Raising the Bar

Trends and Challenges in CMOS FEOL Technology for the 45nm Node and Beyond

Philippe Absil, Thomas Hoffmann, Jorge Kittl, Anne Lauwers,
Anabela Veloso, Hongyu Yu, Malgorzata Jurczak, Serge Biesemans – IMEC

CMOS FEOL technology scaling, traditionally achieved by reducing gate length (L_g) and gate oxide thickness (T_{OX}), is down to a limit beyond which the gate leakage current becomes unacceptable. Strained silicon helps compensate for transistor performance loss from this reduced L_g and T_{OX} scaling. Beyond strain, high-k dielectrics, metal gates, millisecond anneals and new silicides are options for continued scaling. Each presents specific process challenges. This article reviews recent advances in fully silicided gates (FUSI) as an option for metal gate integration to continue transistor performance scaling down to 45nm and below.

Introduction

Traditionally, the industry has faced FEOL process challenges in three key areas:

- Narrowing of the gate dimensions by the use of advanced lithography and dry etch
- Thinning of the gate dielectric
- Control of junctions by ion implantation and short thermal anneal with contacts formed by self-aligned silicidation

The challenges alter dramatically as we move toward increasing transistor innovation at the 65nm node and below. The inversion layer thickness (proportional to T_{ox}) scaling trend from one technology generation to the next has changed when going from the 90nm node to the 65nm node, with a minor reduction of dielectric thickness compensated by a performance gain coming from strain engineering (figure 1). Whether it is achieved with recessed SiGe source/drain (S/D), strained liners or strain memorization techniques, each strain-boosting option presents integration and process control challenges. As we scale transistor dimensions beyond the 65nm node, we must not only maintain the gain obtained from strain boosters, but also adopt new ways to further increase device performance.

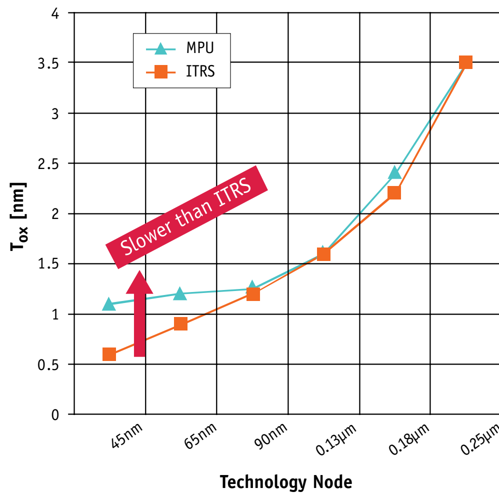


Figure 1: T_{ox} scaling evolution from the 0.25 micron to the 45nm technology node.

Enter Metal Gate Electrodes

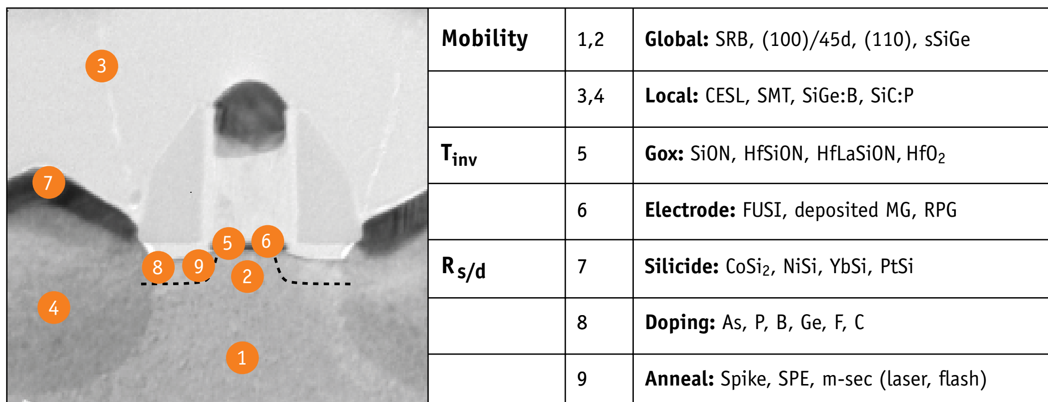
There are many options to increase the MOS transistor drive current (figure 2). First, the mobility may be increased with the use of alternative substrates, recessed S/D or increased stresses of known strain-boosting techniques. The access resistance can also be reduced with new silicides that have reduced contact resistance, and by improved dopant activations obtained by co-implants or advanced anneals. Nevertheless, the prime candidate for achieving the required performance increase for the future generations is the transistor gate, where high-k dielectrics and metal gates offer the opportunity to resume the T_{inv} reduction. During the last five years, many fundamental issues linked to high-k dielectrics have been solved. Using nitrided hafnium-silicates, the industry has found a viable candidate to reduce gate leakage, while maintaining the device performance with respect to SiON dielectrics[1].

