Accelerated evaluation method for the SRAM cell write margin using word line voltage shift

メタデータ	言語: English 出版者: 公開日: 2017-10-03 キーワード (Ja): キーワード (En): 作成者: Makino, Hiroshi, Nakata, Shunji, Suzuki, Hirotsugu, Morimura, Hiroki, Mutoh, Shin'ichiro, Miyama, Masayuki, Yoshimura, Tsutomu, Iwade, Shuhei, Matsuda, Yoshio, 深山, 正幸 , 松田, 吉雄 メールアドレス: 所属:
URL	https://doi.org/10.24517/00008566

This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 3.0 International License.

Accelerated Evaluation Method for the SRAM Cell Write Margin using Word Line Voltage Shift

Hiroshi Makino[1], Shunji Nakata[2], Hirotsugu Suzuki[3], Hiroki Morimura[2], Shin'ichiro Mutoh[2], Masayuki Miyama[3], Tsutomu Yoshimura[4], Shuhei Iwade[1] and Yoshio Matsuda[3]

[1]Faculty of Information Science and Technology/[4]Faculty of Engineering, Osaka Institute of Technology, [1]Hirakata/[4]Osaka, Japan
[2] NTT Microsystem Integration Laboratories, Nippon Telegraph and Telephone Corporation, Atsugi, Japan
[3]Graduate School of Natural Science, Kanazawa University, Kanazawa, Japan

***Abstract*— An accelerated evaluation method for the SRAM cell write margin is proposed based on the conventional Write Noise Margin (WNM) definition. The WNM is measured under a lower word line voltage than the power supply voltage VDD. A lower word line voltage is used because the access transistor operates in the saturation mode over a wide range of threshold voltage variation. The final WNM at the VDD word line voltage, the Accelerated Write Noise Margin (AWNM), is obtained by shifting the measured WNM at the lower word line voltage. The amount of WNM shift is determined from the WNM dependence on the word line voltage. As a result, the cumulative frequency of the AWNM displays a normal distribution. A normal distribution of the AWNM drastically improves development efficiency, because the write failure probability can be estimated by a small number of samples. Effectiveness of the proposed method is verified using the Monte Carlo simulation.**

Keywords-SRAM; write noise margin; Vth fluctuation; word line voltage; write margin distribution

I. INTRODUCTION

The recent progress of process technology has caused various fluctuation problems in the device characteristics due to transistor area reduction. The Vth fluctuation caused by dopant fluctuation has the greatest influence on device characteristics [1-2]. Generally, this dopant induced Vth fluctuation is random and obeys the normal distribution.

The stability of the SRAM cell is greatly affected by Vth fluctuation, because the SRAM cell is usually designed using minimum design rules. Vth fluctuation degrades both the read and write operation stabilities. It has been said that the read operation is usually less stable than the write operation under Vth fluctuation. However, the write operation is also affected by a large Vth fluctuation. In addition, a recent paper reports that write operation failure is more dominant than read operation failure in low power supply voltages [3]. Therefore, an accurate evaluation of the write operation stability is as important as the evaluation of the read operation stability.

Conventionally, the Write Noise Margin (WNM) is used as a metric of write operation stability [4]. Although the WNM is easy to measure, it does not obey a normal distribution because it is not sensitive to Vth variation when the WNM is large [5]. If the write margin obeys the normal distribution, the write margin distribution can be easily estimated by a small number of samples. This drastically improves development efficiency.

In this paper, we propose an accelerated method for evaluating the SRAM cell write margin based on the conventional WNM definition. The WNM is measured at a lower word line voltage than the VDD of the power supply voltage and is calibrated to the WNM of the VDD word line voltage. In the proposed method, the write margin obeys the normal distribution even though it uses the conventional WNM definition.

