Advanced Junction Engineering for 60nm-CMOS Transistors

M. Müller, M. Bidaud
Philips Semiconductors, 850 Rue Jean Monnet, 38926 Crolles Cedex

F. Bœuf, A. Halimaoui, M. Lamy, D. Lenoble, R. Palla, T. Skotnicki
STMicroelectronics, 850 Rue Jean Monnet, 38926 Crolles Cedex

C. Laviron
CEA/LETI, 17 rue des Martyrs, 38054 Grenoble Cedex 9

markus.muller@philips.com

Abstract

In this paper, we evaluate new concepts in the ultra shallow junction engineering such as Ultra Low Energy (ULE) and Plasma Doping Implant (PLAD) and fast ramp-up Spike Annealing after integration into planar 60nm-transistors. Excellent results in terms of SCE and DIBL reduction are obtained for As implant at 1 keV and B PLAD implant for nMOS and pMOS devices respectively. Further improvement can be obtained by using Levitor Spike Annealing. Future work has to be focussed on the optimisation of the transistor performances.

1. Introduction

It has been pointed out that in the deep sub-micron regime, the threshold voltage of planar MOS transistors will be very much determined by the Short Channel Effect (SCE) and the Drain Induced Barrier Lowering (DIBL). As a consequence, the simple increase of the channel doping is no more suitable to readjust the threshold voltage while reducing the gate length. As both SCE and DIBL scale with the ratio extension depth over electric gate length x_j/L_{elec}, the extension depth reduction has therefore been defined to be a major challenge for future CMOS generations [1, 2]. For this purpose, low energy implant techniques are required giving a well defined as-implanted profile as well as an efficient annealing process enabling a high activation rate while suppressing the ion diffusion. Unfortunately, using conventional implant and annealing methods, the extension depth reduction is often paid by a poor activation and thus an increase in the sheet resistance impairing the transistor performances.

Meanwhile, several new ways to overcome this dilemma have been proposed or demonstrated. Among them, Ultra Low Energy (ULE) and Plasma Doping (PLAD) implant have proved to be very promising implant techniques enabling x_j-values below the 30 nm threshold while yielding reasonable sheet resistance values [3-5].

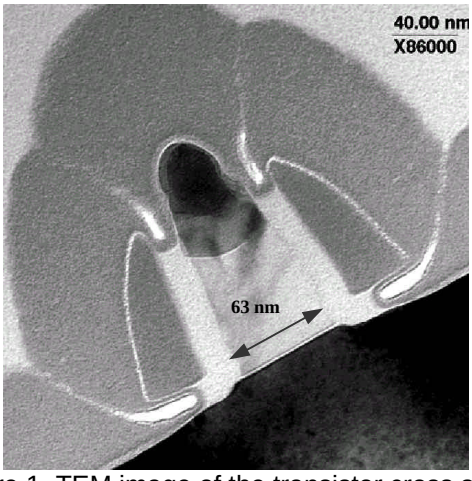

Figure 1. TEM image of the transistor cross section

Especially for light atoms like boron, conventional ion beam implantation reaches its limit in the low energy range due to channelling and beam uniformity problems [6]. This second issue is improved by the ULE technique where a high energetic ion beam is decelerated before

reaching the wafer. The PLAD technique goes a very different way: a plasma cloud is generated near the surface of the wafer. The ionised atoms are then extracted via a pulsed voltage applied on the substrate and are thus accelerated towards the wafer surface.

In order to limit the ion diffusion during the dopant activation process and to maintain thus a well defined junction profile high ramp-up and ramp-down rates are mandatory. Using the ASM-Levitor Spike Annealing system, which is based on a conductive heat transfer by a He gas flow instead of IR radiation for RTA, ramp-up rates up to 800°C/s can be achieved compared to 150°C/s for conventional RTA.

In this paper, we will study the impact of these implant and annealing techniques on several aspects of the CMOS transistor characteristics such as SCE, DIBL and lateral diffusion.

