Performance Enhancements in Scaled Strained-SiGe pMOSFETs With HfSiO$_x$/TiSiN Gate Stacks

Olaiyiwola M. Alatise, Member, IEEE, Sarah H. Olsen, Nick E. B. Cowern, Anthony G. O’Neill, and Prashant Majhi

Abstract—The short-channel performance of compressively strained Si$_{0.77}$Ge$_{0.23}$ pMOSFETs with HfSiO$_x$/TiSiN gate stacks has been characterized alongside that of unstrained-Si pMOSFETs. Strained-SiGe devices exhibit 80% mobility enhancement compared with Si control devices at an effective vertical field of 1 MV·cm$^{-1}$. For the first time, the on-state drain-current enhancement of intrinsic strained-SiGe devices is shown to be approximately constant with scaling. Intrinsic strained-SiGe devices with 100-nm gate lengths exhibit 75% enhancement in maximum transconductance compared with Si control devices, using only ~20% Ge (~0.8% strain). The origin of the loss in performance enhancement commonly observed in strained-SiGe devices at short gate lengths is examined and found to be dominated by reduced boron diffusivity and increased parasitic series resistance in compressively strained SiGe devices compared with silicon control devices. The effective channel length was extracted from I–V measurements and was found to be 40% smaller in 100-nm silicon control devices than in SiGe devices having the same lithographic gate lengths, which is in good agreement with the metallurgical channel length predicted by TCAD process simulations. Self-heating due to the low thermal conductivity of SiGe is shown to have a negligible effect on the scaled-device performance. These findings demonstrate that the significant on-state performance gains of strained-SiGe pMOSFETs compared with bulk Si devices observed at long channel lengths are also obtainable in scaled devices if dopant diffusion, silicidation, and contact modules can be optimized for SiGe.

Index Terms—Dopant diffusion, high-k, metal gates, mobility, parasitic resistance, scaling, strained SiGe.

I. INTRODUCTION

A LOW spin-orbit split-off energy (44 meV) from valence-band degeneracy and a large hole effective mass compared with electrons contribute to the low hole mobility and poor performance of pMOSFETs in bulk silicon compared with nMOSFETs. Using strained silicon–germanium (SiGe) as the channel material has the potential as a major performance booster in pMOSFET devices due to increased hole mobility compared with bulk silicon [1]. However, the scalability of the performance gains induced by compressive strain has remained a concern. SiGe pMOSFETs were investigated in [2], and it was shown that the enhancement in maximum transconductance $g_{\text{MAX}}$ compared with that of bulk Si devices reduced from 50% in 2-µm-gate-length devices to 10% in 0.25-µm-gate-length devices. A similar observation was reported in [3], in which the strain-induced enhancement in $g_{\text{MAX}}$ was shown to reduce from 75% for 10-µm-gate-length devices to 10% for 0.15-µm-gate-length devices. The strained-Si$_{0.72}$Ge$_{0.28}$ pMOSFETs reported in [4] showed that the strain-induced enhancement in $g_{\text{MAX}}$ reduced from 30% for 1.3-µm-gate-length MOSFETs to 15% for 0.3-µm-gate-length MOSFETS. Strained-Si$_{0.72}$Ge$_{0.28}$ pMOSFETs with 85% hole-mobility enhancement were reported in [5], but the drain-current enhancement compared with that of Si control devices was only 55% for 10-µm-gate-length devices and reduced to 15% for 70-nm devices.

Realizing high-performance SiGe pMOSFETs is particularly important in advanced technologies that use heavy halo doping to control short-channel effects and high-k/metal-gate stacks to control gate leakage. Both heavy doping and high-k gate dielectrics reduce channel mobility compared with those that are obtainable in the conventional Si/SiO$_2$ system, so incorporating high-mobility channel materials such as SiGe becomes even more essential. Devices combining high-k gate dielectrics, metal gates, and compressively strained SiGe have consequently received a lot of attention [6]–[14]. In [8], compressively strained-Si$_{0.72}$Ge$_{0.28}$ pMOSFETs with HfO$_2$ gate dielectrics exhibited 65% peak-hole-mobility enhancement compared with Si control devices, but 180-nm-gate-length devices had only 35% drive-current enhancement. Compressively strained Si$_{0.72}$Ge$_{0.28}$ pMOSFETs with HfO$_2$ gate dielectrics and TiN gates in [10] exhibited 65% mobility enhancement compared with bulk Si control devices having the same gate stack. However, the 100% enhancement in drain current observed for 1-µm-gate-length devices was suppressed for 55-nm-gate-length devices when devices were compared at the same gate overdrive voltage. If compressively strained SiGe channels are to be deployed in deep-submicrometer CMOS technology, the reduced gains in drain current and transconductance due to scaling have to be understood and minimized. In this paper, the scalability of compressively strained SiGe pMOSFETs is investigated for devices fabricated with HfSiO$_x$/TiSiN gate stacks.