However, such a dielectric will likely require a metal gate electrode to avoid the compatibility issues of the high-k with doped poly-Si gates that leads, among other things, to unsuitable transistor threshold voltage (V_t). Hence the main challenge is to integrate the metal gate electrodes and fabricate devices such that the V_t levels meet the specifications. The metal gate by itself, independent of the dielectric, is beneficial since it eliminates the so-called "poly-depletion" contribution to the T_{inv}.

The following sections present several metal gate integration approaches with their advantages and disadvantages. The FUSI device integration is then proposed as a practical solution for which process control challenges are briefly discussed.

Gate First or Gate Last?

In theory, the transistor V_t is primarily determined by the work function (Wf) of the gate electrode material present on top of the gate dielectric. Doped poly-Si has traditionally fulfilled this role very elegantly, by tuning the Wf via the use of different species and concentrations to address all device V_t requirements. Although very flexible, poly-Si has the drawback of an increased T_{inv} by about 0.4nm due to the poly-depletion, which is becoming significant (~20% of the total T_{inv}) for state-of-the-art transistors (figure 3).



Mobility	1,2	**Global:** SRB, (100)/45d, (110), sSiGe
	3,4	**Local:** CESL, SMT, SiGe:B, SiC:P
T_{inv}	5	**Gox:** SiON, HfSiON, HfLaSiON, HfO_2
	6	**Electrode:** FUSI, deposited MG, RPG
$R_{s/d}$	7	**Silicide:** $CoSi_2$, NiSi, YbSi, PtSi
	8	**Doping:** As, P, B, Ge, F, C
	9	**Anneal:** Spike, SPE, m-sec (laser, flash)

Figure 2: Bulk CMOS transistor options for performance scaling.

