Abstract—PMOS and NMOS mobility characteristics of the dual channel (strained Si/strained Ge) heterostructure have been reviewed. It is shown that the dual channel heterostructure can provide substantially enhanced mobilities for both electrons and holes. However, germanium interdiffusion from the germanium rich buried layer into the underlying buffer layer could potentially reduce the hole mobility enhancements.

Index Terms—strained-Ge, SiGe, germanium, MOSFET, mobility, strained-Si, pMOSFET, nMOSFET, germanium-diffusion

I. INTRODUCTION

Silicon substrates with extremely good oxide, low power dissipation, high integration levels, good noise immunity and cost-effectiveness occupy a dominant position in microelectronics. However, the low mobilities of holes and electrons in silicon limit its use in high frequency applications. Mobility increase is also desired both in digital CMOS where scaling is reaching its limits and in analog circuits, where high mobility and low noise are key performance metrics.

Strained layers of Si and SiGe offer dramatic mobilities improvements over bulk Si. Relaxed SiGe buffers on Si wafers with low defect densities are ideal platforms for including these layers. Additionally, the technologies for processing these structures are similar to standard Si CMOS processing.

N channel MOSFETs using strained Si (ε-Si) grown on (100) relaxed Si1-xGe x have been fabricated and studied by many researchers and electron mobility enhancements of 1.7 to 2.0 times over bulk Si have been observed1,2. Important variables in determining the enhancements are the strained Si thickness and the degree of strain in the Si layer. The strain levels in the above mentioned heterostructures are from 0.8% to 1.4 % corresponding to a Ge fraction of 20% to 35% in the Si1-xGe x virtual substrate.

Biaxial tensile strain breaks the six-fold degeneracy of Si’s conduction band resulting in reduced intervalley scattering in ε-Si n-MOSFETs and enhanced electron mobility. Additionally, for in-plane transport, electrons sample only the low transverse effective mass (m_t=0.19m_o), resulting in further enhancement. A minimum strain of 0.8%, corresponding to Si grown on relaxed Si0.8Ge0.2, is required for sufficient conduction band splitting to completely suppress intervalley scattering, and little improvement in electron mobility is gained by further increasing the strain3,5.

The enhancements in the PMOS devices vary with the gate overdrive and strain. Biaxial tensile strain also splits the light-hole/heavy-hole degeneracy, though the rate of subband splitting in the valence band is known to be lower than for the conduction band6.

Unlike electrons, both the out-of-plane and in-plane effective mass of holes is reduced by strain. The enhancements in the mobilities can degrade at higher fields8 as seen in the figure 1. With a strain lower than 0.8% the ε-Si PMOSFET loses its enhancements at strong gate overdrive, and in some cases shows worse mobilities than bulk Si. At low vertical fields, the hole wave function tends to be weighted below the surface due to the band offset which acts as a barrier to the holes between the relaxed Si1-xGe x and the ε-Si (Fig 2). The magnitude of the valence band offset (ΔE_v) increases with increasing x. As gate overdrive is increased, the hole wave function can overcome the barrier of the offset and shift toward the surface. The gradual influence of the ε-Si on the valence band results in increasing mobility enhancement with E_{eff} up to ~0.35 to 0.5 MV/cm, where a peak is typically observed. With higher Ge-content in the buffer, the barrier to hole occupation in the ε-Si cap increases, and a stronger vertical field is needed to shift the hole wave function up to the surface7. Beyond 0.5 MV/cm, the mobility...
enhancement in ε-Si p-MOSFETs, regardless of buffer composition, slowly decreases for reasons which are still poorly understood.\textsuperscript{2,5,8}

A biaxial tension greater than 1% in Si can cause the out of plane mobilities to become higher than the in-plane mobilities. This leads to a hole wavefunction which is substantially spread in the vertical direction. Experimental studies show the extent of the hole wavefunction to be more than 5 nm which implies that the hole wavefunction samples the valence band structure of the material below the ε-Si. This implies that even at high vertical fields the properties of holes in the relaxed SiGe below the ε-Si can make a significant contribution to the hole transport.\textsuperscript{9}