0018-9383/$26.00 © 2009 IEEE
Compressively strained SiGe was selectively grown on device active areas with a final thickness of approximately 40 nm. An average Ge composition of 23% in the strained-SiGe layer was determined by secondary-ion mass spectroscopy. Nitrided interfacial layers were used to improve interfacial properties between the high- \( k \) gate dielectric and the MOSFET channel [11], [15]–[18]. The interfacial layers were formed by rapid thermal oxidation, followed by nitridation that resulted in SiGe channels [20], [21]. Nitrided interfacial layers were used to improve interfacial properties [11], [15]–[18]. The interfacial layers were formed by rapid thermal oxidation, followed by nitridation that resulted in SiGe channels [20], [21].

The gate-capacitance and gate-conductance characteristics were measured on 100-\( \mu \)m-2-area MOS capacitors. The midgap \( D_{it} \) values for the Si control and strained-SiGe wafers were 8 \( \times \) 1011 and 3 \( \times \) 1012 \( \text{cm}^{-2} \cdot \text{eV}^{-1} \), respectively. The higher \( D_{it} \) in SiGe devices was expected due to the presence of Ge at the channel/dielectric interface [20], [21]. Nevertheless, both values are comparable with values reported in the literature [11], [22]. TiSiN gates were formed by sputtering. After gate definition, source–drain implants were formed by 10-keV B implantation with a dose of 1.4 \( \times \) 1015 \( \text{cm}^{-2} \). Halo doping at 45° was performed using As ion implantation at an energy of 50 keV and a dose of 6 \( \times \) 1013 \( \text{cm}^{-2} \). Sidewall spacers were subsequently formed, followed by deep source–drain implants using 20-keV B implantation at a dose of 4 \( \times \) 1015 \( \text{cm}^{-2} \). Dopant activation was carried out by rapid thermal annealing at 1000 °C. A self-aligned Ni silicidation process was performed by depositing Ni and annealing at 1000 °C for 30 s. A standard back-end-process completed the fabrication. Fig. 1 shows a TEM image of the processed device and the SiGe channel.

The uniformity of the Si control and SiGe wafers was evaluated by measuring all 1-\( \mu \)m-gate-length devices on both wafers. Fig. 2 shows the distribution of device performance on the wafers in terms of maximum transconductance (\( \mu_{\text{MAX}} \)) measured at 1-V drain voltage (\( V_{\text{DS}} \)). The results are identical for smaller drain biases. The Si control and SiGe wafers exhibited standard deviations of 3% and 5% of the median values of \( \mu_{\text{MAX}} \). Subsequent analysis was performed on median-performing dies for each wafer which are labeled in Fig. 2.

The split \( C−V \) technique with series-resistance correction was used to extract the effective mobilities of strained-SiGe and Si control devices. The inversion charge density was calculated from the integration of gate–channel capacitance, whereas the depletion charge density was calculated from the integration of gate–bulk capacitance [19]. The effective mobility was extracted from 1-\( \mu \)m-gate-length devices. Fig. 3 shows 80% effective-hole-mobility (\( \mu_{\text{EFF}} \)) enhancement for the strained-SiGe device compared with the Si control device at an effective vertical field (\( E_{\text{EFF}} \)) of 1 MV·cm\(^{-1} \). The mobility enhancement due to compressive strain overcomes any mobility reduction caused by the imperfect SiGe/SiO2 interface in the strained-SiGe device [8], [23]–[25], and there is 60% enhancement in the effective hole mobility compared with the universal mobility curve (Fig. 3) at an \( E_{\text{EFF}} \) of 1 MV·cm\(^{-1} \).