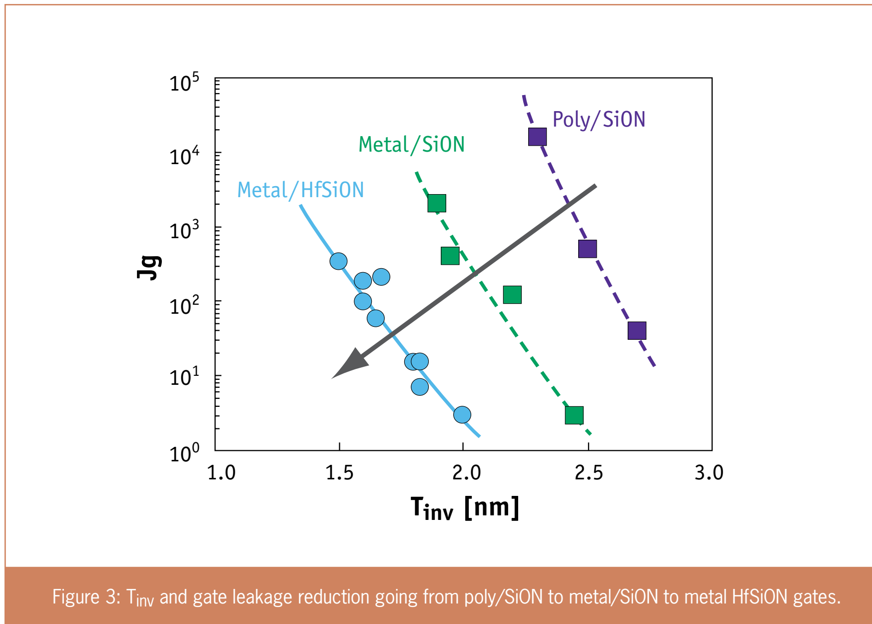


Figure 3: T_{inv} and gate leakage reduction going from poly/SiON to metal/SiON to metal HfSiON gates.

While the performance boost coming from the metal electrodes is well established, the challenge resides now in meeting the appropriate V_t targets. The Wf values for various metals are established and certain materials are identified with "band-edge" properties, i.e. with Wf close to the valence or the conduction band of silicon. Those metals are susceptible to serve the needs for high, regular or even low V_t transistors (HV_t, RV_t and LV_t, respectively)[2].

The question now arises about how to integrate these materials as metal gates.

Two main options exist:

(1) The "gate first approach" where the metal is deposited before the junction formation

(2) The "gate last approach" where the metal insertion is done after[3]

While the first option has the advantage of being the least disruptive from a process flow point of view, most of the metals will produce transistors with unacceptably high V_t, most likely due to material and interface modifications during the high thermal budget treatment (RTA) needed to activate the dopants[4]. Hence, a gate first approach can only use thermally stable metals. On the other hand, a gate last approach requires, as detailed below, additional process steps that result in additional fabrication costs and yield issues.

Deposited Metal or Full Silicidation?

The gate last approach consists in first forming the device channel, junctions and contacts that require a high thermal budget, and then depositing the gate.

Beyond Phase-engineered V_t

Although very elegant due to the relative simplicity of its implementation, the range of applications addressed by the phase engineering of FUSI is currently limited to the HV_t devices with HfSiON gate dielectric. Hence other options have been investigated to push further FUSI gates Wf towards the Si valence and conduction bands. Dopants can be introduced by ion implantation but the Wf variation obtained remains small for SiON dielectrics compared to the shift obtained on SiO_2[10, 11] and non-existing on HfSiON. The Ni can also be alloyed with other metals or the gate poly-Si can be replaced by poly-SiGe. In this way, NiYbSi has been found to be very effective to tune the Wf towards the conduction band on SiON[12] and NiSiGe pushes the Wf towards the valence band only in the case of HfSiON[13]. Alternative metals have also been investigated such as Pt-rich silicide that exhibits pMOS compatible Wf with V_t values meeting LV_t requirements[14]. Opportunities thus exist to modify FUSI gates Wf beyond phase engineering to demonstrate FUSI even for LV_t applications both on HfSiON and SiON (figure A).

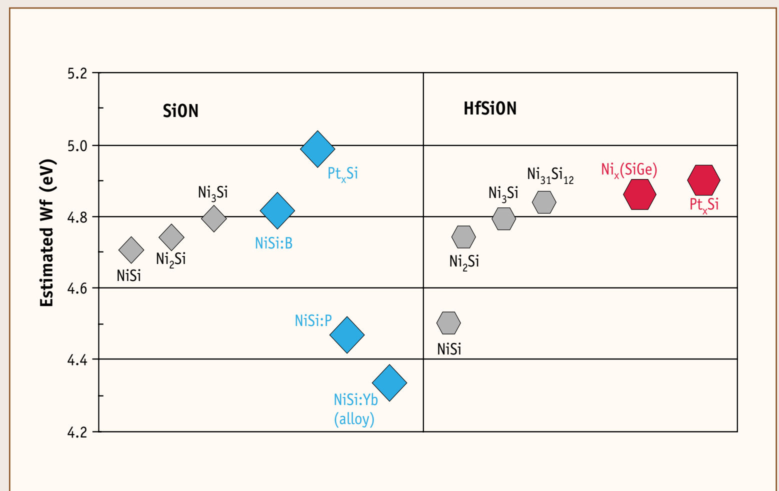


Figure A: Wf tuning options beyond phase engineering for SiON (left) and HfSiON (right).

