Nanoscale Advances

ROYAL SOCIETY OF CHEMISTRY

PAPER

View Article Online
View Journal

Cite this: DOI: 10.1039/d6na00054a

Residual stress modulation as a pathway to reliable multilevel 3D NAND flash storage

Ruifu Zhou, [a] Ik-Jyae Kim, [a] Sejun Park, [b] Hyukchang Kwon [b] and Jang-Sik Lee *[a]

Received 22nd January 2026
Accepted 24th May 2026

DOI: 10.1039/d6na00054a

rsc.li/nanoscale-advances

Three-dimensional (3D) NAND flash memory achieves high density through the vertical stacking of memory cells. However, increasing the number of stacked layers induces significant residual stress, which adversely impacts both structural integrity and electrical characteristics of 3D NAND devices. Thus, effective analysis and management of residual stress in 3D NAND are required. This study employs a Sentaurus technology computer-aided design (TCAD) simulation to examine the effect of residual stress on the reliability of charge-trap 3D NAND. The results show that increasing the compressive stress in the charge-trap nitride (CTN) layer improves data retention through band-edge modulation. To utilize these effects, we propose the insertion of a stress engineering layer (SEL) between the poly-Si channel and surrounding oxide filler. The SEL increases the compressive stress within the CTN layer. Thus, the retention degradation can be reduced. The proposed SEL technique demonstrates the potential to improve the reliability of 3D NAND flash memory by modulating residual stress.

Introduction

Due to limitations in planar NAND flash scaling (*e.g.*, lithographic resolution and cell-to-cell interference), three-dimensional (3D) NAND flash memory with vertically stacked cells has been developed as a high-density, cost-effective data storage solution.[1–3] The memory density in 3D NAND continues to increase as a larger number of word-line (WL) layers are stacked vertically. However, increasing the number of stacking layers also increases mechanical residual stress within 3D NAND, thereby causing physical deformations (*e.g.*, wafer warpage, bowing, and cracking) and adversely affecting electrical characteristics.[1,4] Beyond these structural concerns, the local distribution of residual stress within the stacked structure should also be carefully considered, because it can directly impact critical functional layers. This may further affect electrical characteristics and long-term reliability during operation.[5,6] Therefore, optimizing residual stress within 3D NAND architectures is critical to enable further scaling.[7,8]

Residual stress significantly affects key device metrics such as mobility, on-current, and threshold voltage (V_T).[6,7,9] In silicon channels, residual stress can alter conduction band valley degeneracy, thus modifying effective mass and scattering rates, which, in turn, affect carrier transport.[10,11] Beyond its influence on electrical performance, residual stress can reduce the long-term reliability of memory cells.[12,13] Thus, residual stress simulation has emerged as an important tool for predicting and managing stress throughout the 3D NAND fabrication process.[6,8,14] However, previous studies on mechanical stress in 3D NAND have mainly focused on electrical characteristics such as mobility, on-current, and threshold voltage, whereas the relationship between local residual stress and retention reliability in charge-trap 3D NAND has received limited attention. In particular, the influence of the local charge-trap nitride (CTN) stress state on charge loss and retention degradation remains insufficiently clarified.

In this work, we investigate the residual stress distribution in a process-flow-based 3D NAND structure using Sentaurus technology computer-aided design (TCAD) simulations. We then analyze retention behavior by varying the vertical residual stress applied to the CTN layer. Our results demonstrate that higher compressive stress in the CTN layer improves data retention. Based on this finding, we propose the insertion of a stress engineering layer (SEL) between the poly-Si channel and the surrounding oxide filler to achieve reliable multilevel 3D NAND flash storage. These results highlight the strong potential of residual stress modulation as an optimization approach for ensuring the reliability of 3D NAND memory cells.

