Extended Abstracts of the 2008 International Conference on Solid State Devices and Materials, Tsukuba, 2008, pp. 874-875

B-8-4

Steeper Indium Halo Formation of nMOSFET by Reducing Interstitial Supersaturation with Flash Lamp pre-Annealing and Its Modeling with Atomistic Kinetic Monte Carlo

T. Sawada[2], N. Zographos[4], H. Yoshimura, T. Sanuki, T. Ito[1], T. Itani[1], H. Aikawa, O. Fujii, N. Kariya[3],
M. Oulmane[4], R. Gull[4], Y. Akiyama[3], M. Ikeda[3], M. Iwai and F. Matsuoka
System LSI Div., Process & Manufacturing Eng. Center[1], Toshiba Corp., Semiconductor Company, Toshiba Information Systems Technology Inc.[2],
NEC Electronics Corp.[3], 8, Shinsugita-cho, Isogo-ku, Yokohama, 235-8522, Japan, E-mail: sawadat@semicon.toshiba.co.jp
Synopsys Switzerland LLC[4], Affolternstrasse 52 CH-8050 Zurich, Switzerland

Abstract

By applying flash lamp annealing (FLA) prior to spike RTA, we have successfully suppressed the dose loss and transient enhanced diffusion (TED) of indium for the first time. With steeper indium halo, a saturation drain current (I_{dsat}) increase of 8% for 34 nm gate nMOSFET is demonstrated. Mechanisms of reduced dose loss are investigated by atomistic kinetic Monte Carlo (kMC) simulation. With FLA, initially created end-of-range (EOR) defects evolve into few and big {311} defects and even into loops. As a result, the interstitial supersaturation is suppressed by a factor of 1/100 which, in turn, reduces TED and dose loss of indium at following spike RTA. Furthermore, by combining carbon co-implantation with FLA, maximum indium concentration can be increased to 1×10^{19} cm^{-3}.

Introduction

With scaling of MOSFETs, channel doping increases to mid 10^{18}cm^{-3} at 30 nm gate length (L_{gate}). For such scaled MOSFETs, a steep halo profile (Fig. 1 and Fig. 2) is required to control short channel effect (SCE) as well as a low channel doping to keep a high surface mobility [1]. However, realizing such a profile is difficult for nMOSFET because of the high diffusivity of boron. Indium has been a candidate [2] but it suffers from low solid solubility and out-diffusion during annealing [3]. It is difficult to obtain activation of indium over mid 10^{18}cm^{-3}. Thus, it cannot be applied to advanced nMOSFETs.

Recently, millisecond annealing (MSA) has been investigated due to its capability to combine high activation of dopants with diffusion-less profiles. Halo profile is also reported to be improved by laser spike annealing (LSA) [4]. But application of MSA for improving halo profiles has not been fully investigated in spite of its importance.

In this work, we have investigated the indium dose loss problems and demonstrated the improved indium halo profile by applying FLA prior to spike RTA. In addition, improving I_{dsat} for 34 nm nMOSFET is demonstrated. Furthermore, the mechanism of the suppressed dose loss of indium is clarified with kMC simulation. Finally, carbon co-implantation [5] with FLA is found additively effective to suppress indium dose loss and TED.

SIMS and Electrical Results

The dose dependency of the indium halo profile is investigated by SIMS. Maximum indium concentration outside the extension region is unchanged and is ~4×10^{18}cm^{-3} irrespective of changing indium halo dose as shown in Fig. 3. It indicates that indium shows large dose loss during spike RTA. Increasing implantation dose only results in the increase of indium concentration at the surface. Fig. 4 shows the V_{th} roll-off of nMOSFET with different indium halo doses. With increasing the dose above 4×10^{13} cm^{-2}, V_{th} stops increasing at shorter L_{gate}, because of indium dose loss.

We performed SIMS experiments for investigating the effect of FLA. After indium halo and arsenic extension I/I, SiN as sidewall was deposited at 550 ºC. We made three splits on annealing (Fig. 5). In addition to only performing spike RTA as reference (Flow A), FLA is performed before (Flow B) or after spike RTA (Flow C). The SIMS results (Fig. 6) demonstrate that, by performing FLA before the spike RTA, maximum indium concentration is increased from 4×10^{18} cm^{-3} to 7×10^{18} cm^{-3}.

