Millios™ - millisecond flash anneal for 22nm node

Sing-Pin Tay
Mattson Technology, Fremont, CA, USA

Steve McCoy
Mattson Technology Canada, Vancouver, BC, Canada

mattson
A Partner You Trust. Performance You Value.

Semicon West TechSITE Junction Technology Group Symposium, July 15, 2010

Acknowledgements

- K.L. Lee et al., *IBM Research Division, T.J. Watson Research Center, Yorktown Heights, NY.*
- J. Chan, G. Stuart & D. Camm, *Mattson Technology Canada, Vancouver, BC.*
- B. Peuse & P. Timans, *Mattson Technology, Fremont, CA.*

Outline

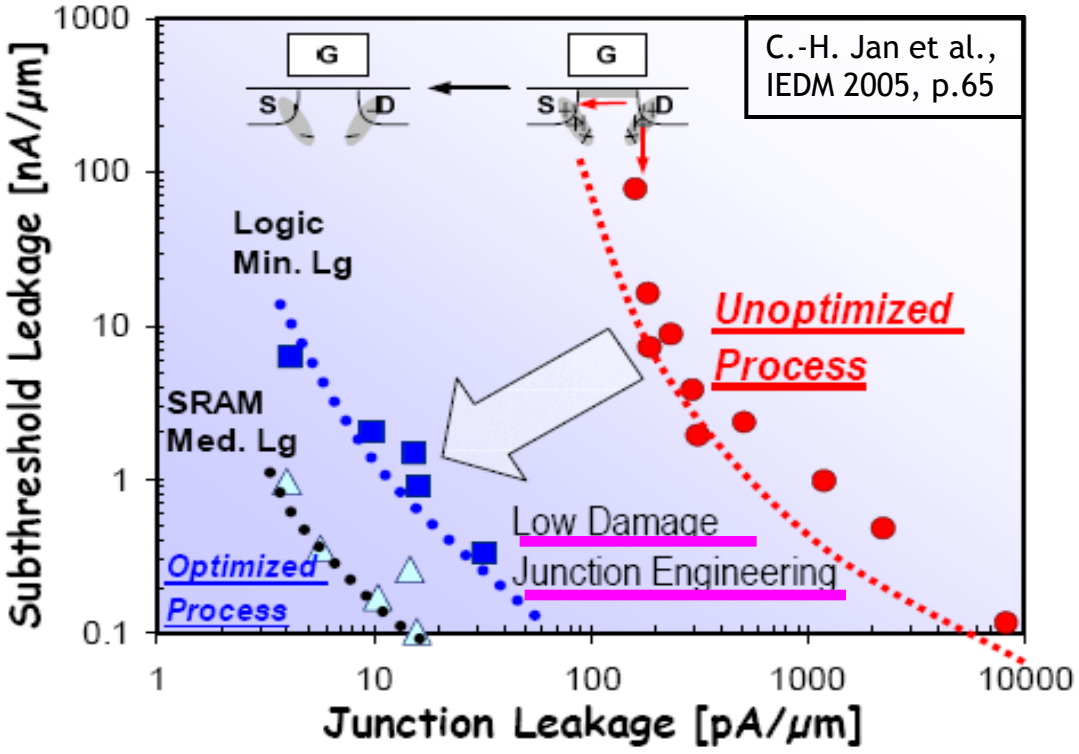
- Introduction
  - Advanced USJ requirements
  - The emergence of millisecond anneal
- MSA for Activation with minimal Diffusion
  - Adequate defect anneal must be maintained
  - Abrupt USJ formation with long ms-flash
- Experimental Results and Discussions
- Conclusions

Advanced USJ Requirements

- Minimize Dopant Diffusion
- Maximize electrical activation
- Maintain adequate defect annealing
  - Junction leakage is a significant issue for low power CMOS
  - Heavy doping in channel/halo contributes to increased junction leakage
- Need to optimize all 3 “dimensions”
  - MSA is emerging as a process solution

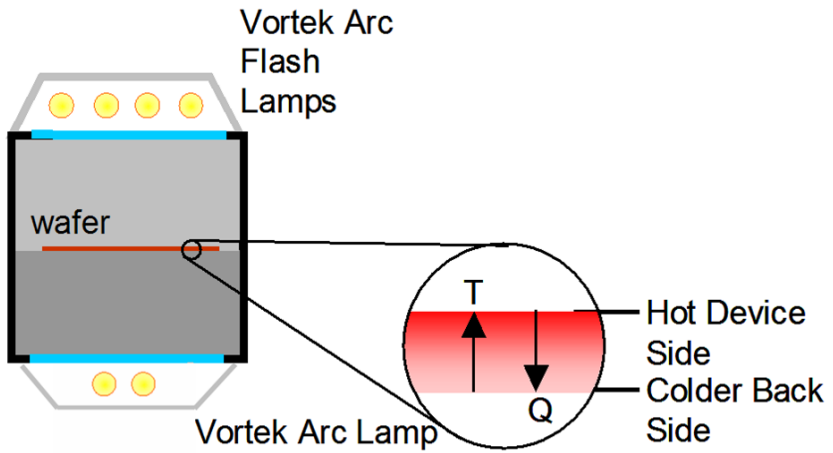
Year of Production	2009	2010	2011	2012	2013	2014	2015
MPU Printed Gate Length (nm)	47	41	35	31	28	25	22
MPU Physical Gate Length (nm)	29	27	24	22	20	18	17
Drain extension X_j (nm) for bulk MPU/ASIC [A]	13	12	10.5	9.5	8.7	8	7.3
Maximum allowable parasitic series resistance for bulk NMOS MPU/ASIC × width (Ω–μm) from PIDS [B]	170	170	160	140	130	110	110
Maximum drain extension sheet resistance for bulk MPU/ASIC (NMOS) (Ω/sq) [B]	650	670	660	680	750	810	900
Contact maximum resistivity for bulk MPU/ASIC (Ω-cm²) [I]	1.60E-07	1.38E-07	8.00E-08	4.00E-08	2.00E-08	1.00E-08	8.00E-09