Results and discussion

Residual stress in 3D NAND flash memory

A ten-layer 3D NAND structure was simulated by Sentaurus TCAD, as illustrated in Fig. 1. Residual stress in 3D NAND stems from the multilayer process sequence and its thermal history; therefore, in our simulations we construct the 3D NAND

[a]Pohang University of Science and Technology (POSTECH), Pohang 37673, Korea. E-mail: jangsik@postech.ac.kr
[b]Device Solutions, Samsung Electronics Co., Hwaseong 18448, Korea

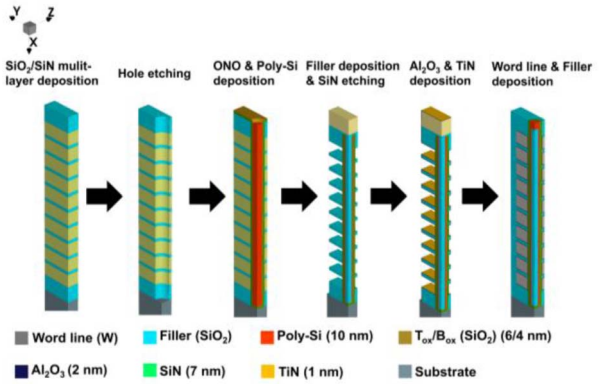

Fig. 1 Simulated 3D NAND flash memory using the technology computer-aided design tool. W, Al_2O_3 + SiO_2, SiN, SiO_2, and poly-Si were used for WL, B_{OX}, CTN, T_{OX}, and the channel, respectively.

structure incorporating the sequential fabrication stages to model the cumulative residual stress arising from the process thermal history. Specifically, the process flow is as follows: sacrificial silicon nitride (SiN) and silicon oxide (SiO_2) multi-layers were deposited on an Si wafer. Next, a hole with a radius of 55 nm was etched, and 4 nm-thick SiO_2 blocking oxide (B_{OX}), 7 nm-thick SiN (CTN), and 6 nm-thick SiO_2 tunneling oxide (T_{OX}) were deposited inside the hole. Then, the hole was filled with a poly-Si channel, and a macaroni structure was formed *via* etching and SiO_2 filling. Subsequently, after the removal of sacrificial SiN layers, a 2 nm-thick Al_2O_3 and a 1 nm-thick TiN metal barrier were sequentially deposited, followed by the deposition of tungsten (W) WL encapsulated in SiO_2. Finally, arsenic-doped (10^{20} cm^{-3}) poly-Si was deposited to establish a bit-line (BL).[2,15] Although this simulation utilizes a ten-layer model, it effectively captures the local stress behaviors governed by the stack configuration and material properties, offering critical insights into more complex, highly-stacked 3D NAND structures.[8,16]

Based on this structure, the stress distribution in each layer was then analyzed. The vertical stress (along the channel direction) was extracted along the cut line AA′ shown in Fig. 2a. The corresponding stress profile across different material layers is depicted in Fig. 2b. Residual stresses in 3D NAND structures occur during thermal processing due to mismatches in the coefficient of thermal expansion (CTE) among constituent materials.[10,17] In our previous work we calibrated the stress model through experiments by depositing 50 nm SiN and 30 nm SiO_2 thin films using chemical vapor deposition (CVD), measuring the residual stress with an FSM-500 TC wafer-curvature system. Key mechanical parameters, including CTE and Young's modulus, were extracted and validated against experimental data.[7] In this work, we continue to use the same calibrated parameters for SiN (CTE: 3×10^{-6}/K) and SiO_2 (CTE: 1×10^{-6}/K), while the properties of other layers are taken from the literature.[18] The CTE, Poisson's ratio (ν), and Young's modulus (E) of each material were used in the simulation to investigate the residual stress, and the corresponding mechanical parameters are summarized in Table S1.[7,18,19] Specifically, Al_2O_3 (CTE: 8.18×10^{-6}/K) and TiN (CTE: 10.8×10^{-6}/K), which have higher CTE values than surrounding materials (*e.g.*, poly-Si (CTE: 2.8×10^{-6}/K) and W (CTE: 4.6×10^{-6}/K)), exhibit greater constrained shrinkage during cooling. As a result, significant tensile stresses are induced within the TiN and Al_2O_3 layers. In contrast, compressive stresses arise in other layers, including approximately −0.2 GPa in the CTN layer, as well as notable compressive stresses in the T_{OX} layer and B_{OX} layer. Such residual stress distributions can potentially influence the band structures and carrier transport properties of the memory cells, thus affecting their overall electrical performance and reliability.[8,16,18] In addition, Fig. S1 presents additional results obtained by varying the CTE of SiO_2 and CTN, showing that the simulated stress distribution is sensitive to the selected mechanical parameters.