2. Integration into 60nm-transistors

The process flow used for our devices was the following: for both pMOS and nMOS devices a 15.5 Å-thick RTO-nitrided oxide has been used on top of which a 1500 Å-thick poly-Si layer has been deposited. Pre-doping and pre-annealing have been carried out for nMOS devices: after a P 4e15 cm^{-2} implant a thermal treatment of 20 min. at 850°C has been used. For the extensions, two splits have been defined: on the pMOS side, we compare ULE (B^+ 5e14 cm^{-2} at 250 eV) implant with PLAD (B^+ 8e14 cm^{-2} at 200 V) implant, whereas for the nMOS devices, As 1e15 cm^{-2} at 1 keV and As 5e14 cm^{-2} at 5 keV are investigated. The pockets are formed by B 2e13 cm^{-2} (nMOS) at 15 keV and P 3e13 cm^{-2} at 30 keV with an additional halo of 2e13 cm^{-2} at 30 keV (pMOS). After a standard spacer formation and S/D junction implant the dopant activation is carried out using either Levitor Spike Annealing or standard RTA (1000°C, 15s). A TEM image showing the cross section of a typical 60nm-transistor is presented in figure 1.

3. Impact on the Roll-off behaviour

Figures 2 and 3 show the evolution of the threshold voltage at saturation V_{th} (V_{drain} = 1 V) as a function of the gate length for the different nMOS and pMOS splits respectively. On the nMOS side we can see a considerable improvement of the roll-off behaviour when using As 1e15 cm^{-2} 1 keV implant instead of As 5e14 cm^{-2} 5 keV. Especially for the low energy implant, Levitor Spike Annealing yields further improvement: the threshold voltage drop appears for smaller gate lengths indicating an increased vertical and lateral diffusion control with respect to standard RTA.

On the pMOS side, we can detect a very good threshold value behaviour in the case of PLAD implant. Very interesting is the influence of Levitor Spike Annealing on the ULE implant technique: a spectacular gain in the roll-off behaviour can be obtained when Levitor Spike Annealing is used yielding even better results than for PLAD. On the other side, the threshold shift for long devices indicates an increase in the poly-depletion. Pre-doping and pre-annealing for pMOS-devices seems therefore necessary.

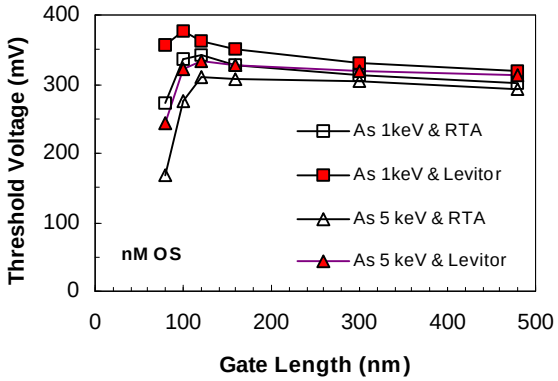

Figure 2. Threshold voltage V_{th} versus gate length for nMOS at V_{drain} = 1 V

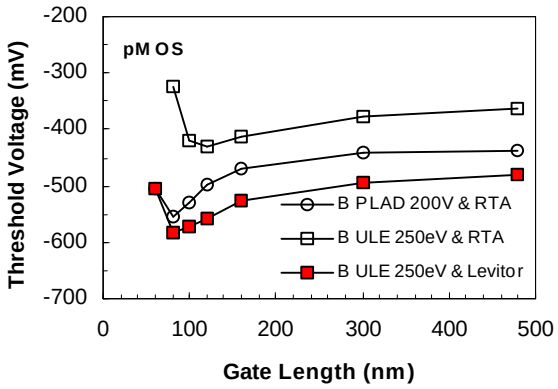

Figure 3. Threshold voltage V_{th} versus gate length for pMOS at V_{drain} = -1 V

For more clarity, we plotted $\Delta SCE/\Delta L_{gate}$ in the interval between 60 and 80 nm as obtained from linear threshold value measurements (cf. figure 4). Figure 5 presents the corresponding DIBL values at 80 nm. Using the analytical approach proposed in reference [1] to calculate the threshold value, we can extract the extension depth x_j as a fitting parameter from the experimental $V_{th}(L)$-curves. The thus obtained x_j-values are added in figure 5 together with the corresponding simulated DIBL-values.