II. DUAL CHANNEL HETEROSTRUCTURES PMOS

Since the hole transport properties in the layer below the ε-Si are important, a way to improve the hole mobility could be the incorporation of a compressively strained layer below the ε-Si. In these devices, a compressively strained Si\textsubscript{x-y}Ge\textsubscript{y} layer is grown upon a relaxed Si\textsubscript{x}Ge\textsubscript{y} buffer (y>x) and capped with tensile ε-Si. The compressive strain in the Si\textsubscript{x-y}Ge\textsubscript{y} layer reduces the inter-valley scattering due to breaking of the valence band degeneracy and also reduces the in-plane and out-of-plane effective masses in a manner analogous to tensile ε-Si. The higher mobility in Ge-rich ε-Si\textsubscript{x-y}Ge\textsubscript{y} occurs because the band structure crosses over from the Si-like X-valley to the Ge-like L-valley at y>0.8 and thus, the ε-Si\textsubscript{x-y}Ge\textsubscript{y} starts exhibiting Ge-like mobilities (\(\mu_{h,Ge}=1900 \text{ cm}^2/\text{Vs}\)) compared to \(\mu_{h,Ge}=450 \text{ cm}^2/\text{Vs}\).

The improved hole mobilities of the dual channel heterostructures as compared to a single channel heterostructure can be seen in the Fig. 4. For all structures presented in Fig. 4, the buried ε-Si\textsubscript{x-y}Ge\textsubscript{y} layer has a Ge-content 30% higher than the relaxed Si\textsubscript{x}Ge\textsubscript{y}. The term “single channel heterostructure” refers to a surface strained Si layer over the relaxed Si\textsubscript{x-y}Ge\textsubscript{y} buffer, the Ge concentration of which is being varied. Thus, we can see that for a constant strain level, an increase in Ge concentration increases the mobilities even at high fields of 0.6 MV/cm. Thus it is believed that the holes are populating the high mobility compressive Ge-rich layer even at higher fields. For all structures in Fig. 4, the ε-Si cap on the top is 7-9 nm and could be reduced so that the holes occupy the bottom compressive layer even at higher fields.\textsuperscript{10}

It can also be extrapolated from the Fig. 5 that the highest hole mobility structure would be the one with Ge as the channel material.\textsuperscript{13} Also, a buffer composition of 50% Ge has the highest mobilities in the investigated structures (Fig. 6).

A. Effect of ε-Ge thickness.

To study the effect of ε-Ge thickness in dual channel heterostructures on hole mobility, \(h_{Ge}\) was varied while holding the strain level and \(h_{Si\text{cap}}\) constant. Theory would indicate that a ε-Ge layer grown on Si\textsubscript{0.3}Ge\textsubscript{0.7} would exhibit higher mobility than one grown on Si\textsubscript{0.5}Ge\textsubscript{0.5}, both because of the lowered in-plane effective mass and the larger valence band offset (type-I band alignment) between the ε-Ge and relaxed Si\textsubscript{x}Ge\textsubscript{y}. However, for the two devices with thin (\(h_{Ge}=6 \text{ nm}\)) ε-Ge layers, the more highly strained sample demonstrates lower \(\mu_{eff}\) across the entire range of \(N_{diss}\). In Ge, compressive strain greatly reduces the vertical effective mass of holes, and the out-of-plane mobility is predicted to exceed the already large in-plane mobility. However, the out-of-plane hole mobilities in ε-Ge are predicted to be much larger than those in ε-Si, as shown in Fig. 7. Thus, the hole wave function can spread into the layers above and below, despite the deep potential well for holes that forms in the ε-Ge. For the sample with thin ε-Ge on Si\textsubscript{0.5}Ge\textsubscript{0.5}, the 2% compressive strain and resultant high out-of-plane mobility cause the hole wave function to penetrate considerably into the ε-Si cap and the lower mobility buffer below. While the same spreading must take place in the thin ε-Ge sample on Si\textsubscript{0.3}Ge\textsubscript{0.7}, the vertical extent of the wave function is lessened due to the lower strain. Increasing compressive strain in the Ge can therefore serve to decrease the overall \(\mu_{eff}\) of the inversion layer if the ε-Ge is not thick enough to confine the hole wave function vertically.\textsuperscript{13}

Increasing \(h_{Ge}\) gave rise to an 80% increase in \(\mu_{eff}\) for the devices on Si\textsubscript{0.3}Ge\textsubscript{0.5} and only a 20% increase for the devices grown on Si\textsubscript{0.5}Ge\textsubscript{0.7} (Fig. 8). \(\mu_{eff}\) increases considerably for the thick ε-Ge layer on Si\textsubscript{0.5}Ge\textsubscript{0.5} because the 12 nm ε-Ge layer allows the large hole wave function to be better confined in the ε-Ge. Conversely, it may be deduced that 6 nm is nearly sufficient to confine the hole wave function when ε-Ge is grown on Si\textsubscript{0.3}Ge\textsubscript{0.7} based upon the relatively small gain in \(\mu_{eff}\) caused by doubling thickness of Ge.