II. CONVENTIONAL WRITE NOISE MARGIN

The circuit of the SRAM write operation is shown in Fig. 1(a). Let us assume that the inverted data are written to the SRAM cell where “1” is stored on the internal node V1 and “0” on the V2. Then, the data “0” and “1” are given on the bit lines BL and /BL, respectively, under the activated word line WL. If the voltages of nodes V1 and V2 are inverted, the write operation is successful. Hereupon, the V1, V2, BL, /BL and WL are also used as the voltages of their nodes. The definition of the conventional Write Noise Margin (WNM) is shown in Fig. 1(b). There are DC characteristic curves of the inverter A (InvA) and the inverter B (InvB) under WL=VDD, BL=0 V and /BL=VDD. The VDD is the power supply voltage. The WNM is defined as the width of the smallest embedded square between the two DC characteristic curves.

Generally, the write margin is a function of the threshold voltage Vth’s of the SRAM cell. If the write margin is linear for the Vth’s, it is expected to obey the normal distribution, allowing us to predict the write margin distribution accurately from a small number of samples. If the write margin distribution follows the normal distribution, the write yield can also be easily estimated [6].

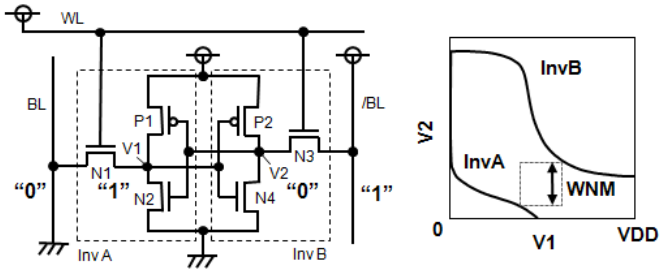

Fig.1. (a) SRAM cell circuit in the write operation (b) Definition of the Write Noise Margin.

The dependence of the WNM on the Vth is examined using the SPICE simulation. The transistor parameter of 45-nm process technology [7] is used with a power supply voltage of VDD=1.0 V. The typical values of threshold voltages are Vthn=0.404 V for the NMOS transistors and Vthp=-0.384 for the PMOS transistors. The transistor sizes are W=55 nm, 83 nm, and 55 nm with L=45 nm for the access, driver, and load transistors, respectively.

The simulation results are shown in Fig. 2. We set ΔVth=0 when the threshold voltages are typical. While the WNM is not linear on the Vth of the access transistor N1, the WNM is almost linear on the Vth's of the other transistors. Non-linearity on the N1 causes the WNM to deviate from the normal distribution [6]. In the lower ΔVth region, the load transistor P1 determines WNM=0. In the higher Vth region, the access transistor N1 determines WNM=0. The slope of the WNM for the N1 notably changes around ΔVth=0.1 V. The WNM is completely linear for ΔVth>0.1V. We call this line the linear section of the WNM for the N1. The point WNM=0 is on this straight line. When ΔVth<0.1V, the slope of the WNM is nearly equal to 0. This means that the WNM is not sensitive to the Vth fluctuation of the N1 when the WNM is large. This is consistent with a previous work [5]. The access transistors affect the WNM only when a large Vth fluctuation occurs. In other words, the WNM distribution has a tail at the side of the small margin. In this case, a large number of samples is needed in order to estimate the distribution. If we estimate the distribution with a small number of samples, almost every sample appears around the WNM for ΔVth=0. This creates a very sharp predicted distribution, resulting in an overestimation of the ΔVth for WNM=0, because the slope of the WNM is nearly equal to 0 around ΔVth=0.

The reason why the WNM has different slopes around ΔVth = 0.1 V can be explained by whether the access transistors are in the saturation mode or in the linear mode when the WNM is evaluated. The dependence of the V1 on the ΔVth of N1 is examined using the SPICE simulation at V2=0 V. The results are shown, together with the WNM, in Fig.3. The dashed line, which is around ΔVth=0.1 V, represents the changing point of the WNM slope. The V1 rapidly increases from that point. This means the operation mode of the N1 changes from a linear mode to a saturation mode around the dashed line. Therefore, a change in the slope of the WNM is related to a change in the operation mode of the access transistor N1.

