Received 7 October 2021; accepted 7 November 2021. Date of publication 15 November 2021; date of current version 24 November 2021.
The review of this article was arranged by Editor S. Menzel.

Digital Object Identifier 10.1109/JEDS.2021.3127953

Analysis and Optimization of Defect Generation Due to Mechanical Stress in High-Density SRAM

KAZUNARI ISHIMARU (Fellow, IEEE), MIZUKI TAMURA, AND OSAMU FUJII

Institute of Memory Technology Research and Development, Kioxia Corporation, Yokkaichi 512-8550, Japan

CORRESPONDING AUTHOR: K. ISHIMARU (e-mail: kazu.ishimaru@kioxia.com)

ABSTRACT Static random-access memory (SRAM) is an essential component for realizing large-scale integration (LSI). The future transition to a 48 V DC supply in datacenters and electric vehicles acting as mobile edge servers will increase the demand for a bipolar-complementary metal-oxide semiconductor-double diffused metal-oxide semiconductor with high-capacity SRAM. When we scaled and optimized an SRAM cell from 130 nm nodes to 90 nm nodes, we observed the generation of crystal defects induced by mechanical stress in the p-channel MOS active area that cannot be explained by previous models. We performed simulations using the finite element method to identify the mechanism. In our results, the edge of the narrow active area showed a large deformation compared to the middle of the active area, which can be attributed to compressive stress from the gate electrode and sidewall. The cell layout and sidewall structure were optimized to suppress this defect generation while satisfying reliability requirements, and the design can be extended to 65 nm nodes.

INDEX TERMS Defect, mechanical stress, SRAM.

I. INTRODUCTION

The advancement of the Internet of Things has led to a data explosion, with the annual data generation predicted to exceed 175 ZB by 2025 [1]. Processing of big data is essential to the development of new applications and businesses, which will require an improvement in computing performance. However, the explosive increase in the power consumption of datacenters has become a significant concern [2]. A typical approach to minimizing power consumption is to reducing the supply voltage, and the introduction of a 48V DC supply for data buses is being studied intensively [3]–[5]. A 48V DC supply has long been used for telecommunication hardware [6] and electric vehicles, which can potentially act as mobile edge servers. Therefore, compatibility with a 48V DC is becoming a requirement for an increasingly wide range of applications. The bipolar-complementary metal-oxide semiconductor-double diffused metal-oxide semiconductor (BiCDMOS) is a technology that combines analog, digital and high voltage devices in a single die [7], and it has been used in the automotive, aerospace and other industrial fields for power management. Silicon carbide (SiC) devices are advantageous for power control at high voltages more than several hundred volts. However, the BiCDMOS is more economical in the range of $<$100 V. Another advantage of the BiCDMOS is that more digital functions can be added by scaling its dimensions and increasing the number of transistors in a single chip. Node with a scale of 180 or 130 nm are mainly used for $<$100 V applications and 90 nm nodes have recently been proposed [8]. The static random-access memory (SRAM) is an essential component that is widely used for large-scale integration (LSI) because it can be fabricated by using a conventional CMOS process. Implementing a large-capacity SRAM in a BiCDMOS platform will make more functions available. In this paper, we propose a design guidelines for a SRAM cell that is suitable for 90 nm nodes and beyond while still satisfying reliability requirements for the BiCDMOS.

II. PRELIMINARY STUDY

A. CELL LAYOUT COMPARISON

We evaluated two SRAM cell layouts suitable for 90 nm nodes of a BiCDMOS, as shown in Fig. 1. Based on past technology trend [9], the target cell size was predicted to be around 1 μm^2. Both cells satisfy the target cell size as per design rules presented in Table 1. Type A is a conventional cell layout that has been used since the advent of 0.25 μm

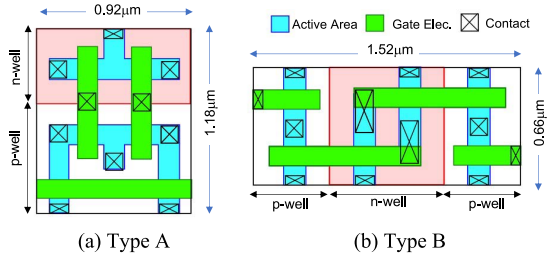

FIGURE 1. SRAM cell layouts for scalability study; (a) conventional (Type A) and (b) low cell height (Type B).