The tendencies are very clear: for nMOS devices, the most promising approach is the combination of As 1 keV implant with Levitor Spike Annealing where we find a x_j value of 33 nm. The positive impact of Levitor Spike

Annealing is less pronounced for the As 5 keV implant. This observation stresses the need for the development of ultra-shallow implant methods when targeting the full exploitation of the potential of Spike Annealing methods. On the pMOS side, both PLAD implant with standard RTA and ULE implant with Levitor Spike Annealing yield shallow junction depth values down to 28 nm.

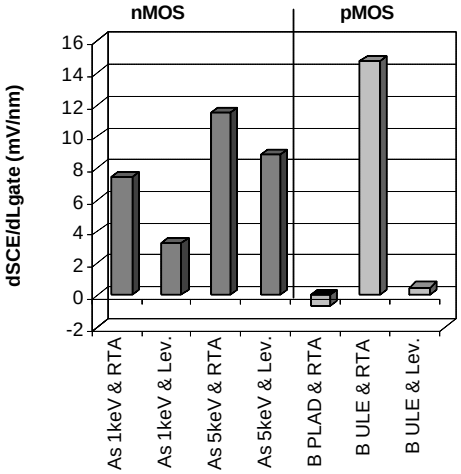

Figure 4. $\Delta SCE/\Delta L_{gate}$ for pMOS and nMOS devices between 60 and 80 nm

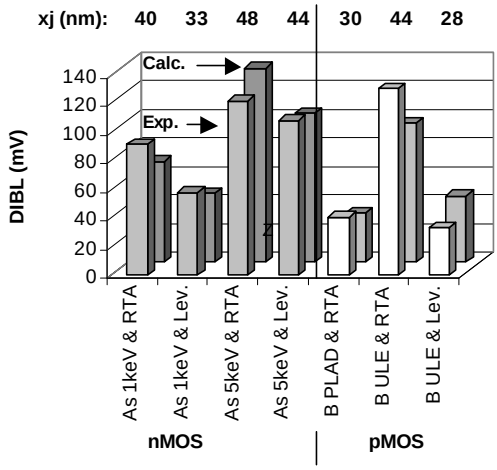

Figure 5. DIBL at 80 nm and V_{drain} = 1V together with the estimated junction depth values

4. Junction Analysis

The analysis of the $V_{th}(L)$-curves indicates a good vertical dopant diffusion control for B ULE and As 1 keV in combination with Levitor Spike Annealing and for B PLAD with RTA process. We like to emphasize that the obtained extension depth values in our transistors are coherent with recently obtained R_s/x_j-results on full sheets for ULE and PLAD implant (cf. Figure 6).

In the following, we want to discuss the complementary parameter, i.e. the lateral diffusion ΔL. This parameter has been extracted from $I_d(V_g, L)$ measurements for nMOS devices. Unfortunately, the threshold values on pMOS devices were not sufficiently adjusted to give reliable values.

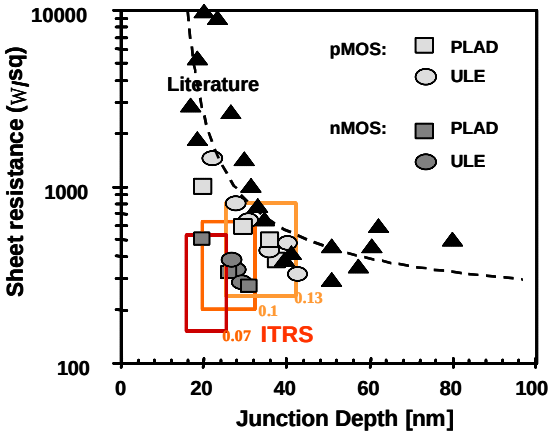

Figure 6. Recently obtained R_s/x_j data points for PLAD and ULE implant in comparison with literature data

Figure 7 demonstrates a reduction of the lateral diffusion of more than 10 nm down to even negative values when reducing the implant energy from 5 to 1 keV despite the higher implant dose. We should note that these results are in contradiction with the observations made in reference [7], where the authors report an enhanced diffusion in the case of As 1 keV implant.

Furthermore, especially for As 1 keV, Levitor Spike Annealing proves to reduce the lateral diffusion efficiently. These findings are very coherent with the previous observations of the DIBL and SCE reductions.