ITRS



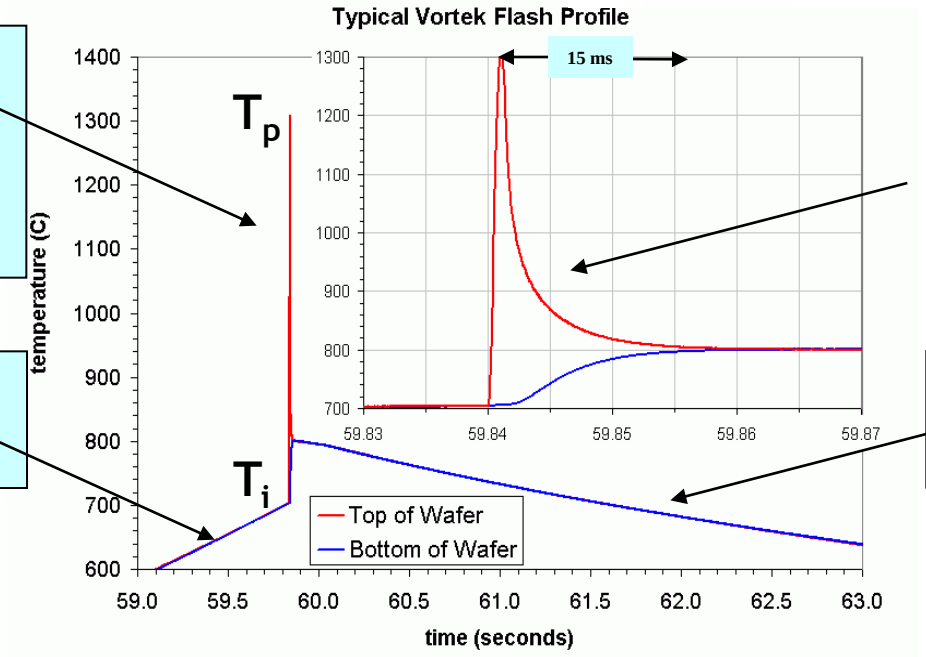
Millisecond Annealing with Millios™ fRTP™

- **Millios™ fRTP™ :**
  - 150 K/s ramp to T_i
  - Pulsed surface heating
  - Millios™ tool provides real-time T measurement and control on front & back of wafer



Fast ramp to T_p by flash heating the surface

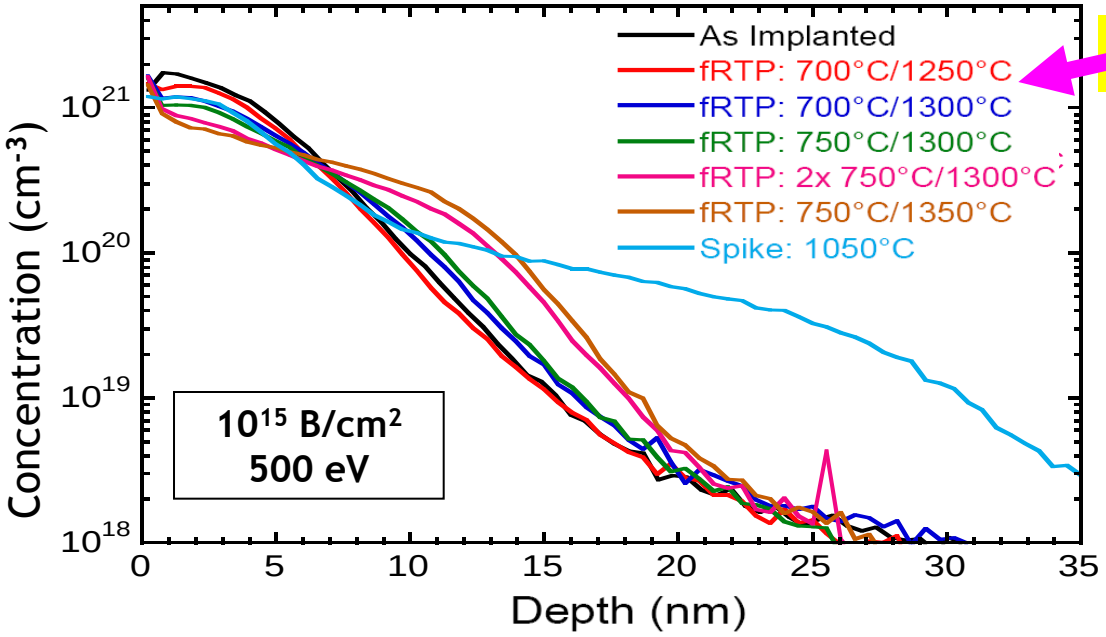
Bulk heating of wafer to T_i

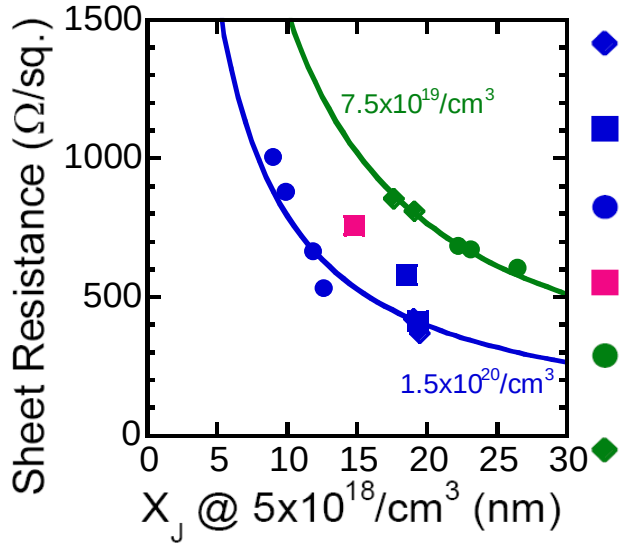


Rapid cooling of surface by thermal conduction to bulk (heat sink)

