Simulation Study of Coulomb Mobility in Strained Silicon

Francesco Driussi and David Esseni, Senior Member, IEEE

Abstract—This paper presents a detailed simulation analysis aimed at assessing and explaining the dependence on the biaxial strain of the Coulomb-limited mobility in n-type silicon MOSFETs. By using a model based on the momentum relaxation time (MRT) approximation, we first show that we can reproduce fairly well a wide set of published experimental data, and then, we use our model to discuss the dependence on the strain of the mobility limited by either interface states or substrate impurities. Different from the experiments, in the simulations, the MRT approach allows us to analyze the different mobility components without resorting to the Matthiessen’s rule, whose use may result in large errors in the extracted mobility components. Our simulations indicate that the interface-state-limited mobility is reduced in strained devices; this is in qualitative agreement with the experiments, and we discuss its interpretation in terms of physically transparent arguments. Our analysis also suggests that the strain-induced changes of the substrate-impurity-limited mobility are instead very small, and we provide a clear interpretation of such a result. Recent experiments, however, have reported a strain-induced improvement of the substrate-impurity-limited mobility, which has been unavoidable extracted by using the Matthiessen’s rule. We argue that the systematic errors produced by the Matthiessen’s rule can help reconcile the simulation and the experimental results.

Index Terms—Coulomb mobility, interface states, Matthiessen’s rule, modeling, scattering, strained Si (s-Si), substrate impurities.

I. INTRODUCTION

WITH THE geometrical scaling that is not as rewarding as in the past in terms of performance improvements, the strained-Si (s-Si) technology has become one of the most important technology boosters. In this respect, s-Si n-type MOSFETs have already demonstrated drain-current enhancements of 20%–30% for sub-50-nm channel lengths [1]–[3], where a tight link still exists between the drain current and the low field mobility [4], [5]. This clear correlation between the strain-induced mobility and $I_{DS}$ enhancements has also been predicted by numerical simulations [6]–[9] and observed in experimental results [1]–[3], [7], [10].

Since the on-current is evaluated at $V_{GS} = V_{DS} = V_{DD}$, much work has been devoted to investigate the effects of the strain on the mobility at high inversion densities and effective fields, where the mobility is essentially limited by the surface-roughness and the phonon scattering mechanisms [7], [8], [11]–[14]. Nevertheless, it is very important to analyze in detail the effect of the strain engineering on the Coulomb-scattering (CS)-limited mobility for several reasons. In fact, the Coulomb or neutral defects are indicated as the possible culprits for the mobility degradation of very short MOSFETs [15], and furthermore, the importance of the CS is unfortunately emphasized in high-$k$ CMOS technologies [16]. Finally, for the Coulomb-limited mobility in s-Si MOSFETs, some inconsistencies between the theoretical predictions and the experiments have been recently pointed out and demand a clarification.

As for the latter point, the behavior of the mobility $\mu_{it}$ and $\mu_{sub}$ limited by interface-state or substrate-impurity scattering, respectively, has been investigated in [17]–[21]. The Monte Carlo simulations reported in [17] suggested that $\mu_{sub}$ should stay essentially the same while $\mu_{it}$ should be improved in the presence of biaxial strain. An insensitivity of $\mu_{sub}$ to the strain has been actually inferred from experiments in [18] (with no explicit discussion for $\mu_{it}$), whereas the recent experiments in [20] and [21] indicate that $\mu_{it}$ is reduced and that $\mu_{sub}$ is improved by the strain. The recent data reported in [19] show a $\mu_{sub}$ modulation even for the case of uniaxial stress. It is worth noticing that the $\mu_{it}$ and $\mu_{sub}$ values have been extracted by using the Matthiessen’s rule in all the aforementioned studies [17], [19]–[21]. The possible inaccuracies related to the Matthiessen’s rule have been pointed out a long time ago [22], [23] and critically reconsidered in [21].

In this paper, the $\mu_{it}$ and $\mu_{sub}$ in biaxially strained n-MOSFETs have been analyzed in detail by means of numerical simulations based on the momentum relaxation time (MRT) approximation, which allows us to determine $\mu_{it}$ and $\mu_{sub}$ without resorting to the Matthiessen’s rule. Starting from a quantitative comparison with the recent experimental data, the mobility is studied for different dopant concentrations, magnitudes of the biaxial strain, and operating temperatures. Our results indicate that, throughout the studied range of parameters, the strain degrades the $\mu_{it}$, whereas it hardly affects the $\mu_{sub}$. All the results are explained by presenting a systematic discussion of the strain-induced subband repopulation and of the features of the CS matrix elements and relaxation times.