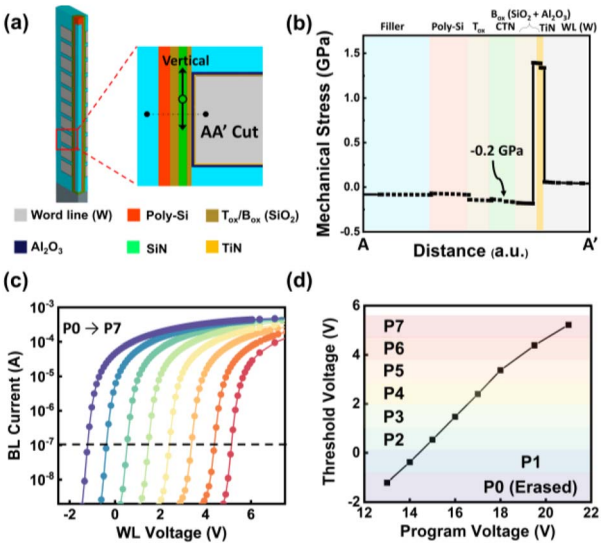

Fig. 2 (a) Cross-sectional view of the ten-layer 3D NAND structure, with the extraction cut line AA′ indicated. (b) Simulated vertical stress distribution across different material layers along the extraction line AA′. (c) Simulated transfer characteristics of the memory cell under various programmed states. (d) V_T as a function of program voltage during ISPP.

I–V characteristics were simulated to examine the V_T of the memory cell at different programmed states. Piezo model (which includes deformation potential), the density-of-states (DOS) effective mass model, and the eSubband effective-mass mobility model were included to capture the effects of residual stress on electrical properties.[16,20] Fig. 2c presents the simulated transfer characteristics, obtained through incremental step pulse programming (ISPP). During the program operation, a voltage pulse with increasing amplitudes (13–21 V) and a fixed duration of 1 ms was applied to the selected WL, while unselected WLs were biased at a pass voltage of 7 V.[7,21] Erasure was conducted by applying a 22 V, 10 ms pulse to the Si substrate.[2] Fig. 2d demonstrates a near-linear increase of V_T with higher program voltages by ISPP.[14,21] In addition, all cells were confirmed to have nearly identical initial V_T values after programming under the same stress condition. Also, it is worth noting that V_T was extracted using the built-in TCAD function, with a reference current of 1×10^{-7} A.

Analysis and mechanisms of residual stress effects on 3D NAND reliability

Next, a controlled vertical compressive stress was applied selectively to the CTN layer to investigate the impact of residual stress on reliability, while all other conditions were kept unchanged so that the observed V_T shifts could be attributed to stress-induced effects. Accordingly, the wide stress range considered here should be interpreted as a controlled simulation condition introduced to probe the stress–retention relationship, rather than as a stress state directly produced by the process-flow-based structure. Retention simulations were performed under varying stress conditions, as shown in Fig. 3a. For each residual-stress condition, the selected memory cell was programmed to the P7 state, and all terminals were then grounded at 85 °C to start the retention phase. The initial V_T may differ among stress conditions due to stress-dependent changes in the effective barrier of the charge-trap layers.[18] Multiple physical models—including non-local Wentzel–Kramers–Brillouin (WKB) tunneling, Poole–Frenkel emission, Shockley–Read–Hall (SRH), and Auger recombination—were employed to capture charge loss.[20] The CTN trap-related parameters adopted in the retention simulation are summarized in Table S2.[20,22–24] These trap-related parameters were kept unchanged for all compared stress states, so that the observed retention difference mainly reflects the residual stress effect. As shown in Fig. 3b, the retention behavior changes markedly under different applied compressive-stress conditions up to 10^8 s. To further interpret the retention trends, band-structure analysis was performed. Charge loss in charge-trap memory involves multiple mechanisms, among which trap-to-band emission significantly contributes to leakage currents.[21,25] According to the Poole–Frenkel theory, electron emission from trap states into the conduction band is highly sensitive to the effective barrier height.[26] A stress-induced shift in the conduction band can alter this effective barrier, thus affecting the electron emission rate and the retention performance.[27] Our simulation results, illustrated in Fig. 3c, support this effect by comparing two residual stress states. Increasing the compressive stress from −0.2 GPa to −2 GPa induces an upward shift in the conduction band, which further leads to improvement in the retention time. Here, the band modulation is used to illustrate the trend of stress-induced changes in the CTN band profile rather than to provide a strictly quantitative prediction of the absolute band shift. To provide a clearer understanding of the factors influencing retention, we further extracted the trapped-charge distribution in the CTN layer after programming, corresponding to the initial state of the retention simulation. As shown in Fig. 3d, the programmed cell exhibits a nearly saturated trapped-electron region within the CTN layer, while the unselected cells contain only a small amount of charge. The extracted results in Fig. 3e show the trapped-electron values obtained from the programmed CTN region after retention under different residual-stress conditions. With increasing vertical compressive stress, the trapped-electron value extracted from the same cut line gradually increases, indicating that a larger amount of stored electrons remains in the CTN and that charge loss after retention is reduced under stronger compressive stress. This tendency is consistent with the retention-simulation results. It also supports the interpretation that residual stress affects the charge-loss behavior and influences the overall retention performance. In addition, we mainly discuss the vertical stress distribution in this paper, because it is more directly related to the dominant charge-loss path across the oxide/CTN layers in the simulation and thus has a clearer impact on the effective barrier and retention behavior. Lateral stress can also influence retention; however, within the scope of this work, we focus on vertical-stress modulation.