Radiative cooling of wafer

Diffusion and Activation with Millisecond Annealing





- fRTP: 10^{15}B/cm^2 500eV c-Si
- fRTP: 10^{15}B/cm^2 500eV PAI
- fRTP: He/B_2H_6 plasma doping
- SPE: 10^{15}B/cm^2 500eV PAI
- Spike Anneal $10^{15}BF_2$/cm^2 1.1keV
- Fast Cool Spike Anneal $10^{15}BF_2$/cm^2 1.1keV

fRTP: Intermed. T/ Peak T

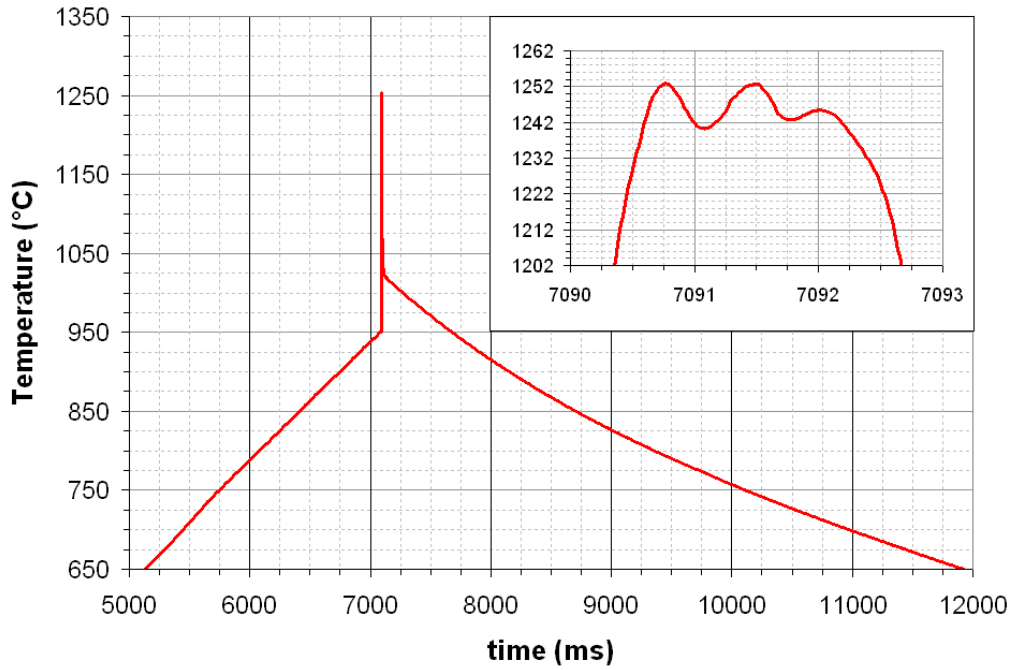
- Minimal diffusion with fRTP, except for highest temperatures
- Concentration-enhanced diffusion ⇒ More abrupt junctions
- High activation can be obtained with minimal diffusion
- Very shallow junctions can be formed, provided that as-implanted profile is shallow enough

P. J. Timans et al., MRS 2006

The Need of a Long ms-Flash Anneal

- For 22 nm technology and beyond where gate pitch is ≤80nm, a difficult trade-off exists:
  - Need to increase SD area for lower contact resistance
  - Need for adequate final spacer width to prevent SD encroachment to SDE, which degrades device SCE performance
- This trade-off tends to drive the use of thinner SDE offset spacer and the corresponding reduction of spike RTA temperature for device centering
- Although subsequent diffusionless millisecond anneal can improve the dopant activation, defects anneal remains inadequate with lower spike RTA temperature
- Process approaches with only diffusion-less MSA offer inadequate defect anneal and profile control to fully optimize device performance
- Looking forward into 15nm node a junction depth of 7-10nm will be needed
  - A new anneal approach to provide nm-scale junction profile control for device centering, improved defects anneal and dopant activation will be required
  - Long pulse seems to be the most promising way to fill the need
- Therefore, we have developed a high temperature (1200-1300°C), long ms-anneal approach (1-2.5 ms) in the Mattson Millios™ system

