

# Program Start Bias Grouping to Compensate for the Geometric Property of a String in 3-D NAND Flash Memory

SUNGJU KIM<sup>1</sup>, SANGMIN AHN<sup>1</sup>, SECHUN PARK<sup>2</sup>, JONGWOO KIM<sup>2</sup>,  
AND HYUNGCHEOL SHIN<sup>1,3</sup> (Senior Member, IEEE)

<sup>1</sup> Department of Electrical and Computer Engineering, Seoul National University, Seoul 08826, South Korea

<sup>2</sup> NAND Design Team, SK Hynix Inc., Icheon 17336, South Korea

<sup>3</sup> Integra Semiconductor Ltd., Seoul, South Korea

CORRESPONDING AUTHOR: H. SHIN (e-mail: hcshin@snu.ac.kr)

This work was supported by SK Hynix Inc.

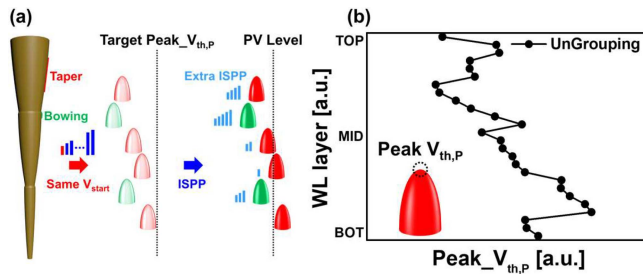
**ABSTRACT** The string (STR) with various geometrical profiles in 3-D NAND flash cause the degradation of program efficiency. This is because the program speed differences among WL layers within the STR are caused by the geometrical properties observed through measurement results. In this work, we propose the method to reduce the program speed differences based on a word-line (WL) grouping in terms of threshold voltage ( $V_{th}$ ) distribution to compensate for the program start voltage ( $V_{start}$ ). To address various geometrical profiles, we consider a flexible compensation method through  $\Delta Peak\_V_{th}$ , i.e., the net amount of movement from the erase to the program state.  $\Delta Peak\_V_{th}$  according to WL layers clearly distinguished the geometrical properties among WL layers, and through this, the linearity of  $\Delta Peak\_V_{th}$  is frequently observed for specific WL layer intervals with taper profile. Utilizing this linearity, we conducted the WL grouping and successfully demonstrated  $V_{start}$  compensation by applying the proposed method to each WL group through the measurement of a commercial 3-D NAND package. Moreover, the reduced WL grouping method is also contrived to relax circuit design complications and evaluated the usefulness of the proposed method.

**INDEX TERMS** 3-D NAND flash, critical dimension (CD), geometrical property, threshold voltage ( $V_{th}$ ) distribution, word line (WL) grouping, start bias ( $V_{start}$ ), program speed.

## I. INTRODUCTION

Owing to the limitations associated with structural scaling down in 2-D planar NAND flash memory, 3-D NAND flash memory with a vertical hole has been developed [1], [2]. Specifically, word-line (WL) stacking technology has emerged as a key process for high-capacity implementation [3], [4]. However, the process difficulty increases as more WLs are stacked. Accordingly, the possibility of forming a non-uniform cell profile, such as a tapered cell or bowed cell, increases [5], [6]. Because of the non-uniformity of these geometric profiles in the string (STR), the critical dimension (CD) of each cell is different, causing capacitance variation not only WL to WL, as well as but also WL to channel [7], [8]. This, in turn,

gives rise to the program speed differences among cells within STR, ultimately leading to WL loading variation in reaching the target program voltage during the same program time [8], [9]. Fig. 1(a) conceptually illustrates the geometric profile of the STR in 3-D NAND Flash memory in terms of threshold voltage ( $V_{th}$ ) distribution. In order to check these geometrical properties, the program speed was measured in terms of the peak of  $V_{th}$  distribution ( $Peak\_V_{th}$ ) which represents the most cells are located. The program start voltage ( $V_{start}$ ), which is the first pulse of the incremental step pulse program (ISPP), was applied at the same magnitude in the STR after the block deep erase. As a result, Fig. 1(b) shows that the peak of the program  $V_{th}$  distribution ( $Peak\_V_{th,P}$ ) points are inconsistent, which means



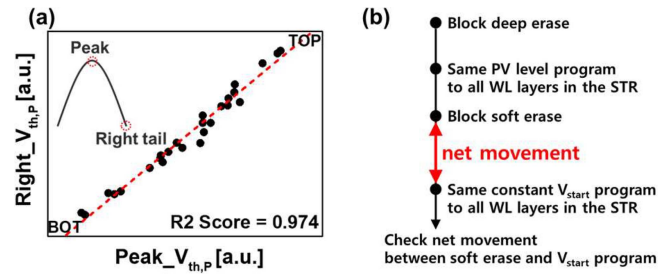
**FIGURE 1.** (a) Conceptual illustration with geometric properties of the STR in 3-D NAND Flash memory. (b) Measurement result of the program speed for each WL layer.

