Extended Abstracts of the 2008 International Conference on Solid State Devices and Materials, Tsukuba, 2008, pp. 234-235

J-2-1 (Invited)

Impact of Random Telegraph Noise (RTN) on Future Memory

Hiroshi Miki

Central Research Laboratory, Hitachi, Ltd.
1-280 Higashi-koigakubo, Kokubunji, Tokyo 185-8601, Japan

1. Introduction

Recently, variations in electrical characteristics of MOSFETs have been recognized as one of the most intrinsic difficulties in future memory technology. In studying these variations, less attention had been paid to temporal changes in charges in gate insulator than other factors like random-dopant fluctuation. Recent studies on non-volatile memories, however, revealed that the charges can have significant impact on reliability of memory cells. Random telegraph noise (RTN) in the threshold voltage (V_T) is the elementary process of the fluctuation; when a MOSFET is small enough to allow a single trap in a device, erratic binary (telegraphic) change of V_T is observed which corresponds to capture and emission of a charge at the single trap.

In this paper, large amplitude of RTN in V_T is reviewed along with its underlying physics, and its areal scaling is explored. Although these are discussed mostly on floating-gate memories, the results are believed to be highly relevant to other V_T-sensitive memories. Selected RTN observations in leakage current are also presented.

2. Threshold voltage RTN

Exceptionally large amplitude of RTN

Waveforms of RTN are described by three parameters; amplitude (ΔV_T) and holding time at each state (τ_c or τ_e) as shown in Fig. 1 (a). The amplitude is of primary importance since it has a close relationship to an error in a memory cell. If one estimates the maximum amplitude by qN_T/C_{ox}, where q, N_T, and C_{ox} are an elementary charge, areal trap density, and capacitance of gate insulator per unit area, respectively, the maximum ΔV_t should be in the order of 10 mV for floating-gate memories. It is therefore expected that RTN has limited effect irrespective of scaling. However, Kurata *et al* found that the amplitude of larger than 100 mV is observed at the tail of their distribution, and expected that the amplitude will increase with scaling-down [1].

To estimate amplitude of RTN, it is of great importance that the amplitudes statistically distribute in a wide range if one measures numerous cells in an array. The distribution of the observed amplitudes has been reported to be described using exponential distribution, whose probability density is given by $\lambda\exp(-\lambda x)$ where λ is a reciprocal of an average amplitude [2][3]. This approximation was qualitatively confirmed by Monte Carlo procedure which assumed *single* trap with discrete atomistic doping in the channel [4]. When the single trap coincides with a current-percolation path induced by the discreteness, this results in a large amplitude of RTN [5][6].

However, the experimentally observed tail distribution is still longer than predicted by the single-trap model, when the higher percentile points are collected with careful procedure [7]. In spite of the fact that probability of finding the extremely large amplitude is much less than 1%, it is a serious problem for large-scale integrated memory products.

The most likely cause of the large amplitude is complex RTN induced by multiple traps in a cell [5]. When two traps are charged and discharged independently, V_T can take four values. Accordingly, n traps in a cell generate 2^n values in V_T as shown in Fig. 1(b). This mechanism produces much larger amplitude than predicted by the single-trap model. It is therefore necessary to take the multiple-trap effect into account to describe the experimental tail distribution. When multiple traps contribute to RTN, the amplitude of the complex RTN is defined as the difference between the highest and lowest levels. If the difference is compared to the maximum in a set of random amplitudes, the distribution of the maximum should be asymptotic to an extreme value distribution. Experimental results were found to follow a Gumbel distribution which is a type of extreme distribution.

These reports on amplitude are all studied on nFET in floating-gate memories. For SRAM application, however, amplitude in pFET is also a considerable concern. Recently it has been revealed that amplitudes of RTN in pFET are much higher than those in nFET as shown in Fig. 2 [8]. Large capture cross section for hole trapping is thought to enable multiple trapping to occur frequently.