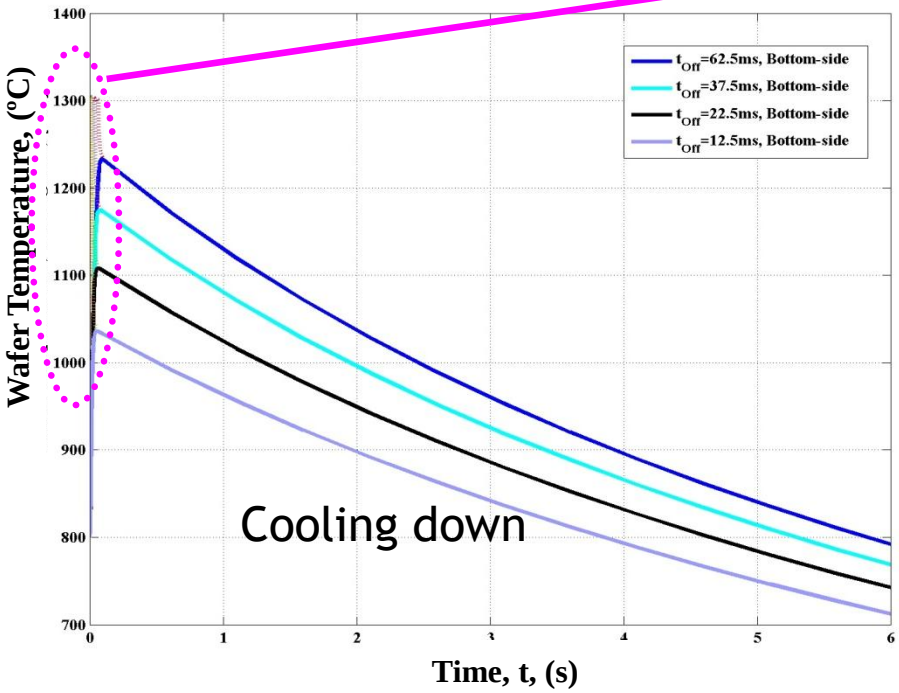
Real-Time Control in a Long ms-Flash

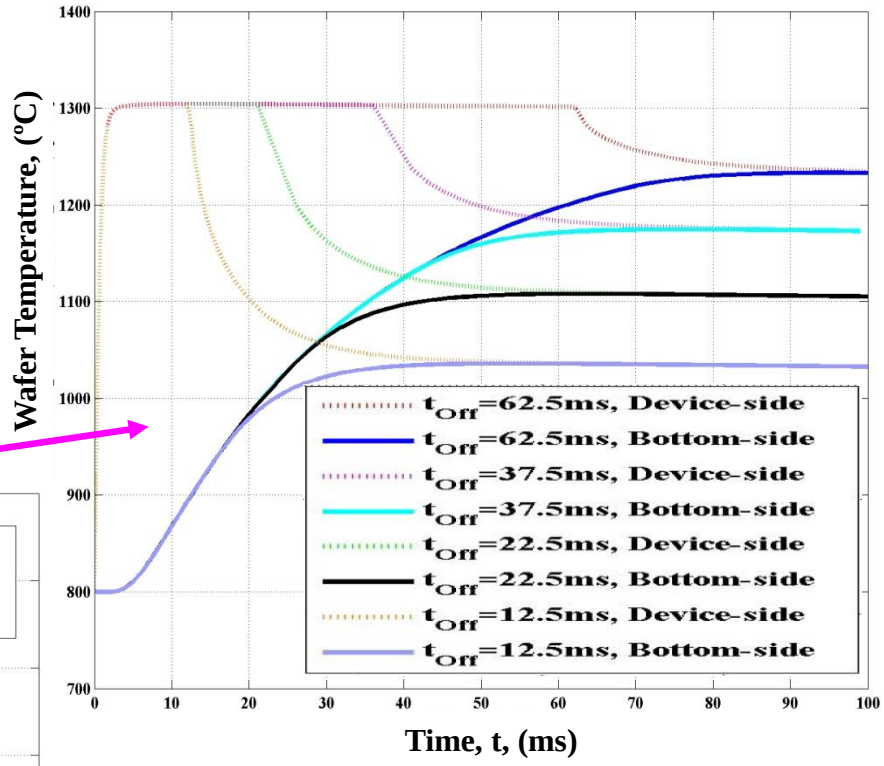


- Real time control of lamps in Millios™ allows for flat top temperate profiles
  - Maximizes activation
  - Reduces diffusion and deactivation during heating and cooling

Practical Range of a Long ms-Flash

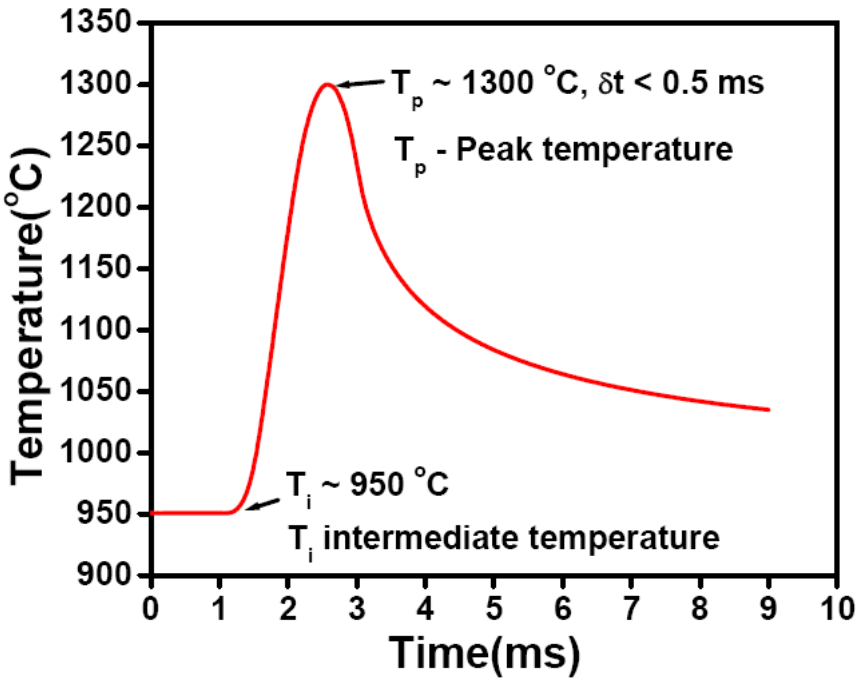
- (Modeling t-T profile) by holding the device side at a constant yet elevated temperature for longer times
- If time at temperature >20ms the bulk of the wafer will approach temperatures where undesirable effects will occur (such as diffusion and deactivation)

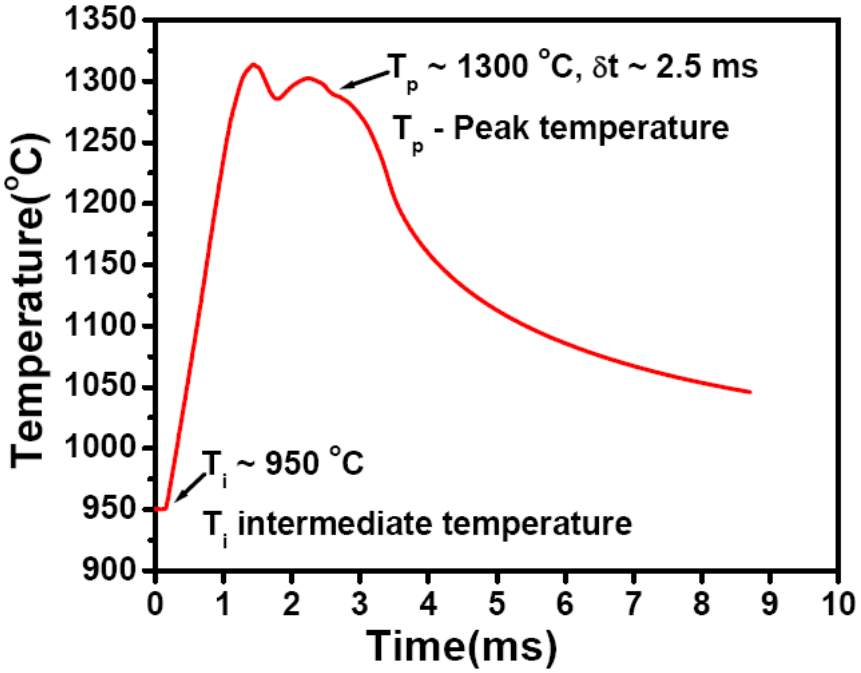




- Practical range of a long ms flash
  - Peak temp up to 1300°C
  - Duration up to 20ms

