Journal of The Electrochemical Society, **157** (5) H497-H500 (2010)
0013-4651/2010/157(5)/H497/4/$28.00 © The Electrochemical Society

Strained Silicon Technology: Mobility Enhancement and Improved Short Channel Effect Performance by Stress Memorization Technique on nFET Devices

Chih-Cheng Lu,[z] Jiun-Jia Huang, Wun-Cheng Luo, Tuo-Hung Hou, and Tan-Fu Lei

Department of Electronics Engineering and Institute of Electronics, National Chiao-Tung University, Hsin-Chu 300, Taiwan

This paper presents a fundamental study of a stress memorization technique (SMT), which utilizes a capping nitride dielectric film to enhance negative channel field-effect transistor (nFET) device performance. SMT strain engineering is highly compatible with current standard complementary metal oxide semiconductor processes without introducing substantial additional complexity. In this work, we report that SMT-strained nFET exhibits a higher transconductance G_{m_lin}, which indicates strain-induced electron mobility enhancement. The nFET short channel effect is also improved by the SMT process. Improved V_t roll-off characteristics manifest itself and are shown to result from retarded junction diffusion as indicated by secondary-ion mass microscopy analysis. Finally, this work demonstrates that when combined with a strained contact etch stop layer (CESL) technique, SMT provides additional strain beyond that provided by the CESL, which results in further improved nFET performance.
© 2010 The Electrochemical Society. [DOI: 10.1149/1.3321948] All rights reserved.

Manuscript submitted October 8, 2009; revised manuscript received January 18, 2010. Published March 15, 2010.

Historically, the number of transistors on an integrated circuit has doubled approximately every 2 years according to the well-known Moore law, a trend which continues in the present-day semiconductor manufacturing technology. To sustain this trend, the device's physical dimensions (width, length, junction depth, etc.) are scaled by approximately 0.7× in the subsequent technology node to achieve a 50% reduction in transistor active area, ultimately resulting in today's complementary metal oxide semiconductor (CMOS) devices with nanometer-scale dimensions. However, the increased channel impurity concentration and the vertical field required to control the short channel effect (SCE) have also been increased to the point where they substantially degrade the carrier mobility. Maintaining the short channel control without sacrificing the drive current performance is one of the most fundamental challenges to be overcome in further scaling today's nanoscale transistors.

Among several approaches, carrier mobility can be enhanced with the use of strain channel techniques. Comparing with the traditional method of using standard bulk silicon, the transistor performance can be boosted when the strain is transferred into the channel. The strained silicon technology was first developed using a virtual substrate to create a strain layer in the CMOS channel region aiming at strain-induced carrier mobility enhancement. It continues to elicit substantial interest for advanced CMOS technology. Several state-of-the-art and low cost strained silicon technologies have been developed and integrated in advanced CMOS fabrication processes. For instance, a highly tensile nitride cap layer used as a contact etch stop layer (CESL) has been widely utilized in advanced metal-oxide-semiconductor field-effect transistor (MOSFET) fabrication technologies owing to the resulting uniaxially tensile mechanical stress created in the negative channel field-effect transistor (nFET) channel region.[1,2] Another strained silicon technology, known as the stress memorization technique (SMT), proposed by Chen et al.,[3] which utilizes a removable tensile nitride cap layer deposited on top of the polygate electrode, further opens up the opportunities for cost-effective and process-friendly methods to achieve a higher level of strain in the advanced CMOS fabrication technology. The SMT improves the nFET performance by depositing a tensile nitride stressor on top of the polygate electrode after source/drain (S/D) implantation. This layer is present during the subsequent dopant activation annealing and is then stripped off after this annealing, resulting in a higher nFET drive current by the process flow shown in Fig. 1. While a detailed understanding of the fundamental mechanism of the SMT process remains the subject of active discussion, recent works have suggested that the improved nFET performance results from the "memorization" in the increased tensile stress in the nFET channel.[3]

This work further demonstrates that, in addition to the aforementioned nFET drive current improvement, SMT also enables better SCE control. At the same time, the SMT nitride cap layer modifies the formation of the defects created by ion implantation and hence alters the diffusion kinetics of the implanted dopants. Furthermore, when combined with the strained CESL technique, SMT provides supplementary strain to the n-channel, elucidating the prospects for aggressively scaled CMOS fabrication featuring strained silicon technology.