TABLE 1. Rules used for cell design.

Layers	Line (nm)	Space (nm)
Active Area	120	140
Gate Electrode	80	190
Contact	120	140
Contact-Gate	-	70
	Type A	Type B
Cell Size (μm^2)	1.0032	1.0856

nodes. The advantage of Type A is the single well (n+/p+) isolation region in the unit cell, which requires a larger isolation width than the diffusion (n+/n+ or p+/p+) isolation width. The aspect ratio (i.e., height to width) of the unit cell is close to one, which is suitable for large cell arrays.

In contrast, Type B cell has two well isolation regions in the wordline direction that increase the cell width. The cell also requires the shared contact to connect both the gate electrode and active area. However, the cell height is lower, which shortens the bitline length and is advantageous for fast access speed.

In terms of manufacturing, a straight pattern for both the active area and the gate electrode is preferable because a small width variation can be realized. The optical proximity correction (OPC) technique [10] adds a sub-lithography pattern (i.e., serif) to the photomask layout to minimize shrinkage. An elongated design such as the gate electrode in Type B shrinks in the longitudinal direction during the patterning process. Such elongated patterns do not affect each other when placed in the longitudinal direction even if the spacing is narrow. On the other hand, when the gate electrodes are arranged orthogonally such as in Type A, they are affected by narrow spacing, and they can shorten the pattern in the worst case, as shown in Fig. 2. We calibrated OPC models according to measurement results and simulated the variability in the wordline gate length for types A and B under various lithography conditions. As shown in Fig. 2, Type B showed better uniformity than Type A. Therefore, we chose Type B as the SRAM cell layout for 90 nm nodes in consideration of future scalability. We optimized the cell layout by using an OPC combined with a manufacturability check [11] and a hotspot fixing system [12]. The active area and gate electrode in the shared contact region had sufficient overlap to avoid open failure.

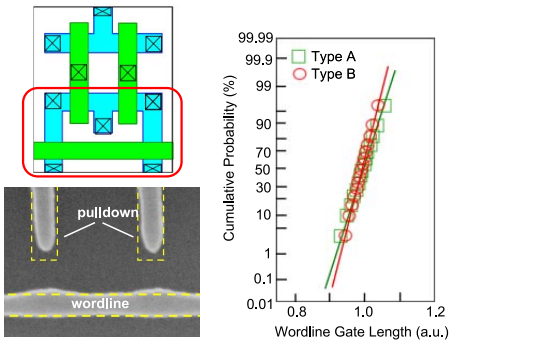

FIGURE 2. SEM photograph of the wordline and pulldown gate resist profile in a Type A cell (left). When the distance between the pulldown gate and wordline is small, the wordline gate shows a wavy shape, although the pulldown gate still shows pattern shortening. The dotted lines in the photos represent the designed layout. Simulated cumulative probability of the wordline width for types A and B (right). Type B shows a better distribution.

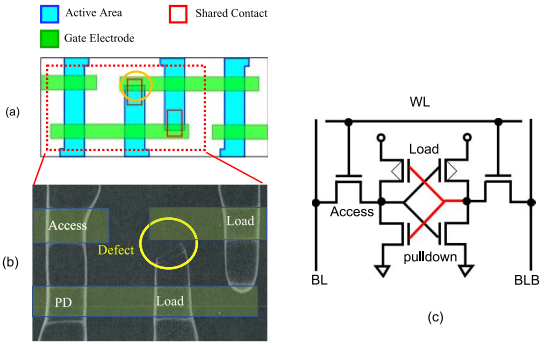

FIGURE 3. (a) Designed SRAM cell layout and (b) SEM photograph after the Wright etching corresponding to single-bit failure. The active area edge of the pMOS load transistor is missing where shared contact was formed. Shared contacts connect both the gate electrode and active area, as indicated by the red line in (c).

B. VALIDATION BY HARDWARE

To validate the cell layout, we fabricated a Type B 256 kbit SRAM cell using 130 nm nodes, which are already in mass production with Type A cells. In a functional test, Type B SRAM cell showed single-bit failures, and failure analysis confirmed a missing active area in the p-channel MOS (pMOS) transistor region as shown in Fig. 3. Several reports have noted that mechanical stress generates crystal defects by the shallow trench isolation or gate electrode [13], [14]. Another cause of defect formation is damage from high-dose ion implantation [15], [16]. However, those phenomena were observed for the n-channel MOS (nMOS) transistor region, and they cannot explain the observed failure in the pMOS region. In addition, previous simulation studies used 2D models and could not accurately analyze the phenomenon of actual structures. Therefore, we carried out a 3D simulations of the mechanical stress by using the finite element simulation tool MARC [17] for failure analysis.