Time-Temp. profile: a short ms-flash vs. a long ms-flash

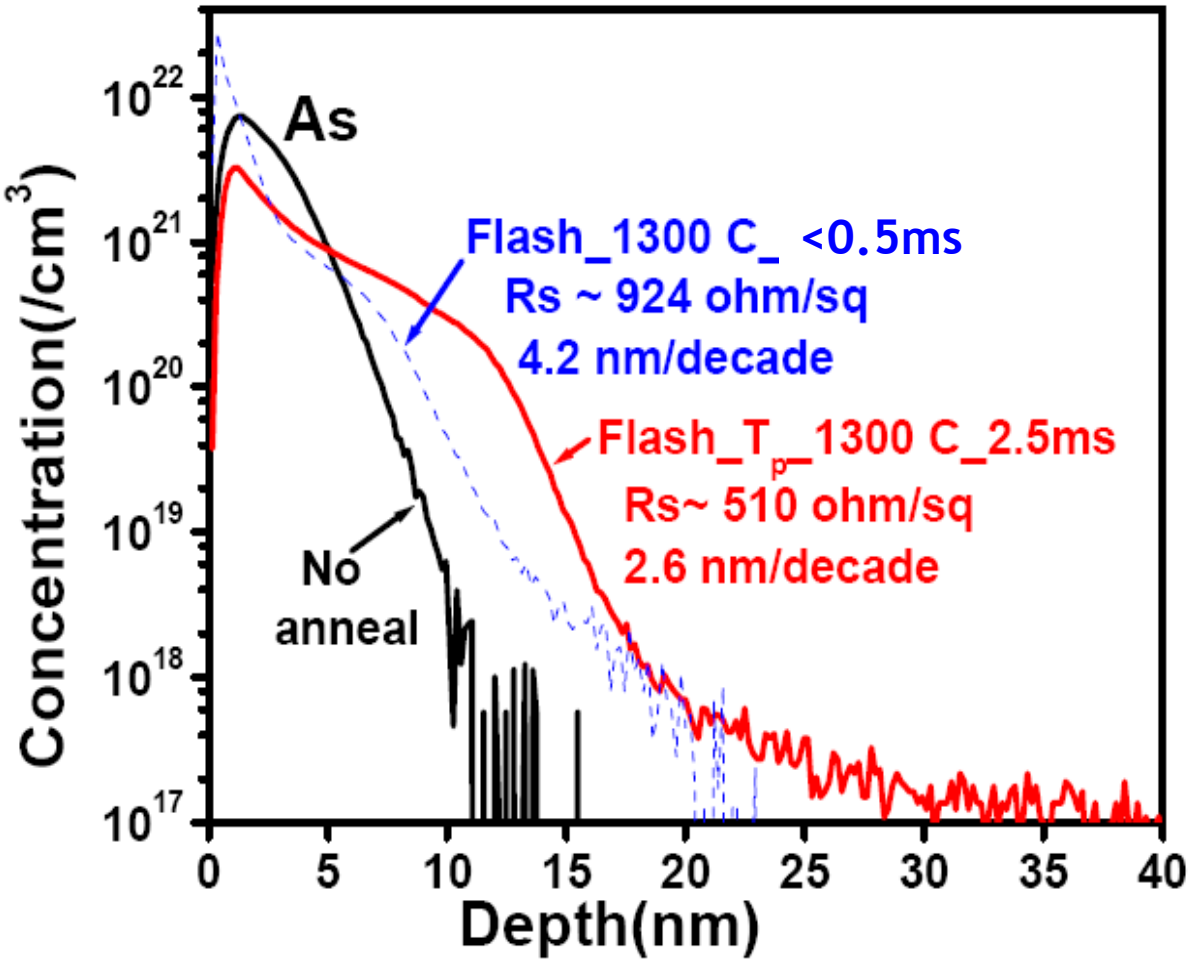




K.L. Lee et al. (IBM) and S. McCoy et al. (Mattson), Int'l Workshop of Junction Technology, May 2010

- These are t-T profiles of blanket wafers
- The surface reflectivity of device wafers changes from 70% to 10% relative to blanket wafers, thus enabling a larger range of (T_p - T_i) in as much as 550°C for long ms-anneal of device wafers

