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

Chih-Cheng Lu, a,1 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: 10.1149/1.33231948] 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) technology. For the SMT-strained nFETs, an 800 Å tensile PE nitride film 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 Å 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 NH3,N2, and SiH4. To facilitate metallization, 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 Å 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. The SMT-strained nFETs, an 800 Å 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 NH3, $N_2$, and SiH4. To facilitate metallization, the SMT nitride film was then wet removed using chemical etchants H3PO4 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). 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-
teristics.

behavior, so as to investigate the SMT-strained nFETs' SCE characteristics compared with that of the non-SMT process. The superior results were in turn correlated with the extracted nFET dopant diffusion profiles from these SIMS measurements 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 characteristics, 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. 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 in the nitride deposition process. The SMT 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 capping nitride deposition, SMT SiN cap introduced during the annealing is retained or "memorized." This 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. 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 Dolland et al. that harder nitride, with lower plasticity 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-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), 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.

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 \(\frac{\Delta V_t}{\Delta V_{ds}}\). 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.

\[
\text{DIBL} = \frac{\Delta V_t}{\Delta V_{ds}} = \frac{V_{t,lin} - V_{t, sat}}{2 - 0.05} \times 100 \text{ mV/V} \tag{1}
\]

The nFET \(V_{t,lin}\) and \(V_{t, sat}\) (extracted from electrically measured \(I_d-V_t\) 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-
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. 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 $1100 ^\circ C$ 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. \cite{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 $5 \times 10^{18}$ 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. \cite{13}

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. \cite{14}

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.
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 $V_g$.

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.$^{13}$ 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, the retarded and shallower 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.$^{23}$ However, the premetall dielectric gap-fill 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.

**Conclusion**

We have investigated the effect 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 lower $V_T$ 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**