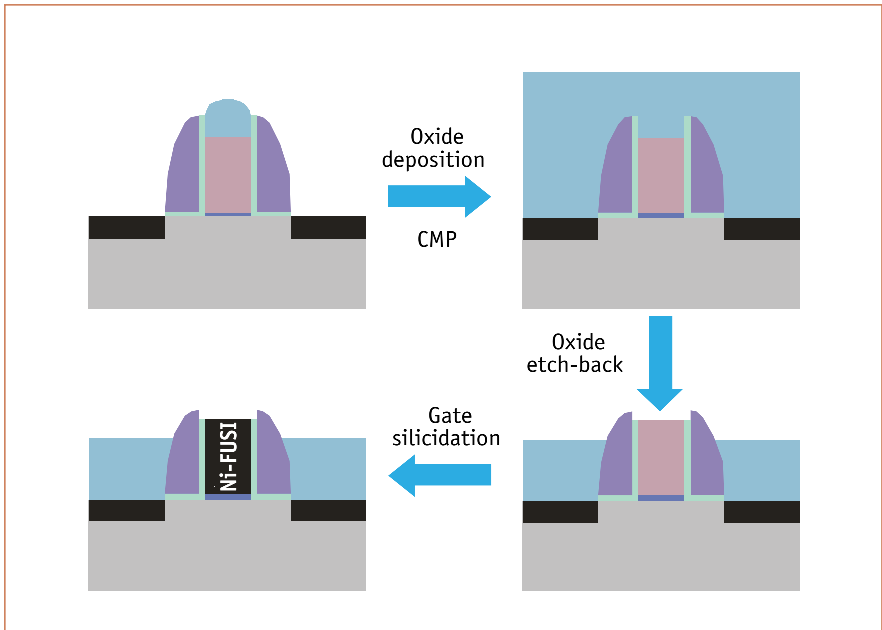


Figure 4: Illustration of a CMP-based FUSI gate integration flow with covered S/D NiSi during the gate silicidation step.

Here again, two options exist:

(1) The poly-Si is removed and the metal is deposited (referred to as the "replacement gate" approach)

(2) The metal is reacted with the poly-Si to form a silicide in contact with the gate dielectric (referred to as the FUSI approach)

The advantages of the latter include the relative simplicity to develop a self-aligned process through selective removal of un-reacted metal. For the replacement gate, however, additional fabrication steps are still needed to keep the metal only in the gate-region.

In recent years, FUSI has attracted considerable attention due to its relative practicality and, in the case of Ni-FUSI, its compatibility with existing processes[5]. An example of such a gate last integration flow is given for the FUSI approach by the use of CMP and etch-back in figure 4.

Ni-FUSI Phase Control is Important

As Ni reacts with Si, it forms a Ni_xSi_y silicide, x and y (i.e. the phase) being determined by the Ni-to-Si ratio and the reaction thermal budget. The possible phases are $NiSi_2$, NiSi, Ni_3Si_2, Ni_2Si, $Ni_{31}Si_{12}$, and Ni_3Si. It has been found that the phase of the silicide determines the Wf of the electrode formed with Ni-rich phases (x>y) closer to the Si valence band (figure 5) when on HfSiON dielectrics[8]. The large Wf increase for Ni-rich phases on HfSiON has been attributed to the un-pinning of the Fermi level[7].

Hence, the Ni-to-Si ratio and the reaction temperature are the key parameters used to control transistor V_t. For the narrow gate, however, the Ni/Si ratio is not very well defined since the Ni present on the spacer can diffuse and the resulting phase will be Ni-rich while NiSi will form in wide-gate devices. This results in an unwanted V_t variation from long to short channel transistors. The use of a two-step RTP process (RTP-1 + selective etch + RTP-2) helps reduce this gate length dependency by controlling the Ni supply by the first anneal step rather than by the deposited thickness (figure 6) [8].