Time constants of RTN and origin of traps

Time constant of RTN is a touchstone to determine whether the noise is RTN, and offers physical insights into the active trap. Holding times in a waveform are not unique but scattered in a wide range. They should follow an exponential distribution since probability of transition from one state to the other is constant.

One intuitive way to assess the distribution is to plot holding times on a Weibull plot, and see if its slope is unity [9]. Power spectrum density with $1/f^2$ roll-off also proves the noise is RTN [10]. Once the noise is verified as RTN, transition probability per unit time is a simple reciprocal of the average holding time at each state. Since the transition occurs through tunneling of a carrier between a trap and a channel, the probability is a function of both spatial and energetic position of the trap.

In particular, the ratio of transition probabilities to and from the trap has a simple relationship, $g\exp(-(E_T - E_F)/kT)$, where g is degeneracy factor, and E_T and E_F are energy levels of the trap and channel, respectively. The energy levels are found to spread in a broad range. This can be attributed to the amorphous structure of the gate insulator [5]. Additionally, when the trap is further away from the channel, both RTN amplitude and transition probability should be lowered. It is therefore expected that the probability strongly depends on the amplitude. These two parameters were, however, found to have a very weak correlation (Fig. 3) [7]. This fact supports the hypothesis that the principal cause of the amplitude distribution is not the trap distance from the channel, but atomistic discreteness of impurities mentioned before.

Impact of future scaling on RTN in threshold voltage

In estimations of areal dependence of RTN, the most optimistic view is that the amplitude is kept constant based on areal charge density model. On the other hand, since the average spatial interval of 10^{10} /cm^2 trapped charge is 100 nm, the average number of trapped carriers in a state-of-the-art device is well below one. Hence the magnitude of trapped charge is at least q even when the device is scaled down, resulting in inverse dependence on area ($\propto 1/LW$). These two dependencies mark the lower and upper boundaries of predictions found in the literature.

In spite of its practical importance, it is not an easy task to obtain the experimental dependence due to its statistical nature. In modeling approaches, there are a few reports discussing statistical nature of amplitude. Sonoda *et al* show that both 99.9 percentile and average of amplitudes are proportional to $1/LW$ [6]. On the other hand, 99.9 percentile has been reported to have as weak as $1/(LW)^{0.24}$ [2].

3. RTN in leakage current

RTN is found in memory cell characteristics other than V_T, and they result in erratic retention failure. One of the phenomena is leakage-current fluctuation observed in relatively thin tunneling oxide in floating-gate memories and called variable stress-induced leakage current (V-SILC) [9] [11]. Binary current levels and holding times which satisfy statistics of RTN prove the fluctuation to be RTN (Fig. 4). The distinguishing feature of V-SILC is independence of its amplitudes on active area; this also demonstrates that the fluctuation is originating from a single defect. Leakage current through *pn* junction is also known to show RTN, called variable junction leakage [12]. The binary leakage current is believed to be a cause of variable retention time in DRAM [13]. The transition corresponds to a change of an oxygen-insertion defect in the silicon lattice between two pseudo static structures [14].

These RTN amplitudes in leakage current are independent of area, resulting in stronger impact if the number of charges stored in a cell is reduced when the cell is scaled down.

4. Conclusions

Random telegraph noise (RTN) observed in scaled memory cells are reviewed. Current-path percolation induced by discrete random dopant configuration induces expanded amplitude distribution, and multiple trap effect further spreads the tail distribution, resulting in unexpectedly large amplitude which upsets stored data in the cell. RTN is also observed in gate or junction leakage current. In all cases presented in this study, the impact of the RTN is expected to increase with scaling down of the cell. Both experimental and theoretical studies on the RTN phenomena will be required for assessment of memory reliability.

Acknowledgements

The author would like to thank Dr. H. Kume, Dr. R. Yamada, T. Ishida and N. Tega of Hitachi, Ltd., and Dr. O. Tsuchiya and Dr. T. Kamohara of Renesas Technology Corp. for their fruitful discussions.