The hole-mobility enhancement is greater than that reported in [13] and [8] where Si0.7Ge0.3 and Si0.8Ge0.2 devices demonstrated 20% hole-mobility enhancement compared with the universal mobility curve at an \( E_{\text{EFF}} \) of 1 MV·cm\(^{-1} \). The mobility enhancement in this paper is also comparable with that reported in [10] for devices having higher Ge contents (28%). Carrier mobilities are affected by the quality of the interface.
ALATISE et al.: PERFORMANCE ENHANCEMENTS IN SCALED STRAINED-SiGe pMOSFETS WITH GATE STACKS

Fig. 4. (a) Drain-current output characteristics of 1-µm-gate-length pMOSFETs measured at gate overdrive voltages \( V_{GS} - V_{TH} \) of 0.5 and 1.0 V. At a drain voltage of 1 V for both gate overdrives, the drain current is increased by 75% for strained-SiGe pMOSFETs compared with Si control devices. (b) Drain-current output characteristics of 100-nm-gate-length pMOSFETs measured at gate overdrive voltages \( V_{GS} - V_{TH} \) of 0.5 and 1.0 V. The Si control device exhibits higher drain current than the strained-SiGe device.

between the gate dielectric and the channel, particularly for high-\( k \)/metal-gate systems. It is known that \( \mu_{EFF} \) can be reduced by increased scattering due to the interaction between the mobile carriers in the channel and the charged traps at the interface, as well as by surface roughness scattering at high vertical fields. Surface passivation and preparation techniques such as nitridation prior to dielectric deposition have been shown to improve \( \mu_{EFF} \) in high-\( k \)/metal-gate devices by reducing \( D_{it} \) [11], [15]–[18]. The inclusion of a nitridation step prior to ALD deposition of high-\( k \) has minimized the impact of \( D_{it} \) on mobility for the devices, as confirmed by \( D_{it} \) measurements on the devices.

Fig. 4 shows the drain current \( (I_{DS}) \) as a function of the drain voltage \( (V_{DS}) \) for 1-µm- and 100-nm-gate-length pMOSFETs at gate overdrive voltages \( V_{GS} - V_{TH} \) of 0.5 and 1.0 V. \( V_{GS} \) is the gate voltage, and \( V_{TH} \) is the threshold voltage. The gate-voltage overdrive is used for \( I_{DS} \) comparison because of the lower \( V_{TH} \) in strained-SiGe devices as a result of reduced bandgap due to the valence-band offset [14]. The \( V_{TH} \) difference between the devices reduces from 290 mV at \( L_{G} = 1 \mu m \) to 50 mV at \( L_{G} = 100 \) nm. This is because \( V_{TH} \) roll-off is evident in Si control devices \( (V_{TH(1\mu m)} = 0.71 \text{ V} \) and \( V_{TH(100 nm)} = 0.51 \text{ V} \)), whereas \( V_{TH} \) remains stable with \( L_{G} \) in strained-SiGe devices \( (V_{TH(1\mu m)} = 0.42 \text{ V} \) and \( V_{TH(100 nm)} = 0.45 \text{ V} \)). The output characteristics in Fig. 4 show that the 75% enhancement in saturation drain current for the 1-µm-gate-length strained-SiGe device compared with the Si control device is lost for 100-nm-gate-length devices. Fig. 5 shows the gate-transfer characteristics for the same devices. The subthreshold slopes are 75 mV/dec (Si) and 83 mV/dec (SiGe) for 1-µm-gate-length devices and 95 mV/dec (Si) and 83 mV/dec (SiGe) for 100-nm-gate-length devices.

Fig. 6 shows the variation in drain-induced barrier lowering (DIBL) with gate length for strained-SiGe and Si control devices. DIBL is 10 mV/V for both Si and strained-SiGe 1-µm-gate-length devices but increases more rapidly in scaled Si devices than in SiGe devices. For 100-nm-gate-length devices, the DIBL values are 70 mV/V for Si pMOSFETs and 45 mV/V for SiGe pMOSFETs.