The FLA+spike RTA scheme is applied to a 45 nm node nMOSFET where EOT of the gate oxide is 1.3 nm and a 60 nm thick gate capping tensile stress liner of 1.8 GPa is used for channel strain enhancement [6]. Steeper indium halo profile results in an improved V_{th} roll-off by 5 nm without degrading I_{dsat}-I_{off} relation (Fig. 7(a) and (b)). In L_{min}-I_{dsat} metric, I_{dsat} at I_{off} of 100 nA/um is increased from 1075 uA/um to 1160 uA/um at L_{gate} of 34 nm and V_{dd} of 1V (Fig. 7(c)). I_{dsat} shows an enhancement of 8%, which clearly indicates the effectiveness of applying FLA prior to spike RTA when using indium halo.

Analysis of indium halo profiles with kMC

In Fig. 8, from (a) to (d), we show XTEM micrographs after halo and extension I/I, SiN deposition, FLA and SpikeRTA, respectively. After I/I, 35 nm thick amorphous layer is created (Fig. 8(a)). It disappears by solid phase epitaxial regrowth during the 550ºC SiN deposition, and {311} defects are developed at EOR region (Fig. 8(b)). After FLA, it is confirmed that several loops are created (Fig. 8(c)) and disappear by SpikeRTA (Fig. 8(d)). Corresponding SIMS results (Fig. 9) show that indium dose is gradually reduced as process proceeds. But peak indium concentration is still maintained to be 1×10^{19} cm^{-3} and 8×10^{18} cm^{-3} after SiN deposition and FLA, respectively. Therefore, it is suggested that the effect of FLA is related to changing EOR defects to large {311} defects and then to loops.

To analyze the suppression of indium dose loss, we modeled the process flows with kMC and monitored the whole extended defect evolution at EOR. kMC accurately reproduces the defect evolution (Fig. 8(e)~(h)) and the final indium SIMS profiles after spike RTA and FLA+spike RTA (Fig. 10). During SiN deposition, simulation shows the formation of a lot of small and irregular {311} clusters, whereas during FLA, these interstitial clusters evolve to fewer and bigger {311} defects and even dislocation loops which is consistent with XTEM results. In addition, interstitials recombine at the surface. As a result, the spike RTA following FLA sees a (about 100 times) lower interstitial super- saturation than the one without FLA (Fig. 11), leading to reduced dose loss of indium. Table 1 is the summary of defect evolution, interstitial supersatiration and indium concentration on Flow A (only Spike RTA) and Flow B (FLA+Spike RTA).

Application of carbon co-implantation with FLA

In addition, we studied the effect of carbon co-implantation in order to improve the indium halo profile. Co-implantation is applied after Halo implantation (Fig. 5). Fig. 12 shows SIMS results of carbon co-implantation after spike RTA. Maximum indium concentration is increased to 8×10^{18} cm^{-3} and dose loss is reduced by carbon co-implantation. kMC simulation, which includes models for interstitial-carbon pair-diffusion and clustering, reproduces well the experimental data. Thus it is confirmed that the capture of interstitials by carbon is the dominant mechanism of suppressing TED of indium. Finally, the combination of FLA and carbon co-implantation for steeper indium halo formation is investigated. The SIMS results (Fig. 13) show that maximum indium concentration reaches 1×10^{19} cm^{-3}. Because the FLA effect and the co-implantation effect are two different mechanisms, their combined effect can be considered to be additive. This indicates that, by combining carbon co-implantation with performing FLA prior to spike RTA, an ideal indium halo profile can be achieved.

Conclusion

By applying FLA prior to spike RTA, performance improvement of sub 30 nm nMOSFET is demonstrated by realizing steeper indium halo profile. TED of indium is suppressed by applying FLA first due to the decrease of interstitial supersaturation. This mechanism is well modeled by atomistic kMC simulation. In addition, increasing indium maximum concentration up to 1×10^{19} cm^{-3} by using both FLA and carbon co-implantation is demonstrated.

References

[1] Y. Taur et al., *IEDM Tech. Dig.*, p. 789, 1998.
[2] K. Miyashita et al., *IEDM Tech. Dig.*, p. 645, 1999.
[3] S. Scalese et al., *J. Appl. Phys.*, Vol. 93, No. 12, p. 9773, 2003.
[4] T. Yamamoto et al., *VLSI Symp. Tech.*, p. 122, 2007.
[5] A. Mineji et al., *IWJT*, p. 84, 2006.
[6] T. Sanuki et al., *IEDM Tech. Dig.*, p. 281, 2007.