References

[1] H. Kurata *et al*, *IEEE J. Solid-state Circuits*, **42**, 1362 (2007).
[2] K. Fukuda *et al*, *in 2007 IEDM Tech. Dig.*, p.169.
[3] C. Compagnoni *et al*, *IEEE Trans. Electron Devices*, **55**, 388 (2008).
[4] A. Ghetti *et al*, *in Proc. 2008 IEEE IRPS*, p.610.
[5] N. Tega *et al*, *in 2006 IEDM Tech. Dig.*, p.491.
[6] K. Sonoda *et al*, *IEEE Trans. Electron Devices*, **54**, 1918 (2007).
[7] H. Miki *et al*, *in Proc. 2007 IEEE IRPS*, p.29.
[8] N. Tega *et al*, *in Proc. 2008 IEEE IRPS*, p.541.
[9] T. Ishida *et al*, *in Proc. 2008 IEEE IRPS*, p.604.
[10] M. J. Kirton and M. J. Uren, *Adv. Phys.*, **38**, 367 (1989).
[11] R. Yamada and T. J. King, *in Proc. 2003 IEEE IRPS*, p.491.
[12] Y. Mori *et al*, *in 2005 IEDM Tech. Dig.,* p.1034.
[13] P. J. Restle *et al*, *in 1992 IEDM Tech. Dig.*, p.807.
[14] K. Ohyu *et al*, *in 2006 IEDM Tech. Dig.*

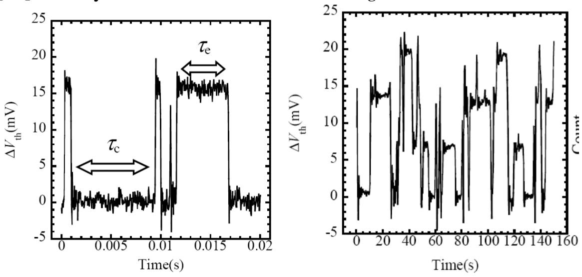

Fig. 1 RTN waveform for (a) single trap and (b) multiple traps [5].

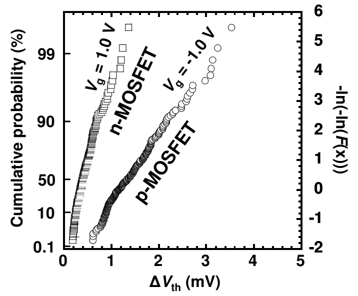

Fig.2 Comparison of RTN amplitude in nFET and pFET [2].

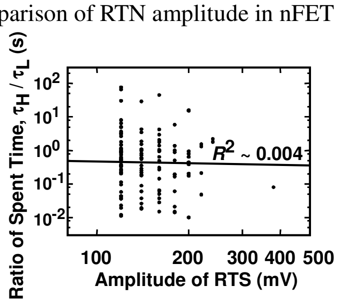

Fig.3 Correlation between amplitude and time constants [7].

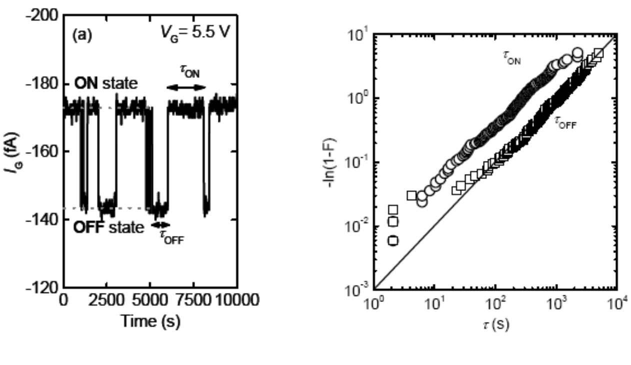

Fig.4 Waveform and distribution of time constant [4].