Once under control for all gate lengths, the phase is used as a tuning parameter to modulate the Wf and hence the V_t. In figure 5, it is shown that on HfSiON

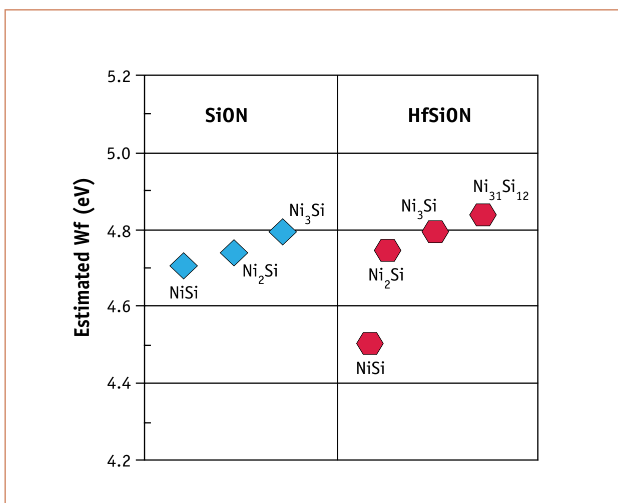


Figure 5: Ni-silicide Wf for monosilicide and Ni-rich phases on SiON (left) and HfSiON (right).

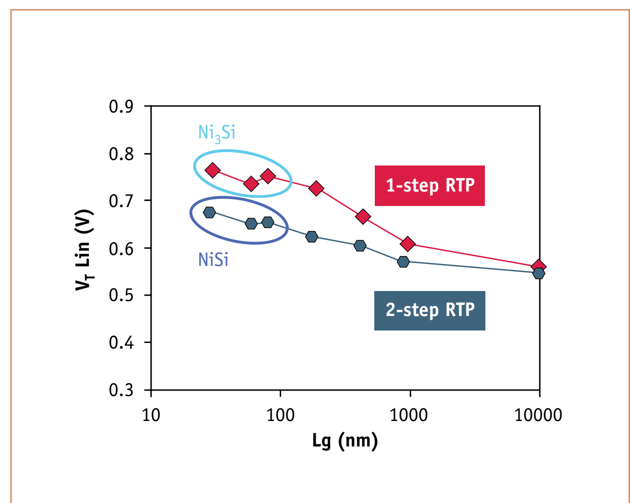


Figure 6: nMOS Vt vs. Lg for one-step and two-step RTP silicidation processes.