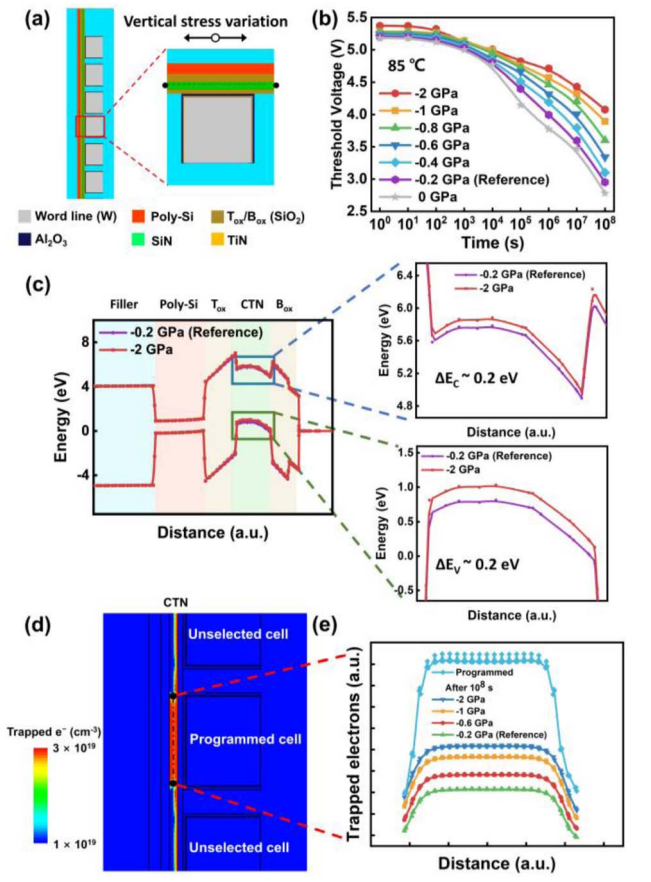

Fig. 3 (a) Schematic diagram of vertical stress along the channel direction. (b) Retention characteristics of the 3D NAND cell under different vertical residual-stress conditions. (c) Band structure variations under different stress conditions. The inset shows a zoomed-in view of the simulated conduction band (top) and valence band (bottom) diagrams. (d) Distribution of trapped electrons (e^-) in the CTN layer of the 3D NAND structure after the program operation. (e) Variation of trapped-electron density along the cut line under different residual stress conditions.

Retention improvement through SEL-induced stress engineering in 3D NAND

Building upon previous findings that increased compressive stress in the CTN layer enhances retention, we propose a mechanical stress engineering strategy to further improve 3D