II. DESCRIPTION OF THE MODEL

The model for the mobility calculation has been described in detail in previous publications [24]–[26]. The intravalley acoustic phonon scattering is included according to the elastic equipartition energy approximation, which makes the scattering mechanism isotropic [25]. The intervalley phonons are
A. Scattering With Coulomb Centers

The determination of CS mobility in our model consists of two steps. We first calculate the scattering potential produced by the point charge by accounting for the screening produced by the electrons in the inversion layer, and then, we determine the MRT and the mobility by accounting for the spatial distribution of the scattering centers [24].

Let us indicate with \( \mathbf{R} = (x, y, z) \) the total real space vector, where \( x = (x, y) \) is the position in the transport plane and \( z \) is the abscissa in the quantization direction, which is positive in silicon and has the origin at the silicon-oxide interface. Hereafter, \( \phi(q, z, z_0) \) denotes the 2-D Fourier component (calculated in the \( x y \) plane) of the CS potential produced by a point charge located at \((0, 0, z_0)\), which is a function of \( z \) and of the magnitude \( q \) of the 2-D wave vector \( \mathbf{q} \). For any given \( z_0 \) and \( q \) values, the \( \phi(q, z, z_0) \) profile accounting for the screening effect of the electrons in the inversion layer was obtained by solving the integral equation stemming from the perturbative approach developed by Stern and Howard [29]. More details concerning the calculation procedure can be found in [24].

Once the screened potential \( \phi(q, z, z_0) \) of the point charge has been determined, the matrix element \( M_{i,j}^{(0)}(q, z_0) \) between the \( i \)th and the \( j \)th subband can be expressed as [24], [30]–[32]

\[
M_{i,j}(q, z_0) = \frac{e}{A} \int dz \phi(q, z, z_0) \xi_i(z) \xi_j(z) \tag{1}
\]

where \( \xi_i(z) \) is the envelope wave function and \( A \) is just the normalization area in the transport plane. All the calculations were performed by using the eigenvalues and eigenfunctions obtained with the self-consistent Schrödinger–Poisson solver described in [33].

The overall squared matrix elements for the CS are finally obtained by summing the product of \( |M_{i,j}(q, z_0)|^2 \) times the appropriate density of Coulomb centers at \( z_0 \) [24], [30]–[32]. The scattering rate is then obtained by using Fermi’s golden rule.

All the mobility calculations presented in this paper have been obtained by accounting for the screening; however, in order to introduce simple expressions for the scattering potential useful for the physical insight, we notice that the unscreened scattering potential \( \phi_{\text{unscr}}(q, z, z_0) \) in silicon (i.e., for \( z > 0 \)) is given by [34]

\[
\phi_{\text{unscr}}(q, z, z_0) = \frac{e}{2q\epsilon_{\text{si}}} e^{-q|z-z_0|} + \frac{\epsilon_{\text{si}} - \epsilon_{\text{ox}}}{\epsilon_{\text{si}} + \epsilon_{\text{ox}}} \frac{e}{2q\epsilon_{\text{si}}} e^{-q(z+|z_0|)} \tag{2}
\]

where \( e \) is the positive electron charge, while \( \epsilon_{\text{si}} \) and \( \epsilon_{\text{ox}} \) denote the oxide and silicon dielectric constants, respectively. Equation (2) is valid for any value of \( z_0 \), and in particular, for the Coulomb centers located at the interface (i.e., \( z_0 = 0 \)), we have

\[
\phi_{\text{unscr}}(q, z, 0) = \frac{e}{q(\epsilon_{\text{si}} + \epsilon_{\text{ox}})} e^{-q|z|}. \tag{3}
\]

The qualitative \( z \) and \( z_0 \) dependence of the screened potential \( \phi(q, z, z_0) \) is very similar to that of the unscreened potential \( \phi_{\text{unscr}}(q, z, z_0) \).