In the AC write operation of a SRAM cell, the access transistor N1 is always in the saturation mode at the beginning of the operation because the V1 is not less than the WL. The write failure occurs when the N1 stays in the saturation mode during the write operation. Therefore, the write margin should be evaluated in the saturation mode of the access transistor N1. Contrary to the actual AC write operation, the conventional WNM is evaluated in the linear mode of the access transistor when the write margin is large. The conventional WNM definition is not considered adequate to evaluate the stability of a SRAM cell.

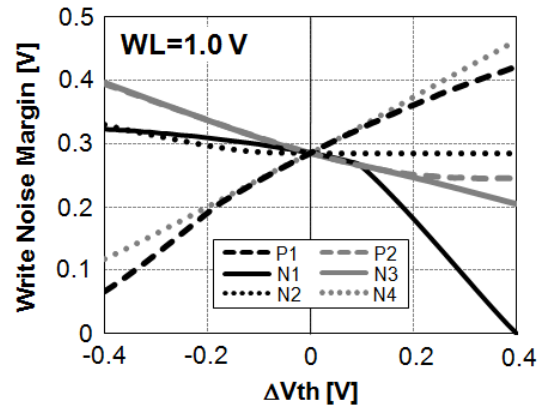

Fig. 2. Dependence of the WNM on the Vth in the conventional definition at WL=VDD.

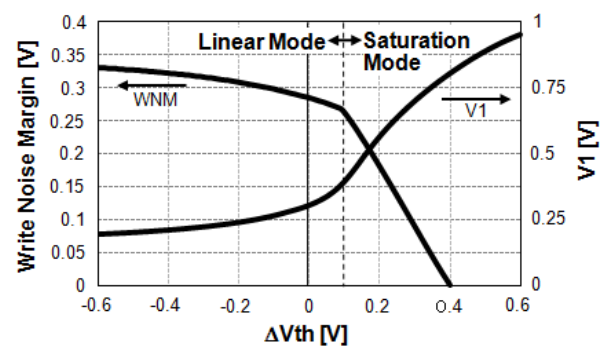

Fig. 3. The voltage of V1 at the fixed voltage of V2=0 V for the threshold voltage variation of the access transistor in the write operation.

III. Accelerated Evaluation Method

In this section, we propose an accelerated evaluation method for the SRAM cell write margin based on the conventional WNM definition. In the proposed method, the access transistor is forced to operate in the saturation mode by lowering the word line voltage from the VDD. The WNM is measured under a lower word line voltage. The word line voltage is chosen within a range where the access transistor operates in the saturation mode. The WNM at the word line voltage of the VDD is calibrated from the WNM at a lower word line voltage. This calibrated WNM is called the Accelerated WNM (AWNM).

First, we measure the dependence of the WNM on the word line voltage. The WNM given by the SPICE simulation is shown in Fig. 4. The power supply voltage is VDD=1.0 V. The solid line represents the simulation results. The WNM is linear for word line voltages of less than 0.9 V. The slope change of the WNM at the word line voltage of 0.9V corresponds to the change in the operation mode of the access transistor N1 around the threshold voltage of ΔVth=0.1 V in Fig. 3. The N1 operates in a saturation mode in WL<0.9 V. The dashed line represents the extrapolated line. The extrapolated value of the WNM is 0.35 V which, in the proposed method, is the WNM at WL=1.0 V. This is the AWNM. The slope of the WNM for the WL voltage in the

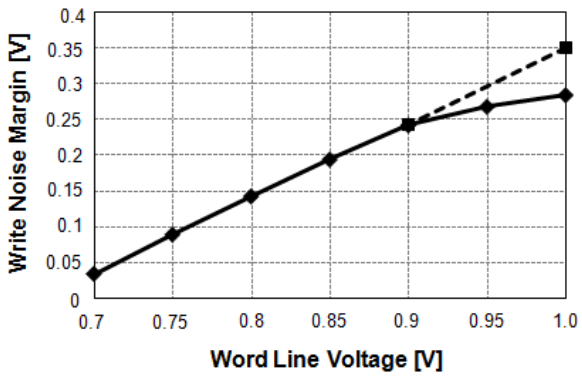

Fig. 4. Dependence of the WNM on the word line voltage.

linear section is used in the calibration of the measured WNM, as shown later.