Experimental and Characterization

The devices were fabricated in the National Nano Device Laboratories (NDL) on 6 in. (150 mm) silicon wafers using a conventional MOSFET process flow that includes local oxidation of silicon isolation, gate oxide, spacer, S/D implantation, thermal annealing, and Al metallization processing, as shown in Fig. 1. All the nFET devices characterized in this study had 3 nm thick gate oxide thermally grown in a vertical furnace and a 150 nm thick poly-Si layer as the gate electrode. After self-aligned spacer formation, a 100 A plasma-enhanced (PE) TEOS oxide was deposited before the S/D implantation to reduce the N^+ region junction depth. Before the S/D activation annealing, a similar SMT process similar to that of Ref. 3 was employed for the SMT-strained nFET while skipping this SMT nitride capping process for the nonstrained nFET control devices. For the SMT-strained nFETs, an 800 A tensile PE nitride film was deposited at a relatively low temperature of 300°C over the entire device structure including the polygate and S/D region. The main reaction gases utilized for the SMT nitride formation were NH_3, N_2, and SiH_4. To facilitate metallization, the SMT nitride film was then wet removed using chemical etchants H_3PO_4 and diluted 100:1 HF. Next, passivation was performed by 300 nm thick TEOS oxide, followed by patterning and etching of and Al metallization.

After the completion of the metallization processes, the wafers were annealed at 400°C in a forming gas ambient, and the fabricated device wafers were characterized electrically. The nominal device dimension evaluated was a gate length of 0.4 μm and a width of 1 μm. The device characteristics were measured by a semiconductor parameter analyzer (HP 4156A, plus Agilent ICS Software). The threshold voltage V_t was determined by the constant drain current method when the drain current was set to 10^{-7} W/L (ampere).[4] The linear and saturation mode threshold voltages, V_{t_lin} and V_{t_sat}, were extracted at drain voltages of 50 mV and 2 V, respectively.

The identical S/D implantation condition was chosen for the pro-

[z] E-mail: ritexluu@yahoo.com.tw

Downloaded on 2014-04-24 to IP 140.113.38.11 address. Redistribution subject to ECS terms of use (see ecsdl.org/site/terms_use) unless CC License in place (see abstract).

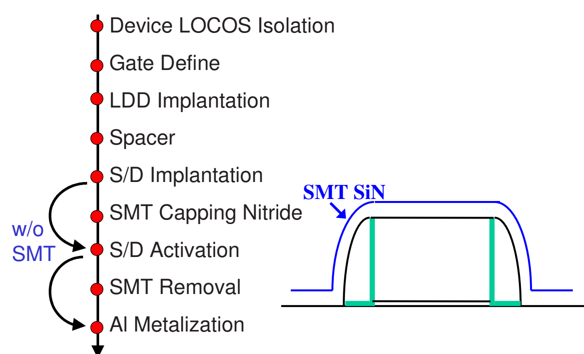

Figure 1. (Color online) SMT process flow. The SMT process deposits a nitride layer over the gate before annealing. The presence of this nitride layer during the subsequent S/D annealing and recrystallization of amorphized regions acts to increase tensile strain in the n-type metal-oxide semiconductor channel. Then, this nitride layer is removed, but the strain by the nitride cap introduced during the annealing is retained or "memorized."

cess on both blanket and device wafers. The diffusion profiles of arsenic and phosphorus on blanket wafers were studied by secondary-ion mass spectrometry (SIMS) analysis and were used to quantify device junction depth and to study the arsenic and phosphorus diffusion. The dopant diffusion profiles from these SIMS results were in turn correlated with the extracted nFET V_t roll-off behavior, so as to investigate the SMT-strained nFETs' SCE characteristics.