B. Calculation of the Relaxation Time and the Mobility

As for the calculation of the MRT \( \tau_m \), we first notice that, in our calculations, we embraced a parabolic isotropic electron energy dispersion. Within the parabolic band approximation, the isotropic dispersion is correct for the unprimed subbands (which have a quantization mass \( m_z = 0.916 m_0 \) and a density of state mass \( m_d = \sqrt{m_x m_y} = 0.19 m_0 \)), whereas it neglects the anisotropy of the primed subbands (which have \( m_z = 0.19 m_0 \) and \( m_d \approx 0.417 m_0 \)) [35]. The unprimed subbands are two-time degenerate and are correspondingly denoted as \( \Delta_2 \); similarly, the primed subbands are denoted as \( \Delta_4 \) because they are four-time degenerate. The approximation of the isotropic energy dispersion is not expected to introduce significant errors in the determination of the low field mobility in silicon inversion layers [36], and it is very useful in the calculation of MRTs.

Our model employs an isotropic formulation of the phonon scattering mechanisms, so that an explicit expression for the relaxation time \( \tau_m \) is easily obtained [24]. As far as the anisotropic scattering mechanisms, the expression for the relaxation time \( \tau_m \) of the intersubband transitions in the subband \( i \) is given by

\[
\frac{1}{\tau_{m,i}(E)} = \frac{m_i^{(i)}}{\hbar^3} \int_0^\pi d\theta |M_{i,j}|^2(q)(1 - \cos \theta). \tag{4}
\]

Equation (4) has been very frequently used in the literature [31], and \( \theta \) denotes the angle between the before and after scattering wave vector, so that the magnitude \( q \) of the exchanged wave vector is given by \( q = 2k \sin(\theta/2) \). In the isotropic energy model, the magnitude \( k \) of the wave vector is \( k = \sqrt{2m_d (E - \varepsilon_i)}/\hbar \), where \( \varepsilon_i \) is the subband bottom energy and \( (E - \varepsilon_i) \) is the kinetic energy in the subband.

The calculation of the relaxation time is much more complicated for intersubband transitions. In fact, in this case, one cannot find an explicit expression for the relaxation time in a given subband, but rather an implicit definition that links the relaxation time of the different subbands [24], [36]. As discussed in detail in [24], since our model uses parabolic isotropic bands, the relaxation times are independent of the direction of the wave vector \( \mathbf{k} \). This simplification allows us to calculate the relaxation times for Coulomb and surface-roughness scattering by appropriately accounting for the coupling between the \( \tau_m \) values of the different subbands produced by the intersubband transitions [24].

When different scattering mechanisms are included in the calculations, the total MRT \( \tau_{m,i}^{(\text{tot})} \) is obtained by summing the
Inverses of the MRTs. Then, for a parabolic energy model, the average value of the relaxation time necessary for the mobility calculation is simply given by [31]

\[ \langle \tau_{m,i} \rangle = \frac{\int_{E_i}^{\infty} (E - E_i) \tau_{m,i}^{(tot)}(E) f_0(E)[1 - f_0(E)] dE}{K_B T \int_{E_i}^{\infty} f_0(E) dE} \]  

(5)

where \( f_0(E) \) is the equilibrium Fermi–Dirac occupation function.

The mobility \( \mu_i \) in the subband \( i \) and the effective mobility \( \mu_{eff} \) in the inversion layer are finally expressed as

\[ \mu_i = \frac{e \langle \tau_{m,i} \rangle}{m_{c,i}} \quad \mu_{eff} = \frac{1}{N_{inv}} \sum_i N_i \mu_i \]  

(6)

where \( m_{c,i} \) denotes the effective conduction mass in the plane of transport, while \( N_i \) and \( N_{inv} \) denote the inversion density in the subband \( i \) and in the total inversion layer, respectively. The effective mobility \( \mu_{eff} \) is thus the average of the subband mobilities weighted by their corresponding electron densities.

It is understood that, in (6), the sum over the subband \( i \) includes both the unprimed and the primed subbands; for simplicity of notation, we did not introduce an index for the valleys. The effective conduction mass is given by \( m_c = 2(m_{c,1}^{-1} + m_{c,2}^{-1})^{-1} \), and it evaluates to 0.19 \( m_0 \) for the \( \Delta_2 \), i.e., the unprimed subbands, and to 0.315 \( m_0 \) for the \( \Delta_4 \), i.e., the primed subbands.

