14.2

Gate Postdoping to Decouple Implant/Anneal for Gate, Source/Drain, and Extension: Maximizing Polysilicon Gate Activation for 0.1 μm CMOS Technologies

H. Park, D. Schepis, A.C. Mocuta, M. Khare, Y. Li, B. Doris, S. Shukla, T. Hughes, O. Dokumaci, S. Narasimha, S. Fung, J. Snare, B.H. Lee, J. Li, P. Ronsheim, A. Domenicucci, P. Varekamp, A. Ajmera, J. Sleight, P. O'Neil, E. Maciejewski, and C. Lavoie*

IBM Microelectronics Semiconductor Research and Development Center, 2070 Route 52, Zip E40, Hopewell Junction, NY 12533
*IBM, T.J. Watson Research Center, Yorktown Heights, NY 10598

Abstract

We present a systematic study on maximizing polysilicon gate activation for aggressively scaled 0.1 μm CMOS technologies. The fundamental limit of gate activation due to poly depletion effect was investigated in terms of gate implant/anneal condition and sequence, poly grain size, dopant penetration and activation. For the first time, we achieved significant improvement in CMOS performance by developing a novel process of "gate postdoping" to decouple implant and anneals for gate, source/drain, and extension. The method successfully reduces the poly depletion effect and thus the equivalent gate oxide thickness in inversion by up to ~2 Å, improving CMOS on-currents by 9 ~ 33% over a conventional process.

Introduction

In today's CMOS technologies with polysilicon gate, improving gate dopant activation is crucial to our pursuit to obtain the thinnest effective oxide thickness. Recent studies emphasize gate depletion and dopant penetration as the main challenges in continued use of poly gate in deep sub-μm regime [1-3]. Poly-SiGe [3,4] and metals are explored as new gate materials. So far, however, the maximum activation of conventional poly gate has not been explored systematically. To maximize gate activation for a given gate stack, we should consider (i) total dose of dopants implanted in the poly; (ii) anneal conditions and sequence; and (iii) structural limits imposed by the scaling, such as dopant penetration through thinner oxide, salicidation, and spacer thickness. In the self-aligned gate process, simply increasing source/drain/gate implant dose degrades short channel performance due to lateral encroachment of the S/D dopants towards the channel during activation anneals. FIG. 1 shows that scaling the spacer thickness and increasing the single S/D/G implant dose cause off-current degradation in CMOS with L_{gate} of 0.1 μm, due to the lateral diffusion of dopants. It is necessary to selectively increase the gate doping and activation by decoupling it from the S/D doping. Predoping the gate by implanting the deposited poly before poly stack etch has been a method to decouple the implant dose. In this paper, we propose a new method of "Postdoping" the gate: Selectively implanting the gate by blocking the S/D regions with a planarized buffer material after gate poly stack definition (FIG. 2.) The gate postdoping not only decouples the dose between S/D and gate, but also allows us to decouple anneal steps for CMOS optimization, and maximizes the gate activation. FIG. 3 illustrates an example of the postdoping steps at different stages of a conventional integration flow. We discuss new results evaluating the gate postdoping and the predoping.

Gate Predoping

The gate stack consists of 1500 Å of LPCVD poly deposited on ~15 Å of gate oxide formed by N_2O oxidation on an SOI substrate. Gate predoping was tested on NFET by implanting the deposited poly with phosphorus at 3×10^{15} cm^{-2} followed by preanneal at 1000°C, S/D phosphorus implant, final RTA at 850 °C ~ 1000 °C, and Co silicidation. FIG. 4 (a) shows the low temperature deactivation effects on effective oxide thickness (Tinv) from gate inversion capacitance at 1.2 V. Although the higher doping concentration in the poly reduces Tinv by ~1.7 Å at 1000 °C, the Tinv increases significantly at lower temperatures. In FIG. 4 (b,c), TEM shows that poly grain size after the reactivation at 900 °C is much smaller than that after 1000 °C. For PFET, predoping was followed by S/D boron implant, RTA at 1000 °C, and additional low temperature RTA steps. FIG. 5 shows that Tinv decreases due to the predoping. However, boron deactivation in the poly during the RTA at 700 °C ~ 900 °C degraded PFET Tinv by ~0.5 Å. An additional RTA at 1000 °C, 1 sec reactivates the poly, but it causes boron penetration as shown in FIG.5 (b). To minimize boron penetration with aggressively scaled gate oxide, the stage where the gate doping occurs in the CMOS process should be later than poly deposition. Despite its apparent simplicity, gate predoping poses challenges in obtaining good poly profiles with n- and p-doped poly RIE, and in optimizing NFET and PFET with minimum dopant penetration and deactivation.