Results and Discussion

SMT process mechanism and associated nFET device improvement.— The linear transconductance, G_{m_lin}, using the SMT process is compared with that of the non-SMT process. The superior G_{m_lin} with around 10% enhancement is shown in Fig. 2. These results suggest that the SMT process induces a permanent stress in the channel region to enhance nFET carrier mobility and results in the improvement of the device performance, which agrees with Chen et al.[3] In addition, the strain can be retained and is attributed to the stress memorization effect even with the removal of the nitride stressor film. The SMT strain results from the residual stress memorized in the channel after nitride removal, rather than the initial mechanical stress created after nitride deposition as in the conventional CESL technique.[1,2] Many studies with the SMT process[3,5-7] suggest that the process involves the following fundamental elements: S/D n-type dopant implantation, amorphized silicon implanted region, capping nitride deposition, S/D thermal annealing, and amorphous silicon transformation into recrystallized silicon with plastic silicon

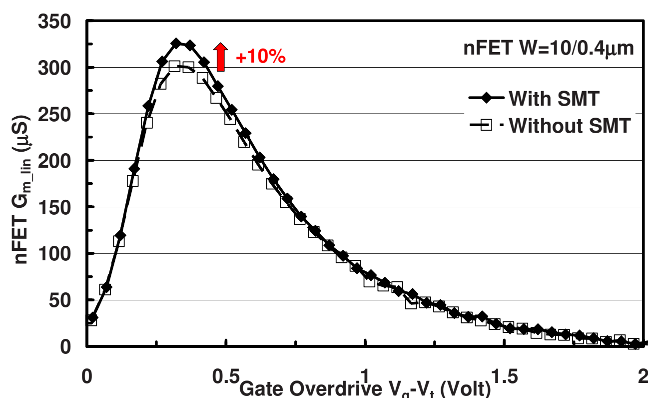

Figure 2. (Color online) Measured 10 × 0.4 µm nFET G_{m_lin}. There is an ~10% improvement in the peak value of G_{m_lin} with the adoption of SMT. The nFET S/D is implanted with arsenic.

deformation. While the precise physics underlying the stress memorization is still unclear, recent studies have given some insight into the problem. Adam et al.[5] proposed a phenomenological model to understand the stress memorization effect. A plastic deformation model is used to simulate such irreversible shape change in the polysilicon gate. Their simulation results show that the polysilicon gate undergoes a structure deformation via a transition from the elastic to the plastic states during the S/D anneal. Consequently, this plastic deformation generates a tensile strain in the channel lateral direction and a compressive strain in the channel vertical direction and can boost electron mobility according to the silicon piezoresistance model.[8,9]

Moreover, the capping nitride film also affects the poly-Si stack deformation by the SMT process, as depicted in Ref. 5, by suppressing an outward expansion of the poly-Si gate. This scenario is also supported by Miyashita et al.[6] in which the correlation between nitride film porosity and the resulting channel stress are analyzed. In Miyashita et al.'s work, the gate polysilicon is amorphized by S/D dopant implantation and then capped with the SMT nitride layer. During the S/D annealing, the amorphized silicon expands as it recrystallizes, but because it is constrained by the capping nitride and sidewall spacers, compressive vertical stress is thus generated and applied to the channel underneath. It has also been suggested by Ortolland et al.[7] that harder nitride, with lower porosity and higher capping strength, is more effective to conduct the memorized stress to the channel. In addition, the gate polysilicon's volume can further expand after implantation with high atomic mass n-type species such as arsenic or phosphorus. Because the rigid capping tensile-strained nitride film acts as a boundary layer to suppress poly-Si volume outward expansion during the recrystallization of amorphous poly-Si upon rapid thermal annealing (RTA),[3,5] a tensile strain along the channel direction is generated in the device itself. This is also consistent with results reporting that a rigid capping nitride film plays a major role in the SMT strain generation process.[7]

Improved nFET SCE with the SMT process.— In addition to the enhancement of nFET transconductance and mobility by permanent strain memorization, as shown in Fig. 2, the variations in threshold voltages with and without the SMT process were further investigated. As the drain voltage is increased to switch the nFET operation from the linear region to the saturation region, the depletion region of the p–n junction between the drain and body increases in size and extends under the gate. Thus, the drain assumes a greater portion of the burden of balancing depletion region charge, leaving a smaller burden for the gate. Therefore, the device threshold voltage decreases with shorter channel length conditions, a phenomenon commonly referred to as "V_t roll-off" or SCE. Another index commonly used to represent the SCE is drain-induced barrier lowering (DIBL), as predicted based on Eq. 1 by $\Delta V_t/\Delta V_{ds}$. Because of the smaller V_{t_sat} for shorter length nFET devices, DIBL increases as the channel length is reduced. This suggests that the channel built-in potential or the barrier height for electrons would be decreased when the channel length is highly scaled, thus sharply increasing the off-state leakage. Hence, to suppress SCE is an important task for manufacturing highly scaled device geometries

$$\text{DIBL} = \frac{\Delta V_t}{\Delta V_{ds}} = \frac{V_{t_lin} - V_{t_sat}}{2 - 0.05} 1000 \left(\frac{\text{mV}}{\text{V}}\right) \quad [1]$$