Although the accelerated evaluation method gives a good linearity for the WNM, the value of the WNM itself is small when compared to the WNM at WL=1.0V. This value is calibrated to the WNM at WL=1.0 V. In the accelerated evaluation method, the AWNM at WL=1.0V, the AWNM ($WL_{1.0}$), is obtained from the WNM at a low word line voltage, WNM(WL_m), as:

$$AWNM(WL_{1.0}) = WNM(WL_m) + \alpha(WL_{1.0} - WL_m), \quad (1)$$

where α is the slope of the WNM for the WL voltage in the linear section.

Fig. 5 shows the dependence of the WNM on ΔVth at the word line voltage of 0.8V. A negative value is defined as the maximum length of an embedded square in the crossed butterfly curves. This means that the data are not inverted. In Fig. 5, the WNM for the N1 is linear around ΔVth=0. In Fig. 6, the dependences of the AWNM and the WNM on the Vth's are shown for the transistor N1. The solid line represents the AWNM and the dashed line represents the WNM. The thin solid line is an extrapolated line from the slope of AWNM at ΔVth =0V. The most important point is that the extrapolated line gives WNM=0 correctly. The threshold voltage ΔVth's giving AWNM=0 and WNM=0 are the same for the most influential transistor, N1.

IV. Monte Carlo Simulation

The proposed method is verified using the Monte Carlo simulation. The Vth's are assumed to obey the normal distribution with a variance of σ_{Vth} =50 mV and the means of Vthn=0.404 V and Vthp=-0.384 V. In the Monte Carlo simulation, we make the Vth's of six SRAM cell transistors independently change at random. The number of samples is 100,000. For simplicity, we set the same variance for all of the transistors.

In this simulation, we use Vth's of typical values, but the actual threshold voltages of SRAM cell transistors are not known when real devices are measured. By measuring the dependence of the WNM on the word line voltage in several samples, it is possible to determine a word line voltage that causes the access transistor to operate in the saturation mode. We should exclude samples with a very low Vth, where the access transistor operates in the linear mode in the WNM

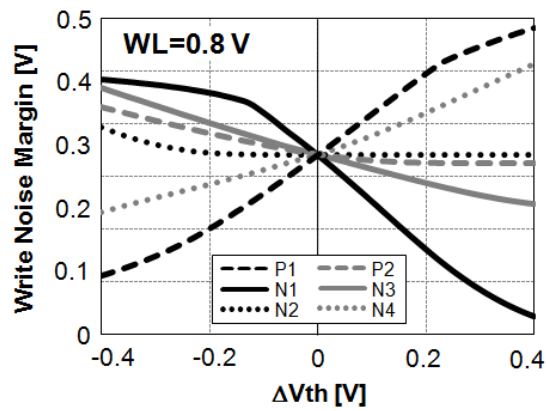

Fig. 5. Dependence of the WNM on ΔVth under WL=0.8 V.

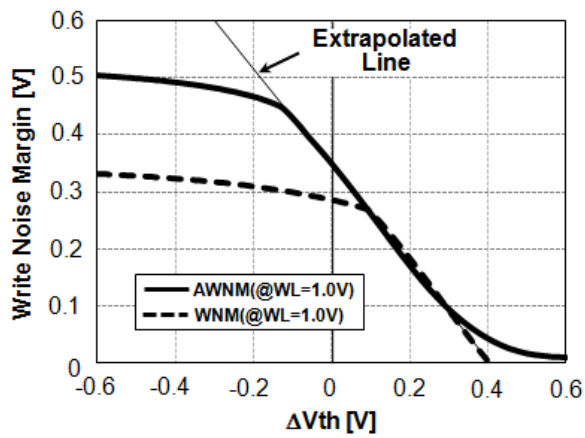

Fig. 6. The AWNM and the WNM at WL=1.0V for the access transistor N1.

measurements, when determining the word line voltage. The excluded samples do not influence the measurements, because the probability of encountering such devices is very small.