The maximum transconductance measured in SiGe and Si control devices at a drain voltage of 1 V are presented for a range of lithographic gate lengths \( (L_G)^{’s} \) in Fig. 7(a). Performance enhancements for SiGe devices are shown in Fig. 7(b) and have been shown to reach 80% compared with that for Si control devices. In agreement with other reports [2]–[5], [11], Fig. 7(b) shows that the enhancements in \( g_{MAX} \) for strained-SiGe devices decrease as the lithographic gate length is reduced. For \( L_G \) that is below 250 nm, no enhancement is evident, and for 100-nm-gate-length devices, the Si control outperforms the SiGe devices by approximately 20%.
Technological nodes, the graphic gate lengths are diminished at gate lengths that are below devices compared with that of Si devices that are evident at large lithographic gate lengths for Si and SiGe devices. The performance gains of SiGe devices were confirmed using the TCAD process simulator TSUPREM4. The boron implant dose, implant energies, arsenic halo implant energies, tilt angles, doses, and activation anneal temperature–time cycles used in TSUPREM process simulation were identical to those used in device fabrication. Diffusivity data were taken from [28], and strain was calculated as $(1 - a_{Ge}/a_{Si})x$, where $a_{Ge}$ is the lattice constant of Ge, $a_{Si}$ is the lattice constant of Si, and $x$ is the Ge mole fraction in the SiGe layer. Raman spectroscopy showed that the average compressive strain in the SiGe channel was about 0.8%, which is close to the theoretical strain value for 20% Ge. The diffusivity of boron in compressively strained Si$_{0.77}$Ge$_{0.23}$ is $\sim$0.2 $D_{O}$, where $D_{O}$ is the diffusivity of boron in unstrained Si [28]. As a first approximation, taking $\Delta L \sim \sqrt{Dt}$, where $D$ is the dopant diffusivity and $t$ is the diffusion time, boron will diffuse approximately 55% less in SiGe devices compared with Si control devices. This difference in diffusion leads to a shorter metallurgical channel length $L_{MET}$ in bulk Si devices. The metallurgical channel length $L_{MET}$ is defined as the lateral distance between the source and drain over which the substrate arsenic doping is higher than the source/drain doping. $L_{MET}$ correlates with the effective channel length and the lithographic gate length. Fig. 8(a) shows a 2-D profile of B doping contours in a simulated 100-nm processed device, whereas Fig. 8(b) shows the lateral cross section of B between the source and drain 5 nm below the MOSFET surface. The metallurgical channel length is extracted at this position and is found to be approximately 30 nm for the Si device and 40 nm for the SiGe device. The effective channel lengths extracted from $I$–$V$ data are larger than the metallurgical channel length predicted by TCAD simulation. This is expected because the lateral straggle of the junction implants [29] and nonabrupt source–drain junction profiles effectively shortens the chemical length between the source and the drain. For an ideal junction profile (infinitely abrupt with no lateral straggle), $L_{MET}$ is larger than $L_{EFF}$ due to the sheet resistivity only being modulated by the gate voltage inside the metallurgical channel. However, for a nonabrupt junction profile, there is a fraction of the channel with the junction implant straggle in accumulation, thereby causing $L_{EFF}$ to be larger than $L_{MET}$.

Both experimental methods and TCAD process simulations show a shorter effective and metallurgical channel length for Si devices compared with SiGe devices due to reduced boron diffusivity in compressively strained SiGe. These results also explain the increased $V_{TH}$ roll-off for Si devices and Fig. 6, where
DIBL is seen to be lower in short-channel strained-SiGe devices than in Si control devices. Since DIBL relates to $L_{EFF}$ through a negative exponential, a small difference in $L_{EFF}$ causes a significant difference in DIBL [30]. The lower DIBL for strained-SiGe devices therefore correlates, as expected, with the stable $V_{TH}$ roll-off, which is usually due to DIBL. Hence, the reduced B diffusion in compressively strained SiGe results in longer effective channel lengths and better electrostatic integrity.

The $g_{m\text{MAX}}$ data in Fig. 7 are reevaluated using the effective channel length calculated by the “shift-and-ratio” method and are shown in Fig. 9. Using the effective channel length demonstrates that strained-SiGe devices can offer performance enhancements for all channel lengths that are greater than 150 nm. In contrast, when the devices were analyzed in terms of their lithographic gate length (Fig. 7), performance gains only appeared possible in SiGe devices if the gate length exceeded 175 nm.

