Spike anneal optimization for digital and analog high performance 0.13µm CMOS platform

E. Josse, F. Arnaud, F. Wacquant, D. Lenoble, O. Menut, E. Robilliart

STMicroelectronics, Central R&D, Crolles, France
Outline

- Motivation for spike anneal
- Impact on 0.13 µm pMOSFET
- Impact on 0.13 µm nMOSFET
- CMOS optimization for digital and analog applications
- Conclusions
Motivation for spike anneal: thermal ramps profile

- **High peak temperature**: maximize dopant activation
- **Aggressive ramp-up**: minimize dopant diffusion
- **Aggressive ramp-down**: minimize dopant deactivation
Motivation for spike anneal: simulated 2D MOSFET profiles

Standard RTA: 1025°C 15s

Spike RTA at 1113°C

Simulated: 0.13 µm pMOSFET with either a standard RTA (1025°C 15s) or a spike RTA at 1113°C

→ motivation for spike RTA: reducing Boron diffusion

→ challenge: optimizing the spike RTA temperature and ramps to obtain both reduced Boron diffusion and high level of Boron and Arsenic activation
Motivation for spike anneal: experimental

**Context**: 0.13 µm CMOS flow (Lg=0.11 µm and tox=20Å)

- nMOS junction: As at 2 keV + double pocket (Indium + Bore)
- pMOS junction: BF2 at 4 keV + double pocket (Arsenic + Phosphorus)

→ Demo done with an industrial Lamp-based RTP (AMAT CENTURA XE+)

→ Various spike anneal T° and ramps conditions are compared with a standard RTA 1025°C 15s:

<table>
<thead>
<tr>
<th>Ramp: up / down [°C/s]</th>
</tr>
</thead>
<tbody>
<tr>
<td>75 / 75</td>
</tr>
<tr>
<td>1075°C</td>
</tr>
<tr>
<td>1113°C</td>
</tr>
<tr>
<td>1150°C</td>
</tr>
</tbody>
</table>
Outline

- Motivation for spike anneal
- Impact on 0.13 µm pMOSFET
- Impact on 0.13 µm nMOSFET
- CMOS optimization for digital and analog applications
- Conclusions
Impact on 0.13 µm pMOSFET : Vt-L plots

1st order player is temperature :
Aggressive ramps are not required

Compared to reference RTA :
T=1150°C : SCE not corrected
T=1075°C : SCE inverted = RSCE
T=1113°C : flat Vt-L profile

→ Major impact of the final RTA temperature on Vt-L plots at pMOS side :
Boron diffusion is driven by the spike temperature
Impact on 0.13 µm pMOSFET: Boron activation

No impact from ramp-up and down (not presented here)

Compared to reference RTA:
- T=1150°C: higher B activation
- T=1075°C: lower B activation
- T=1113°C: same B activation

→ Major impact of final RTA temperature step on Boron activation:
  T=1113°C is the best trade-off between Boron diffusion and activation
Impact on 0.13 \( \mu \text{m} \) pMOSFET: Boron penetration

C-V plot: zoom on gate depletion

- Spike RTA 1113°C compared to reference RTA 1025°C 15s:
  - Both gate depletion and Boron penetration reduced: improved gate stack
Outline

- Motivation for spike anneal
- Impact on 0.13 µm pMOSFET
- Impact on 0.13 µm nMOSFET
- CMOS optimization for digital and analog applications
- Conclusions
Impact on 0.13 µm nMOSFET: Vt-L plots

- No significant impact of the final RTA on Vt-L plots at nMOS side: Arsenic diffusion is not driven by the spike temperature.
Impact on 0.13 um nMOSFET: Arsenic activation

No impact from ramp-up and down (not presented here)

Compared to reference RTA:
- T=1150°C: higher As activation
- T=1075°C: lower As activation
- T=1113°C: same As activation

→ Major impact of final RTA temperature step on Arsenic activation:
  T=1150°C is the best trade-off between Arsenic diffusion (none) and activation
Major impact of final RTA temperature step on performance:

- $T=1150^\circ C$: reduced gate depletion and enhanced performance
- $T=1113^\circ C$: gate depletion and performance are the same as with ref. RTA

Within wafer uniformity: full sheet Rs mapping

RTA 1025°C 15s

Spike RTA at 1113°C

1 color = 1°C variation

→ Activation uniformity is not significantly degraded when compared to the standard RTA 1025°C 15s, except on the extreme edge of the wafer.
Outline

- Motivation for spike anneal
- Impact on 0.13 µm pMOSFET
- Impact on 0.13 µm nMOSFET
- CMOS optimization for digital and analog applications
- Conclusions
### CMOS optimization: spike RTA selection

**nMOS** | **pMOS**  
---|---  
**Temp. [°C]** | **activation** | **diffusion** | **activation** | **diffusion**  
**1075°C** | - | = | - | ++  
**1113°C** | = | = | = | +  
**1150°C** | + | = | + | =  

*compared to reference RTA 1025°C 15s: + better / = same / - worse*

→ **1113°C**: CMOS optimum for SCE control enhancement  
→ **1150°C**: CMOS optimum for performance enhancement  
→ Aggressive ramps not needed (1st order is temperature): 75°C/s is enough
CMOS optimization: device optimization (1/2)

Device optimization: for both digital and analog applications

Reference RTA 1025°C 15s

- **Digital (short channel):**
  - strong pocket dose needed

- **Analog:**
  - poor voltage gain (long channel)
  - poor matching (short channel)

Spike RTA at 1113°C

- **Digital (short channel):**
  - low pocket dose needed

- **Analog:**
  - improved voltage gain (long channel)
  - improved matching (short channel)
Spike RTA at 1113°C allows As pocket dose lightening (from 4.2e13 to 1.9e13)
→ Flat pMOS Vt-L is obtained thanks to Boron diffusion freeze-out
→ Voltage gain for long channel is significantly improved
→ No impact on pMOS performance nor on nMOS performance and SCE
Outline

Motivation for spike anneal

Impact on 0.13 µm pMOSFET

Impact on 0.13 µm nMOSFET

CMOS optimization for digital and analog applications

Conclusions
Conclusions

→ It is demonstrated that using a spike RTA instead of a classical few-seconds RTA improves the trade-off between Boron activation and diffusion

→ Under our process condition, the prevalent player for this trade-off is the spike temperature. « Slow » ramps (75°C/s) are then enough to take the full advantage of the spike RTA concept

→ Based on the spike RTA integration within a 0.13 µm CMOS flow, we demonstrated that the MOSFET architecture can be optimized for both digital and analog applications: this is made possible because the Boron diffusion freeze-out is traded for a lighter pocket dose