The dependence of the WNM on the word line voltage is shown for the first ten samples in Fig. 7. The slopes α=ΔWNM/ΔWL are almost the same for the ten samples in each linear section. The word line voltage of 0.8 V was chosen from these data. The cumulative frequency scaled by the variance σ is shown in Fig. 8. The straight line of the cumulative frequency means a normal distribution of the write margin. Lines I, II, and III are the WNM at WL=0.7 V, 0.8 V, and 1.0 V, respectively. For reference, the data at WL=0.7 are also shown. Line IV is the AWNM corresponding to WL=1.0 V calibrated from line I, that is, the WSNM at WL=0.7 V. Line V is the AWNM corresponding to WL=1.0 V calibrated from line II, the AWNM at WL=0.8 V. For the calibration, we used α=1.060, which is the mean value of the ten samples. The mean values of the write margins are summarized in Table I. The two means of the AWNM at WL=1.0 V, μ_{AWNM}'s, calibrated from the WNM at WL=0.7 V and WL=0.8 V match well. The AWNMs at WL=1.0 V calibrated from I and II almost overlap, which shows the universality of the proposed method.

The slope of the cumulative frequency of the WNM at WL=1.0 V changes at about WNM=0.2 V. Obviously, the

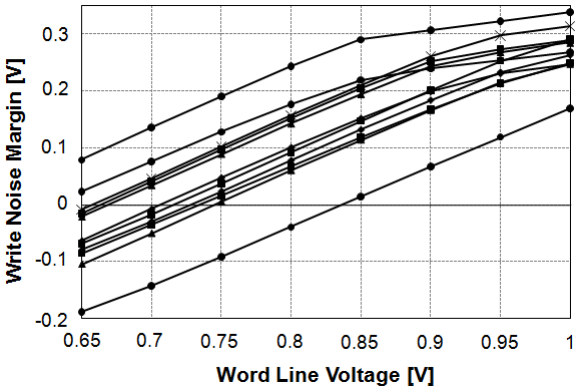

Fig. 7. Dependence of the WNM on the word line voltage for the first ten samples in the Monte Carlo simulation.

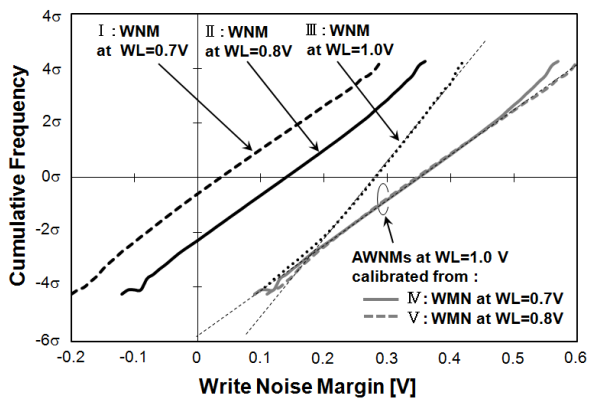

Fig. 8. The cumulative frequency of the AWNM and the WNM. Lines I, II and III are the WNM at WL=0.7 V, 0.8 V, and 1.0 V, respectively. Line IV and V are the AWNMs at WL=1.0 V calibrated from lines I and II, respectively.

conventional WNM does not obey the normal distribution. Therefore, the WNM at WL=1.0 V gives a small write failure probability if the probability is estimated from the slope in the neighborhood of ΔVth=0. On the other hand, the cumulative frequency of the WNM at WL=0.7 V and WL=0.8V is a straight line. As a result, the AWNM is also straight and obeys the normal distribution.