By including one single scattering mechanism, (5) and (6) can be naturally used to calculate the mobility limited by the single mechanism at the study. This is in fact how we calculated the \( \mu_{it} \) and \( \mu_{ab} \) values discussed in Section IV, without resorting to the Matthiessen’s rule.

III. COMPARISON BETWEEN SIMULATIONS AND EXPERIMENTS

In Fig. 1, the model previously described is used to analyze the experimental mobility data for relaxed Si and s-Si reported in [20] and [21]. In order to simulate the mobility of s-Si devices, we followed the same procedure used in [6] and [7], where the valley splitting induced by strain is calculated as a function of the Ge content of the SiGe virtual substrate [7, eq. (6)], and the rms value of the surface-roughness spectrum (\( \Delta_{SR} \)) has been reduced to fit the mobility data at large effective fields (\( E_{eff} \)). The other model parameters have the same values as in [7, Table 1].

Fig. 1(a) shows that our simulations (symbols) well reproduce the experimental \( \mu_{eff} \) values (lines) for different substrate doping concentrations ranging from \( N_{inv} = 3 \times 10^{16} \) to \( 2.2 \times 10^{18} \) cm\(^{-3} \) (doping values are taken from [20] and [21]).

Fig. 1(b) shows the experimental mobility data (taken from [20, Fig. 6]) for relaxed and strained transistors before and after a gate current stress that is able to produce a significant amount of surface interface states. The stress clearly reduces the experimental mobility in the case of both relaxed Si and s-Si. The stress-induced mobility degradation is reproduced fairly well by the simulations by using interface-state densities of \( N_{it} = 5 \times 10^{10} \) eV\(^{-1}\)cm\(^{-2} \) and \( N_{it} = 5 \times 10^{11} \) eV\(^{-1}\)cm\(^{-2} \) for the virgin and the stressed devices, respectively. In this case, \( N_{it} \) has been used as a fitting parameter. The same couple of \( N_{it} \) values is used for both the relaxed-Si and the s-Si cases.

It is worth pointing out that the simulated strain-induced mobility improvements at large \( N_{inv} \)’s in Fig. 1 are essentially due to the smaller \( \Delta_{SR} \) value used for surface-roughness scattering in s-Si [6], [7]. However, the discussion of the \( \mu_{it} \) and \( \mu_{ab} \) carried out in the following sections is totally independent of the phonon-limited or the surface-roughness-limited mobilities, because \( \mu_{it} \) and \( \mu_{ab} \) are directly calculated with the MRT approach, without resorting to the Matthiessen’s rule.

IV. ANALYSIS OF THE COULOMB-LIMITED MOBILITY

The previous section illustrated a good consistency between our simulations and the experiments for both the relaxed and the strained cases and for different channel-doping and interface-state concentrations. Such an agreement legitimates the use of our model to analyze how the strain affects the Coulomb-limited mobility, even because, owing to the MRT approach, it is possible to directly determine the \( \mu_{it} \) and \( \mu_{ab} \) values, without resorting to the Matthiessen’s rule (whose reliability will be further discussed in Section V).
A. Interface-State-Limited Mobility

Fig. 2 shows the simulated $\mu_{\text{hit}}$ for a small channel-doping concentration ($N_A = 10^{16}$ cm$^{-3}$) for the relaxed as well as different strain conditions. We observe a systematic reduction of the $\mu_{\text{hit}}$ with the strain, which is consistent with the experimental data of [21].

In order to explain the results in Fig. 2, Fig. 3(a) shows, for $N_{\text{inv}} = 10^{12}$ cm$^{-2}$, the $\mu_{\text{hit}}$ mobility for the two lowest $\Delta_2$ unprimed subbands and the two lowest $\Delta_4$ primed subbands versus the Ge content of the virtual substrate. As it can be seen, for all the strain conditions, the mobility is significantly smaller in the $\Delta_2$ than in the $\Delta_4$ subbands, and it is the smallest in the lowest $\Delta_2$ subband. This result may seem surprising since the $\Delta_2$ subbands have a smaller conduction mass; however, it will be convincingly explained later on in this section.

