Benefit of NMOS by Compressive SiN as Stress Memorization Technique and Its Mechanism

Chia-Chun Liao, Tsung-Yu Chiang, Min-Chen Lin, and Tien-Sheng Chao

***Abstract*—In this letter, we certify that the compressive SiN capping layer has more potential than the tensile layer for fabrication using the stress memorization technique to enhance NMOS mobility. The mechanism that we have proposed implies that the conventional choice of the capping layer should be modulated from the point of view of stress shift rather than using the highest tensile film.**

***Index Terms*—Contact etch-stop layer (CESL), poly amorphization implantation (PAI), strain, stress memorization technique (SMT).**

I. Introduction

AS THE scaling down of CMOS reaches its fundamental limits, coupling strain in the channel to enhance carrier mobility for continuing roadmap demand has been intensely studied [1]. For biaxial strain, a relaxed SiGe buffer layer was introduced with a complex process [2]. For uniaxial strain, a number of techniques, including SiGe source/drain [3], contact etch-stop layer (CESL) [4], and the stress memorization technique (SMT), have been introduced [5]. Of the aforementioned strain techniques, SMT is simpler to fabricate, is compatible with conventional processes, and may be combined with other strain technique advantages, making it both useful and attractive.

In recent approaches, several investigations have focused on optimization of the SMT technique by manipulating the geometrical structure and material features to obtain higher mobility. For example, SiN_x offers a higher Young's modulus than SiO_x. Hence, using SiN_x as spacer or capping layer could confine an expanded poly-Si gate firmly and sufficiently suppress lateral expansion to attain higher tensile strain coupling. The phenomenon of material expansion also plays a critical role in the SMT process. For poly-Si implantation, the deeper the R_p projection range using the same dosage, the more obvious the stress memorization [6], [7]. With a heavier implant source and the amount of dose aid, the poly-gate condition would be more amorphized [8]. More effective recrystallization would thus occur, and volume swelling could be enhanced significantly. However, the SiN intrinsic-stress impact on the SMT technique and its mechanism still need clarification. As the poly-Si pitch shrinks in high-density static random access memory circuits, the thickness of the capping layer that we can use is reduced, meaning that the mobility enhancement would be lowered with each successive generation [9]. Therefore, choosing the appropriate material and taking advantage of the limits to the thickness of the capping layer to create the highest mobility are important issues.

In this letter, we explore a variety of SiN as the capping layer to execute SMT. We find that the material with the greatest potential is a compressive SiN layer rather than the tensile SiN layer that is typical of the conventional process for CESL and SMT. Among the intrinsic characteristics of the SiN layer, the key factor is the amount of stress shift rather than the initial or final stress. The corresponding mechanism is also clarified.

II. Experimental Procedure

After the RCA cleaning process, a 3-nm gate oxide was thermally grown on 6-in wafers in a vertical furnace. An undoped poly-Si gate (2000 Å) was deposited, and As^+ (50 keV with $5 \times 10^{15}/cm^2$) was adopted to enable the maximum poly-amorphization-implantation (PAI) effect. After extension and spacer formation, four different silicon nitride (SiN) capping layers (1500 Å) were deposited by PECVD, modulating the flow rate of N_2 or SiH_4 to confine the poly-Si gates. The power was set at 100 W, the pressure was 137 Pa, and the temperature was 300 °C. To obtain a different intrinsic stress of SiN, the flow rates of N_2 were set to 50, 100, and 1000 sccm, with $NH_3 = 6$ sccm and $SiH_4 = 50$ sccm across all three values. The samples were labeled N_2-low, N_2-med, and N_2-high, respectively. In addition, a SiH_4-high sample in which the flow rate of SiH_4 was 100 sccm, although NH_3 and N_2 were 6 and 50 sccm, respectively, was prepared for comparison. After RTA at 900 °C for 30 s, the source/drain and poly-Si gate were activated, and the channel strain induced simultaneously. Subsequently, the capping layers were removed, leaving residual strain in the channel, and standard MOSFET fabrication steps were followed to complete the final devices with channel lengths of 0.4 μm.

