Advanced CMOS device technologies for 45 nm node and below

A. Veloso*, T. Hoffmann, A. Lauwers, H. Yu, S. Severi, E. Augendre, S. Kubicek, P. Verheyen, N. Collaert, P. Absil, M. Jurczak, S. Biesemans

IMEC, Kapeldreef 75, B-3001 Leuven, Belgium

Received 13 October 2006; accepted 24 November 2006
Available online 29 January 2007

Abstract

We review and discuss the latest developments and technology options for 45 nm node and below, with scaled planar bulk MOSFETs and MuGFETs as emerging devices. One of the main metal gate (MG) candidates for scaled CMOS technologies are fully silicided (FUSI) gates. In this work, by means of a selective and controlled poly etch-back integration process, dual work-function Ni-based FUSI/HfSiON CMOS circuits with record ring oscillator performance (high-VT) are reported (17 ps at VDD = 1.1 V and 20 pA/μm I off), meeting the ITRS 45 nm node requirement for low-power (LP) CMOS. Compatibility of FUSI and other MG with known stress boosters like stressed CESL (contact-etch-stop-layer with high intrinsic stress) or embedded SiGe in the pMOS S/D regions is validated. To obtain MuGFET devices that are competitive, as compared to conventional planar bulk devices, and that meet the stringent drive and leakage current requirements for the 32 nm node and beyond, higher channel mobilities are required. Results obtained by several strain engineering methods are presented here.

© 2007 NIMS and Elsevier Ltd. All rights reserved.

Keywords: CMOS scaling; Metal gates; FUSI; Strain boosters

1. Introduction

Over the past four decades, MOSFET scaling has been the major driving force for unprecedented high productivity and performance gains as quantified in Moore's law. In recent years, scaled device performance has been compromised with fundamental material limits, requiring the introduction of new materials (Fig. 1). A good example is the gate stack, with the high gate tunneling leakage currents and loss of gate capacitance due to depletion of poly-Si for conventional poly-Si/SiON gate stacks. Hafnium-based dielectrics such as HfO2 and HfSINO have emerged as the leading candidates to replace SiON for sub-45 nm technologies [1]. Due to fundamental limitations such as Fermi-level pinning, both high-k dielectrics and metal gates (MG) are now expected to be introduced at the same node [2]. Fig. 2 shows that, e.g., only MG/high-k stacks can meet the 45 nm ITRS requirement for high V_T gate stacks. In the near term, the use of strained Si channels is counted on to provide the needed boost to devices performance. In the long term, the continued scaling is expected to require the introduction of dual gate (DG) devices that allow an excellent control of short-channel effects and better carrier transport as the channel doping is

*Corresponding author. Tel.: +32 16 28 17 28; fax: +32 16 28 17 06.
E-mail address: Anabela.Veloso@imec.be (A. Veloso).

1468-6996/$ - see front matter © 2007 NIMS and Elsevier Ltd. All rights reserved.
doi:10.1016/j.stam.2006.11.018

Fig. 1. Technology options for continued CMOS scaling.
reduced. Hitherto, one of the most successful realizations of the DG structures is the MuGFET (multiple gates FET)/FinFET (single gate FET) [3].

Table 1
Three main metal gate integration schemes

|       | MIPS | FUSI | RPG |
|-------|------|------|-----|
| Dielectric | GF   | GF   | GL  |
| Electrode  | GF   | GL   | GL  |
| Pros      | Scalability | Low complexity | Low thermal budget |
| Cons      | High thermal budget | Process control (phases) | Complexity |

MIPS: Metal Inserted Poly-Si Stack, FUSI, RPG: Replacement Gate. GF and GL stand for Gate first and Gate last, respectively.

Fig. 2. Gate leakage (J_G) vs. Tinv for different gate stacks.

Fig. 3. Work function (WF) modulation of Ni-FUSI on SiON and HfSiON by means of implantation, Ni-FUSI phase control or alloying.

Fig. 4. Simulated Idsat vs. gate WF for different target Ioff (Lg = 45 nm). Channel doping is adjusted to match Ioff_target for each WF.

Fig. 5. Performance and Lgmin relative change for pMOS poly vs. FUSI gates for different RTA conditions. FUSI enables thermal budget lowering w/o poly depletion increase penalty.
2. Scalable planar bulk CMOS

For metal gates CMOS planar integration, three main integration schemes (Table 1) have been proposed: Metal Inserted Poly-Si Stack (MIPS), Fully Silicided Gates (FUSI) and Replacement Gate (RPG) [4–7]. To date, results for dual work function (WF) CMOS circuits have been published for FUSI gates only [6,8]. Presently, a major challenge continues to be obtaining dual MG electrodes with appropriate WF. For FUSI gates, ion implantation of B, P, As, or Sb dopants was demonstrated to be effective in modulating the WF of NiSi-FUSI on SiO₂, with limited success on SiON or HfSiON [9,10]. In these cases, as shown in Fig. 3, Ni-silicide phase control (NiSi, Ni₂Si, Ni₃Si, Ni₃₁Si₁₂) [5,10], alternative ion implantation (Yb, Al), or alloying (NiYb) [11] can provide substantial WF modulation, as required for use in different applications. The optimum WF for a given application is set by a trade-off between too high channel doping (mobility limited) and too low channel doping (short-channel limited). Fig. 4 shows that, for low leakage targeted devices, ~450–500 meV ΔWF(n-p) is required.

Fig. 6. Sheet resistance vs. junction depth (experimental data; p-type) for different annealing techniques: SPE Re-growth, Laser, Flash, Spike and Cocktail Spike annealed junctions.

Fig. 7. nMOS (left) and pMOS (right) benchmark vs. JG, at fixed Ioff. Unless indicated, no stress booster is added. Published data from [4,21–23].