there are the program speed differences among WL layers in the STR. This signifies that WL loading variation cause  $Peak\_V_{th,P}$  to deviate from the target  $Peak\_V_{th,P}$ , as shown in Fig. 1(a). Furthermore, the relatively small  $Peak\_V_{th,P}$  point requires extra ISPP to pass through the program verify (PV) level. As this extra ISPP increases, both program time and energy consumption keep increasing [10], [11]. As a result, the program performance is degraded. In order to improve program performance, the development of a method to compensate for the program speed differences among WL layers is imperative. In previous research, Kang et al. introduced a variable-pulse scheme to cope with WL loading variation by controlling program pulse duration [8]. While certainly a commendable approach, for STRs with various geometrical properties, this strategy could lead to an overall increase in program time, thus potentially degrading program efficiency. Furthermore, while presenting a methodology to improve WL loading variation among WL layers, it did not exhibit actual improved outcomes. Hence, we propose an alternative method to improve WL loading variation by controlling program pulse amplitude instead of its duration. This approach is designed to address the RC delay challenge and subsequently mitigates the program speed differences among WL layers. Additionally, to make it adaptable to various geometrical profiles, we introduce WL grouping methodology. Based on this methodology, we successfully compensated for the program speed differences among WL layers within WL group. Consequently, we present the measurement results after applying the proposed method.

## II. GROUPING METHODOLOGY AND START BIAS COMPENSATION METHOD

### A. WL GROUPING METHODOLOGY

Within a single Die, non-uniformity of the CD is caused by the process variation [12]. Thus, various geometrical profiles exist between the block (BLO)s within plane (PLA)s that comprise the die [13]. To overcome these challenges, we have developed a flexible compensation method applicable to various geometrical profiles. Therefore, we need to accurately identify the geometrical properties in the STR. As shown in Fig. 1(b), it is possible to confirm the differences in program speed among WL layers, but it is difficult to identify clear geometrical properties. The program speed is physically

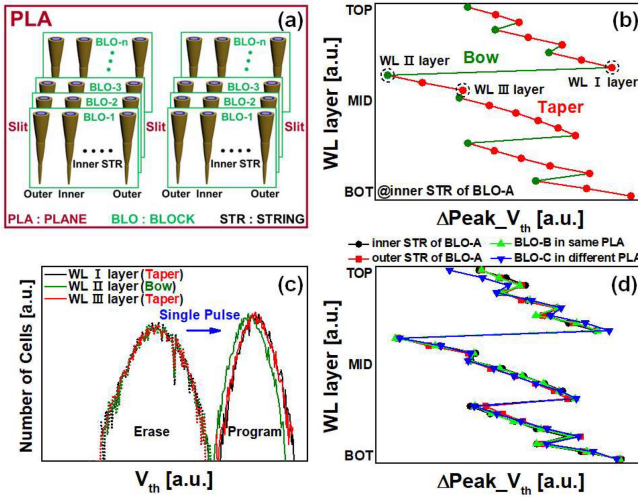


**FIGURE 2.** (a) The relation between  $Peak\_V_{th,P}$  and  $Right\_V_{th,P}$  for each WL layer at relatively low program voltage. (b) The flexible measurement method of the net movement reflecting the geometrical properties.

dependent on RC delay and electric field concentration caused by the CD variation [9], [14], [15]. This implies the net movement, induced by applying the single program pulse with the same magnitude to all WL layers reflects the geometrical properties of the corresponding WL layer. Consequently, the amount of the net movement from the erase state to the program state can reflect the geometrical properties. We utilize  $Peak\_V_{th}$  as the reference point to measure the amount of the net movement because a page unit consists of cells from different STRs within the same WL layer [16]. Thus,  $Peak\_V_{th}$  reflects the average geometrical properties of the different STRs within the page. Moreover, it is appropriate to regard the program speed of  $V_{th}$  distribution as the average program speed of the most numerous cells within it. However, in practice, in the case of multi-level cells,  $V_{start}$  needs to be determined based on the right tail ( $Right\_V_{th,P}$ ) to prevent the overshoot problem of fast cells. Fig. 2(a) shows the trend between  $Peak\_V_{th,P}$  and  $Right\_V_{th,P}$  for each WL layer at relatively low program voltage, such as  $V_{start}$ . This indicates no problem in defining the  $V_{start}$  based on  $Peak\_V_{th,P}$ . Measurements were conducted following the approach in Fig. 2(b) to consider the initial erase state reflecting various geometrical profiles. Particularly,  $Peak\_V_{th,P}$  of each WL layer in the STR was aligned as closely as possible to the same PV level before the soft erase in order to focus on feasible geometrical properties [16], [17]. For analyzing the amount of the net movement in terms of  $Peak\_V_{th}$ , Eq. (1) was used to calculate  $\Delta Peak\_V_{th}$ , which is the net travel amount from the peak of the erase  $V_{th}$  distribution ( $Peak\_V_{th,E}$ ) to  $Peak\_V_{th,P}$ . For all WL layers,  $Peak\_V_{th,E}$  was measured after soft erase, and then,  $Peak\_V_{th,P}$  was measured after the single  $V_{start}$  program with the same magnitude.