The nFET V_{t_lin} and V_{t_sat} (extracted from electrically measured I_d-V_g transfer characteristics) for both SMT-strained and nonstrained control devices are plotted as a function of the transistor channel length in Fig. 3, and the calculated DIBL is shown in Fig. 4. Evidently, there is less roll-off of the threshold voltage for the split with SMT compared to the non-SMT-strained split, and thus the DIBL is improved with the adoption of the SMT process. SMT provides nFET a higher immunity to SCE because of the reduction in the threshold voltage roll-off and the improved DIBL performance. This is another major advantage of SMT nonextensively reported in previous literature: The improved SCE immunity enables devices fab-

Downloaded on 2014-04-24 to IP 140.113.38.11 address. Redistribution subject to ECS terms of use (see ecsdl.org/site/terms_use) unless CC License in place (see abstract).

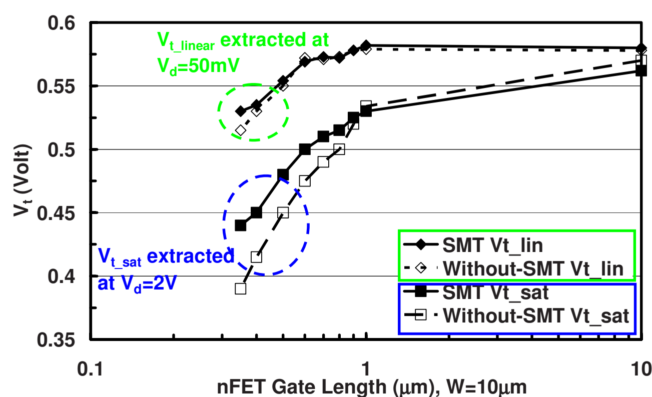

Figure 3. (Color online) Measured $W = 10$ μm nFET V_t as a function of channel length; SMT improves nFET SCE, as indicated by less V_t roll-off.

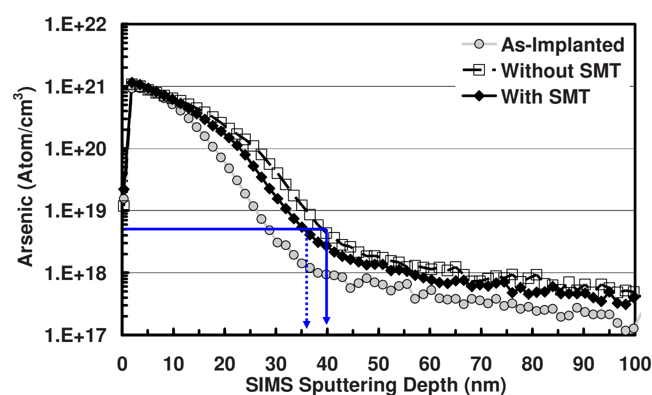

Figure 5. (Color online) SIMS analyses for arsenic implanted wafers. The SIMS diffusion profiles suggest that SMT retards arsenic diffusion upon thermal annealing, resulting in 10% reduction in X_j.

ricated with reduced pocket implantation dose (reducing mobility degradation by coulomb scattering in the channel and silicon lattice damage) and increased N^+ S/D dosage (reducing conducting resistance). Both of these techniques, when properly employed, can further enhance the SMT-strained nFET drive current under comparable SCE conditions.