III. Results and Discussion

Fig. 1 shows the value of transconductance divided by capacitance to make the strain contribution to mobility clearer. We found that the N_2-low split shows the highest enhancement (27%), while the N_2-high split exhibits the lowest enhancement (10%). In addition, the SiH_4-high sample also shows a 23%

Manuscript received November 25, 2009; revised January 3, 2010. First published February 22, 2010; current version published March 24, 2010. This work was supported by the National Science Council, Taiwan, under Contract NSC-97-2221-E-009-153-MY3. The review of this letter was arranged by Editor C.-P. Chang.

The authors are with the Department of Electrophysics, National Chiao Tung University, Hsinchu 30010, Taiwan (e-mail: forgethard@hotmail.com; TYChiang0528@gmail.com; yihsin0914@hotmail.com; tschao@mail.nctu.edu.tw).

Color versions of one or more of the figures in this letter are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/LED.2010.2041524

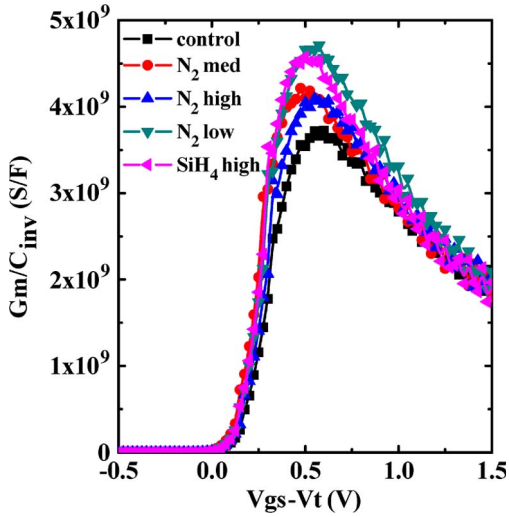

Fig. 1. $Gm/C_{\rm inv}$ for NMOSFET with different capping layer SiNs (1500 Å) in the linear region with a channel length that is equal to 0.4 μm.

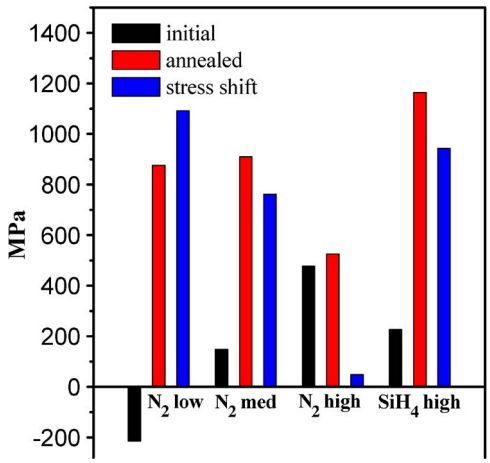

Fig. 2. Initial strain, after annealing strain, and strain shift by the annealing process of different N_2 and SiH_4 fluxes.

improvement. Fig. 2 shows the stress characteristics of all capping layers for comparison as initial stress, final stress, and stress shifts (defined as final stress—initial stress). The different SiN splits have a comparable etching rate in hot H_3PO_4 solution, and thus, the films have analogous density.

This investigation was focused on a compressive SiN layer rather than TEOS oxide, or a tensile SiN layer was used to execute SMT. After the annealing process, all SiN splits became more tensile. As shown in Fig. 1, the compressive SiN (N_2-low) that we used offered the greatest enhancement of mobility instead of the highest tensile capping layer, as is conventionally used. This result is contradictory to the usual practice of optimizing the tensile capping layer to enhance NMOS mobility.

