Available online at www.sciencedirect.com
ScienceDirect
Solid-State Electronics 51 (2007) 828–837
SOLID-STATE ELECTRONICS
www.elsevier.com/locate/sse

Design and optimization of a buried channel PMOS integrable in a $Si_{1-x}Ge_x$ BiCMOS process

P. Khare [a,b,*], D. Schroder [b], R. Sampson [a]

[a] STMicroelectronics, 1000 E. Bell Road, Phoenix, AZ 85022, USA
[b] Department of Electrical Engineering and Center for Solid State Electronics Research, Arizona State University, Tempe, AZ 85287-5706, USA

Received 25 February 2006; received in revised form 25 February 2007; accepted 16 March 2007
Available online 23 May 2007

The review of this paper was arranged by Prof. Y. Arakawa

Abstract

A strained $Si_{1-x}Ge_x$-channel PMOSFET, fully integrable in a standard $Si_{1-x}Ge_x$ BiCMOS process is proposed. It uses an n^+-polysilicon emitter of the bipolar transistor as the gate, and has a p-type $Si_{1-x}Ge_x$ channel, making it a truly buried channel device. It uses only two extra masks and a few minor changes in the baseline BiCMOS flow, keeping the additional processing cost low. The basic device design is systematically optimized to get the best compromise among hole confinement in $Si_{1-x}Ge_x$, layer stability and channel hole mobility. Optimization was done using Ge fraction dependent process and device simulation parameters for the first time. A novel, semi-analytical approach to model the effect of Ge fraction on hole current was used. The impacts of Si cap layer thickness, $Si_{1-x}Ge_x$ film thickness, channel doping and Ge fraction were investigated. During process and device design, care was taken to ensure that the $Si_{1-x}Ge_x$ layer remains strained as well as the standard available devices are not impacted by additional process steps. The optimized device shows a 2× gain in transconductance and in drain current over the standard PMOS device available in the baseline process.

Keywords: Silicon–germanium; Buried channel; PMOSFET; BiCMOS; Simulation; Hole mobility

1. Introduction

Strained $Si_{1-x}Ge_x$ layers have been extensively pursued to improve the performance of bipolar and CMOS technologies, as strain offers device designers the ability to modify band gap and carrier mobilities [1–4]. While strained $Si_{1-x}Ge_x$ HBT (heterojunction bipolar transistor) has now become the mainstream bipolar technology, progress in the CMOS arena has been relatively slow until now. One of the main reasons behind this is the focus on strained Si technologies that offer an attractive alternative at 90 nm node and beyond [5,6]. Strained Si technology is highly complex as it requires good quality relaxed SiGe layers, which are difficult to grow because of high density of threading dislocations, non-planarity and high surface roughness [2], making it cost effective only for sub-100 nm technology nodes.

For industrialized BiCMOS processes already using $Si_{1-x}Ge_x$ layers in bipolar devices, PMOS channel mobility can be enhanced by having those compressively strained $Si_{1-x}Ge_x$ layers in the channels. This reduces the asymmetry between NMOS and PMOS, and thus the CMOS performance can be improved in 0.35 and 0.25 μm processes without moving to the next more expensive technology node. Since enhancement in electron mobility is not as much as enhancement in hole mobility by using strained SiGe, PMOSFETs benefit more with SiGe than NMOSFETs.

Several researchers have worked on developing $Si_{1-x}Ge_x$-channel PMOSFETs [7–11]. However, all of them

* Corresponding author. Address: STMicroelectronics, 1000 E. Bell Road, Phoenix, AZ 85022, USA. Tel.: +1 602 485 2164; fax: +1 602 485 2955.
E-mail address: prasanna.khare@st.com (P. Khare).

0038-1101/$ - see front matter © 2007 Elsevier Ltd. All rights reserved.
doi:10.1016/j.sse.2007.03.007

use undoped or n-type doped channels with p^+-polysilicon gates, making them conventional surface channel PMOSFETs. The equilibrium band diagram of such devices is shown in Fig. 1a. They all used SEG (selective epitaxial growth) for $Si_{1-x}Ge_x$ epitaxy, which is more complex than a non-selective epitaxial growth. Additionally, only one or two combinations of $Si_{1-x}Ge_x$ thickness, Ge mole fraction, cap layer thickness and doping concentrations were used in the fabrication.