$$\Delta Peak\_V_{th} = Peak\_V_{th,P} - Peak\_V_{th,E} \quad (1)$$

Fig. 3(a) shows the physical organization of the 3-D NAND Flash memory measured in this paper. Fig. 3(b) shows  $\Delta Peak\_V_{th}$  according to the WL layers of the inner STR that is located far away from the slit in BLO-A, which is any BLO. Considering the program speed variation caused by the geometrical properties, it is expected that an increase in  $\Delta Peak\_V_{th}$  from the top WL to the bottom WL layer would

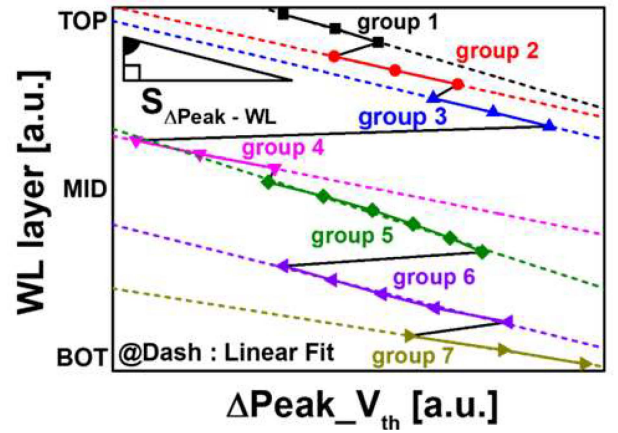


**FIGURE 3.** (a) Physical organization of 3-D NAND flash memory. (b)  $\Delta\text{Peak}_{V_{th}}$  of each WL layer in the inner STR of BLO A when  $V_{start}$  with the same magnitude is applied to all WL layers. (c) Measured  $V_{th}$  distribution reflecting geometrical properties at WL I, II, and III layers. (d) Data proving the generality of  $\Delta\text{Peak}_{V_{th}}$  through the STR of different BLOs in other PLAs.

result from the STR having a tapered profile with a relatively thin CD. Meanwhile, a decrease in  $\Delta\text{Peak}_{V_{th}}$  would result from the STR having a bowed profile with a relatively thick CD. To confirm the geometrical properties predicted in Fig. 3(b), the program speeds for the WL I, II, and III layers located near bow property were measured. Then, the  $\text{Peak}_{V_{th,E}}$  points for the WL I, II, and III layers were aligned by regulating the erase voltage, thereby minimizing the initial state variation, as shown in Fig. 3(c). As a result,  $\text{Peak}_{V_{th,P}}$  at the WL II layer with a bowed profile shows less movement than  $\text{Peak}_{V_{th,P}}$  at the WL I and III layers with a tapered profile. It is confirmed that  $\Delta\text{Peak}_{V_{th}}$  reflects the geometrical properties of the STR well. The trend of  $\Delta\text{Peak}_{V_{th}}$  also shows the same linear tendency at the outer STR located close to the slit in BLO-A as well as the inner STR of BLOs in the same and different PLAs, as shown in Fig. 3(d). It is predicted that there will be linearity because the CD reduction by the non-ideal etching profile appears in a linear form. To compensate for this geometrical property, the program speed differences among WL layers should be reduced, which can be achieved by adjusting the linearly increased  $\Delta\text{Peak}_{V_{th}}$  to the same amount. Therefore, we decided on a linear interval where the STR has a tapered profile as the criterion for the WL group to compensate for the program speed.