Retarded dopant diffusion and shallower junction depth X_j with SMT process.— One of the major challenges for MOSFET scaling in the sub-130 nm regime is the formation of ultrashallow junctions (USJs). The International Technology Roadmap for Semiconductors indicates that future generations of CMOS technology should have S/D extension junctions that are <15 nm deep, with sheet resistance $R_s \sim 1000\ \Omega/\square$, to keep pace with historical improvements in high performance logic devices.[10] This USJ requirement stems from the need to suppress SCEs, and dictates a very limited thermal annealing budget to suppress dopant diffusion. To form the S/D regions of the MOSFET transistor with low sheet resistance, it is necessary to ensure low parasitic resistance by adopting a high annealing temperature to maximize S/D dopant activation. The S/D annealing in the present-day advanced CMOS processes is typically performed at a temperature between 900 and 1100°C to activate the impurity implanted regions. However, during annealing to activate the implanted dopants under such high temperature, the n-type dopants can diffuse into the crystalline semiconductor substrate such that the junction depth X_j increases.[11,12]

To understand the root cause of the improved SCE performance shown in Fig. 3 and 4, the same S/D dopant implantation conditions and activation annealing process steps were applied to the 6 in. Si(100) blanket wafers. n-Type dopants, arsenic, and phosphorus concentration profiles inside the silicon substrate were characterized by SIMS analysis. In the following study, the junction depth X_j (depth in nanometers) is taken to be the depth at which the n-type dopant concentration falls to 5×10^{18} atoms/cm^3, as measured from the SIMS profiles.

The SIMS analyses for arsenic and phosphorus implanted wafers, before and after activation anneal, are shown in Fig. 5 and 6, respectively. Clearly, S/D thermal activation annealing results in junctions deeper than the as-implanted profiles. The high ($\sim 5 \times 10^{15}$ atoms/cm^2) S/D implant doses applied in modern MOSFET devices nowadays are sufficient to amorphize the silicon surface region.[11] Consequently, the subsequent S/D dopant activation annealing, adopted either by high temperature furnace or RTA annealing, results in substantial transient enhanced diffusion (TED) for both arsenic and phosphorus, culminating in junction depths considerably deeper than the as-implanted profiles, as shown in Fig. 5 and 6. The origin of this enhanced dopant diffusion could be the annealing of ion-implantation damage in the silicon substrate and the formation of an amorphous silicon surface layer by high dose arsenic ($>2 \times 10^{14}$ cm^{-2}) implantation, as suggested by Fair et al.[11]

Nevertheless, reduced diffusion is seen for the arsenic and phosphorus profiles after activation annealing with the adoption of the SMT process. More specifically, there is a 10 and 13% X_j reduction for arsenic and phosphorus by the addition of the SMT process under the same implantation and activation annealing conditions. The SIMS data shown in Fig. 5 and 6 imply that a similar retarded dopant diffusion can also be obtained for the lateral or longitudinal direction underneath the spacer and gate on a real nFET structure.

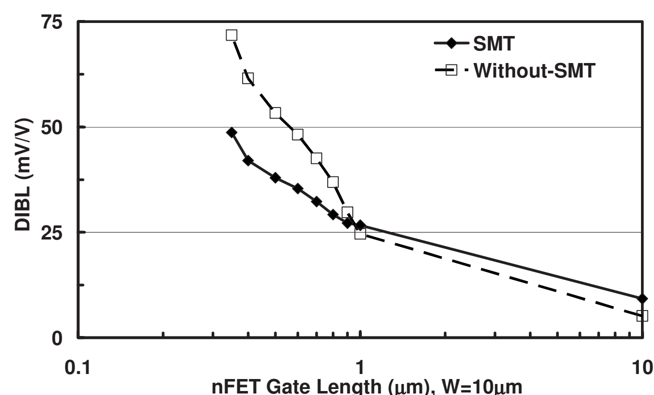

Figure 4. Calculated $W = 10$ μm nFET DIBL. SCE improvement by SMT is indicated by improved short channel DIBL.

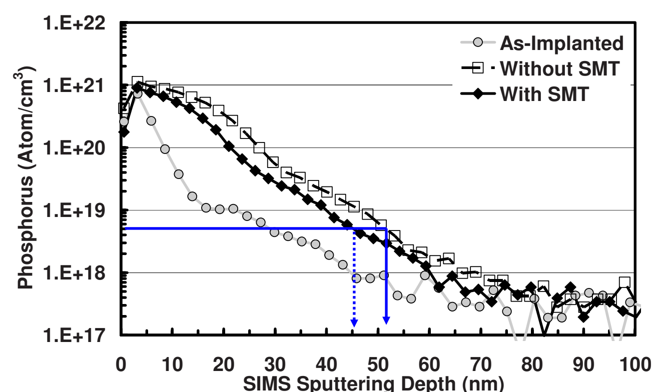

Figure 6. (Color online) SIMS analyses for phosphorus implanted wafers. The SIMS diffusion profiles suggest that SMT retards phosphorus diffusion upon thermal annealing, with 13% reduction in X_j.