NAND flash reliability based on thermal mismatch. In a 3D NAND structure, multiple materials are sequentially deposited and stacked. The thermal mismatch among these materials plays an important role in determining the overall stress distribution in the 3D NAND stack.[8,28] Therefore, inserting an additional layer whose CTE is significantly different from the original materials can effectively change the overall stress distribution. By tuning the SEL properties, the stress in targeted regions can be deliberately engineered. Specifically, we introduce an SEL with high CTE (10×10^{-6}/K) deposited after the poly-Si channel formation, located between the SiO_2 filler and the poly-Si channel (Fig. 4a). In this work, we also studied different SEL thicknesses (10 and 20 nm) to evaluate their impact on stress modulation. Notably, the oxide layer is employed here as a conceptual SEL to validate the stress-engineering strategy; hence, a specific commercial material was not designated in this study. Nevertheless, materials such as tetragonal ZrO_2 and CeO_2 can exhibit CTE values comparable to the value assumed here,[29–31] suggesting physical feasibility and motivating future integration studies. However, practical implementation of such a layer would still require further process optimization since the insertion of an additional layer may increase fabrication complexity and affect device geometry as well as electrical characteristics. In addition, the high CTE of the SEL may influence the global stress balance of the full stack, particularly in ultra-high-stack configurations; therefore, its thickness and material properties should be carefully optimized to avoid excessive wafer bowing or warpage during future process integration. At the local device level considered in this work, the higher expansion tendency of the SEL induces tensile stress upon thermal contraction, thereby increasing compressive stress in the adjacent CTN layer (Fig. 4b). This enhanced compressive stress reduces charge leakage, thus improving overall retention characteristics. The retention improvement observed after SEL insertion therefore indicates that the moderate increase in CTN compressive stress generated by the proposed structure can still contribute to charge-loss suppression, consistent with the stress-dependent retention trend discussed above.

The results of retention simulations at 85 °C for different programmed states (P7, P5, and P2; Fig. 4c–e) demonstrate that a 20 nm SEL induces stronger compressive stress in the CTN layer and leads to slower V_T degradation during retention, but it also causes noticeable changes in the initial V_T values,[18] particularly at the lower programmed state (P2). In contrast, a 10 nm SEL achieves a more balanced improvement in retention while minimally affecting initial V_T values. To further evaluate the electrical impact of SEL insertion beyond retention characteristics, the program/erase speed and on-current were also analyzed. Using the same program/erase conditions as those reported in the previously published work,[1,7] Fig. S2 shows that SEL thickness has a negligible effect on the programming speed and only a minor influence on the erase speed. In addition, Fig. S3 shows that the on-current gradually decreases with increasing SEL thickness because the channel compressive stress becomes stronger. This reduction in on-current can also narrow the practical read margin during sensing, even when the retention behavior is improved. Consequently, implementing an SEL necessitates a careful trade-off between enhanced retention and electrical drawbacks, such as V_T instability and reduced read current. Among the configurations studied, the 10 nm SEL offers a favorable balance, improving retention with minimal impact on program/erase speed, V_T stability, and on-current. Practical 3D NAND devices and fabrication processes involve additional material and process complexities that are not fully captured in the present simulations. Nevertheless, the simulations identify a consistent link between increased CTN compressive stress and reduced retention degradation, supporting SEL-induced residual-stress modulation as a strategy for improving multilevel storage reliability in 3D NAND flash memory.

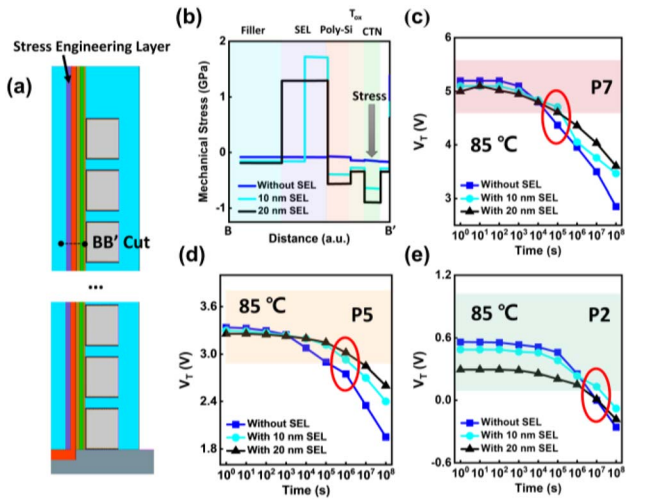

Fig. 4 (a) Schematic diagram of the 3D NAND structure after introducing the SEL layer. (b) Vertical residual stress distribution with and without the SEL layer along the cut line BB′. (c–e) V_T evolution during retention for multilevel cells with and without SEL insertion.