## B. PROGRAM SPEED COMPENSATION METHOD AMONG WL LAYERS

The linearly increased  $\Delta\text{Peak}_{V_{th}}$  can be adjusted to the same amount by reducing the program speed differences among WL layers. So, we controlled the program pulse amplitude by linearly reducing  $V_{start}$  from the top WL to the bottom WL layer within the group at the same program



**FIGURE 4.** Total of seven WL groups based on the linear interval in Fig. 2 (b).

duration. The linearly controlled  $V_{start}$  was calculated for each WL group through a linear equation, and the slope ( $= S$ ) of the linear equation was calculated using Eq. (2).

$$S_{\text{Start-WL}} = \frac{V_{start,f} - V_{start,i}}{f - i} = \gamma \times \frac{S_{\Delta\text{Peak-WL}}}{S_{\Delta\text{Peak-Start}}} \quad (2)$$

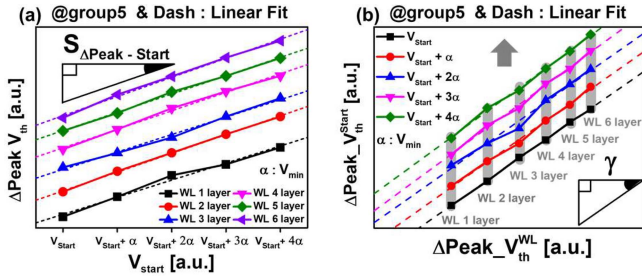
where the constants  $i, f$  are WL layer numbers that respectively correspond to the first and last numbers within the group. In Eq. (2), the relationship of  $V_{start}$  according to the WL layer is derived into three variables based on the chain rule. The variables  $S_{\Delta\text{Peak-WL}}$ ,  $S_{\Delta\text{Peak-Start}}$ , and  $\gamma$  are respectively

$$S_{\Delta\text{Peak-WL}} = \frac{\Delta\text{Peak}_{V_{th,f}}^{WL} - \Delta\text{Peak}_{V_{th,i}}^{WL}}{f - i} \quad (3)$$

$$S_{\Delta\text{Peak-Start}} = \frac{\Delta\text{Peak}_{V_{th,f}}^{\text{Start}} - \Delta\text{Peak}_{V_{th,i}}^{\text{Start}}}{V_{start,f} - V_{start,i}} \quad (4)$$

$$\gamma = \frac{\Delta\text{Peak}_{V_{th,f}}^{\text{Start}} - \Delta\text{Peak}_{V_{th,i}}^{\text{Start}}}{\Delta\text{Peak}_{V_{th,f}}^{WL} - \Delta\text{Peak}_{V_{th,i}}^{WL}} \quad (5)$$

where  $\Delta\text{Peak}_{V_{th}}^{WL}$  and  $\Delta\text{Peak}_{V_{th}}^{\text{Start}}$  refer to  $\Delta\text{Peak}_{V_{th}}$  according to the WL layer and  $V_{start}$  in Fig. 4 and Fig. 5(a), respectively. Fig. 4 shows the total of seven WL groups based on the linear intervals. Figs. 5(a) and 5(b) exhibit consistent linearity both within each WL layer and between WL layers in group 5 within the range of  $V_{start}$  being regulated. This regulated  $V_{start}$  is a single program pulse increased by the minimum voltage ( $V_{min}$ ), representing the lowest voltage resolution that is measurable range using our equipment. Moreover, the linearity of group 5, as presented in Fig. 5(a) and Fig. 5(b), also shows the same tendency in other groups (see the Appendix). Given the linearity observed in the measurement data from Fig. 4 and Fig. 5,  $S_{\Delta\text{Peak-WL}}$ ,  $S_{\Delta\text{Peak-Start}}$ , and  $\gamma$  can be determined through linear fitting. Here,  $S_{\Delta\text{Peak-Start}}$  is calculated as the average value of the linear fitting slope for each WL layer, as shown in Fig. 5(a). Similarly,  $\gamma$  is also calculated as the average value of the linear fitting slope between WL layers in Fig. 5(b). Through

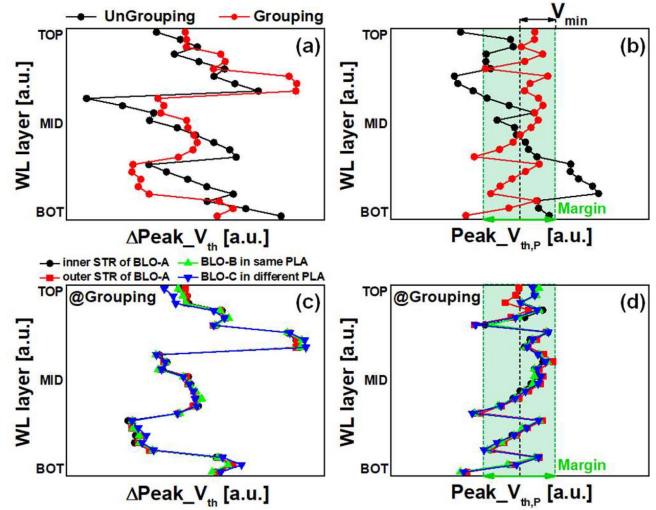


**FIGURE 5. (a) Linearity of each WL layer in group 5. (b) Linearity between WL layers in group 5 according to increasing  $V_{start}$ .**