Fig. 9 shows that, for effective channel lengths that are below 150 nm, Si control devices still outperform strained-SiGe devices. The series resistance ($R_{SD}$) was also extracted using the “shift-and-ratio” method [29] and was found to be 60% higher in SiGe devices compared with Si control devices (80 $\Omega$ compared with 50 $\Omega$). The higher $R_{SD}$ in strained-SiGe devices was confirmed by silicide-sheet-resistance ($R_{SH}$) measurements on test structures, which showed that $R_{SH}$ was approximately 100% higher in SiGe devices (8.2 $\Omega$/□ in SiGe and 4.2 $\Omega$/□ in Si control). Since the silicide anneal process was optimized for bulk Si and the pMOSFETs have a 40-nm SiGe surface channel layer, high-resistance nickel germanosilicides will have formed due to the presence of Ge [31]–[33]. It was shown in [32] that the $R_{SH}$ of Ni-silicided Si$_{0.75}$Ge$_{0.25}$ was 3 $\Omega$/□ at 500 °C but increased abruptly at anneal temperatures above 800 °C. This increase was attributed to the segregation of Ge at the grain boundaries of nickel germanosilicide during the interfacial reactions between nickel and SiGe. In [33], nickel-silicided Si$_{0.7}$Ge$_{0.3}$ and Si$_{0.8}$Ge$_{0.2}$ films showed minimum $R_{SH}$ of 3.9 and 3.5 $\Omega$/□, respectively, at 400 °C. These values are 50% lower than the $R_{SH}$ on the SiGe wafers in this paper, and this is likely to be due to the higher annealing temperature used (1000 °C). However, under appropriate annealing conditions, NiSiGe can offer improved $R_{SH}$. NiSiGe can also improve contact resistance because of reduced barrier height and higher boron activation compared with NiSi [34], [35].

Process optimization of the silicidation temperature–time cycle for SiGe alloys can prevent the formation of such
high-resistance films, so it is valid to investigate the intrinsic device performance without parasitic resistances. The intrinsic drain current was calculated by correcting for the series resistance using the formula

\[ I_{DS0} = I_{DS} (1 - I_{DS} R_{SD}/V_{DS})^{-1} \]  

where \( I_{DS0} \) is the intrinsic drain current. Equation (1) is derived from the strong-inversion MOSFET square-law model taking source/drain series resistance into account [19], [36]–[38].

The resulting intrinsic maximum transconductance data are shown in Fig. 10(a). The ON-state performance of strained-SiGe devices is now found to be improved compared with Si control devices down to effective gate lengths of 100 nm. The difference between the intrinsic and measured transconductances increases as the gate length reduces due to the increasing impact of series parasitic resistance, which becomes a larger proportion of the total channel resistance in scaled geometries. Comparing the intrinsic \( g_{mMAX} \) in Fig. 10(a) with the measured \( g_{mMAX} \) in Fig. 9(a) shows that approximately 50% of \( g_{mMAX} \) is lost in scaled SiGe pMOSFETs. Fig. 10(b) shows that the intrinsic enhancement in \( g_{mMAX} \) of SiGe devices reaches 80% at long channel lengths and reduces by only 15% as the effective channel length is scaled from 1 \( \mu \)m to 100 nm. The 65% enhancement in \( g_{mMAX} \) at short channel lengths is the highest reported to date for SiGe devices using low Ge contents (∼20%).

In [10], no enhancement in drain current compared with Si control devices was reported for 55-nm strained-Si\(_{0.7}\)Ge\(_{0.28}\) pMOSFETs, and only 15% enhancement in linear transconductance compared with Si control devices was achieved in 130-nm Si\(_{0.68}\)Ge\(_{0.32}\) pMOSFETs [3]. In [39], 50-nm-gate-length strained-Si\(_{0.65}\)Ge\(_{0.35}\) pMOSFETs exhibited 35% drive-current enhancement compared with Si control devices, and in [5], 13% drive current enhancement compared with Si control devices was reported in 50-nm strained-Si\(_{0.7}\)Ge\(_{0.28}\) devices. Our new results suggest that significantly larger performance gains in strained-SiGe pMOSFETs are realizable in deep-submicrometer CMOS technology nodes than previously demonstrated if processing is optimized to take account of the modified dopant diffusion and parasitic series resistance in the SiGe material system.