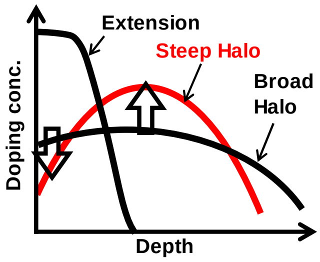

Fig. 1: Schematic 1D profile of broad and steep halo. Steep halo is suitable for advanced MOSFET.

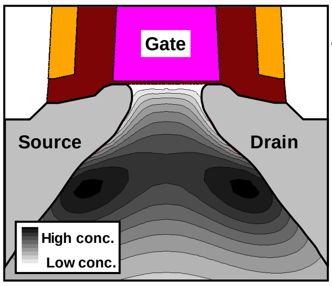

Fig. 2: The ideal halo profile to suppress SCE without degrading surface mobility.

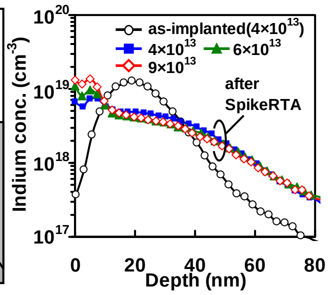

Fig. 3: SIMS data of indium halo profiles. Indium I/I dose dependence is shown.

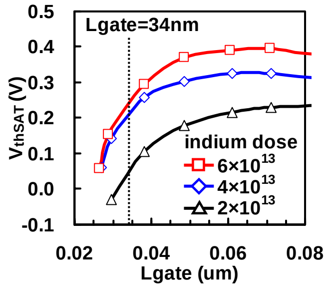

Fig. 4: V_{th} roll-off of nMOSFETs with different indium halo doses.

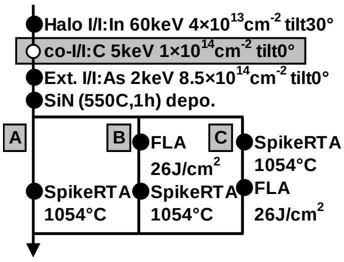

Fig. 5: Process flows for studying the effect of FLA on indium halo. FLA is done before or after spike RTA for flow B and C, respectively. Co-I/I is applied after Halo implantation.

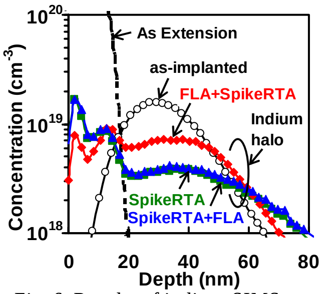

Fig. 6: Results of indium SIMS profiles after the process of Fig. 8. FLA prior to spike RTA (flow B) suppresses indium dose loss and TED.

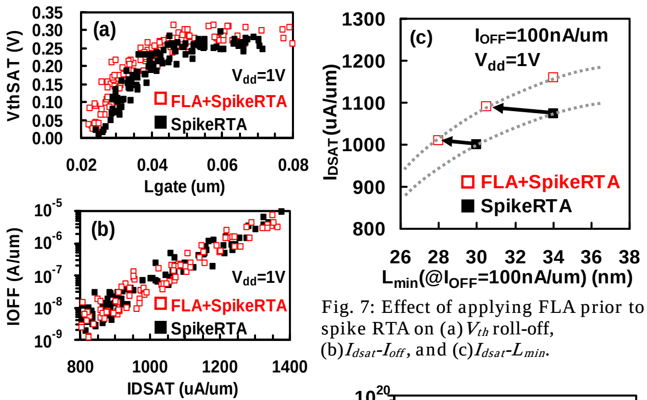

Fig. 7: Effect of applying FLA prior to spike RTA on (a) V_{th} roll-off, (b) I_{dsat}-I_{off}, and (c) I_{dsat}-L_{min}.