III. PROCESS FLOW AND SIMULATION SETUP

We investigated the process flow for the step that caused high stress. Fig. 4 shows the process flow for fabricating a

- Trench isolation
- Gate Oxide formation (750 °C 10min) --- a
- Poly-Si deposition (600 °C 30min) --- b
- Gate Electrode patterning
- Post Etch Oxidation (950 °C 10min.) --- c
- n-LDD formation (900°C 10sec) --- d
- p-LDD formation (800 °C 10sec) --- e
- Sidewall formation (750 °C 30min) --- f
- Source/Drain formation (1050 °C 10sec) --- g
- Contact formation
- Mettalization

FIGURE 4. **Process flow for SRAM fabrication: (a-g) process steps that may generate crystal defects.**

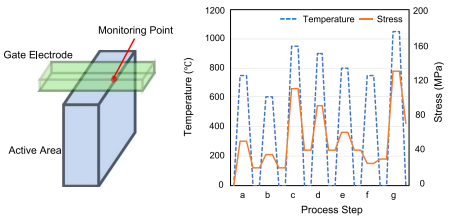

FIGURE 5. **Mechanical stress monitoring location (left) and simulated stress values at process *a* – *g* (Fig. 5) with the temperature profile (right).**

TABLE 2. **Material parameters.**

Parameters	Si	SiO_2	SiN
Young Modulus (GPa)	130	75	300
Poison Ratio	0.2	0.25	0.28

SRAM cell. Process steps that may generate crystal defects are labeled a-g. Table 2 presents the Young's modulus and Poison's ratio for each material used in this simulation, such as Si, SiO_2, and SiN. In the simulation, shear stress was observed at the top corner of the active area under the gate electrode, as shown in Fig. 5. Correlations were found between the amount of stress and process temperature. After gate electrode formation, the source/drain formation had the largest stress. Therefore, we focused on this process step.

Before detailed analysis, we evaluated the effect of structure fidelity on the simulation results. Despite the use of OPC, the actual device has a rounded shape. To reflect the actual device structure precisely, a complicated mesh structure would be required that would increase the simulation time. The top corner of the active area is rounded to prevent kink effects in the transistor [18]. We compared the von Mises stresses value with and without rounding the top corner. The radius of the rounded corner was 20 nm, and the other dimensions are shown in Fig. 6.

The von Mises stress value with a rounded corner shows some fluctuations caused by mesh generation. However, the difference between stresses with and without a rounded corner was less than 5%, which is acceptable for comparing stress along the channel. Thus, we carried out simulations without rounding the corner. Fig. 7 plots the contours of the von Mises stress with (a) the designed layout and (b) the actual device shape shown in Fig. 3 (b). For a fair comparison, the number of elements was set to 572 for the designed

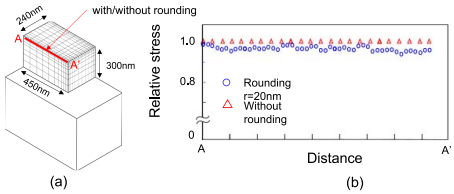

FIGURE 6. **(a) Simulated device structure (active area only) and (b) stress at the top corner of the active area with/without corner rounding. The stress was normalized against the stress without corner rounding.**

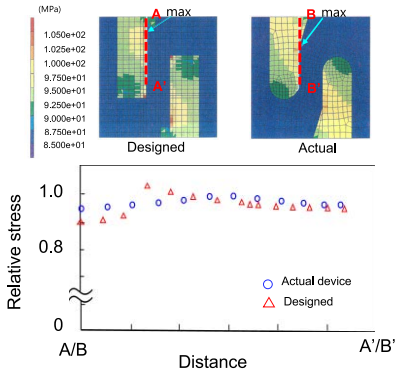

FIGURE 7. **Stress of the designed pattern and actual pattern along the active area directions (dotted lines *A* – *A'* and *B* – *B'*).**