The changes produced by the strain in the $\mu_{\text{hit}}$ of the subbands are relatively modest compared with the $\mu_{\text{hit}}$ differences among the subbands. The subband occupation, instead, is remarkably varied by the strain. In this respect, Fig. 3(b) shows that, in the relaxed case, the subband carrying the largest $N_{\text{inv}}$ is the lowest $\Delta_4$ subband rather than the lowest $\Delta_2$ subband [13], [14].

This is because, at the relatively low inversion density $N_{\text{inv}} = 10^{12}$ cm$^{-2}$, the lower occupation of the $\Delta_4$ due to the larger eigenvalue is more than compensated by the almost four times larger density of the states. The strain-induced splitting between the $\Delta_2$ and the $\Delta_4$ subbands drastically reduces the occupation of the $\Delta_4$ subbands, and Fig. 3(b) shows that the relative population in the $\Delta_4$ subbands becomes vanishing for a progressively larger Ge content.

Returning to the $\mu_{\text{hit}}$ values in the $\Delta_2$ and $\Delta_4$ subbands, Fig. 4 shows the explanation why $\mu_{\text{hit}}$ is the smallest in the lowest $\Delta_2$ subband. The figure shows the scattering potential $\phi(q, z, 0)$ produced by a point charge located at the silicon-oxide interface for a wave vector $q = 0.2$ nm$^{-1}$ comparable with the electron thermal wave vector $k_{\text{th}} = \sqrt{2m_e k_B T/\hbar}$; for the $\Delta_2$ subbands, we have $m_e = 0.19m_0$; hence, $k_{\text{th}} \approx 0.36$ nm$^{-1}$. In the same figure, we also see the squared magnitude of the wave functions for the $\Delta_2$ and the $\Delta_4$ lowest subbands obtained with the Schrödinger–Poisson solver [33]. By recalling the formulation of the scattering matrix element given in (1), Fig. 4 shows very neatly that the lowest $\Delta_2$ has the largest matrix element with the scattering potential produced by an interface charge (as confirmed by the numerical calculations—not shown). In fact, the large quantization mass results in a stronger confinement of the wave function toward the silicon-oxide interface.

The proximity of the wave function to the interface is critical for the scattering with the interface states, and therefore, the $\Delta_2$ subbands are unfavorable with respect to the $\Delta_4$ subbands. Such an effect results in a smaller $\mu_{\text{hit}}$ for the $\Delta_2$ subbands despite their smaller effective transport mass $m_c$.

The fact that the strain-induced $\mu_{\text{hit}}$ degradation is essentially due to the repopulation of the $\Delta_2$ subbands is further confirmed by the mobility behavior as a function of the temperature, as
temperature-induced repopulation of the lowest degradation with decreasing temperatures is again due to the shown in Fig. 5. In fact, the simulations in Fig. 5(a) reveal (a) of inversion density Fig. 6. Simulated mobility limited by substrate impurities as a function of temperature. At low T, essentially all the electrons belong to the Δ₂ subband in both relaxed Si and s-Si. what was observed for the μₜᵢ in Fig. 3; however, the μₚₛₛ difference between the lowest Δ₂ and the lowest Δ₄ subbands is small for Nᵢᵥᵣ below approximately 10¹² cm⁻². Fig. 7(a) also shows that the μₛₛ in the subbands is not much affected by the strain. The aforementioned considerations explain why the strain-induced subband repopulation shown in Fig. 7(b) has an overall modest effect on the μₛₛ. Moreover, since, in the relaxed Si, only approximately 30% of the inversion charge is in the lowest Δ₂ subband (a smaller fraction with respect to Fig. 3 due to the larger doping concentration and the correspondingly larger confining electric field), this effect of the repopulation is further reduced [14]. Therefore, as shown in Fig. 6, the μₛₛ is very similar in the relaxed-Si and s-Si devices.

B. Substrate-Impurity-Limited Mobility

Fig. 6 shows that the simulated μₛₛ is very similar in the strained and unstrained devices for both different inversion densities and different doping concentrations. The μₛₛ is slightly larger in the strained case, but the difference is very small. We also verified that this result holds for different strain magnitudes, i.e., for different Ge contents in the virtual substrate (not shown).