The extrapolated $-\mu_{WNM}/\sigma_{WNM}$'s are shown in Table II. Although the two values of AWNMs agree well, the value of WNM does not match them in this simulation range. We can estimate the write failure probability precisely not from the WNM but from the AWNM. The write failure probability P_{WF} is obtained when considering both the cases of "0" writing and "1" writing as [6]:

$$P_{WF} = 2\int_{-\infty}^{-\mu_{WM}/\sigma_{WM}} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) dx \qquad (2)$$

The value $-\mu_{WM}/\sigma_{WM}$ is the cumulative frequency at AWNM=0 in the scale of σ, corresponding to the $-\mu_{WNM}/\sigma_{WNM}$ in Table II.

V. Conclusion

We have proposed an accelerated evaluation method for the SRAM cell write margin based on the conventional WNM definition. The WNM is measured under a lower word line voltage than the VDD of the power supply voltage and the access transistor is forced to operate in the saturation mode. The Accelerated Write Noise Margin (AWNM), the WNM at WL=VDD in this method, is obtained by shifting the WNM at the lower word line voltage. The degree of shift is determined from the pre-measured WNM dependence on the word line voltage. The cumulative frequency of the AWNM is linear for the AWNM, which means a normal distribution of the AWNM. A normal distribution of the AWNM dramatically improves development efficiency. The effectiveness of the proposed accelerated evaluation method for the write margin is verified using the Monte Carlo simulation.

Table I. Mean values of the WNM and AWNM.

	μ_{WNM}	μ_{AWNM}
WL=0.7 V	0.036	0.35
WL=0.8 V	0.14	0.35
WL=1.0 V	0.28	

Table II. Extrapolated values of $-\mu_{WM}/\sigma_{WM}$ in Fig. 8.

	$-\mu_{WNM}/\sigma_{WNM}$
WNM	-8.58* / -6.08**
AWNM at WL=1.0 V from WL=0.7 V	-5.72
AWNM at WL=1.0 V from WL=0.8 V	-5.78

*) The extrapolated value of a straight line with the slope of the WNM around the mean value of 0.28 V.
**) The extrapolated value of a straight line with the slope of the WNM around the mean value of 0.15 V.

References

[1] 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.

[2] P. A. Stolk, F. P. Widdershoven and D. B. M. Klaassen, "Modeling statistical dopant fluctuations in MOS transistors", *IEEE Trans. on Electron Devices*, vol. 45, no. 9, pp. 1960-1971, 1998.

[3] O. Hirabayashi, A. Kawasumi, A. Suzuki, Y. Takeyama, K. Kushida, T. Sasaki, A. Katayama, G. Fukano, Y. Fujimura, T. Nakazato, Y. Shizuki, N. Kushiyama, and T. Yabe, "A process-variation-tolerant dual-power-supply SRAM with 0.179um^2 cell in 40 nm CMOS using level programmable wordline driver," *ISSCC Dig. Tech. Papers*, pp. 458-459, Feb. 2009.

[4] A. Bhavnagarwala, S. Kosonocky, C. Radens, K. Stawiasz, R. Mann, Q. Ye and K. Chin, "Fluctuation limits & scaling opportunities for CMOS SRAM cells," *IEEE IEDM Dig. Tech. Papers*, pp. 659-662, Nov., 2005.

[5] K. Takeda, H. Ikeda, Y. Hagihara, M. Nomura, and H. Kobatake, "Redefinition of write margin for next-generation SRAM and write-margin monitoring circuit," *IEEE ISSCC, Digest of Tech. Papers*, pp.630-631, Feb. 2006.

[6] H. Makino, S. Nakata, H. Suzuki, S. Mutoh, M. Miyama, T. Yoshimura, S. Iwade, and Y. Matsuda, "Re-examination of SRAM cell write margin definitions in view of predicting distribution," *Trans. on Circuits and Systems II: Brief Express*, vol. 58, issue 4, pp.230-234, April 2011.

[7] Visit http://www.eas.asu.edu/~ptm/ for "45nm PTM HP model: V2.1," Predictive Technology Model.