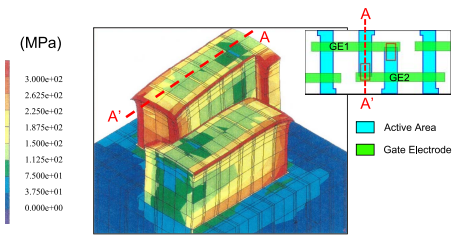

FIGURE 8. **Maximum resolved shear stress with displacement (200× enhancement) of the active area during source/drain activation annealing.**

layout and 582 for the actual device. Both cases had similar stress distributions, and the maximum stress was observed near the concave corner. The stresses along with the top corner of the active area are indicated by the dotted lines $A-A'$ and $B-B'$. Some differences were observed in the maximum stress near the concave corner. However, the overall difference in other regions was less than 5%, which is sufficient for a stress comparison. Based on the above results, we decided to use a rectangular shape without rounded corners for the following simulation.

IV. SIMULATION RESULTS

A. MECHANICAL STRESS AND DEFECT GENERATION MODEL

Fig. 8 plots the contour of the maximum resolved shear stress in the active area of the pMOS region during the source/drain annealing process. The displacement value (multiplied 200 times for easy observation) is also shown.

The corner region of the active area had a stress of more than 300MPa, which is large enough to generate crystal

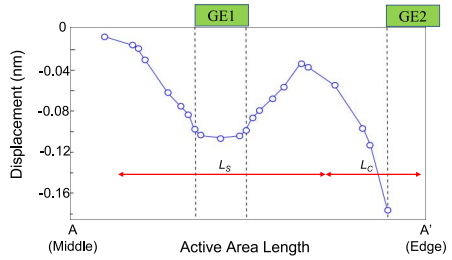

FIGURE 9. Displacement of the active area (dotted line $A - A'$ in Fig. 8).

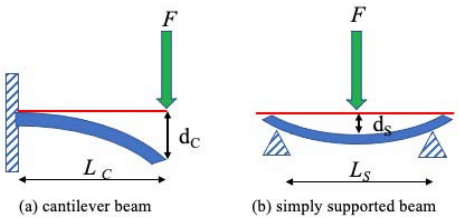

FIGURE 10. Models for explaining the mechanical stress and deflection under the gate electrode shown in Fig. 9: (a) cantilever beam structure for under GE2; (b) simply supported beam structure under GE1.

defects [14]. In addition, the active area edge showed a large displacement from the top surface in the trench depth direction. This region overlaps with the gate electrode, which indicates that mechanical stress by the gate electrode may be a root cause for this displacement. Fig. 9 shows the displacement of the active area under the gate electrode along $A - A'$ in Fig. 8. The gate electrode clearly applied compressive stress to the active area. GE1 is in the middle of the active area, while GE2 is at the edge of the active area. Both the stress and displacement were in the same direction, and the displacement was about 1.8 times larger at GE2 than at GE1. This difference can be attributed to the deflection of the beam structure. Fig. 10 shows a model that explains the deflection under the gate electrode. The cantilever beam structure reflects the deflection of the active area edge under GE2. The simply supported beam structure reflects the deflection of the active area under GE1. The displacements d_C and d_S can be calculated with the following formulas [19]:

$$d_C = \frac{4FL_C^3}{EI} \tag{1}$$

$$d_S = \frac{FL_S^3}{4EI} = \frac{2FL_C^3}{EI} \quad (L_S = 2L_C) \tag{2}$$

F: Load
E: Young's modulus
I: Moment of inertia of area
L_C, L_S: beam length
d_C, d_S: displacement

where F is the load, E is the Young's modulus, I is the moment of inertia of the area, L_C and L_S are the lengths of the cantilever and simply supported beams, respectively, and d_C and d_S are the corresponding displacements, respectively. L_S is about twice of L_C. If equation (2) is rewritten with L_C, d_C becomes twice of d_S. This is close to the simulation result, which found a correction factor of 1.8. The active areas of nMOS transistors in the SRAM cell are connected and considered to have a beam structure with both ends fixed. The deflection of a beam with both ends fixed beam, d_B, is given by

$$d_B = \frac{FL_B^3}{16EI} = \frac{FL_C^3}{2EI} \quad (L_B = 2L_C) \tag{3}$$

Thus, it is eight times smaller than d_C. This is why defects were only generated in the pMOS region. Therefore we concluded that the beam model can describe this phenomenon and is applicable to estimating the stress effect.