Comparison of As junction formation with a short and long-ms flash

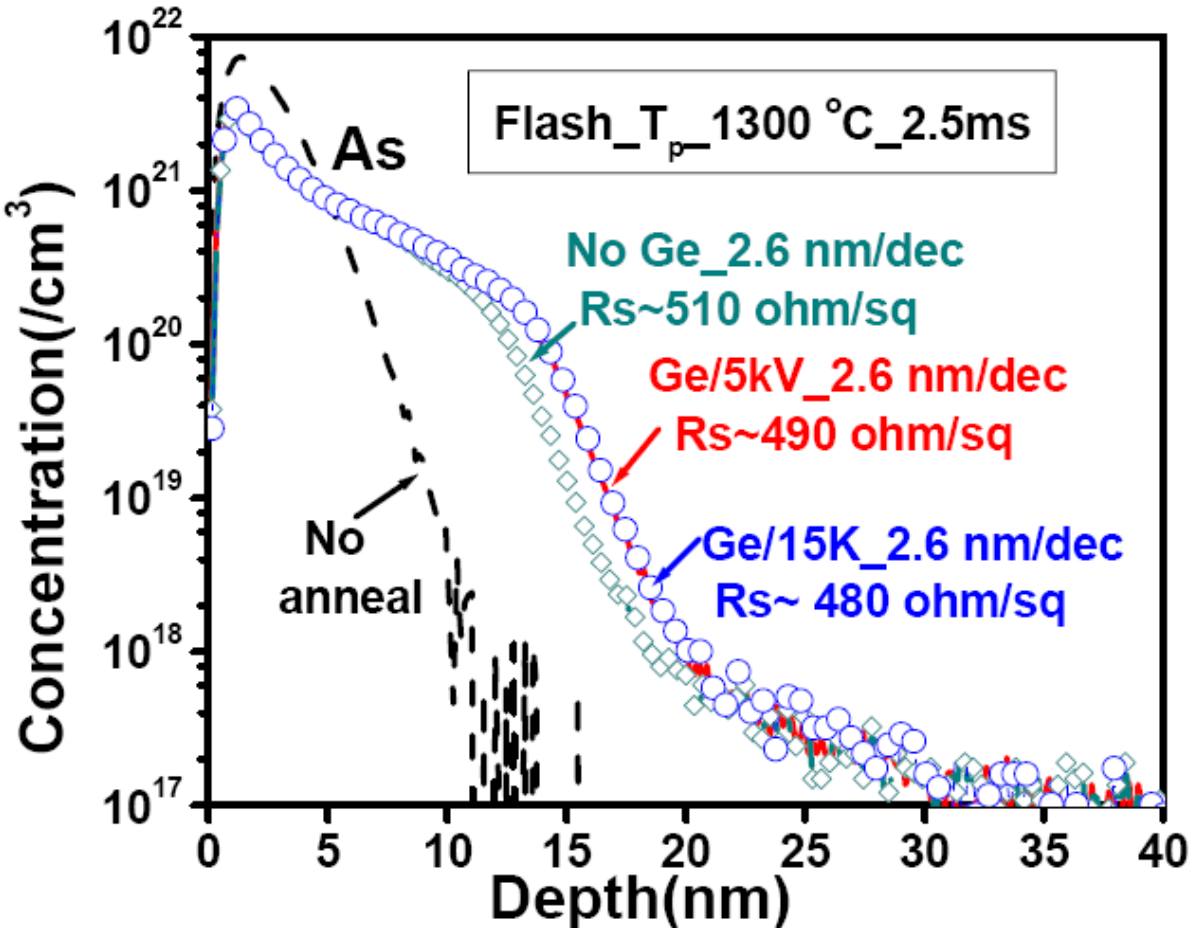


A long ms-flash enables a more box-like abrupt junction with lower Rs

With long ms-flash, conc'n-enhanced diffusion is observed to the extent similar to the As diffusion in c-Si ➔ improved damage re-growth

K.L. Lee et al. (IBM) and S. McCoy et al. (MTSN), IWJT, May 2010

Comparison of As junction formation with different Ge PAI (0, 5kV, 15kV) and long-ms Flash



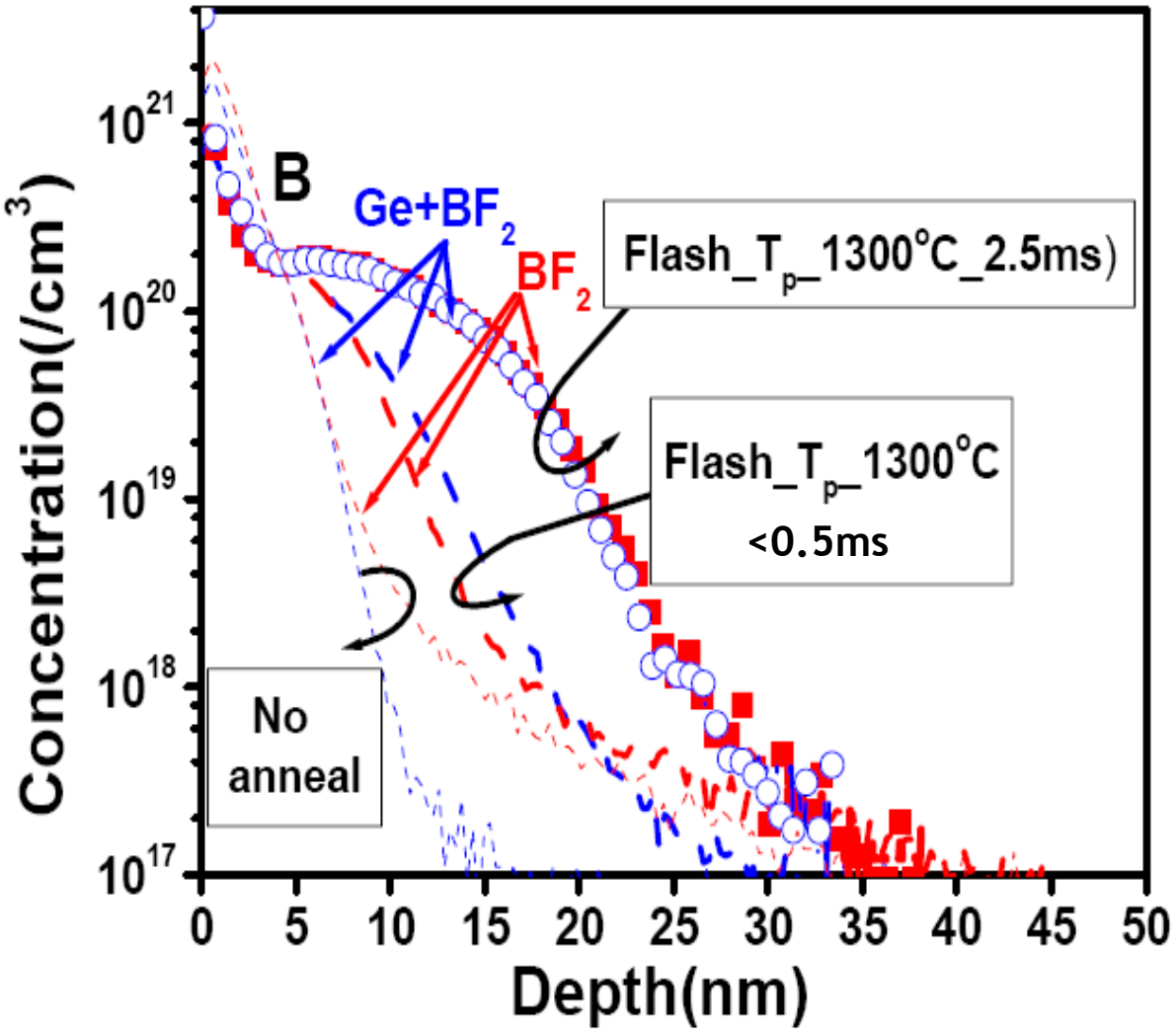
Again, long ms-flash enables a box-like abrupt junction with similar Rs indep. of Ge PAI