To understand this result, also for the μₛₛ, we performed an analysis based on the mobility in the different subbands and on the subband repopulation. In this respect, Fig. 7(a) shows that the unstrained μₛₛ (filled symbols) is somewhat larger for the lowest Δ₂ than it is for the lowest Δ₄ subband. This is an opposite behavior with respect to what was observed for the μₜᵢ in Fig. 3; however, the μₛₛ difference between the lowest Δ₂ and the lowest Δ₄ subbands is small for Nᵢᵥᵣ below approximately 10¹² cm⁻². The Δ₄ subbands feature a smaller effective conduction mass mₑ. Such an mₑ difference is the main explanation for the larger phonon-limited mobility in the Δ₂ subbands discussed in detail in [25]. However, the μₛₛ advantage for the Δ₂ subbands shown in Fig. 7(a) is small compared with the phonon-limited mobility. This is a delicate point for the interpretation of the μₛₛ results, because, if the advantage of the lowest Δ₂ subband were larger, then the strain-induced subband repopulation could result in a more appreciable mobility improvement than that shown in Fig. 6.

In order to further analyze this latter point, we return to the main constituents of the Coulomb-limited mobility discussed in Section II. The form of the CS potential suggests that the matrix elements are larger for smaller magnitudes q of the exchanged wave vector. As already said in the discussion of (4), for an intrasubband transition, we have

\[ q = \frac{2\sqrt{2m_e(E - \varepsilon_i)}}{\hbar} \sin(\theta/2) \]
Fig. 8. Calculated MRT $\tau_m$ for the lowest $\Delta_2$ and the lowest $\Delta_4$ subband as a function of the kinetic energy $E_k$. At the thermal energy $E_T = K_B T$, the $\tau_m$ limited by substrate impurities is larger for the $\Delta_2$ valleys compared with the $\Delta_4$ valleys.

10^{12} \text{ cm}^{-2}$, the electron gas is not appreciably degenerate, and the average energy in each subband is $K_B T$, so that (7) indicates that, statistically speaking, the $q$ values are smaller in the $\Delta_2$ subbands than in the $\Delta_4$ subbands, because the former have a smaller value for $m_d$ (i.e., 0.19 $m_0$ instead of 0.417 $m_0$).

The picture at small $q$ values is complicated by the screening, which tends to reduce the scattering rates particularly at small $q$ values [24]. However, the scattering events with small $q$ and, thus, small $q$ values are less effective in the relaxation of the momentum, as clearly pointed out by (4), so that, overall, the smaller $m_d$ of the $\Delta_2$ subbands is expected to yield smaller relaxation times, at least in the energy range of most practical importance, namely, around $K_B T$.

The reduction in the $\tau_m$ values of the $\Delta_2$ lowest subband is confirmed by the numerical calculations performed at room temperature and shown in Fig. 8. These smaller $\tau_m$ values tend to counteract the better effective conduction mass of the $\Delta_2$ with respect to the $\Delta_4$ subbands. Such an antagonist effect with respect to the conduction mass is responsible for the modest improvement of the $\mu_{\text{sub}}$ in $\Delta_2$ with respect to $\Delta_4$ subbands, which, in turn, results in a very modest strain-induced mobility increase.

V. DISCUSSION AND CONCLUSION

The simulated results in Fig. 2 are consistent with the experiments of [21] in showing a $\mu_{\text{it}}$ degradation in strained devices, and this conclusion cannot be disputed on the basis of the possible inaccuracies in the experimental $\mu_{\text{it}}$ due to the use of the Matthiessen’s rule.

In fact, the Matthiessen’s rule yields a systematic underestimate of $\mu_{\text{it}}$ [11], [22], [23], and the smaller is the error, the larger is the reference mobility curve with respect to the $\mu_{\text{it}}$ [21]. At small inversion densities, the reference curve is essentially the phonon-limited mobility, which is significantly improved in strained devices. Consequently, the underestimate of $\mu_{\text{it}}$ due to the Matthiessen’s rule is larger in relaxed devices than it is in strained devices. These arguments demonstrate that the $\mu_{\text{it}}$ degradation in strained devices cannot possibly be an artifact of the Matthiessen’s rule, but instead, the $\mu_{\text{it}}$ degradation is likely to be underestimated in the experiments because of the extraction procedure [21].