Thus, we can see that a large enough ε-Ge thickness is required for obtaining the highest hole mobilities in these dual channel heterostructures. The thickness should be large enough to accommodate the vertical extent of hole wavefunction which is determined by the amount of strain in the Ge channel.

The above discussion on the effect of thickness of compressive Ge layer on hole mobility points in the
direction that the hole wavefunction penetrates the underlying buffer layers as well as top Si layer. The penetration of the hole wavefunction into the relaxed Si1-xGe, below is not desirable because it provides a very low mobility channel for the holes and hence reduces the overall mobility. Thus, it is expected that higher hole mobilities could be extracted if the hole wavefunction penetration in the underlying buffer is reduced so that the hole is confined to the high mobility Ge layer. Similarly, it has been determined experimentally that a thin top Si layer (thickness of around 3 nm) would provide the highest hole mobility because the hole wavefunction would remain in the higher mobility Ge layer below. Reducing the thickness of top Si (hSi_cap) to below 3 nm decreases the hole mobility for reasons that are unclear.

III. DUAL CHANNEL HETEROSTRUCTURES NMOS

Since the dual channel provides the highest hole mobilities, the electron mobilities in such structures also become important for commercial CMOS applications.

The band structure of the dual channel heterostructure provides a quantum well for the electron in the surface as seen in the Fig 3. Thus, the fraction of the electron wavefunction that exists in the Si cap will be high due to the band offset but could be varied by changing hSi_cap. A thick layer of Si should cause all of the electron wavefunction to remain in the top ε−Si and not penetrate in the buried channel. However, in a thin layer of Si the electron wavefunction would tend to remain partially in part in the buried channel as well.

We saw earlier that one factor in the improvement of the hole mobility in dual-channel p-MOSFETs was the fact that the hole mobilities are higher in the Ge (Ge-rich) buried channel. We might expect a similar kind of behavior in the NMOS case. However, bulk Ge’s conduction band minima lie in <111> directions while Si’s lie in <100> directions. Therefore, it is very possible that electrons near the ε-Ge/ε-Si hetero-interface may undergo strong X-valley to L-valley scattering reducing the mobilities. Fig. 10 shows that the above mentioned inter-valley scattering might be occurring as explained later. The dual channel heterostructure devices are compared against a control ε-Si on 25%Ge virtual substrate and a top Silicon cap of 10 nm and a strain of 1%.

We see that the mobility degradation in the heterostructure with buried ε-Ge is much higher as compared to one with buried ε-Si0.2Ge0.8 layer (Fig. 10). Heterostructure with ε-Ge is grown on Si0.8Ge0.6 virtual substrates while the one with ε-Si0.2Ge0.8 is grown on ε-Si0.6Ge0.4. Thus these layers have exactly same compression in the buried layer and similar ε-Si thickness (3 nm). We expect that the inter-valley scattering is responsible for this mobility degradation. Thus, buried ε-Ge, which has conduction band minima in the <111>, would tend to show a larger scattering rate as compared to ε-Si0.2Ge0.8 which possesses a band structure similar to Si.

Regardless of the mechanism of mobility degradation, Fig. 9 shows that the only way to achieve high electron mobility in a dual-channel heterostructure is to completely isolate the electron wave function from the buried compressive layer. This is made possible by increasing the Si cap thickness to roughly 6-7 nm. This would, however, reduce the hole mobility enhancements as discussed earlier. A solution to this problem could be an extra oxidation step for PMOS which would reduce the thickness of top Si layer to an optimum value of 3 nm. An optimum structure for symmetric hole and electron mobilities has been suggested elsewhere.