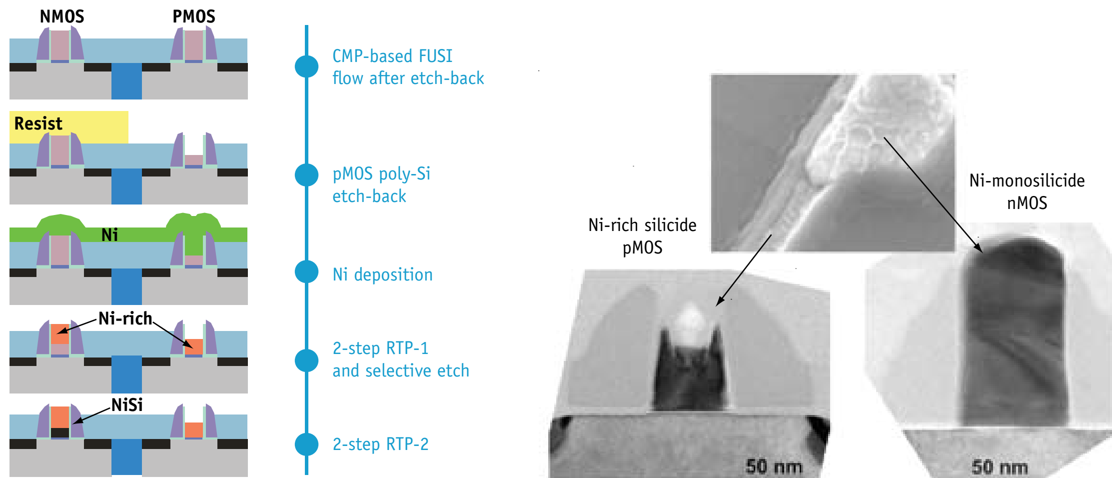


Figure 7: Integration flow for CMOS implementation of phase engineered dual Wf Ni-FUSI devices (left) and TEM cross-section of pMOS and nMOS (right).

gate dielectric, Ni-rich silicides have Wf values compatible with HV_t device requirements for pMOS while Ni monosilicide meets the same requirements for nMOS.

Dual Wf FUSI Ring Oscillators

A practical implementation of phase engineering to control the Wf selectively on pMOS and nMOS is illustrated in figure 7, presented by Lauwers et al. at IEDM 2005[9].

From the Wf vs. phase observations, the appropriate phases can be obtained if higher Ni-to-Si ratio is achieved on pMOS compared to nMOS. Since the two-step RTP process is thermally limited, the Ni-to-Si ratio cannot be modulated by the Ni thickness. However, the poly-Si thickness available for the reaction can beselectively tuned by an etch-back in pMOS regions. The simultaneous silicidation of p- and n-MOS results in Ni-rich Si FUSI and NiSi FUSI, respectively. This implementation delivers a functional metal gate-based ring oscillator with $V_{t\text{-sat}}$ of 0.4V and 0.5V for p- and n-MOS, respectively.

FUSI Process Control

As discussed earlier, the control of the phase is critical to fabricate devices with the targeted V_t values. First, the amount of Ni to react is determined by the RTP-1 temperature that impacts the phase formation (NiSi for nMOS). For a temperature that is too low, not enough Ni reacts, leading to a partial silicidation; for a temperature that is too high, extra Ni reacts and forms an unwanted Ni-rich phase. Figure 8 illustrates the determination of the temperature range for which NiSi forms in nMOS transistors (i.e. process window, PW) by electrical characterization. For temperatures below the lower bound of the PW, the capacitance equivalent thickness (CET) increases, indicating the presence of un-reacted poly-Si. For temperatures above the upper bound of the PW, narrow devices have a higher V_t.

Besides the RTP-1 temperature, the control of the poly-Si thickness and the spacer height at the time of silicidation are also important to achieve the desired phase in transistors. While it is obvious that the Ni-to-Si ratio is directly affected by the poly-Si thickness, the impact of spacer height is more subtle: In the case of a recessed spacer below the poly-Si top surface, the Ni deposited on the poly-Si sidewalls will also be available for reaction and an undesirable Ni-rich phase may be formed in nMOS devices.

In the CMP-based flow proposed in figure 4, the poly-Si available for

NiSi phase control is critical to achieve targeted V_t values, needing attention in terms of process control/inline metrology