Gate Postdoping

We first investigated the critical dose of phosphorus required to maximize n-type gate activation. FIG. 6 shows Tinv as a function of P dose and postanneal temperature. It shows that with a total P dose of 1 ~ 1.5×10^{16} cm^{-2}, NFET gate activation can be maximized even at 950 °C. With doses above this range, however, silicide formation on poly becomes difficult, and gate resistance can be degraded. We found that $CoSi_2$ formation temperature is strongly affected by the increased doping level in silicon (FIG. 7.) Considering the limitations in dose, RTA temperature and time, we performed gate postdoping by blocking the S/D regions after reducing spacer thickness by 25%. FIG. 8 shows the distribution of Tinv vs. gate tunneling current measured on two groups of wafers; Compared with the case of no postdoping, Tinv at a constant tunneling current is reduced by ~2 Å and ~0.7 Å for NFET and PFET, respectively, by postdoping with phosphorus (1×10^{16} cm^{-2}) and boron (7.5×10^{15} cm^{-2}.) The gate postdoping successfully improved poly activation. As a result, FIGS. 9 and 10 show for example that I_{on} is improved by ~9% for NFET, and by ~33% for PFET, at a constant I_{off} of 100 nA/μm.

Unlike predoping, postdoping can introduce dopants into the poly at different stages. The efficiency of gate activation varies with the initial size and structure of poly grains which evolve during subsequent steps. To estimate the initial grain size effect, we preannealed a deposited poly at 1025 °C before the postdoping, and compared it with un-preannealed poly. FIG. 11 shows that the initial grain size increased from ~200 Å to ~500 Å due to the preanneal. In FIG. 12 (a), SIMS profiles show that the preanneal reduces the final concentration of phosphorus in the poly. The different poly grain boundaries seem to cause dose loss at the poly interfaces. FIG. 12 (b) shows that the smaller size of initial poly grains before postdoping leads to reduction of NFET Tinv by ~1 Å. It indicates that grain size control for short/narrow channel devices is a critical factor in maximizing sub-0.1μm CMOS performance.

Conclusion

For maximizing gate activation in sub-0.1 μm CMOS devices, we introduced gate postdoping as the most efficient method to activate the poly gate by decoupling both implant and anneal for gate and S/D. We achieved tremendous improvement in device performance by applying the new method to 0.1 μm node CMOS technologies.

Acknowledgments

The authors would like to thank Y. Taur, S. Crowder, R. Purtell, P. Gilbert, K.P. Muller, I. Yang, E. Leobandung, J. Welser, P. Agnello, and G. Shahidi for the valuable discussions and support.

References

[1] E. Josse and T. Skotnicki, *IEDM Tech. Dig.*, p. 661, (1999).
[2] J. Schmitz *et al.*, *Proc. ESSDERC*, p. 156, (1999).
[3] B. Yu *et al.*, *IEEE Trans. Elect. Dev.* Vol. 45, no.6, p.1253, (1998).
[4] K. Uejima *et al.*, *IEDM Tech. Dig.*, p. 445, (2000).

 0-7803-7312-X/02/$17.00 ©2002 IEEE

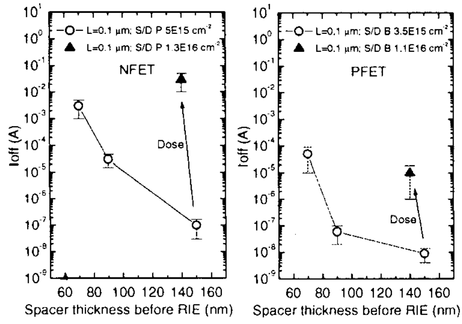

FIG. 1. Off-state currents in NFET and PFET with 0.1 μm gate length as a function of spacer thickness (as-deposited) and S/D implant dose.

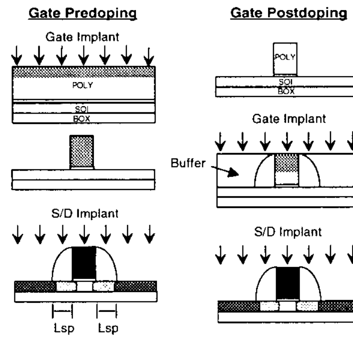

FIG. 2. Schematic diagram of gate postdoping compared with gate predoping.

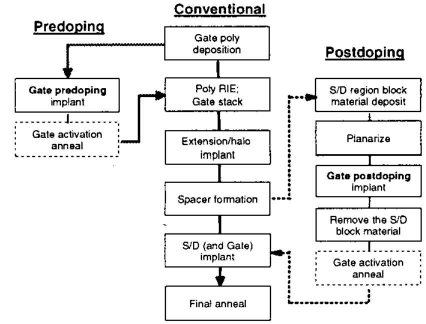

FIG. 3. Various sequences of gate postdoping, compared with predoping.

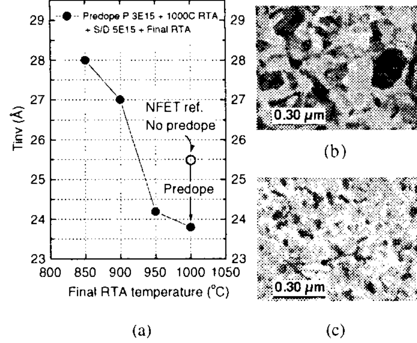

FIG. 4. (a) NFET Tinv (@1.2V) as a function of anneal temperature after gate predoping; (b) Plan-view TEM of final poly grains after RTA at 1000 °C and $CoSi_2$; (c) After RTA at 900 °C and $CoSi_2$.