Downloaded on 2014-04-24 to IP 140.113.38.11 address. Redistribution subject to ECS terms of use (see ecsdl.org/site/terms_use) unless CC License in place (see abstract).

Therefore, the lateral encroachment of n-type S/D implants under the gate can be reduced to form a shallower n-type S/D junction. Moreover, the undesired device SCE can also be improved by reducing the junction depth to lessen the drain voltage impact onto the channel potential, enabling the channel potential to be primarily controlled by the gate voltage V_g.[13]

In brief, the SIMS results in Fig. 5 and 6 suggest that the SMT process retards the n-type dopant diffusion and the lowered junction depth X_j can be obtained when compared with the wafers processed without SMT. This is consistent with previous findings that SMT enables the fabrication of nFETs with improved SCE, as shown in Fig. 3 and 4.

Proposed model for retarded diffusion with SMT.— Thermal annealing of wafers implanted by high dose arsenic ($>2 \times 10^{14}$ cm^{-2}) is believed to produce large transient concentrations of point defects that may continue to exist for the duration of the annealing.[11] Such point defects are generally highly mobile at high temperatures. They also increase the mobility of the atoms of the host lattice and thus cause a solid-state diffusion. The concentration and specific nature of point defects (impurity atoms), their migration parameters, equilibrium or nonequilibrium conditions, and the atomic mechanisms of diffusion determine the final diffusion profile in a specific solid-state experiment. At the same time, such point defects may effectively lengthen the duration of the diffusivity transient region and give rise to lowered dopant diffusivity upon thermal activation, generating the aforementioned retarded n-type dopant diffusion behavior in the silicon substrate itself.[14] In addition, it has also been proposed and experimentally confirmed that point defects do retard n-type dopant diffusion upon thermal annealing.[15] Such point defect generated via either shallow end-of-range damage or small clusters of point defects during the annealing can dominate the whole TED process. For example, Kong et al.[15] observed retarded arsenic diffusion by specific implant conditions that created a vacancy-rich region overlapped with the principle dopant region. Two possible mechanisms, interstitial-vacancy recombination and dopant clustering, have been proposed and analyzed based on their diffusion retardation.[15]

With a similar concept, the retardation of n-type dopant diffusion by the SMT process during high temperature thermal annealing demonstrated in our work is likely caused by the generation of point defects in the S/D regions of the transistor after implantation and capping nitride deposition. Such excess interstitials or point defects form a retarding gradient during activation annealing and would oppose n-type dopant diffusion. Hence, retarded and shallower dopant diffusion profiles are formed after SMT processing, improving the nFET SCE characteristics. Thus, in this paper, the fabrication of USJs on nFET devices is demonstrated with proposed SMT engineering and associated point defect manipulation approaches.

nFET device strain superposition by CESL and SMT.— Strained channel effect from CESL techniques can be enhanced by increasing the CESL nitride thickness;[1,2] however, the premetal dielectric gapfill capability limits this approach at highly scaled geometries. Therefore, combining more than one stressor to further boost the device current is another promising field to be explored. A further experiment in our work addressed this by fabricating nFET devices with a process integrating both SMT and CESL techniques. Performance gains from SMT and uniaxial 800 A tensile CESL nitride onto transistor channels are demonstrated to be additive, as indicated by the nFET output characteristics shown in Fig. 7.

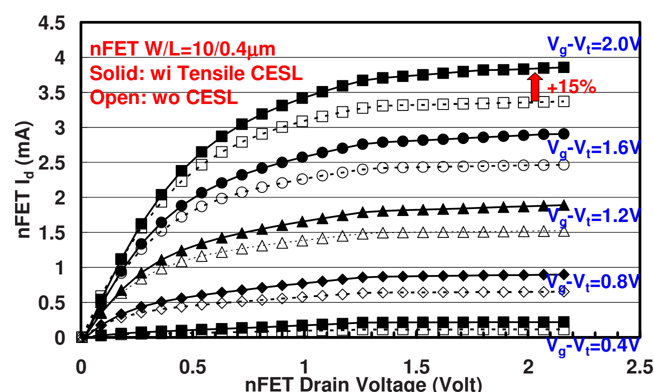

Figure 7. (Color online) Measured 10×0.4 µm nFET I_d-V_d output characteristics. There is a further ~15% I_d improvement with the adoption of 800 A tensile CESL SiN at gate overdrive = 2.0 V and V_d = 2.0 V.