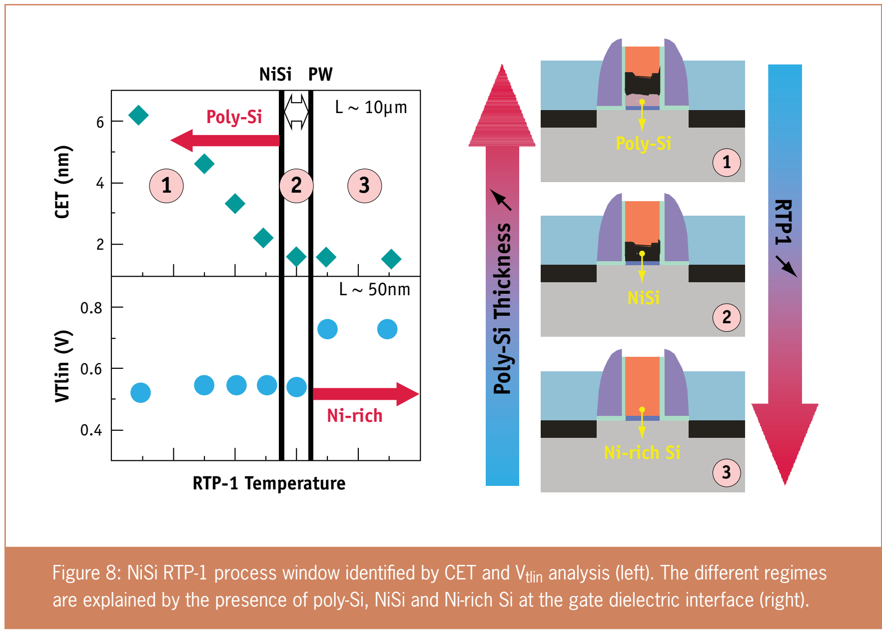


Figure 8: NiSi RTP-1 process window identified by CET and V_{tlin} analysis (left). The different regimes are explained by the presence of poly-Si, NiSi and Ni-rich Si at the gate dielectric interface (right).

silicidation is determined by the as-deposited thickness and the erosion occurring during the hard mask dry etch-back step. Similarly, the spacer recess depends on the spacer formation etch and the erosion during subsequent steps. Most of the erosion occurs during the gate hard mask dry etch back, due to the exposure of the spacers and the poly-Si once the oxide is removed, while an over-etch step must be applied to take into account process non-uniformities.

Figure 9 shows a wafer map of the oxide to be removed above the poly-Si gate after the CMP step. The area with the thinnest post-CMP oxide experiences more erosion, hence the available Si and recessed spacers shift the balance of the Ni-to-Si ratio towards Ni-rich phase formation. This is correlated with the nMOS device V_t showing higher value (signature of the presence of Ni-rich phase) where the oxide thickness before etch-back was thinner.

The stringent requirements to control these parameters thus necessitate the development of new inline metrology techniques capable of measuring poly-Si thickness and spacer height in gate lengths down to 30nm, since those quantities are likely to be gate-length dependent.

Conclusions

Just as strain was introduced at the 90nm node and widely adopted at the 65nm technology node to continue transistor performance scaling, metal gates may fulfill this role for the 45nm node and beyond. Among the many options available to integrate them in transistors, the practicality of phase or dopant-engineered Ni-FUSI makes this the most mature approach, with functional ring oscillators demonstrated for HV_t applications. The control of the NiSi phase (i.e. the Ni-to-Si ratio and thermal treatment) is found to be critical to achieve the targeted V_t values, and requires particular attention in terms of process control and inline metrology. Beyond the phase engineering of the Wf, FUSI options exist to lower the transistors' V_t to levels suitable for LV_t applications either with dopant, alloys of NiSi or alternative metals. Although many concerns remain, including yield and reliability issues, the FUSI approach is arguably the most practical way to integrate metal gates in advanced CMOS.

The FUSI approach is arguably the most practical way to integrate metal gates in advanced CMOS.