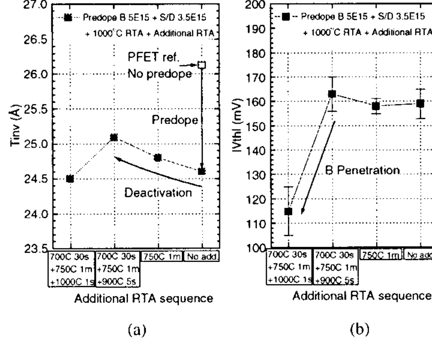

FIG. 5. (a) Tinv of PFET predoped with boron, annealed at 1000 °C, followed by low thermal anneals shown on x-axis; Long-channel Vth of the predoped PFET in the linear regime.

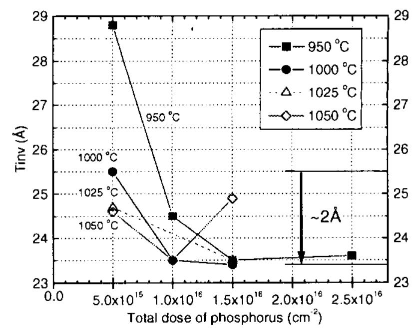

FIG. 6. NFET Tinv vs. total dose of phosphorus implanted in the gate poly, annealed at different temperatures. The gate postdoping maximized gate activation leading to reduction of Tinv by ~2 Å.

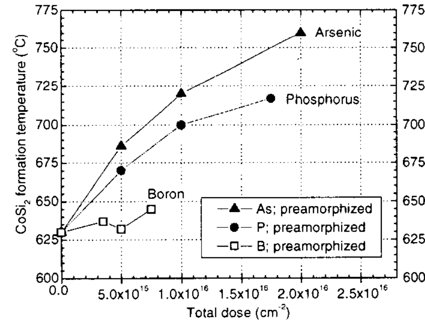

FIG. 7. Effects of doping level in silicon on cobalt disilicide ($CoSi_2$) formation temperature. The phase formation was measured by *in situ* x-ray diffraction.

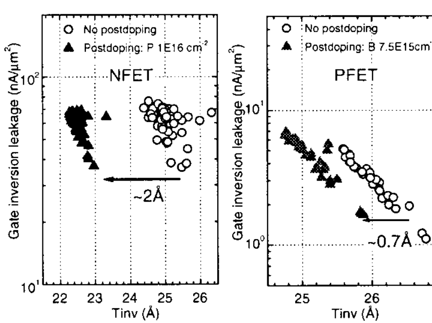

FIG. 8. Tinv vs. gate tunneling current for NFET and PFET devices with and without gate postdoping.

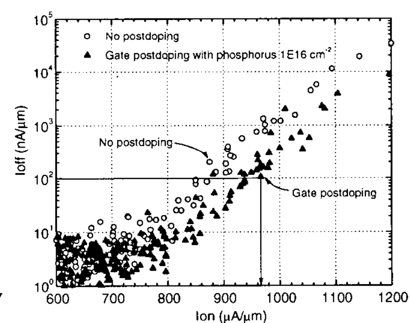

FIG. 9. NFET Ion vs. Ioff improved by the gate postdoping with phosphorus S/D (@V_{dd}=1.2V.)

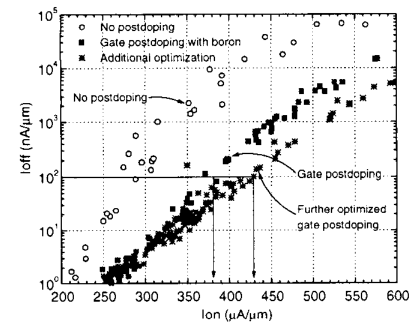

FIG. 10. PFET Ion vs. Ioff improved by gate postdoping, and then further improved by an additional doping/thermal optimization (@V_{dd} = 1.0V.)

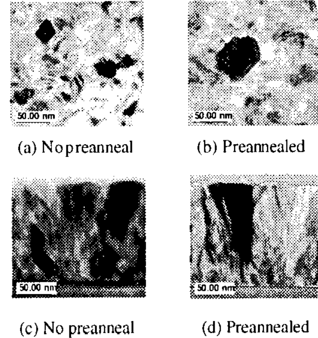

FIG. 11. (a, b) PTEM of poly with and without preanneal at 1025 °C, 15 sec after the deposit; (c, d) XTEM of the above samples.

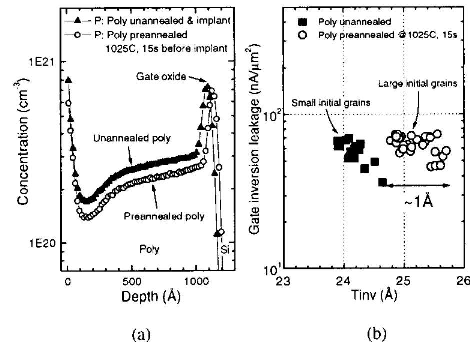

FIG. 12. (a) SIMS profiles of phosphorus distribution in gate poly with different grain sizes; (b) The initial grain size effect on the NFET gate activation (Tinv.)