IV. GROWTH PROCEDURE AND PROCESSING

All films were grown in an ultra-high vacuum chemical vapor deposition (UHVCVD) system with a base pressure of around 1e-9 torr and is equipped with SiH4, GeH4, and Si2H6 source gases. The reactor is a hot-walled, load-locked system with a quartz growth tube, and up to 10×6” wafers can be loaded simultaneously. Temperatures and growth pressures could be varied from 900°C -350°C and 1-30mTorr. Growth begins with depositing relaxed compositionally graded SiGe buffers, grown at high temperatures (900°C – 750C depending on the composition of the buffer). All compositionally graded layers were graded at 10% Ge/µm (2% jumps of 2000 Å) and capped with 0.5-1.5 µm uniform composition layers. Some samples were doped in situ to concentrations of 1×1016 cm-3 using B2H6 or PH3 dopant gases. The elevated growth temperatures maximizes the strain relaxation and dislocation glide velocity while the optimized grading rate prevents the dislocation nucleation. From previous experience these virtual substrates have threading dislocation densities of 107 – 108 cm-2, regardless of final composition. Deposition of device layers followed virtual substrate growth.

The previous discussion about the dual channel heterostructure shows that the structures would have layers with different levels of strain. Thus, these layers have a tendency to deviate from planarity due to strain-driven surface diffusion. However, a planar structure is desired for the MOSFET applications. A specific procedure for growing these films was adopted in order to suppress surface instabilities. The tensile layers have a much reduced tendency to undulate as compared to the compressive layers. Thus, a strained...
silicon layer grown on Si rich (60%-100% Si) virtual substrate will not have a tendency to undulate at temperatures around 650°C. However, if the strain in the Si layer is more than 2.5% then the layer could show ramp like undulations. However, 3-dimensional strained Si morphologies can be suppressed at all levels of strain by growing the layers at low temperatures.

In a dual channel heterostructure the layer of Ge (or the Ge-rich layer) deposited on the relaxed Si,Ge, has a compressive strain and hence is much more prone to islanding surface morphology. The undulations which can form at around 450°C because of strain-driven surface diffusion can scatter holes and hence reduce mobility enhancements of the holes. Hence, the Ge-rich layer must be grown at even lower temperatures (350-450°C) to suppress islanding. The temperature for growth is chosen based on the lattice mismatch experienced by the Ge.

Since the kinetics of silane decomposition are very slow at the temperatures required to grow the Ge channel a novel procedure was used to grow the tensile Si cap without causing undulations to the compressive Ge layer. The Si cap is grown by flowing silane while raising the growth temperature. As the Si layer is initiated at low temperature surface diffusion is halted preventing undulations of the Ge-rich buried channel. As the temperature increases further the silane decomposition kinetics increases allowing growth of Si cap in a reasonable amount of time. Thus, growth of the top Silicon layer could be carried out at temperatures required for silane decomposition without introducing undulations in the bottom compressive layers.

The MOSFET processing was performed in the Integrated Circuits Lab of the MIT Microelectronics Technology Laboratory. In order to facilitate the investigation of many device structural variants, the short-flow MOSFET process was utilized. In this process, large geometry (~100-200µm) ring transistors are fabricated with a single lithography step.

The short-flow process, despite the large MOSFET geometries, is useful for extracting data relevant to short channel devices. Specifically, the MOSFETs can be biased in the gate region with a very high vertical field, thus simulating the nature of the channel in a short channel device. We note here that no other characteristics of these devices were optimized. Essentially, the short flow MOSFET process allows the expedient exploration of Si/SiGe heterostructure channels under large vertical electric fields.

The short-flow MOSFET process began with a modified RCA clean, which incorporated a piranha clean in place of the traditional SC-1 clean in order to prevent rampant etching of the surface Si device layer. The remainder of the RCA clean was followed by deposition of the 3000Å thick SiO₂ gate dielectric via low pressure chemical vapor deposition (LPCVD) at 400°C. This low temperature oxide (LTO) was chosen as the gate dielectric not only to preserve a low thermal budget for the process, but also because a thick oxide is required for this short-flow process, as described below.

