Extended Abstracts of the 2004 International Conference on Solid State Devices and Materials, Tokyo, 2004, pp. 180-181

B-3-1 (Invited)

Integrating Diffusionless Anneals Into Advanced CMOS Technologies

R. Surdeanu, R. Lindsay[1,*], S. Severi[1], A. Satta[1], B.J. Pawlak, A. Lauwers[1], C.J.J. Dachs, K. Henson[1], S. McCoy[2], J.C. Gelpey[2], X. Pages[3]

Philips Research Leuven
Kapeldreef 75, B-3001, Leuven, Belgium
Phone: +32 1628 8319 E-mail: radu.surdeanu@philips.com
[1]IMEC, Kapeldreef 75, B-3001, Leuven, Belgium
[2]Vortek Industries, Vancouver, Canada
[3]ASM International, Bilthoven, The Netherlands

1. Introduction

Over the past years, diffusionless activation has been the subject of extensive studies within the topic of ultra-shallow junctions (USJ) for advanced CMOS devices [1-7]. Main drive for this research is the necessity for high, metastable dopant activation (above the solid solubility in crystalline silicon) in the USJ region. Translated into USJ properties, for the 45 nm CMOS technology node, the ITRS roadmap [8] requires the junction depth, X_j, to be between 7 nm and 12 nm, and the sheet resistance, R_{sheet}, below 830 Ω /sq., for pMOS USJ, and below 390 Ω/sq. for NMOS. Two approaches are able to fulfill these stringent requirements: (*i*) solid phase epitaxy re-growth (SPER) at low temperature of a doped amorphous region [1-3,7] and (*ii*) high temperature anneals, for extremely short time, using arc lamps (such as flashRTP) or sub-melt laser [4-7]. While very attractive owing to the outstanding USJ properties they provide, both approaches have their specific integration challenges into a CMOS flow, such as residual defects, junction leakage, lateral straggle, gate electrode and gate dielectric integrity and channel doping de-activation.

In this paper we will present NMOS and PMOS device results, addressing these issues for SPER and flash annealing.

2. Experimental

Implantation and anneal

For all experiments 200mm [100] p-type and n-type Si wafers were used. Implants were done on the Applied XR80 Leap low energy implanter. Anneals were done on the ASM Levitor™ for the spike and SPER, and in the Vortek flash RTP, fRTP™, system for flash annealing [4].

All devices were unichannel PMOS or NMOS, containing STI, 100 nm pre-doped polysilicon as a gate material and heavily nitrided gate oxides, with an Effective Oxide Thickness (EOT) of 1.7 nm. Pockets were typically annealed by spike anneal however tests were made without a specific anneal after pocket. To maintain the junctions' characteristics the process flow was designed such that, apart from the USJ activation step, all temperature steps after gate patterning are kept below 500ºC (i.e. minimize dopant diffusion and de-activation). NiSi was used for the silicidation step, in order to benefit from the lower thermal budget (RTA2 < 450ºC) and lower contact resistance compared to $CoSi_2$.

3. Results

Integration of SPER USJ

In the SPER process, dopants in a pre-amorphized (e.g. with Ge implantation) silicon region are activated during the crystalline re-growth at low temperatures (550ºC ≤ T ≤ 700ºC). The advantages of this method are mainly two-fold: the dopants in the amorphous silicon are activated

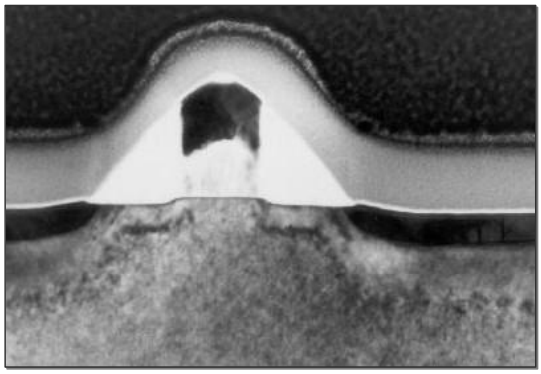

Fig. 1. XTEM of a 50nm transistor with SPER extensions

to levels well above the equilibrium solubility in Si; and, owing to lower processing temperature, insignificant dopant diffusion [1-3,7]. By tuning the implant conditions, the USJ can be confined to the pre-amorphized region and have a box-like highly-active carrier profile, shallower than the as-implanted profile. Furthermore, a low temperature SPER process is beneficial for advanced gate stacks where high-k and metal gate thermal stability is of great concern.

Figure 1 shows a XTEM picture of a 50 nm transistor with both SPER extensions and deep source/drain junctions. The residual End-Of-Range damage is clearly visible, edging the junction region. These defects can increase the junction leakage by trap-assisted tunneling due to the introduction of trap levels in the forbidden gap in the depletion region, but, on the other hand, they may deactivate the background doping and therefore reducing the band-to-band tunneling, BBT, thereby improving the overall junction leakage. The graph in Fig. 2a indeed shows the interplay of these effects for a SPER PMOS device.

[*] now with Infineon Technologies

Additionally, it is clear from Fig. 1 that there is only little overlap of the junction under the gate due to the negligible lateral straggle of the amorphization implant. This can lead to higher spreading resistance, and as a consequence reduce the on-state current of the device, as illustrated in Fig.2b, where both NMOS and PMOS SPER devices cannot match electrically the I_{on} - I_{off} performance of the spike references. However, the lack of overlap is not the only reason for the reduced on-current. This is most likely a result of several other effects as well: gate depletion due to polysilicon gate damage created by the pre-amorphization implant, access and contact resistance. The introduction of metal gates could give the necessary improvement by solving the gate depletion issue and offering more flexibility in tuning the SPER USJ.