➔Rapid dissolution of Ge PAI EOR damages and regrowth during long ms-flash

➔Final As junction profile dominated by As diffusion in c-Si

K.L. Lee et al. (IBM) and S. McCoy et al. (MTSN), IWJT, May 2010

Comparison of B junction formation with short and long ms-flash anneal



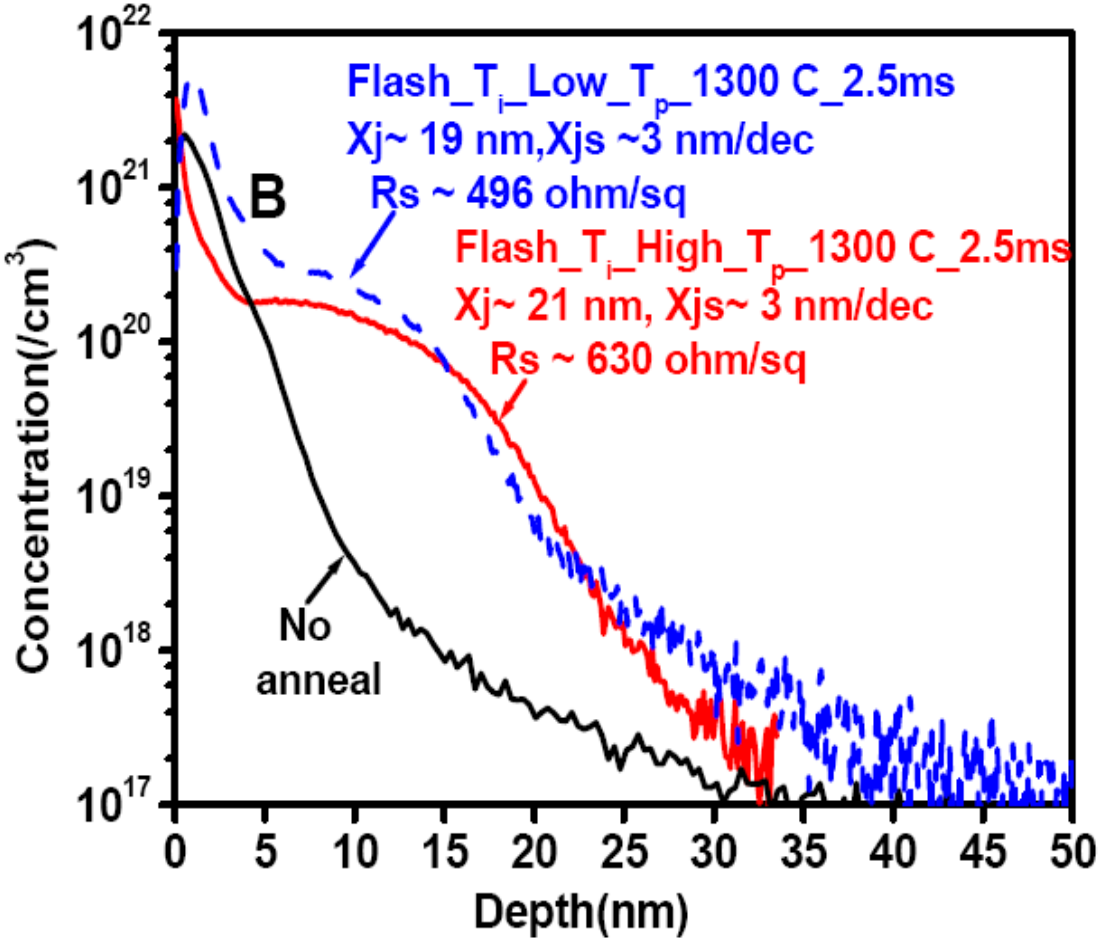
K.L. Lee et al. (IBM) and S. McCoy et al. (MTSN), IWJT, May 2010

Again, long ms-flash enables a more box-like abrupt junction

For short ms-flash the EOR of Ge PAI provides excess interstitials to enhance diffusion

For long ms-flash final B junction is same with or w/o Ge PAI ➔ Rapid dissolution of Ge PAI EOR damages and regrowth