Fig. 10(b) shows that the intrinsic performance enhancement of short-channel strained-SiGe pMOSFETs compared with bulk Si is 15% lower than observed in long-channel devices. This reduction may be due to self-heating arising from the low thermal conductivity of SiGe [40], strain loss with scaling, or the increased impact of halo doping on mobility at short gate lengths. One of the dominating factors behind the compromised drain current and transconductance enhancement in scaled strained-Si devices fabricated on relaxed SiGe virtual substrates is self-heating [41]. The devices in this paper comprise a 40-nm SiGe surface layer on a Si substrate rather than a thin Si layer on a thick SiGe layer, so the impact of self-heating is expected to be considerably lower. AC-drain-conductance measurements, which remove self-heating effects [42], were carried out on 100-nm strained-SiGe devices. Fig. 11 shows that there is an increase in drain current compared with dc conditions of only ∼3% for the SiGe device when measurements were carried out at 10 MHz. Therefore, self-heating is not a performance-limiting mechanism in scaled strained-SiGe pMOSFET devices.

Potential variations in mobility with gate length may also impact performance gains in scaled devices. Unintentional fluctuations in channel strain with gate-length scaling due to stresses induced during epitaxial growth, from the silicide and
trench isolation may counteract the intentional channel strain. Strain relaxation in SiGe layers at the source/drain regions may also result from defects caused by ion implantation. These defects, together with the loss of strain, can cause additional carrier scattering, thereby contributing to mobility reduction. Electrically, the impact of these defects and strain relaxation will be manifested as increased series resistance, and its relative impact would also increase as the gate length is scaled. Since strain relaxation at nanoscale dimensions cannot be characterized electrically, it is not possible to accurately separate its effect from that of increased series resistance due to silicide sheet resistance. At present, the limited availability of nanoscale strain characterization techniques prevents a full understanding of the impact of various strain contributions in deep-submicrometer CMOS.

IV. Summary

The scalability of high-performance strained-SiGe pMOSFETs with HfSiOx/TiSiN gate stacks has been examined. The intrinsic performance of short-channel SiGe devices has been shown to exceed that of coprocessed Si control devices by as much as 65% in 100-nm-gate-length devices, whereas extrinsic gains are completely diluted for drawn gate lengths that are below 175 nm. Performance enhancements exceeding 70% in both long- and short-channel devices are the highest reported gains to date compared with bulk Si using just 20% Ge in the channel region. The dominating factors behind the compromised performance gains observed in the measured short-channel strained-SiGe devices are found to be differences in the effective channel length due to strain-altered dopant diffusion and increased parasitic series resistance. Self-heating due to the low thermal conductivity of SiGe was analyzed and shown to have a smaller impact on scaled device performance (~3%). The intrinsic potential of devices was assessed by correcting for the differing effective channel lengths and source–drain series resistance. After correction, 65% of the long-channel strain-induced performance enhancement was maintained at scaled geometries. This paper suggests a major underestimation of the potential of SiGe for advanced technology nodes. The effective channel length of SiGe devices was 40% larger than that of Si control devices for 100-nm lithographic gate lengths due to suppressed boron diffusivity from the source/drain regions in SiGe. This difference was confirmed by TCAD simulations that showed a 30% increase in the metallurgical channel length for 100-nm-gate-length SiGe devices. The source–drain series resistance was also 70% higher in SiGe devices due to silicidation being optimized for bulk Si. While electrical results are convincing, direct comparisons between \( L_{\text{EFF}} \) and \( R_{\text{SD}} \)-matched strained-SiGe and strained-Si pMOSFETs would be even more reliable in evaluating the scalability of performance enhancements in compressively strained SiGe devices. However, this comparison is only possible if the respective thermal processes of devices are customized. This paper shows that strained-SiGe pMOSFETs are scalable and are suitable for deep-submicrometer CMOS technology nodes if series resistance and dopant diffusion can be controlled.

References