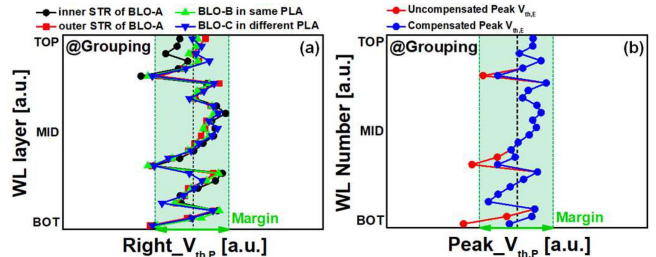
this procedure, we obtain  $S_{Start-WL}$  for each group, and then calculate the corresponding  $V_{start}$  for each WL layer.

### III. MEASUREMENT RESULTS AND DISCUSSION

As shown in Fig. 6,  $\Delta Peak\_V_{th}$  compensated for the taper property was measured by applying the linear  $V_{start}$  calculated based on Eq. (2). Fig. 6(a) shows that  $\Delta Peak\_V_{th}$  points tend to converge to a similar amount for each WL group. However,  $\Delta Peak\_V_{th}$  points do not converge completely in the same amount due to the approximately applied  $V_{start}$  by the limitation in resolution of our measurement equipment. It implies the measurement equipment with lower resolution can apply more detailed  $V_{start}$ , enabling a more completely convergence in the same amount. Therefore, the proposed linear  $V_{start}$  calculated based on Eq. (2) can sufficiently compensated for the taper property in each WL group. With this compensation of  $\Delta Peak\_V_{th}$ , as shown in Fig. 6(b),  $Peak\_V_{th,P}$  points of each WL layer are well gathered within the margin with a range of  $\pm V_{min}$  away from the target  $Peak\_V_{th,P}$ . This means that, since the program step voltage is larger than  $V_{min}$ ,  $Peak\_V_{th,P}$  points within the margin can be sufficiently compensated for. To confirm the generality of Eq. (2), a measurement was conducted on the STR of BLOs in the same and different PLAs by applying the linear  $V_{start}$ . As a result, Fig. 6(c) shows that the taper property for each WL group in Fig. 3(d) is also compensated for with the same tendency, thus resulting in  $Peak\_V_{th,P}$  points being well gathered into the margin, as shown in Fig. 6(d). However, the effectiveness of our method in controlling  $V_{start}$  for each group must be validated through  $Right\_V_{th,P}$  grouping result as presented in Fig. 7(a). In Fig. 7(a), it is evident that  $Right\_V_{th,P}$  points are well gathered within the margin, thereby affirming the validity of our approach in compensating for the program speed differences among WL layers through  $\Delta Peak\_V_{th}$  compensation. Nevertheless, in specific WL layers, there are points for both  $Peak\_V_{th,P}$  and  $Right\_V_{th,P}$  that fail to gather within the margin. This result is attributed to a significant mismatch in  $Peak\_V_{th,E}$  within a particular WL group, as erase WL grouping was not applied [18]. Since the focus of this paper is on program WL grouping, this work was conducted based on the situation in which erase WL grouping was excluded. But, in practice, both WL groupings



**FIGURE 6. (a)  $\Delta Peak\_V_{th}$  of each WL group compensated for geometrical property. (b)  $Peak\_V_{th,P}$  gathered at the target  $Peak\_V_{th,P}$  by the linear  $V_{start}$ . (c), (d) Data proving the generality of compensation for geometrical property of each WL group and linear  $V_{start}$ .**

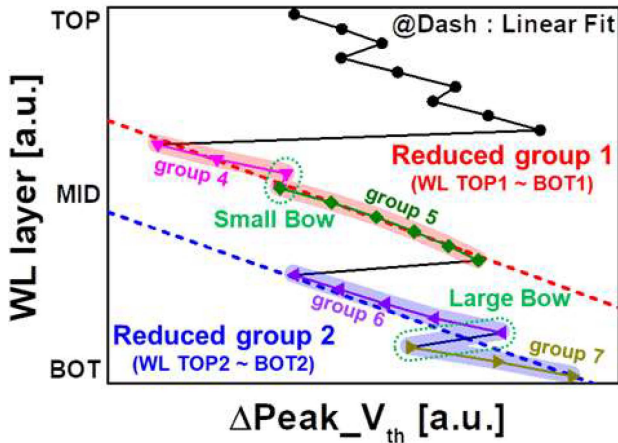


**FIGURE 7. (a) Verification of  $Right\_V_{th,P}$  grouping through  $\Delta Peak\_V_{th}$  compensation. (b) Complete gathering of  $Peak\_V_{th,P}$  by compensating initial bow property.**