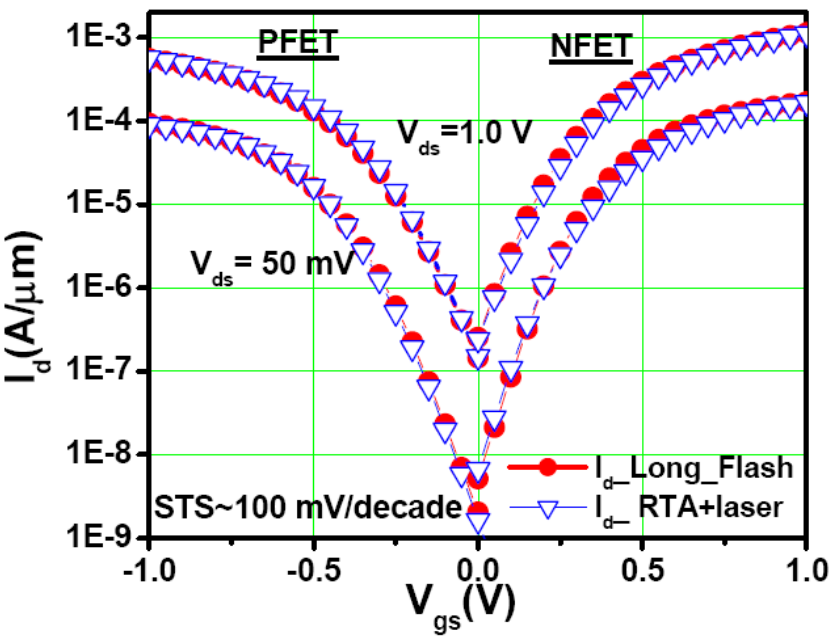
Comparison of B junctions with different energy, doses and flash Ti temperatures

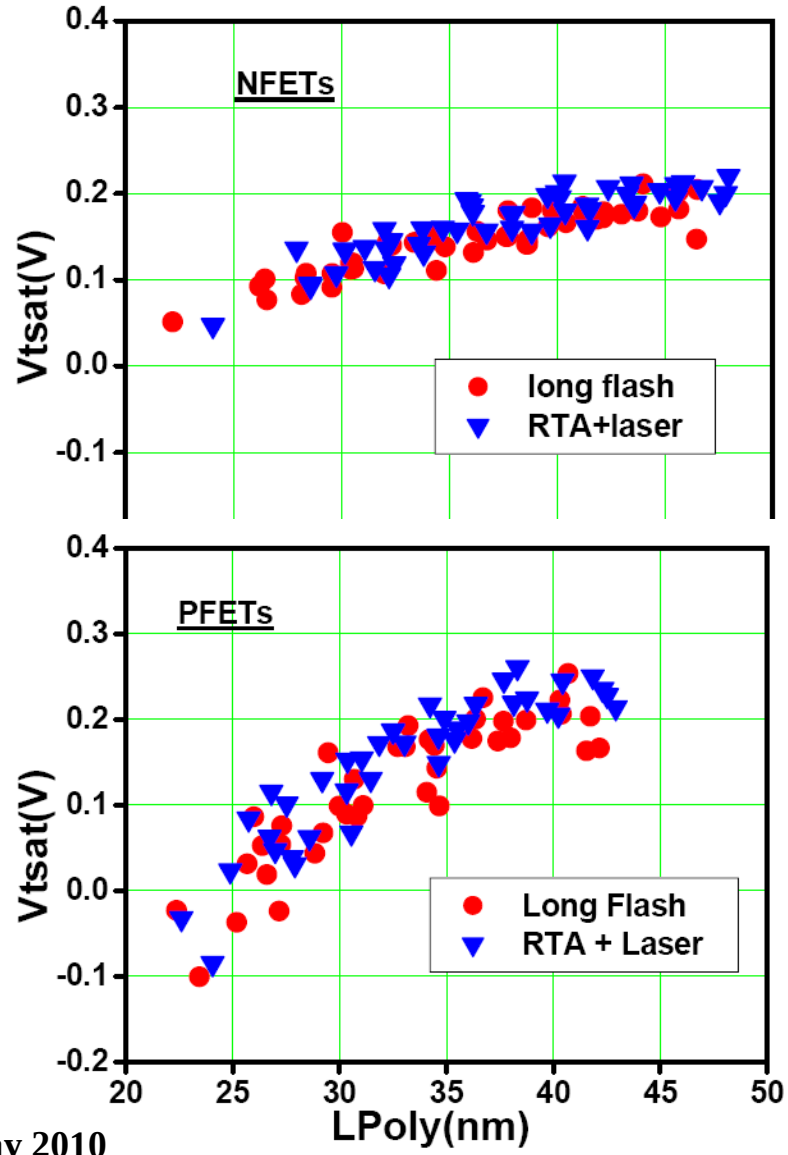


- Using higher BF_2 I/I energy & dose, and an absorber on blanket I/I wafer to simulate device wafer reflectivity, long ms-flash with lower T_i at T_p of 1300°C was enabled and compared to one with higher T_i
- Lower T_i case yields shallower Xj (19 vs 21 nm), lower Rs (496 vs 630 ohm/sq) and similar junction abruptness 0f ~3nm/dec

K.L. Lee et al. (IBM) and S. McCoy et al. (MTSN), IWJT, May 2010

I_d vs V_{gs} of PFETs and NFETs at Vds of 50 mV & 1V





- Long ms-flash implemented in a SOI process with a Poly/SiON gate stack was compared with a "combo process" of Spike-RTA + Laser Anneal
- High perf. P- & N-FETs with matching I_d-V_{gs} and similar $V_{t(sat)}$ roll-off have been obtained for both annealing schemes

K.L. Lee et al. (IBM) and S. McCoy et al. (MTSN), IWJT, May 2010

Conclusions

- High temp long ms-flash anneal has been employed to improve defect annealing while maximizing dopant activation with minimal diffusion
- Mattson Millios™, with the long ms-flash anneal (2.5ms) capability, has been demonstrated to better activate dopants in sub-20 nm N+ & P+ SD with more box-like abrupt junctions compared to shorter ms-flash anneal (<0.5ms)
  - By means of this, high performance SOI CMOS has been achieved with single long ms-flash and matches well with CMOS created with a "combo" of spike RTA + laser anneal
- Mattson Millios™ capabilities are being developed to enable a long ms-flash anneal, up to a practical limit of 20ms with peak temp of 1300°C
  - The Millios™ is capable of measuring & controlling both the top & bottom side temp in real time over a wide range of temperature, extending down to 250°C & up to 1400°C