This latter statement is also supported by the comparison of Fig. 2 of this paper with the data of [21] (in particular, see [21, Fig. 9]). In this respect, we first notice that the absolute values of the simulated $\mu_{\text{it}}$ are well larger than the corresponding experimental values. We believe that this discrepancy is mainly due to the systematic errors produced by the Matthiessen’s rule used in the experiments, because the simulated and measured values of the total effective mobility are instead very similar [see Fig. 1(b)]. Moreover, it is also worth noticing that the strain-induced mobility degradation is larger in our simulations than it is shown in [21, Fig. 9]; we think that this happens mainly because, in the experiments, the Matthiessen’s rule underestimates the mobility difference between the relaxed and the strained devices.

As for the substrate-impurity-limited mobility, our simulations have shown a very modest improvement of $\mu_{\text{sub}}$ in strained devices, because the lower conduction mass of the $\Delta_2$ valleys is compensated by a larger scattering rate with respect to the $\Delta_4$ valleys. In particular, the strain-induced $\mu_{\text{sub}}$ changes are smaller than those experimentally shown in [21, Fig. 7]. We think that the quantitative discrepancy is mainly due to the use of the Matthiessen’s rule in the experimental data.

In this respect, it is useful to remember that, in order for the Matthiessen’s rule to be accurate, it is necessary that 1) the carriers are confined in a single subband [11], [21] and 2) the different scattering mechanisms contributing to $\mu_{\text{eff}}$ have the same energy dependence of the MRT $\tau_m$ [22], [23].

It has been argued that the errors of the Matthiessen’s rule related to the subband occupation are negligible in the case of strained devices because the strain-induced valley splitting confines the carriers in the $\Delta_2$ subbands [see Fig. 7(b)] [21]. The authors of [21] also noticed that such errors are expected to be small even for the unstrained silicon if the relaxation time $\tau_m$ of the $\Delta_4$ subbands is lower than the one of $\Delta_2$ subbands. However, Fig. 8 shows that $\tau_m$ is slightly larger in $\Delta_4$ than in $\Delta_2$ valleys, which can lead to significant errors in the $\mu_{\text{sub}}$ values extracted with the Matthiessen’s rule.

Aside from the issue of the subband occupation, the accuracy of the Matthiessen’s rule for the extraction of the Coulomb-limited mobility is seriously hampered by the fact that the Coulomb and the phonon scattering mechanisms have a very different energy dependence of the relaxation time. For a given subband, the phonon scattering rate is roughly proportional to the density of the allowed final states [24], [25]. This implies that the relaxation time $\tau_m$ typically tends to reduce for increasing energy, at least in the range of small energy values which determine the mobility, namely, up to 3 or 4$K_B T$. Fig. 8 shows that, for the CS, instead, the relaxation time tends to increase with the kinetic energy inside the subband. This is essentially due to the dependence of the scattering potential on the magnitude $q$ of the wave vector [see (2) and (3)], which favors scattering events with small $q$ values, which, in turn, correspond to electrons with small kinetic energies. The screening effect tends to reduce the scattering rates for very small $q$ and energy values; however, the energy dependence in the range of practical interest is the one shown in Fig. 8, where the screening effect is, in fact, accounted for.
The aforementioned arguments suggest that errors in the mobility extracted by using the Matthiessen’s rule can be unfortunately large even for the s-Si case, particularly at room temperature [22]. In order to quantitatively address this point, we applied the Matthiessen’s rule to our simulated $\mu_{\text{eff}}$ curves in order to emulate the procedure of $\mu_{\text{sub}}$ extraction used in the experiments [21]. Fig. 9 shows the $\mu_{\text{sub}}$ values obtained from the simulated $\mu_{\text{eff}}$ in Fig. 1(a). As it can be seen, the $\mu_{\text{sub}}$ values extracted by the Matthiessen’s rule (solid lines) are remarkably smaller than the correct values (dashed lines) and much closer to the experimental data reported in [21, Fig. 10].

The comparison of the curves in Fig. 9 shows, by using simulative data, the large errors produced by the use of the Matthiessen’s rule [22], [23]. Furthermore, the most important thing for the purpose of this paper is that such errors are quite different for the relaxed-Si and the s-Si cases. This is the origin of the apparent $\mu_{\text{sub}}$ improvement for s-Si shown in Fig. 9 (solid lines), which is not observed in the exact results for $\mu_{\text{sub}}$ in Fig. 6. For our simulation results, it is possible to certainly state that the strain-induced $\mu_{\text{sub}}$ enhancement shown in Fig. 9 is an artifact of the Matthiessen’s rule.