Conclusion

We have investigated the effects of utilizing tensile-strained nitride deposited as an SMT stressor to enhance nFET device performance. The SMT improves nFET performance and transconductance because of the enhancement of electron mobility by the memorized strain, and a higher on-current to off-current ratio (I_{on}/I_{off}) can be obtained for highly scaled nFET devices. Furthermore, the SMT strain technique suppresses nFET SCEs, enables less V_t roll-off, and improves DIBL because of a reduced junction depth X_j, as demonstrated by the SIMS profiles. Finally, we demonstrate that the strain enhancements from SMT and conventional tensile CESL nitride are mutually additive and can enhance transistor performance even further when both techniques are integrated.

Acknowledgment

The authors thank all the team members in the National NDL for their excellent support of device fabrication and electrical characterization and for many stimulating technical discussions.

National Chiao Tung University assisted in meeting the publication costs of this article.

References

1. A. Shimizu, K. Hachimine, N. Ohki, M. Koguchi, Y. Nonaka, H. Sato, and F. Ootsuka, *Tech. Dig. - Int. Electron Devices Meet.*, **2001**, 433.
2. H. S. Yang, R. Malik, S. Narasimha, Y. Li, R. Divakaruni, P. Agnello, S. Allen, A. Antreasyan, J. C. Arnold, K. Bandy, et al., *Tech. Dig. - Int. Electron Devices Meet.*, **2004**, 1075.
3. 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, et al., *Dig. Tech. Pap. - Symp. VLSI Technol.*, **2004**, 56.
4. D. K. Schroder, *Semiconductor Material and Device Characterization*, 3rd ed., p. 225, John Wiley & Sons, New York (1998).
5. L. S. Adam, C. Chiu, M. Huang, X. Wang, Y. Wang, S. Singh, Y. Chen, H. Bu, and J. Wu, in *International Conference on Simulation of Semiconductor Processes and Devices (SISPAD) 2005*, IEEE, pp. 139–142 (2005).
6. T. Miyashita, T. Owada, A. Hatada, Y. Hayami, K. Ookoshi, T. Mori, H. Kurata, and T. Futatsugi, *Tech. Dig. - Int. Electron Devices Meet.*, **2008**, 1.
7. C. Ortolland, P. Morin, C. Chaton, E. Mastromatteo, C. Populaire, S. Orain, F. Leverd, P. Stolk, F. Boeuf, and F. Arnaud, *Dig. Tech. Pap. - Symp. VLSI Technol.*, **2006**, 78.
8. S. E. Thompson, G. Sun, Y. S. Choi, and T. Nishida, *IEEE Trans. Electron Devices*, **53**, 1010 (2006).
9. S. E. Thompson, M. Armstrong, C. Auth, M. Alavi, M. Buehler, R. Chau, S. Cea, T. Ghani, G. Glass, T. Hoffman, et al., *IEEE Trans. Electron Devices*, **51**, 1790 (2004).
10. International Technology Roadmap for Semiconductors, 2006 Update.
11. R. B. Fair, J. J. Wortman, and J. Liu, *J. Electrochem. Soc.*, **131**, 2387 (1984).
12. Y. Sasaki, K. Itoh, and T. Sei-ichi, *Jpn. J. Appl. Phys., Part 1*, **28**, 1421 (1989).
13. Y. Taur, *Fundamentals of Modern VLSI Devices*, Cambridge University Press, New York (1998).
14. X. Qi, U.S. Pat. 0,054,164 A1 (2005).
15. N. Kong, S. K. Banerjee, T. A. Kirichenko, S. G. H. Anderson, and M. C. Foisy, *Appl. Phys. Lett.*, **90**, 062107 (2007).

Downloaded on 2014-04-24 to IP 140.113.38.11 address. Redistribution subject to ECS terms of use (see ecsdl.org/site/terms_use) unless CC License in place (see abstract).