The LTO was followed by deposition of 500Å of polysilicon at 560°C to complete the gate stack. The LTO and polysilicon on the wafer backside were etched away to facilitate backside electrical contact. Ring transistors were then patterned on the wafers in the only lithography step of the process and etched in an Applied Materials Precision 5000. Wafers were then subjected to a 15 second dip in buffered oxide etchant (BOE) to underetch the gate polysilicon and form a large “T-gate” geometry. Uniform BF₂ or As ion implants were performed at 35keV to dope the source, drain, and gate contact regions. Four identical implants were used, each after a 90 degree rotation of the wafer, to prevent shadowing effects and ensure implantation of the entire source/drain region of the ring transistors. The implant activation was done for 30 min at 600°C. It is to be ensured that the processing temperatures are not very high because the thermal budget of the structure is not very high.

Blanket Ti/Al metallization was then performed via e-beam deposition at perpendicular incidence. Due to the extreme geometry of the “T-gate” FET structure, breaks occur in the metal which isolate the source, gate, and drain regions without further lithography, much like a traditional liftoff process. A backside contact was formed via deposition of 1µm of Al. A 30 minute 400°C sinter in forming gas (5% H₂ in N₂) completed the FET fabrication.

V. GERMANIUM INTERDIFFUSION PROBLEM IN THE SILICON GERMANIUM HETEROSTRUCTURES

A problem for the processing of the dual channel heterostructure is the very low thermal budget of this structure. It was shown earlier that the difference in the Germanium concentration between the underlying SiGe buffer and the buried Ge-rich layer determines the hole mobility enhancements in a dual channel heterostructure. The enhancements also increase if the Ge concentration in the buried channel increases. Thus, it means that for a constant buffer concentration increasing the Ge content in the buried channel increases the enhancements.

However, with an increase in the Ge concentration the thermal budget of the structure drastically decreases. Si and Ge are completely miscible in the solid state and have a lens shaped phase diagram. The melting temperature of the Ge is 940°C. Thus, the maximum temperature at which the films could be processed can go down significantly because of the low melting temperature of the SiGe films.

The diffusion coefficient of Ge also increases with the
increase in Ge concentration in the different layers of the dual channel heterostructure. Thus, increasing the Ge concentrations in the buried layer for larger hole mobility enhancements would lead to a problem of increased interdiffusion of Ge among the layers.

The diffusion of Ge can occur both through point defects and lattice site exchanges. Below 925°C point defect dominated diffusion takes place because of its low activation energies. It is believed that at temperatures below 925°C Ge diffuses in the Si lattice primarily through vacancy mechanism. The activation energies for Ge diffusion into the SiGe show a generally decreasing trend with an increase in the Ge concentration. Thus, the activation energies for diffusion of Ge decrease from 4.7eV in pure silicon to 3.2eV in 50% SiGe. Thereafter, the activation energy appears to remain constant with further increases in germanium concentration. Accordingly, the diffusion coefficient of Ge in Si$_{1-x}$Ge$_x$ spans 4 orders of magnitude depending on the Ge content of the Si$_{1-x}$Ge$_x$.

From the results of ref. 19 it is expected that diffusion of Ge into underlying layers during high processing and growth temperatures would change the concentration and thickness of the compressively strained Ge rich layer. Thus, the interdiffusion can cause reduction in mobility enhancements from the reduced mobilities in the lower Ge concentration layer and also because of strain relaxation in the buried layer. It could also lead to a degraded interface quality between the SiGe/Si interface and hence potentially increasing the Coulombic-scattering of carriers.

Thus, the structure has a low thermal budget since Ge diffusion from the middle layer into the underlying layers can occur at temperatures which the structure faces during device fabrication stages. These high temperatures can be encountered during the growth of thermal oxide or activation of ion implants. Since, these layers are grown at lower temperatures (around 350-550°C), interdiffusion during the growth steps is fairly limited. The maximum temperature that a particular heterostructure can tolerate depends on the specific concentrations of the layers. Increasing the Ge concentration in both the relaxed buffer and the buried channel dramatically raises the hole mobility enhancement, but simultaneously lowers the temperature that these layers can withstand without diffusing significantly.

### CONCLUSIONS

The dual channel heterostructure combines a surface strained Si layer and a compressed Ge rich layer. Thus an optimized structure can be attained which would give the highest possible mobilities for both holes and electrons. One such structure suggested elsewhere is a strained Si/strained Ge structure grown on Si$_{0.5}$Ge$_{0.5}$.