In our opinion, the $\mu_{\text{sub}}$ enhancements for strained devices observed in the experiments of [21] are similarly due, to a large extent, to the inaccuracies in the $\mu_{\text{sub}}$ determination due to the unavoidable use of the Matthiessen’s rule. This interpretation, on the one hand, helps to reconcile our simulation results with the experiments of [21] and, on the other hand, underlines that the Matthiessen’s rule should not be considered a quantitatively reliable procedure for the extraction of the different mobility components from the measured mobility characteristics. In this sense, we think that the mobility components obtained from the measurements by using the Matthiessen’s rule (such as the CS mobility) should not be considered as experimental results; in fact, the extraction procedure relies on heavily restricting modeling assumptions which, unfortunately, are not satisfied in most practical cases.

We thus believe that a quantitative analysis and interpretation of the experimental mobility curves demands a more theoretically sound basis than the Matthiessen’s rule. In this sense, once the model has been validated by a comparison with the experiments, the different mobility components can be distinguished without resorting to the Matthiessen’s rule.

**Acknowledgment**

The authors would like to thank Prof. L. Selmi for the many helpful discussions.

**References**


Francesco Driussi received the Laurea and Ph.D. degrees in electronic engineering from the University of Udine, Udine, Italy, in 2000 and 2004, respectively, where he worked on the reliability of Flash memory devices.

From October 2002 to September 2003, he held a student position at Philips Research Leuven, Leuven, Belgium. In 2005, he became a Research Associate with the University of Udine, where he is currently with DIEGM. He is also with IU.NET, Bologna, Italy. His research activities are mainly in the field of nonvolatile memory cell reliability, and in particular, his interests have been in the characterization of device degradation and gate dielectric reliability. In particular, he has worked on substrate-enhanced hot-electron phenomena, and he has investigated on their implications for Flash EEPROM devices. More recently, he worked on the development of physical and statistical models for the study of SILC in large Flash memory arrays and of the oxide trap generation and distribution. He is also working on the characterization and modeling of SONOS memory cells and of the trapping properties of silicon nitride. Part of his current activities are also in the field of experimental characterization and modeling of the carrier mobility in MOSFET devices featuring strained silicon and high-κ gate stacks.

David Esseni (S’98–M’00–SM’06) received the Laurea and Ph.D. degrees in electronic engineering from the University of Bologna, Bologna, Italy, in 1994 and 1998, respectively.

During 2000, he was a Visiting Scientist at Bell Labs–Lucent Technologies, Murray Hill, NJ. Since 2005, he has been an Associate Professor with the DIEGM, University of Udine, Udine, Italy. He is also with IU.NET, Bologna, Italy. His research interests are mainly focused on the characterization, modeling, and reliability of MOS transistors and nonvolatile memories (NVMs). In the field of NVMs, he has worked on the low-voltage and the substrate-enhanced hot-electron phenomena and on several aspects of Flash EEPROM memories, including innovative programming techniques and reliability issues related to the statistical distribution of stress-induced leakage current. Starting from year 2000, he has been involved in the field of advanced or innovative CMOS devices. In particular, he has experimentally investigated low field mobility in ultrathin SOI MOS transistors and then started an activity of semiclassical transport modeling in advanced n-MOS and p-MOS transistors. In this field, his research interests also include quantization models beyond the effective mass approximation as well as modeling and characterization of stress effects on the CMOS technologies.

Dr. Esseni served as a member of the technical committee of the International Electronic Devices Meeting in 2003 and 2004. He is currently in the technical committee of the European Solid-State Device Research Conference and the IEEE International Reliability Physics Symposium and is a member of the Technology Computer Aided Design Committee of the IEEE Electron Devices Society. He is the Associate Editor for the IEEE TRANSACTIONS ON ELECTRON DEVICES (TED) and has been one of the Guest Editors of a Special Issue of the IEEE TRANSACTIONS ON ELECTRON DEVICES devoted to the simulation and modeling of nanoelectronics devices.