Conclusion

In this study, residual stress distributions induced by thermal processing were simulated through Sentaurus TCAD. Retention characteristics were then evaluated under different compressive stress conditions within the CTN layer. The results show that higher compressive stress in the CTN layer is associated with improved charge retention, consistent with stress-induced modulation of the CTN band profile and reduced thermally assisted charge loss. Building on these results, we propose inserting a high-CTE SEL between the SiO_2 filler and poly-Si channel as a potential stress-engineering strategy. Within the investigated range, the 10 nm SEL provides a relatively balanced improvement in retention while limiting the associated electrical penalty. The simulation results suggest that this strategy enhances compressive stress within the CTN, thus showing the potential to improve retention performance. The results highlight residual stress modulation as a promising approach for improving the reliability of advanced 3D NAND flash memory.

Conflicts of interest

There are no conflicts to declare.

Data availability

The data supporting the findings of this study are available within the article. Additional raw data and/or processed datasets are available from the corresponding author upon reasonable request.

Supplementary information (SI): simulation parameters and additional supporting simulation results. See DOI: https://doi.org/10.1039/d6na00054a.

Acknowledgements

This work was supported by Samsung Electronics Co., Ltd. (IO201215-08198-01). This research was also supported by the National Research Foundation (NRF) funded by the Korean government (MSIT) (No. RS-2026-25539124), by K-CHIPS (RS-2024-00405179, 24048-15TC) funded by the Ministry of Trade, Industry & Energy (MOTIE, Korea), and by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (RS-2025-25410048). In addition, this research was partially supported by the BK21 FOUR funded by NRF. The Sentaurus TCAD simulator was provided by the Electronic Design Automation (EDA) tool program of IC Design Education Center (IDEC) in Korea.