B. CELL LAYOUT OPTIMIZATION FOR 90NM NODE

Process variability affects the overlap length between the active area and gate electrode, such as misalignment and pattern shortening. To realize a robust cell design, we carried out detailed simulations for layout optimization. Mechanical stress is known to generate crystal defects and degrade the transistor characteristics and reliability. Therefore, these reliability issues need to be considered for SRAM cell optimization. Several studies have reported on the impact of mechanical stress on transistor characteristics and reliability. Time-dependent gate oxide breakdown (TDDB) can be caused by a mechanical stress of more than 200MPa [20], [21]. Both the hot carrier reliability and bias temperature instability are degraded by compressible stress under the gate electrode [22]–[24]. However, these degradations depend on the amount of stress, and a stress of less than 200PMa does not affect reliability [25]–[28]. Because the mechanical stress in the active area increased about 5% with a 10% reduction in dimensions [13], the maximum mechanical stress should be at least 15% less than 200MPa for possible extension to 65 nm nodes, which are 30% smaller than 90 nm nodes. Thus, the maximum allowable stress should be less than 170MPa. The active area edge of the initially designed cell had a 60 nm overlap length with the gate electrode. However, the active area shrank in the longitudinal direction because of pattern shortening, and the actual overlap length was about 50 nm. Because the gate length was 80nm, the edge of the active area was located at about the middle of the gate length. According to the cantilever beam model, this is close to the conditions that cause the maximum deflection. We investigated the effect of the overlap length on the mechanical stress at the active area under the gate electrode. Fig. 11 shows a schematic of the overlap structure. The overlap length d was varied from −60 nm to +120 nm. Fig. 12 plots the contours of the resolved shear stress for each overlap length based on the simulation results. For easy observation, only the active area is shown. The displacement in the depth direction is also

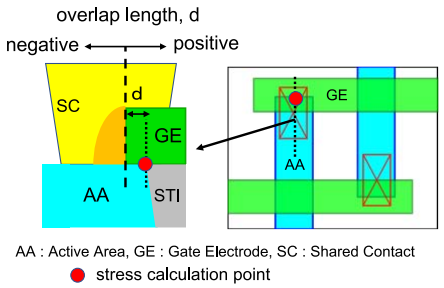

FIGURE 11. Schematic image of AA–GE overlap. The overlap length d was varied from −60 nm to +120 nm.

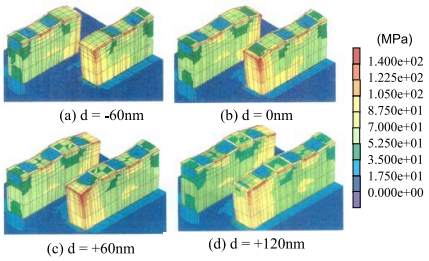

FIGURE 12. Simulated mechanical stress for each AA–GE overlap amount with active area deflection.

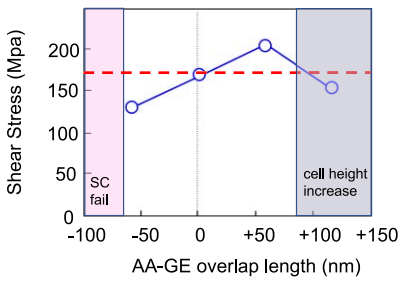

FIGURE 13. Effect of overlap length on the shear stress at the active area under the gate electrode shown in Figs. 12 (a)–(d)

shown, as given in Fig. 8. A high stress was observed at the edges of the active area where defects were observed, and its region increased as the overlap length was increase from −60 nm to +60 nm. However, increasing in overlap length to +120 nm reduced the high-stress region at the active area edge. This is because the edge of the active area separated from the gate electrode, which shifted from a cantilever structure to a simply supported beam structure, as shown in Fig. 10. Fig. 13 shows the effect of the overlap length on the maximum resolved shear stress at the top corner of the active area under the gate electrode (red circle in Fig. 11). The stress clearly increased with an increasing overlap length and then decreased at an overlap length of +120 nm, which means that the active area passed under the gate electrode. An overlap length of more than 80 nm increased the cell height, and an overlap length of less than −60 nm overlap (away from the gate electrode edge) caused open failure of the shared contact. Therefore, the maximum allowabole overlap length is between −60 nm to 0 nm to ensure that the stress remains less than 170MPa. The above simulation was carried

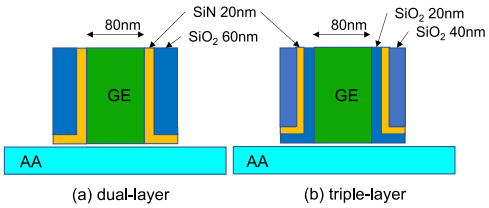

FIGURE 14. Sidewall configurations used for simulations: (a) dual-layer and (b) triple-layer structures.

