabe, H. Kurata, K. Tokami, Y. Ikeda, S. Kamohara, and R. Yamada, "Quantitative Analysis of Random Telegraph Signals as Fluctuations of Threshold Voltages in Scaled Flash Memory Cells," in *Proc. IRPS.*, pp. 29–35, 2007.

[9] T. Ishida, N, Tega, Y. Mori, H. Miki, T. Mine, H. Kume, K. Torii, M. Muraguchi, Y. Takeda, K. Shiraishi, R. Yamada, "A new insight into the dynamic fluctuation mechanism of stress-induced leakage current, " in *Proc. IRPS*, pp. 604–609, 2008.

[10] H. Miki, M. Yamaoka, N. Tega, Z. Ren, M. Kobayashi, C. P. D'Emic, Y. Zhu, D. J. Frank, M. A. Guillorn, D.-G. Park, W. Haensch, and K. Torii, "Understanding Short-term BTI Behavior through Comprehensive Observation of Gate-voltage Dependence of RTN in Highly Scaled High- / Metal-gate pFETs," in *VLSI Tech. Dig.*, pp. 148–149, 2011.

[11] Hiroshi Miki, Naoki Tega, Zhibin Ren, Christopher P. D'Emic, Yu Zhu, David J. Frank, Michael A. Guillorn, Dae-Gyu Park, Wilfried Haensch, and Kazuyoshi Torii, "Hysteretic Drain-Current Behavior Due To Random Telegraph Noise in Scaled-down FETs with High-k/Metal-gate Stacks, " in *IEDM Tech. Dig.*, pp. 620–623, 2010.

[12] H. Miki, N. Tega, M. Yamaoka, D. J. Frank, A. Bansal, M. Kobayashi, K. Cheng, C. P. D'Emic, Z. Ren, S. Wu, J-B. Yau, Y. Zhu, M. A. Guillorn, D.-G. Park, W. Haensch, E. Leobandung, and K. Torii, "Statistical Measurement of Random Telegraph Noise and Its Impact in Scaled-down High-κ/Metal-gate MOSFETs," in *IEDM Tech. Dig.*, pp. 450–453, 2012.

National conventions (First author)

[1] 手賀直樹　三木浩史　鳥居和功 「ランダム・テレグラフ・ノイズによるトランジスタ特性ばらつき」2009年（平成21年）秋季 第70回応用物理学会学術講演会 シリコンナノエレクトロニクスの新展開.

Review articles (First author)

[1] 手賀直樹　三木浩史 「微細MOSFETにおけるランダム・テレグラフ・ノイズの影響」応用物理 第77巻　第1号　p41–44 (2008).