References

1 J. Jang, H. S. Kim, W. Cho, H. Cho, J. Kim, S. I. Shim, Y. Jang, J. H. Jeong, B. K. Son, D. W. Kim, K. Kim, J. J. Shim, J. S. Lim, K. H. Kim, S. Y. Yi, J. Y. Lim, D. Chung, H. C. Moon, S. Hwang, J. W. Lee, Y. H. Son, U. I. Chung and W. S. Lee, *Presented in Part at Symp. VLSI Technol. Dig. Tech. Pap.*, 2009.
2 H. Tanaka, M. Kido, K. Yahashi, M. Oomura, R. Katsumata, M. Kito, Y. Fukuzumi, M. Sato, Y. Nagata, Y. Matsuoka, Y. Iwata, H. Aochi and A. Nitayama, *Presented in Part at Symp. VLSI Technol. Dig. Tech. Pap.*, Kyoto, Japan, 2007, pp. 12–14.
3 S. Whang, K. Lee, D. Shin, B. Kim, M. Kim, J. Bin, J. Han, S. Kim, B. Lee, Y. Jung, S. Cho, C. Shin, H. Yoo, S. Choi, K. Hong, S. Aritome, S. Park and S. Hong, *Presented in Part at 2010 IEEE International Electron Devices Meeting (IEDM)*, San Francisco, California, USA, 2010.
4 S. H. Lee, *Presented in Part at 2016 IEEE International Electron Devices Meeting (IEDM)*, San Francisco, CA, USA, 2016.
5 A. Kruv, A. Arreghini, D. Verreck, M. Gonzalez, G. Van den Bosch, I. De Wolf and M. Rosmeulen, *IEEE Trans. Electron Devices*, 2020, **67**, 4891–4896.
6 K. B. Kim, Y. T. Oh and Y. H. Song, *J. Korean Phys. Soc.*, 2017, **70**, 1041–1048.
7 E. K. Jang, I. J. Kim, C. A. Lee, C. Yoon and J. S. Lee, *IEEE J. Electron Devices Soc.*, 2022, **10**, 104–108.
8 J. Lee, D. G. Yoon, J. M. Sim and Y. H. Song, *Electronics*, 2021, **10**, 2632.
9 Y. Namkoong, H. J. Yang and Y. H. Song, *J. Nanosci. Nanotechnol.*, 2017, **17**, 5055–5060.
10 A. Moridi, H. H. Ruan, L. C. Zhang and M. Liu, *Int. J. Solids Struct.*, 2013, **50**, 3562–3569.
11 M. Horstmann, A. Wei, T. Kammler, J. Höntschel, H. Bierstedt, T. Feudel, K. Frohberg, M. Gerhardt, A. Hellmich, K. Hempel, J. Hohage, P. Javorka, J. Klas, G. Koerner, M. Lenski, A. Neu, R. Otterbach, P. Press, C. Reichel, M. Trentsch, B. Trui, H. Salz, A. Schaller, H. J. Engelmann, O. Herzog, H. Ruelke, P. Hübler, R. Stephan, D. Greenlaw, M. Raab, N. Kepler, H. Chen, D. Chidambarrao, D. Fried, J. Holt, W. Lee, H. Nii, S. Panda, T. Sato, A. Waite, S. Luning, K. Rim, D. Schepis, M. Khare, S. F. Huang, J. Pellerin and L. T. Su, *Presented in Part at 2005 IEEE International Electron Devices Meeting (IEDM)*, Washington, DC, USA, 2005.
12 E. Ungersboeck, S. Dhar, G. Karlowatz, V. Sverdlov, H. Kosina and S. Selberherr, *IEEE Trans. Electron Devices*, 2007, **54**, 2183–2190.
13 R. Arghavani, N. Derhacobian, V. Banthia, M. Balseanu, N. Ingle, H. M'Saad, S. Venkataraman, E. Yieh, Z. Yuan, L. Q. Xia, Z. Krivokapic, U. Aghoram, K. MacWilliams and S. E. Thompson, *IEEE Trans. Electron Devices*, 2007, **54**, 362–365.
14 K. L. Wang, G. Du, Z. Y. Lun and X. Y. Liu, *IEEE J. Electron Devices Soc.*, 2019, **7**, 150–157.
15 J. Kim and H. Shin, *J. Comput. Electron.*, 2025, **24**, 57.
16 D. Kim, K. Nam, C. Park, H. You, Y. Kim, S. Park and R. H. Baek, *IEEE Trans. Electron Devices*, 2024, **71**, 5354–5360.
17 Y. M. Wang, L. B. Wang, J. X. Gu, X. Q. Yan, J. R. Lu, S. L. Dou, Y. Li and L. Wang, *Nanomaterials*, 2022, **12**, 4262.
18 D. Kim, K. Nam, C. Park, H. You, M. S. Park, Y. Kim, S. Park and R. H. Baek, *Solid State Electron.*, 2024, **216**, 108927.
19 T. W. B. Riyadi, D. Setiadhi, A. D. Anggono, W. A. Siswanto and H. H. Al-Kayiem, *Forces Mech.*, 2021, **4**, 100042.
20 Synopsys, Sentaurus Device User Guide, Mountain View, CA, 2024.
21 A. Goda, *Electronics*, 2021, **10**, 3156.
22 J. Park, G. Yoon, D. Go, J. Kim and J. S. Lee, *IEEE Access*, 2021, **9**, 118794–118800.
23 K. Nam, C. Park, J. S. Yoon, H. Yun, H. Jang, K. Cho, H. J. Kang, M. S. Park, J. Sim, H. C. Choi and R. H. Baek, *Nanomaterials*, 2022, **12**, 1808.
24 J. H. You, H. W. Kim, D. H. Kim, T. W. Kim and K. W. Lee, *Presented in Part at 2011 International Conference on Simulation of Semiconductor Processes and Devices (SISPAD 2011)*, Osaka, Japan, 2011.
25 Y. Cai, Y. X. Luo, E. F. Haratsch, K. Mai and O. Mutlu, *Presented in Part at IEEE Int. Symp. High Perform. Comput. Archit.*, 2015.
26 H. Schroeder, *J. Appl. Phys.*, 2015, **117**, 215103.
27 S. Kim, K. Lee, C. Woo, Y. Hwang and H. Shin, *IEEE Trans. Electron Devices*, 2020, **67**, 5472–5478.
28 Z. Q. Tian, G. Zhang, Y. H. Huang, S. Z. Wang and S. Liu, *IEEE Trans. Electron Devices*, 2025, **72**, 193–198.

29 M. D. Mathews, B. R. Ambekar and A. K. Tyagi, *J. Nucl. Mater.*, 2000, **280**, 246–249.
30 M. Mishra, P. Kuppusami, S. Murugesan, C. Ghosh, R. Divakar, A. Singh and E. Mohandas, *Mater. Chem. Phys.*, 2015, **162**, 592–607.
31 M. R. Rao, N. G. Krishna and S. C. Lingam, *High Temp.*, 2022, **60**, 761–767.