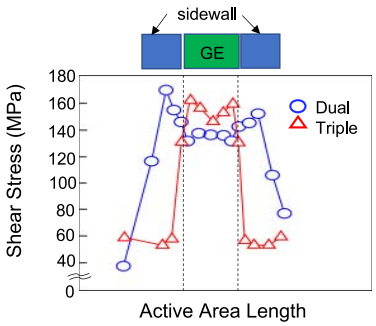

FIGURE 15. Effect of the sidewall structure on mechanical stress under the gate electrode.

out without a sidewall structure because only the effect of the overlap length was being considered. Because the sidewall structure of the transistor is known to affect the channel stress [29], we investigated the effect of the sidewall structure to finalize the cell layout. Fig. 14 shows schematic images of two sidewall structures. The original process had a dual-layer structure consisting of SiN/SiO_2 (Fig. 14 (a)). The sidewall material affects the transistor reliability [30], so we considered a triple-layer structure consisting of SiO_2/SiN/SiO_2 (Fig. 14 (b)) for comparison. The total sidewall width was the same for both structure at 80 nm. Fig. 15 shows the simulation results for both sidewall structures. The triple-layer structure showed a maximum stress at the gate edge region rather than the center of the gate electrode, and the stress decreased drastically in the sidewall region. In contrast, the dual-layer structure showed a maximum stress in the sidewall region. Moreover, the stress in the channel region was 10% less than that for the triple-layer sidewall structure. For the dual-layer structure, the compressive stress from the SiN film appears to have increased the stress under the sidewall. This compensated for the stress by the gate poly-Si, which reduced the mechanical stress in the channel region. In contrast, for the triple-layer structure, the SiO_2 film between the gate poly-Si and SiN reduced the stress from the SiN film, and the stress from poly-Si was dominant in the active area. Fig. 16 shows the effect of the overlap on the mechanical stress at the active area edge for both sidewall structures. In terms of process robustness, the optimal design for an SRAM cell with 90 nm nodes is an active area edge that is 30 nm away from the gate electrode edge with triple-layer sidewalls because this design can be extended to 65 nm nodes. A 1 Mbit SRAM was fabricated to confirm the robustness of the optimized cell layout. Fig. 17 shows the cumulative

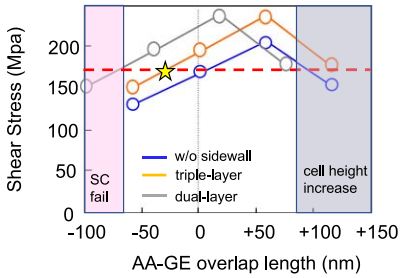

FIGURE 16. **Effect of the overlap length on the shear stress for both dual-layer and triple-layer sidewall structures.**

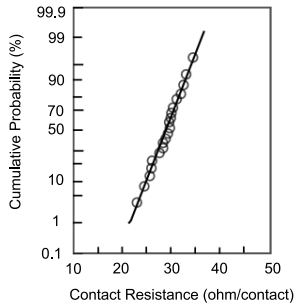

FIGURE 17. **Cumulative probability of the shared contact resistance.**

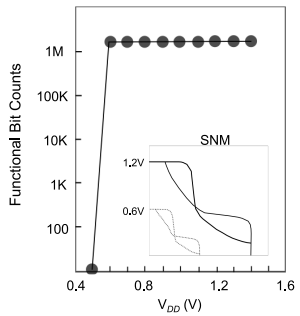

FIGURE 18. **Functional bit count vs SNM of a fabricated 1 Mbit SRAM.**

distribution of the shared contact resistance, which indicates a tight distribution and no open failure. Fig. 18 plots the functional bit count against the static noise margin (SNM). The SRAM could operate with a supply voltage as low as 0.6 V with a sufficient SNM.