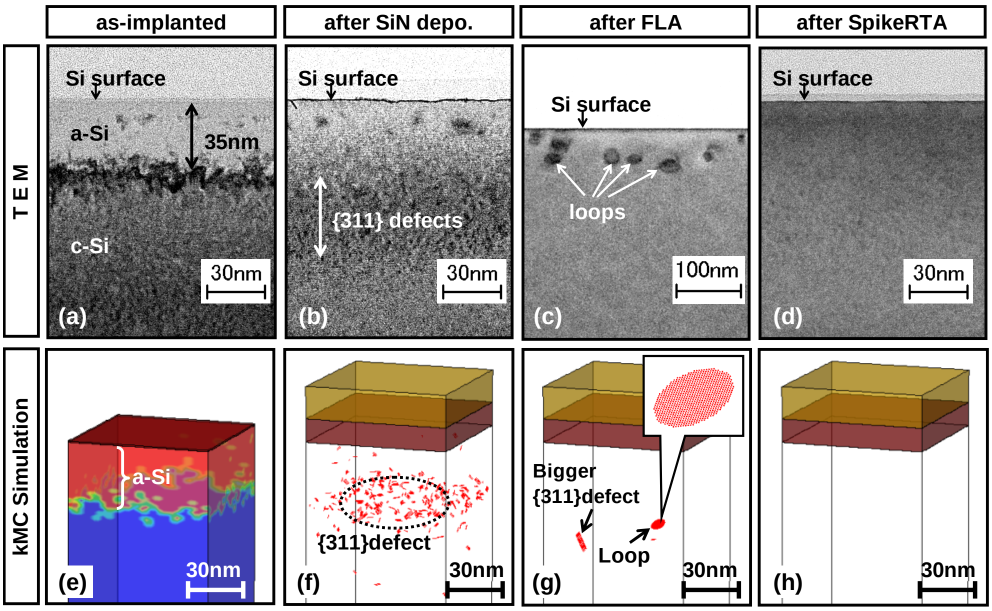

Fig. 8: XTEM of SIMS samples ((a)~(d)) and kMC simulation ((e)~(h)) result with Flow B (FLA+Spike RTA). The images after halo/ext I/I, after SiN deposition, after FLA, and after SpikeRTA are shown, respectively.

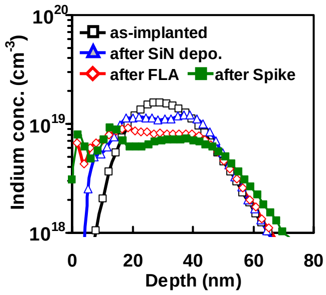

Fig. 9: Indium SIMS profiles of as-I/I, after SiN depo., after FLA and after Spike RTA.

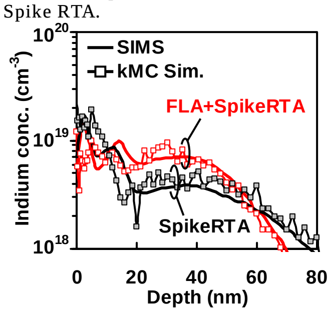

Fig. 10: Indium SIMS profile and kMC simulation for flow A (only spike RTA) and flow B (FLA+spike RTA).

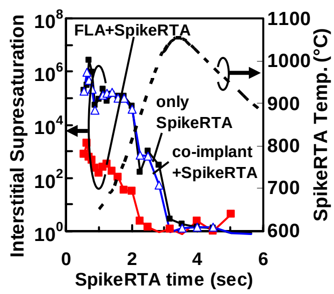

Fig. 11: Interstitial super-saturation during spike RTA for each process.

Flow	A	B
Process scheme	only SpikeRTA	FLA+ SpikeRTA
Defect type before SpikeRTA	{311} defect	Loop
Interstitial Supersaturation before SpikeRTA	1.0E+05	1.0E+03 (×0.01)
Indium peak conc. after SpikeRTA	4.0E+18	7.0E+18 (×1.75)
Indium surface conc. after SpikeRTA	1.7E+19	7.9E+18 (×0.46)

Table 1: Summary of defect evolution, interstitial supersaturation and indium concentration on Flow A (only Spike RTA) and Flow B (FLA+Spike RTA).

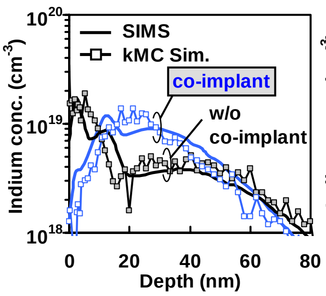

Fig. 12: Indium SIMS profiles and kMC simulation for cases with and without carbon co-implant.

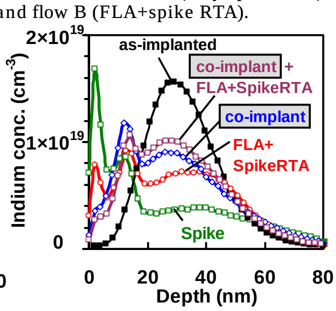

Fig. 13: SIMS results of combining carbon co-implant with FLA on indium halo.