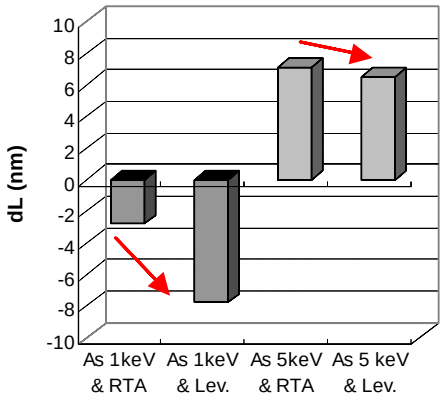

Figure 7. Lateral diffusion ΔL as a function of the As implant and the annealing technique

5. Impact on the performance

This first demonstration lot was not optimised for such efficient suppression of junction and poly-gate diffusion as that resulting from the Levitor Spike Annealing. As a consequence, the absolute performance values of our transistors are rather mediocre. Nevertheless, our data allows us to give the relative trends.

An important issue in the process flow for sub-micron CMOS transistors is the simultaneous control of the junction depth and of the gate poly-depletion during the annealing processes. The problem of insufficient dopant activation and diffusion into the poly-silicon can be overcome by pre-doping and pre-annealing. In our process flow, these steps have been integrated for nMOS devices.

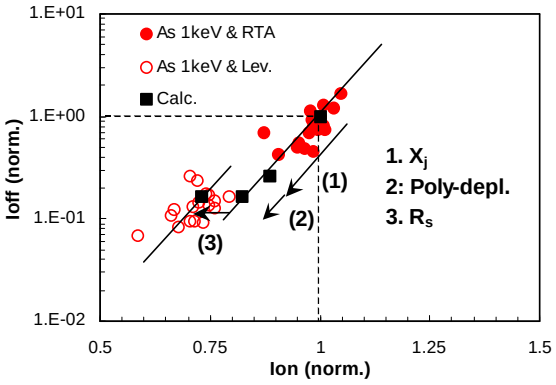

Figure 8. Impact of Levitor Spike Annealing on the performance of 80 nm nMOS devices

To evaluate the impact of the Levitor Spike Annealing technique, we compare the performances for both annealing processes for 80 nm-nMOS transistors (cf. figure 8). For better understanding, the experimental data points have been reproduced using the equations given in reference [1]. According to these simulation results, the performance shift can be mainly attributed to the threshold shift caused by the extension depth reduction. We can see that the increase in the poly-depletion is largely suppressed in our process flow. The changes in the series resistance remain acceptable as well. A careful adjustment of both the pre-doping and pre-annealing conditions and the spike annealing temperature is expected to give further improvement on these issues and will be done in future experiments.

7. Conclusion

In this paper we demonstrated that extension depths below 30 nm can be realized in planar transistors with 60 nm gate lengths by combining As 1 keV implant with Levitor Spike Annealing for nMOS and by using B PLAD 200 V or B ULE at 250 V with Levitor Spike Annealing for pMOS devices. Our results show that these junction engineering techniques are powerful tools for an efficient reduction of the SCE, DIBL and the lateral diffusion under the gate, which is mandatory for the manufacturing of deep sub-micron CMOS devices. Future work will be focussed on the adjustment of the pre-doping and pre-annealing conditions as well as on the Spike Annealing temperature for performance improvement.

8. Acknowledgments

A part of the results has been generated in the framework of the IST HUNT and ARTEMIS projects.

8. References

[1] T. Skotnicki, Proceedings ESSDERC 2000, pp. 19-33.

[2] Q. Xu, et al., IEEE Vol. 48, 2001, pp. 1412-1420

[3] D. Lenoble, Proceedings ESSDERC 2000, pp. 392-395.

[4] D. Lenoble, Proceedings ESSDERC 2000, pp. 400-403.

[5] V.M.H. Meyssen et al., Proceedings ESSDERC 2000, PP. 396-399.

[6] Handbook of Ion Implantation Technology, J. F. Ziegler, 1992, Elsiever Science Publishers B. V., P. 119

[7] R. Kasnavi et al., Proceedings VLSI 2000, pp. 112 -113