The thickness of the top silicon would have to be reduced by an extra etching/oxidation step for the p-MOSFET. However, the thermal budget of these structures is expected to be very low and therefore, these structures would have to be fabricated using temperatures lower than those employed in standard silicon processing. Thus, for these structures to be commercially relevant, it is important to improve the thermal budget to which these layers can be exposed without significant losses in mobility enhancement.

### REFERENCES


Figure 1: Hole mobility enhancement vs $E_{\text{eff}}$ for $\varepsilon$-Si $p$-MOSFETs grown upon $\text{Si}_x\text{Ge}_y$ buffers with $x=0.35$ to 0.5. Adapted from Leitz et al. (Ref. 8). The point of peak enhancement is pushed to higher $E_{\text{eff}}$ with higher $x$.

Figure 2: Band alignment of strained Si grown on relaxed $\text{Si}_1-x\text{Ge}_x$. Increasing $x$ relaxes $\text{Si}_1-x\text{Ge}_x$.

Figure 3: Schematic band alignment of a dual-channel heterostructure. Higher $y$ in the compressive layer leads to a deeper well for holes. In $p$-MOSFETs based on these structures, large gate overdrives force holes towards the $\text{SiO}_2$/Si interface.

Figure 4: Hole mobility enhancement over Cz-Si at $E_{\text{eff}} = 0.6$ MV/cm for single-channel and dual-channel devices grown on the same relaxed buffer compositions. The buried layer Ge content in all of the dual-channel devices is 30% higher than the relaxed buffer. Even at high $E_{\text{eff}}$ dual-channel devices demonstrate far larger enhancements than single-channel devices.

Figure 5: Effective hole mobilities of dual-channel heterostructure $p$-MOSFETs (Ref. 20). The difference in Ge content between the virtual substrate and compressive layer, $y-x$, is held constant at 0.3.

Figure 6: Hole effective mobility vs $N_{\text{inv}}$ of $\varepsilon$-Si / $\varepsilon$-Ge dual-channel heterostructures grown on $\text{Si}_1-x\text{Ge}_x$ virtual substrates with $x = 0.5$ to 1.

Figure 7: Calculated in-plane and out-of-plane hole mobilities in $\varepsilon$-Ge. The extremely large out-of-plane hole mobilities lead to a hole wave function that can significantly penetrate into both the Si cap above and the relaxed $\text{Si}_x\text{Ge}_y$ below.

Figure 8: Hole effective mobility vs $N_{\text{inv}}$ for the samples described in the paper. The use of a thicker $\varepsilon$-Ge layer gave rise to a 20% increase in $\mu_{\text{eff}}$ for the samples grown on $\text{Si}_x\text{Ge}_y$, and an 80% increase for the samples grown on $\text{Si}_x\text{Ge}_y$. 

% Ge in buried layer (dual-channels only)

% Ge in relaxed buffer

Mobility Enhancement over Cz-Si

Bulk Si
80% Ge
75% Ge
70% Ge
60% Ge
Strained Si

12 nm $\varepsilon$-Ge on $\text{Si}_1\text{Ge}_{0.8}$
12 nm $\varepsilon$-Ge on $\text{Si}_1\text{Ge}_{0.7}$
6 nm $\varepsilon$-Ge on $\text{Si}_1\text{Ge}_{0.8}$
6 nm $\varepsilon$-Ge on $\text{Si}_1\text{Ge}_{0.7}$
Cz-Si

Effective mobility (cm$^2$/Vs)

$N_{\text{inv}}$ per cm$^2$
Figure 9: Electron effective mobility vs $N_{inv}$ of $\varepsilon$-Si / $\varepsilon$-Ge dual-channel $n$-MOSFETs grown on Si$_{0.5}$Ge$_{0.5}$ with $h_{Si}$ cap varied. Only the sample with $h_{Si}$ cap = 7.5 nm matches the mobility of the $\varepsilon$-Si control.

Figure 10: Electron effective mobility vs $N_{inv}$ for single-channel and $\varepsilon$-Si / $\varepsilon$-Ge dual-channel $n$-MOSFETs grown on Si$_{0.4}$Ge$_{0.6}$. $h_{Si}$ cap = 3 nm for both devices.

Figure 11: Short-flow MOSFET process