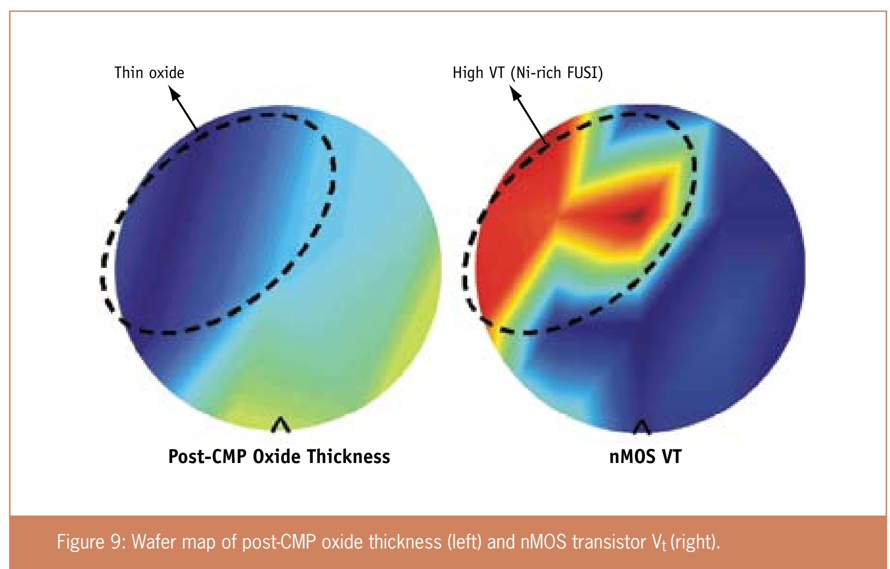


Figure 9: Wafer map of post-CMP oxide thickness (left) and nMOS transistor V_t (right).

References

1. M. Takayanagi, et al., Technical Digest IEDM, p.903, 2005.
2. P. Mahji, et al., ICICDT International Conference, p.69, 2005.
3. L. Colombo, et al., Future Fab International, p.120, issue 18, 2005.
4. J.K. Schaeffer, et al., Technical Digest IEDM, p.287, 2004.
5. P. Ranade, et al., Technical Digest IEDM, p.227, 2005.
6. J. Kittl, et al., IEEE Electron Device Lett., vol.27, no.1, p.34, 2006.
7. C.C. Hobbs, et al., Electron Devices, IEEE Transactions, no 51, p.971, 2004.
8. J. Kittl, et al., Symposium on VLSI Technology, p.72, 2005.
9. A. Lauwers et al., Technical Digest IEDM, p.661, 2005.
10. J. Kedzierski, Technical Digest IEDM, p.315, 2003.
11. C. Cabral, Symposium on VLSI Technology, p.184, 2004.
12. H. Yu et al., Technical Digest IEDM, p.645, 2005.
13. H. Yu et al., Technical Digest IEDM, p.653, 2005.
14. C.S. Park, Symposium on VLSI Technology, p.48, 2005.

Author Biographies

Philippe Absil received his M.S. in physics engineering in 1997 from Université Libre de Bruxelles in Belgium, and his Ph.D. in electrical engineering in 2000 from University of Maryland in the field of microphotonics. In 2000, he joined Little Optics, involved in advanced microphotonics components fabrication. In 2003, he returned to Belgium and joined IMEC where he now manages the Bulk-Si CMOS FEOL integration research program. E-mail: philippe.absil@imec.be

Thomas Hoffmann, IMEC vzw, Belgium; thomash@imec.be

Anne Lauwers, IMEC vzw, Belgium; anne.lauwers@imec.be

Anabela Veloso, IMEC vzw, Belgium; anabela.veloso@imec.be

Hongyu Yu, IMEC vzw, Belgium; hongyu.yu@imec.be

Malgorzata Jurczak, IMEC vzw, Belgium; malgorzata.jurczak@imec.be

Serge Biesemans, IMEC vzw, Belgium; serge.biesemans@imec.be

Jorge Kittl, Assignee from Texas Instruments at IMEC; jorge.kittl@imec.be

YMS Taiwan
August 23 – Ambassador Hotel, Hsinchu, Taiwan

YMS Singapore
August 25 – Raffles The Plaza Hotel, Raffles City Convention Center, Singapore

YMS Shanghai
August 29 – Ramada Plaza, Shanghai, China

YMS Beijing
September 1 – JinJiang Hotel, Beijing, China

YMS Japan
December 7 – New Otani, Makuhari, Japan

Attend KLA-Tencor's Yield Management Seminar Series in Asia-Pacific & China

2006

For more details and registration for KLA-Tencor events, please visit **www.kla-tencor.com/events**