Based on the experimental results, we postulate a model (Fig. 3) to clarify the mechanism. Previous research suggests that we should use a material with a higher Young's modulus as a spacer to confine the expanded poly gate, and use the PAI technique to make the grain recrystallize effectively. In this letter, we thoroughly investigate the capping layer and postulate another mechanism for enhancing mobility. N_2-rich and N_2-low possess Young's modulus values of 166.562 and

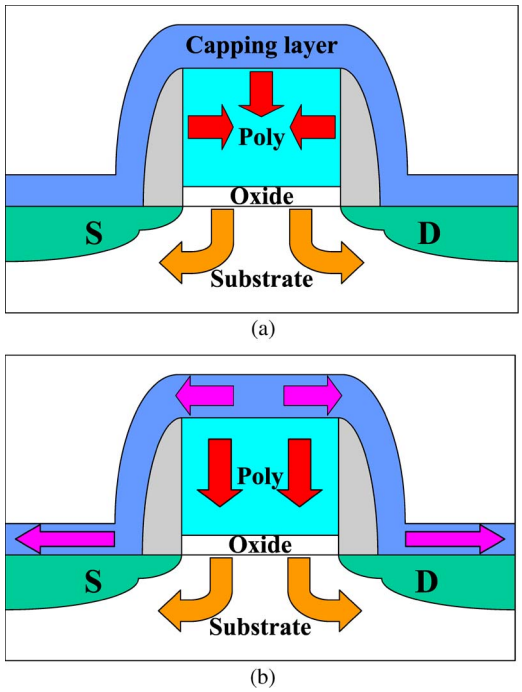

Fig. 3. (a) Traditional thermal poly-Si expansion mechanism. (b) New model with the mechanically induced strain mechanism for the capping layer.

149.386 GPa, respectively. In conventional SMT, as shown in Fig. 3(a), the poly-Si gate in the RTA process expands its volume by an abrupt increase in temperature and tends to be more compressive due to confinement of volume by the spacer and capping-layer wrapping. Subsequently, the poly-Si gate expands its volume into the channel region during the annealing process. The channel eventually becomes vertically compressive and subsequently tends to be tensile. However, N_2-rich does not show a higher mobility than that of the N_2-low one, which is not consistent when Young's modulus is considered solely. Our experimental results show that the capping layer also plays an obvious role, in addition to the poly-Si thermal expansion condition. Because N_2-low and SiH_4-high splits possessing higher stress shifts have higher mobility, the benefit of mobility enhancement is the split with the highest stress shift instead of the original or final most tensile capping layer. It is found that the film with the highest stress shift tends to compress the channel more during the annealing process, as shown in Fig. 3(b). This is consistent with the model that the larger the shift of stress, the higher the vertical stress shift [4]. The highest stress shift in vertical results in the largest compressive strain gain vertically, making the channel more tensile horizontally and enhancing the electron mobility in return. Consequently, not only the volume confinement during annealing as traditional viewpoint but also the intrinsic-stress shift of the capping layer have an influence on the SMT technique.

The mechanisms that we explore, along with the results that we have obtained, indicate the existence of a greater variety of choice for the optimization of SMT. Because the compressive SiN layer that we use possesses the greatest stress-shift potential, the choice of materials for optimization is now broader. Fig. 4 shows the $I_{\rm DS}$–$V_{\rm GS}$ characteristics. All devices display a similar subthreshold swing in the $V_g = 0.05$-V condition,

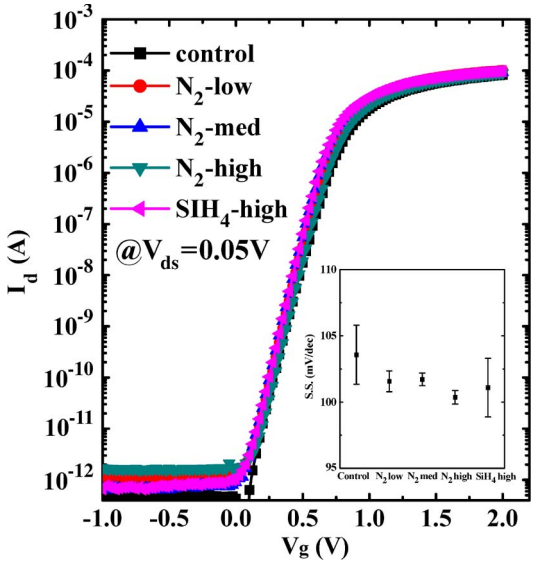

Fig. 4. I_d–V_g for different SiNs at $V_g = 0.05$-V characteristics. The inset shows the subthreshold swing for different SiNs.

which implies an identical interface at the oxide/substrate for capping-layer optimization.