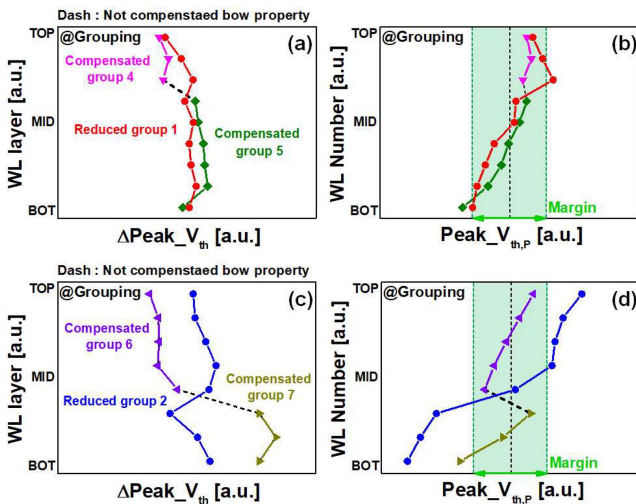
are applied. Therefore, in order to demonstrate the result of applying our program grouping method after erase grouping, we compensated for significant mismatched  $Peak\_V_{th,E}$  among WL layers in Fig. 6(d). As a result, Fig. 7(b) shows that  $Peak\_V_{th,P}$  points are completely gathered in the margin. And we confirmed that  $Right\_V_{th,P}$  points also showed the same result. This indicates that utilizing our program WL grouping method after erase WL grouping can lead to a more efficient compensation of  $Peak\_V_{th,P}$  with the expectation that the differences in  $Peak\_V_{th,P}$  among WL layers within the margin will further decrease. Further the development of erase WL grouping method which reflects erase physics for geometrical properties will be carried out in the future.

### IV. REDUCED GROUPING METHODOLOGY AND MEASUREMENT RESULTS

As the number of WL groups increases, a problem arises in that the circuit design becomes complicated. Therefore, we considered a method for reducing the number of the WL groups as shown in Fig. 8 by merging between neighboring WL groups. The reduced group 1 consists of group 4

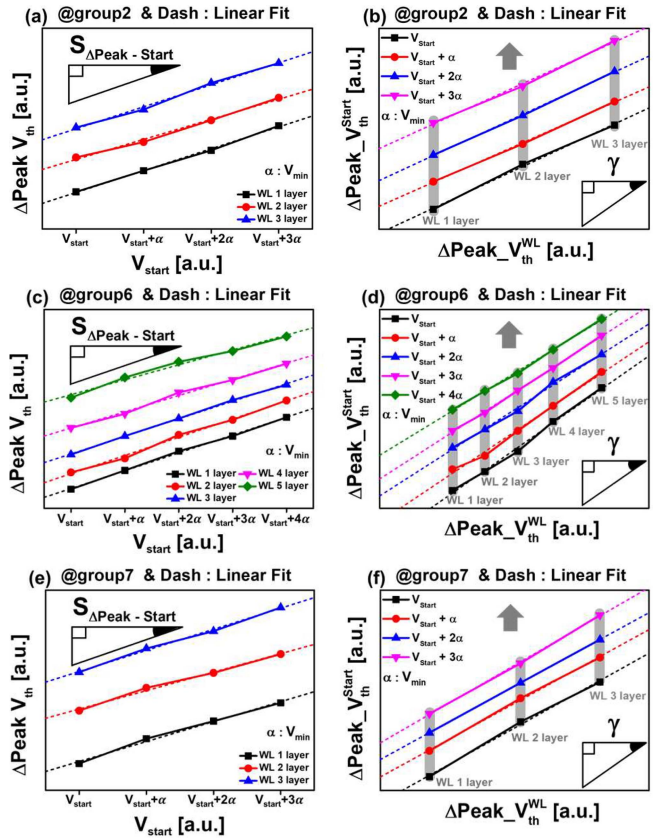


**FIGURE 8.** Reduced WL grouping between neighboring WL groups. Reduced group 1 (Group 4 + Group 5) has a small bow property and reduced group 2 (Group 6 + Group 7) has a large bow property.