V. CONCLUSION

Finite element simulations were performed to analyze the mechanical stress induced by defects generated in SRAM cells. Compressive stress by the gate electrode was found to cause a large deflection in the active area, and its effect was enhanced at the active area edge. This phenomenon can be described by models of cantilever and simply supported beams. The sidewall structure modulates the mechanical stress and affects the optimum overlap length between the gate electrode and the active area. For a maximum allowable mechanical stress of 170 MPa, an overlap length between −60 nm to −30 nm with a triple-layer sidewall structure is recommended for 90 nm nodes. A 1 Mbit SRAM with an overlap length of −30 nm was fabricated and demonstrated a tight shared contact resistance distribution and functional bit count even down to $V_{dd} = 0.6$ V. The SRAM cell, considering the proposed failure model, can be scaled down to the 65 nm nodes. However, this is a prediction by the simulation and needs to be verified by actual hardware. Although the maximum mechanical stress is kept below 170 MPa, the hot carrier effect and bias temperature instability are also need to be evaluated for future work.

REFERENCES

[1] K. Ishimaru, "Future of non-volatile memory -from storage to computing," in *Tech. Dig. Int. Electron Devices Meet.*, 2019, pp. 12–17.
[2] A. Shehabi, S. J. Smith, and J. Koomey, "Data center growth in the United States: Decoupling the demand for services from electricity use," *Environ. Res. Lett.*, vol. 13, no. 12, 2018, Art. no. 124030.
[3] X. Lyu *et al.*, "Composite modular power delivery architecture for next-gen 48V data center applications," in *Proc. 1st Workshop Wide Bandgap Power Devices Appl. Asia*, 2018, pp. 343–350.
[4] R. Das and H.-P. Le, "A regulated 48V-to-1V/100A 90.%-efficient hybrid converter for POL applications in data centers and Telecommunication Systems," in *Proc. IEEE Appl. Power Electron. Conf. Expo.*, Mar. 2019, pp. 1997–2001.
[5] T. B. Lazzarin, C. Bharatiraja, and B. Lehman, "Direct 48V to 1V step-down DC-DC converter," in *Proc. IEEE Int. Conf. Power Electron. Drives Energy Syst.*, Dec. 2018, pp. 1–6.
[6] F. Bodi and E. H. Lim, "Criteria for emerging telecom and data center powering architectures," in *Proc. Int. Telecommun. Energy Conf.*, 2012, pp. 1–9.
[7] K. Hamada, F. Kawai, T. Kushida, and A. Kawahashi, "60V BiCDMOS device technology for automotive applications," in *Proc. Conf. Rec. IAS Annu. Meet.*, vol. 2, 1995, pp. 986–990.
[8] H. Fujii *et al.*, "A 90nm bulk BiCDMOS platform technology with 15–80V LD-MOSFETs for automotive applications," in *Proc. 29th Int. Symp. Power Semicond. Devices IC's*, May 2017, pp. 73–76.
[9] Y. Fukaura *et al.*, "A highly manufacturable high density embedded SRAM technology for 90 nm CMOS," in *Tech. Dig. Int. Electron Devices Meet.*, 2002, pp. 415–418.
[10] S. Tanaka, S. Inoue, T. Kotani, K. Izuha, and I. Mori, "Impact of OPC aggressiveness on mask maufacturability," in *Proc. Photomask Next-Gener. Lithogr. Mask Technol. X*, vol. 5130, 2003, p. 23.
[11] T. Kotani, S. Tanaka, S. Nojima, K. Hashimoto, S. Inoue, and I. Mori, "Yield-enhanced layout generation by new design for manufacturability (DfM) flow," in *Proc. Design Process Integr. Microelectron. Manuf. II*, vol. 5379, May 2004, p. 128.
[12] S. Kobayashi, S. Kyoh, T. Kotani, S. Tanaka, and S. Inoue, "Automated hot-spot fixing system applied to the metal layers of 65-nm logic devices," *J. Micro/Nanolithogr. MEMS MOEMS*, vol. 6, no. 3, Sep. 2007, Art. no. 031010.
[13] K. Ishimaru *et al.*, "Mechanical stress induced MOSFET punch-through and process optimization for deep submicron TEOS-O3 filled STI device," in *Dig. Tech. Paper Symp. VLSI Technol.*, 1997, pp. 123–124.
[14] M. L. Polignano, G. P. Carnevale, I. Mica, and C. Pastore, "Mechanical stress and defect formation in device-processing: Validity of the numerical models for mechanical stress calculation," *IEEE Trans. Electron Devices*, vol. 54, no. 5, pp. 1108–1114, May 2007.
[15] S. Ikeda, H. Ohta, H. Miura, and Y. Hagiwara, "Mechanical stress control in a VLSI-fabrication process: A method for obtaining the relation between stress levels and stress-induced failures," *IEEE Trans. Semicond. Manuf.*, vol. 16, no. 4, pp. 696–703, Nov. 2003.
[16] P. Ferreira, R.-A. Bianchi, F. Guyader, R. Pantel, and E. Granger, "Elimination of stress induced silicon defects in very high density SRAM structures," in *Proc. 31st Eur. Solid-State Device Res. Conf.*, 2001, pp. 427–430.
[17] "MSC MARC." Accessed: Nov. 1, 2021. [Online]. Available: https://www.mscsoftware.com/product/marc
[18] K. Ishimaru *et al.*, "Trench isolation technology with 1 μm depth n- and p-wells for a full-CMOS SRAM cell with a 0.4 μm n/sup +p/sup +/ spacing," in *Proc. VLSI Technol. Symp.*, vol. 54, 1994, pp. 97–98.