IV. Conclusion

In summary, we have found that the stress-shift quantity rather than the initial- or final-stress amount should be considered in SMTs. The initial compressive SiN capping layer has a higher stress-shift potential than all tensile SiN capping-layer splits in CESL and SMT adoption, and the largest enhancement in mobility is achieved. This implies that the traditional choice of capping layer should be widened to include the stress shift with the highest split. Finally, we clarified the physical mechanism to explain the capping-layer impact on SMT. Based on this investigation, we believe that the limit thickness of the capping layer can be used to attain the highest mobility enhancement for the scaling pitch demand.

Acknowledgment

The authors would like to thank the National Nano Device Laboratories for the process support and the Nano Facility Center, National Chiao Tung University.

References

[1] P. R. Chidambaram, C. Bowen, S. Chakravarthi, C. Machala, and R. Wise, "Fundamentals of silicon material properties for successful exploitation of strain engineering in modern CMOS manufacturing," *IEEE Trans. Electron Devices*, vol. 53, no. 5, pp. 944–964, May 2006.

[2] J. S. Lim, S. E. Thompson, and J. G. Fossum, "Comparison of threshold-voltage shifts for uniaxial and biaxial tensile-stressed n-MOSFETs," *IEEE Electron Device Lett.*, vol. 25, no. 11, pp. 731–733, Nov. 2004.

[3] M. Li, K. H. Yeo, Y. Y. Yeoh, S. D. Suk, K. H. Cho, D. W. Kim, D. Park, and W. S. Lee, "Experimental investigation on superior PMOS performance of uniaxial strained ⟨110⟩ silicon nanowire channel by embedded SiGe source/drain," in *IEDM Tech. Dig.*, 2007, pp. 899–902.

[4] S. Orain, V. Fiori, D. Villanueva, A. Dray, and C. Ortolland, "Method for managing the stress due to the strained nitride capping layer in MOS transistors," *IEEE Trans. Electron Devices*, vol. 54, no. 4, pp. 814–821, Apr. 2007.

[5] K. Ota, K. Sugihara, H. Sayama, T. Uchida, H. Oda, T. Eimori, H. Morimoto, and Y. Inoue, "Novel locally strained channel technique for high performance 55 nm CMOS," in *IEDM Tech. Dig.*, 2002, pp. 27–30.

[6] C. H. Chen, T. L. Lee, T. H. Hou, C. L. Chen, C. C. Chen, J. W. Hsu, K. L. Cheng, Y. H. Chiu, H. J. Tao, Y. Jin, C. H. Diaz, S. C. Chen, and M. S. Liang, "Stress memorization technique (SMT) by selectively strained-nitride capping for sub-65 nm high-performance strained-Si device application," in *VLSI Symp. Tech. Dig.*, 2004, pp. 56–57.

[7] H. Ohta, N. Tamura, H. Fukutome, M. Tajima, K. Okabe, A. Hatada, K. Ikeda, K. Ohkoshi, T. Mori, K. Sukegawa, S. Satoh, and T. Sugii, "High performance sub-40 nm bulk CMOS with dopant confinement layer (DCL) technique as a strain booster," in *IEDM Tech. Dig.*, 2007, pp. 289–292.

[8] T. Miyashita, T. Owada, A. Hatada, Y. Hayami, K. Ookoshi, T. Mori, H. Kurata, and T. Futatsugi, "Physical and electrical analysis of the stress memorization technique (SMT) using poly-gates and its optimization for beyond 45-nm high-performance applications," in *IEDM Tech. Dig.*, 2008, pp. 1–4.

[9] T. Y. Lu, C. M. Wang, and T. S. Chao, "Enhancement of stress-memorization technique on nMOSFETs by multiple strain-gate engineering," *Electrochem. Solid-State Lett.*, vol. 12, no. 1, pp. H4–H6, 2008.