**FIGURE 9.** Comparison of compensation of geometrical properties for two reduced groups in terms of (a), (b) small bow property and (c), (d) large bow property between neighboring WL groups. And compensated groups are results of detailed WL grouping, as depicted in Fig. 6.

and 5. Similarly, the reduced group 2 consists of group 6 and 7. The linear  $V_{start}$  calculated based on Eq. (2) regarding the two reduced WL groups was applied to each reduced WL group. As a result, in Fig. 9(a), the reduced group 1 shows effective convergence between compensated group 4 and 5, as depicted in Fig. 6(a). Consequently, most of  $Peak_{V_{th,P}}$  points are well gathered into the margin. In contrast, Fig. 9(c) illustrates that the reduced group 2 do not converge as effectively as the reduced group 1, and as shown in Fig. 9(d), most of  $Peak_{V_{th,P}}$  points are not gathered into the margin. This indicates the limitation of the linear approximation in addressing the large degree of bowing between neighboring WL groups. In other words, when the degree of bowing between the adjacent WL groups is small, it is possible to apply a reduced WL grouping as shown in Fig. 9(b). However, there is a trade-off that



**FIGURE 10.** (a) Linearity of each WL layer in group 2. (b) Linearity between WL layers in group 2. (c) Linearity of each WL layer in group 6. (d) Linearity between WL layers in group 6. (e) Linearity of each WL layer in group 7. (f) Linearity between WL layers in group 7 according to increasing  $V_{start}$ .

gathers less at the target  $Peak_{V_{th,P}}$  than the detailed WL grouping such as groups 4,5, and 6,7. Therefore, it is crucial to determine the degree of bowing to which the reduced WL grouping is applicable, considering the acceptable trade-off. Taking this consideration into account, when applying the reduced WL grouping in Fig. 4, it is possible to reduce the number of WL groups from 7 to a minimum of 6 and a maximum of 4 groups. Thus, the reduced WL grouping is advantageous in its flexibility to enable selective merging of neighboring groups according to the acceptable trade-off.

## V. CONCLUSION

This paper proposed the method to compensate the geometrical properties by reducing the program speed differences among WL layers through controlling program voltage. To develop the flexible method capable of considering the various geometrical profiles of the STRs, we utilized  $\Delta Peak_{V_{th}}$ .  $\Delta Peak_{V_{th}}$  clearly distinguished the geometrical properties of each WL layer within the STR and exhibit frequently the linearity attributed to taper property according to WL layers. After conducting the WL grouping based on these linear intervals, we successfully compensated for the taper property of  $\Delta Peak_{V_{th}}$  through our linear  $V_{start}$  equation.

Through the compensation of  $\Delta\text{Peak\_V}_{\text{th}}$ , we observed that most of  $\text{Peak\_V}_{\text{th,p}}$  points gathered within the margin. This means that our proposed method is valid to controlling  $V_{\text{start}}$  in response to each WL layer, as  $\text{Peak\_V}_{\text{th,p}}$  and  $\text{Right\_V}_{\text{th,p}}$  show the same tendency according to the WL layer within  $V_{\text{start}}$  range. Moreover, to solve the problem in which the circuit design is complicated by the number of WL groups, the reduced WL grouping method was contrived. As a result, when the degree of bowing between neighboring WL groups is sufficiently small, then the reduced WL grouping will be applicable. Although a trade-off that gathers less at the target  $\text{Peak\_V}_{\text{th,p}}$  occurs, it is possible to selectively apply the reduced WL grouping according to acceptable trade-off. Through this flexibility, the reduced WL grouping method can alleviate the circuit complexity. Finally, the WL grouping method proposed in this paper reflects program physics for geometrical properties. Therefore, the flexibility applicable to various geometrical profiles of STRs will play a very important role in efficient WL grouping.

## APPENDIX

See the Fig. 10.