[19] J. M. Gere and B. J. Goodno, *Mechanics of Materials*, 8th ed. Stamford, CT, USA: Cengage Lern., 2013.
[20] H. Miura, S. Ikeda, and N. Suzuki, "Effect of mechanical stress on reliability of gate-oxide film in MOS transistors," in *Tech. Dig. Electron Devices Meeting*, 1996, pp. 743–746.
[21] E. Morifuji, "Impact of mechanical stress on hot-carrier lifetime and time-dependent dielectric breakdown in downscaled complimentary metal-oxide-semiconductor," *Jpn. J. Appl. Phys.*, vol. 48, no. 2R, 2009, Art. no. 021206.
[22] A. Hamada, T. Furusawa, N. Saito, and E. Takeda, "A new aspect of mechanical stress effects in scaled MOS devices," *IEEE Trans. Electron Devices*, vol. 38, no. 4, pp. 895–900, Apr. 1991.
[23] J. R. Shih, J. J. Wang, W. Ken, Y. Peng, and J. T. Yue, "The study of compressive and tensile stress on MOSFET's I-V, C-V characteristics and it's impacts on hot carrier injection and negative bias temperature instability," in *Proc. IEEE Int. Rel. Phys. Symp.*, 2003, pp. 612–613.
[24] Z. Q. Teo, D. S. Ang, K. S. See, and P. Z. Yang, "Effect of mechanical strain on the NBTI of short-channel p-MOSFETS: Role of impact ionization," in *Proc. IEEE Int. Rel. Phys. Symp.*, 2009, pp. 1019–1022.
[25] R. Degraeve, G. Groeseneken, I. De Wolf, and H. E. Macs, "The effect of externally imposed mechanical stress on the hot-carrier-induced degradation of deep-sub micron nMOSFET's," *IEEE Trans. Electron Devices*, vol. 44, no. 6, pp. 943–950, Jun. 1997.
[26] A. Shickova *et al.*,"Negligible effect of process-induced strain on intrinsic NBTI behavior," *IEEE Electron Device Lett.*, vol. 28, no. 3, pp. 242–244, Mar. 2007.
[27] D. P. Ioannou and G. La Rosa, "Mechanical stress effects on p-channel MOSFET performance and NBTI reliability," in *Proc. IEEE Int. Rel. Phys. Symp.*, 2014, pp. 1–4.
[28] A. Clausner *et al.*, "Analysis of 28 nm SRAM cell stability under mechanical load applied by nanoindentation," in *Proc. IEEE Int. Rel. Phys. Symp.*, 2018, pp. 5B.11–5B.16.
[29] K.-Y. Lim *et al.*, "Impact of gate sidewall spacer structures on DRAM cell transistors under Fowler-Nordheim and gate-induced drain leakage stress conditions," in *Proc. IEEE Int. Rel. Phys. Symp.*, 2004, pp. 485–488.
[30] Y. Sambonsugi and T. Sugii, "Hot-carrier degradation mechanism and promising device design of nMOSFETs with nitride sidewall spacer," in *Proc. Annu. Rel. Phys.*, 1998, pp. 184–188.