FUSI, as any MG, allows decoupling source/drain from gate activation, hence allowing a significant reduction of the junction anneal thermal budget (RTA) [8]. As
illustrated for pMOS devices in Fig. 5, when decreasing RTA, a smoother performance roll-off is obtained for FUSI vs. poly-Si gate devices, as the latter are impacted by the poly-depletion effect in addition to a series resistance increase. For further scaling, Fig. 6 shows a comparison of results obtained with different annealing techniques targeting very shallow, abrupt and low resistive junctions. Very attractive results are obtained with cocktail spike anneals but, in the long term, flash and laser anneals appear as most promising. A key integration issue is seen to be the limited thermal stability of most high-k materials or metals (in MIPS), adding constraints on the thermal budgets linked with dopants activation. In this regard, FUSI or RPG schemes offer more flexibility. Fig. 7 shows a benchmark (with gate leakage normalization) of FUSI/HfSiON devices against other gate stacks, after device optimization (reduced RTA, halo and HDD retargeting). As expected, FUSI is intrinsically matched to deposited MG and it significantly outperforms poly-Si gate devices. A thorough evaluation of performance, process control, VT targeting, and reliability of MGs (realized with FUSI process) [8,12] enabled obtaining a record high VT ring oscillator speed (17 ps at Ioff = 20 pA/μm for VDD = 1.1 V), in line with low power needs for the 45 nm node (Fig. 8).

To obtain strain-enhanced mobility channels, several methods are currently available. In particular, the use of a stressed CESL, compressive for pMOS and tensile for nMOS, and of embedded SiGe in the S/D regions for pMOS have received a lot of attention lately [13–17]. Viability of these strain-enhancement approaches for future technology nodes requires, apart from scalability issues [16], also a successful combination with advanced gate stacks (high-k and MG). Fig. 9 shows the performance improvement obtained for MG (MIPS)/high-k devices with different Si1−xGex splits, combined with a compressive CESL, illustrating the additive nature of the two stressor ARTICLE IN PRESS
techniques. Also for FUSI gate option, its compatibility with stressed CESL (Fig. 7) or SiGe S/D [18] is validated.

3. MuGFETs

The success of MuGFET technology is largely due to the fact that it can re-use 70% of the processing developed in conventional bulk CMOS, with special attention required for FIN and gate patterning, and junctions formation. For MuGFETs, since $V_T$ cannot be set by channel doping for narrow FIN devices, WF modulation with MG is the preferred solution. Fig. 10 compares the optimum WF calculated for planar bulk to FinFET devices for high-performance (HP) applications ($I_{off}$ target = 100 nA/μm). ~300 meV $\Delta WF(n-p)$ is extracted for the DG devices, with a 20% loss in performance estimated for a mid-gap WF (n and p). For low standby power applications (LSTP), optimum WF is mid-gap. Integration of a single MG seems thus a more viable option for MuGFETs (Fig. 11).

In MuGFET devices, the difference in crystal orientation between the top channel and the sidewalls affects differently electron and hole mobilities. As for bulk, several stress techniques can be used to compensate and boost the performance. Fig. 12 shows the drive improvement obtained by implementation of a compressively strained SiGe layer in the S/D regions of a p-type MuGFET, which leads both to a reduction in series resistance and hole mobility improvement [19]. Fig. 13 shows the improvement resulting from the combined use of Strained-Si-On-Insulator substrate (SSOI) and tensile CESL, for an n-type MuGFET [20].

References

[1] H. R. Huff et al., IEEE IWGI Tech. Digest, 2001, p. 2.
[2] C. Hobbs et al., IEEE VLSI Tech. Digest, 2003, p. 9.
[3] B. Yu et al., IEEE IEDM Tech. Digest, 2002, p. 251.
[4] S. C. Song et al., IEEE VLSI Tech. Digest, 2006, p. 16.
[5] J. A. Kittl et al., IEEE VLSI Tech. Digest, 2005, p. 72.
[6] A. Lauwers et al., IEEE IEDM Tech. Digest, 2005, p. 646.
[7] S. Yamaguchi et al., IEEE VLSI Tech. Digest, 2006, p. 192.
[8] T. Hoffmann et al., IEEE VLSI Tech. Digest, 2006, p. 194.
[9] M.A. Pawlak, et al., IEEE Electron Dev. Lett 27 (2) (2006) 99.
[10] J.A. Kittl, et al., IEEE Electron Dev. Lett 27 (1) (2006) 34.
[11] H. Yu et al., IEEE IEDM Tech. Digest, 2005, p. 630.
[12] A. Veloso et al., IEEE VLSI Tech. Digest, 2006, p. 116.
[13] T. Ghani et al., IEEE IEDM Tech. Digest, 2003, p. 978.
[14] P. R. Chidambaram et al., IEEE VLSI Tech. Digest, 2004, p. 48.
[15] F. Nouri et al., IEEE IEDM Tech. Digest, 2004, p. 1055.
[16] G. Eneman et al., IEEE VLSI Tech. Digest, 2005, p. 22.
[17] P. Verheyen et al., IEEE IEDM Tech. Digest, 2005, p. 907.
[18] P. Ranade et al., IEEE IEDM Tech. Digest, 2005, p. 227.
[19] P. Verheyen et al., IEEE VLSI Tech. Digest, 2005, p. 194.
[20] N. Collaert et al., IEEE VLSI Tech. Digest, 2006, p. 194.
[21] T. Watanabe et al., IEEE IEDM Tech. Digest, 2004, p. 507.
[22] C.-H. Jan et al., IEEE IEDM Tech. Digest, 2005, p. 65.
[23] M. Takayanagi et al., IEEE IEDM Tech. Digest, 2005, p. 903.