## REFERENCES

- [1] H. Aochi, "BiCS flash as a future 3D non-volatile memory technology for ultra high density storage devices," in *Proc. IEEE Int. Memory Workshop*, 2009, pp. 1–2, doi: [10.1109/IMW.2009.5090581](https://doi.org/10.1109/IMW.2009.5090581).
- [2] H. Tanaka et al., "Bit cost scalable technology with punch and plug process for ultra high density flash memory," in *Proc. IEEE Symp. VLSI Technol.*, 2007, pp. 14–15, doi: [10.1109/VLSIT.2007.4339708](https://doi.org/10.1109/VLSIT.2007.4339708).
- [3] J. H. Kim et al., "Highly manufacturable 7th generation 3D NAND flash memory with COP structure and double stack process," in *Proc. Symp. VLSI Technol.*, 2021, pp. 1–2.
- [4] C. W. Yoon, "Trend, hurdle, and core technology for next generation high performance storage system," in *Proc. IEEE Asian Solid-State Circuits Conf.*, 2021, pp. 64–65.
- [5] M. Fan, R. Ranjit, A. Thurber, and D. Engelhard, "High resolution profiles of 3D NAND pillars using x-ray scattering metrology," in *Proc. 35th Metrol., Insp., Process Control Semicond. Manuf.*, 2021, pp. 120–126, doi: [10.1117/12.2585217](https://doi.org/10.1117/12.2585217).
- [6] W. Sun, H. Ohta, T. Ninomiya, and Y. Goto, "High-voltage CD-SEM-based application to monitor 3D profile of high-aspect-ratio features," *J. Micro-Nanolithogr. MEMS MOEMS*, vol. 19, no. 2, pp. 1–1, May 2020, doi: [10.1117/1.jmm.19.2.024002](https://doi.org/10.1117/1.jmm.19.2.024002).
- [7] A. Tilson and M. Strauss, "STEM/EDS metrology and statistical analysis of 3D NAND devices," in *Proc. IEEE Int. Symp. Phys. Failure Anal. Integr. Circuits (IPFA)*, 2018, pp. 1–5, doi: [10.1109/IPFA.2018.8452545](https://doi.org/10.1109/IPFA.2018.8452545).
- [8] D. Kang et al., "7.1 256Gb 3b/cell V-NAND flash memory with 48 stacked WL layers," in *Proc. IEEE Int. Solid-State Circuits Conf. (ISSCC)*, San Francisco, CA, USA, 2016, pp. 130–131, doi: [10.1109/ISSCC.2016.7417941](https://doi.org/10.1109/ISSCC.2016.7417941).
- [9] W. Jeong et al., "A 128 Gb 3b/cell V-NAND flash memory with 1 Gb/s I/O rate," *IEEE J. Solid-State Circuits*, vol. 51, no. 1, pp. 204–212, Jan. 2016, doi: [10.1109/JSSC.2015.2474117](https://doi.org/10.1109/JSSC.2015.2474117).
- [10] C. Huang, M. Yang, E. Yang, T. H. Yang, and K. C. Chen, "Within-wafer CD variation induced by wafer shape," in *Proc. SPIE*, 2016, pp. 1055–1063, doi: [10.1117/12.2216048](https://doi.org/10.1117/12.2216048).
- [11] V. Mohan et al., "Modeling power consumption of NAND flash memories using flashpower," *IEEE Trans. Comput.-Aided Des. Integr. Circuits Syst.*, vol. 32, no. 7, pp. 1031–1044, Jul. 2013, doi: [10.1109/TCAD.2013.2249557](https://doi.org/10.1109/TCAD.2013.2249557).
- [12] H. Maejima et al., "A 512Gb 3b/Cell 3D flash memory on a 96-word-line-layer technology," in *Proc. IEEE Int. Solid State Circuits Conf. (ISSCC)*, San Francisco, CA, USA, 2018, pp. 336–338, doi: [10.1109/ISSCC.2018.8310321](https://doi.org/10.1109/ISSCC.2018.8310321).
- [13] Y. Pan, G. Dong, and T. Zhang, "Error rate-based wear-leveling for NAND flash memory at highly scaled technology nodes," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 21, no. 7, pp. 1350–1354, Jul. 2013, doi: [10.1109/TVLSI.2012.2210256](https://doi.org/10.1109/TVLSI.2012.2210256).
- [14] Y. Seok Oh et al., "Impact of etch angles on cell characteristics in 3D NAND flash memory," *Microelectron. J.*, vol. 79, pp. 1–6, Sep. 2018, doi: [10.1016/j.mejo.2018.06.009](https://doi.org/10.1016/j.mejo.2018.06.009).
- [15] R. Micheloni, S. Aritome, and L. Crippa, "Array architectures for 3-D NAND flash memories," in *Proc. IEEE*, vol. 105, no. 9, pp. 1634–1649, Sep. 2017, doi: [10.1109/JPROC.2017.2697000](https://doi.org/10.1109/JPROC.2017.2697000).
- [16] M. Raquibuzzaman, M. M. Hasan, A. Milenkovic, and B. Ray, "Layer-to-layer endurance variation of 3D NAND flash memory," in *Proc. IEEE Int. Rel. Phys. Symp. (IRPS)*, Dallas, TX, USA, 2022, pp. 1–5, doi: [10.1109/IRPS48227.2022.9764441](https://doi.org/10.1109/IRPS48227.2022.9764441).
- [17] S. Sakib, A. Milenkovic, and B. Ray, "Flash-DNA: Identifying NAND flash memory origins using intrinsic array properties," *IEEE Trans. Electron Devices*, vol. 68, no. 8, pp. 3794–3800, Aug. 2021, doi: [10.1109/TED.2021.3087454](https://doi.org/10.1109/TED.2021.3087454).
- [18] D. Kang et al., "13.4 A 512Gb 3-bit/cell 3D 6th-generation V-NAND flash memory with 82MB/s write throughput and 1.2Gb/s interface," in *Proc. IEEE Int. Solid-State Circuits Conf. (ISSCC)*, San Francisco, CA, USA, 2019, pp. 216–218, doi: [10.1109/ISSCC.2019.8662493](https://doi.org/10.1109/ISSCC.2019.8662493).