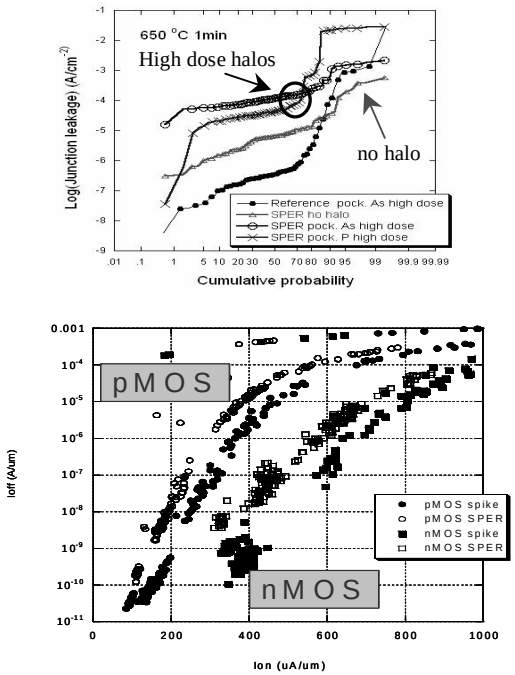

Fig. 2. (a) Area junction leakage for PMOS devices and (b) Ioff vs. Ion curve for NMOS and PMOS SPER devices (open symbols) compared to reference conventional spike devices (closed symbols)

Integration of flash RTA USJ

In the flash annealing process, the temperature profile vs. time of the spike anneal is remarkably sharpened by using Ar or Xe ARC lamps with a discharge time of < 1ms. In such a way it is possible to heat up the surface of a wafer to around 1300 ºC within the discharge time and the heat subsequently flows into the bulk by conduction. Cooling down by conduction allows ramp downs of ~10^6 ºC/s, resulting in a high activation level characteristic of the peak temperature and insignificant dopant diffusion. In flash annealing the wafer is normally brought to an intermediate temperature ranging between 500-800 ºC either as a soak (>10s) or spike anneal before the flash [4-7].

Also for this diffusionless method, similar topics of concern as for SPER process are encountered. Additionally, pattern dependency effects on junction activation may be a potential issue for flash annealing. However, in Ref. [7], we showed that these effects are minimal when compared to conventional spike anneals. Figure 3a shows that the area junction leakage for fRTA transistors is at least 2 orders of magnitude higher compared to the spike reference. An intermediate temperature of 800 ºC improves the nMOS leakage by around 1 order of magnitude as compared to the 700 ºC case most likely due to a less abrupt junction and fewer residual defects. By further reducing the pocket dose the BBT is reduced and therefore the junction leakage can be brought close to the spike reference level. By choosing appropriate conditions, for NMOS the transistor performance was improved up to 15% compared to the conventional spike devices, as shown in Fig. 3b.

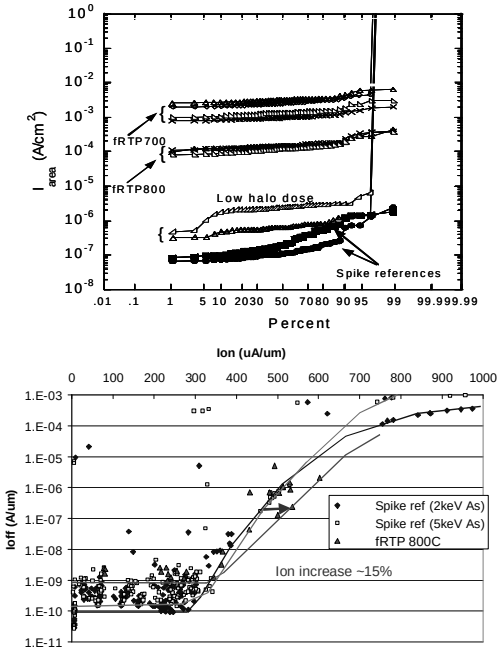

Fig. 3. (a) Area junction leakage and (b) Ioff vs. Ion curve for NMOS fRTP devices compared to reference conventional spike. The intermediate temperature for fRTP is indicated in the figures.

3. Conclusions

We have investigated several integration aspects for low temperature (SPER) and high temperature (flash RTP) diffusionless anneals. No fundamental integration showstoppers are identified for the two approaches. Flash annealing seems easier to integrate, however further improvement is needed. For SPER, metal gates might be the solution to a large part of CMOS integration concern issues.

References

[1] K. Tsuji et. al, *IEDM Tech. Dig.* 1999
[2] B.J. Pawlak et al. *IIT 2002, ECS* 2003
[3] R. Lindsay et al. *Mat. Res. Soc. Symp. Proc.* 2003
[4] J.C. Gelpey et al., *Proceedings ECS* 2003
[5] A. Mokhberi et al., *IEDM Tech. Dig.* 2003
[6] A. Satta et al., *Mat. Res. Soc. Symp. Proc.* 2004
[7] R. Lindsay et. al, *Proceedings ECS* 2004
[8] ITRS2003: http://public.itrs.net/