Several p-channel MODMOSFETs have also been reported in literature [12–14]. The difference between MOSFETs and MODMOSFETs is that MODMOSFETs use a modulation doping layer (δ-doped boron layer) below the $Si_{1-x}Ge_x$ channel and the channel is kept undoped. Holes *spill* from the δ-doped layer into the $Si_{1-x}Ge_x$ channel because of large valence band offset. Because of the absence of impurity scattering in the channel, there is a further enhancement in the mobility, in addition to the strain-induced enhancement, giving a very high channel hole mobility. The δ-doped layer may be separated from the $Si_{1-x}Ge_x$ layer by means of a thin undoped Si buffer (or spacer) layer. MODMOSFETs have several advantages over MOSFETs, including higher hole mobility [12], better hole confinement [15], higher subthreshold slope because of reduction in the number of holes in the channel at zero gate bias, and higher drive current [2]. However, placing a narrow, highly-doped boron layer immediately below the channel places severe restrictions on the thermal processing since the $Si_{1-x}Ge_x$ channel should remain undoped [14]. The maximum thermal cycle used in the MODMOSFETs described in [14] was 800 °C for 20 s. Another boron profile has also been suggested that can be annealed at 970 °C for 5 s. This constraint is too severe to meet in 0.35 μm processes, where S/D RTA budgets are generally much higher.

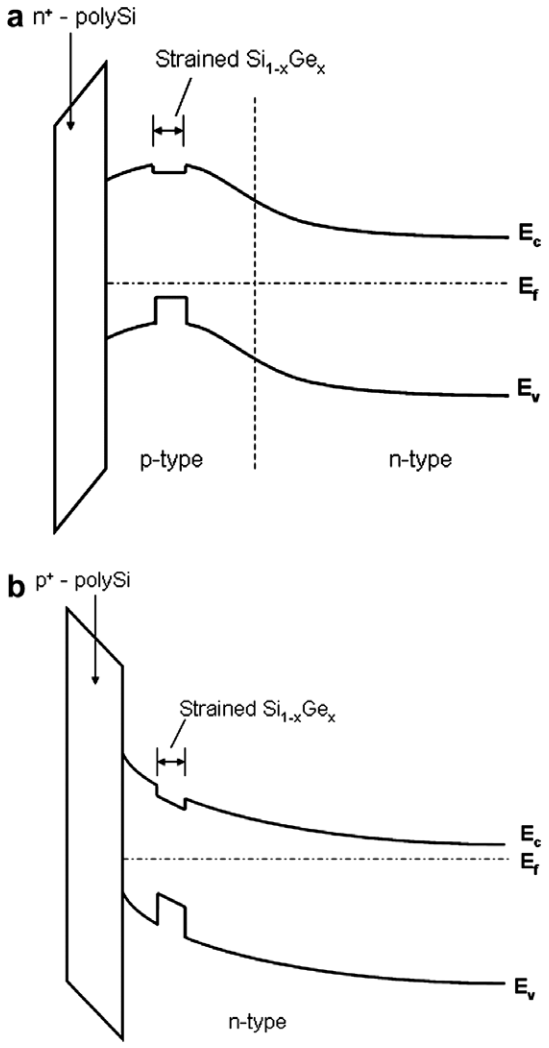

Fig. 1. Schematic equilibrium band diagrams of (a) conventional surface channel $Si_{1-x}Ge_x$ PMOSFET and (b) proposed buried channel $Si_{1-x}Ge_x$ PMOSFET.

In this paper, we show how a higher mobility PMOSFET can be economically integrated into an existing 0.35 μm $Si_{1-x}Ge_x$ BiCMOS process using non-selective SiGe epitaxy. The proposed device is a buried channel device with p-type doping in the channel and using the same n^+-polysilicon layer for the gate as is used to form the n^+-emitter of the bipolar transistor. The design parameters of the feasible structure and the optimization criteria are presented first, followed by the optimization results obtained by quasi-2D process and device simulations. Process simulation of $Si_{1-x}Ge_x$ PMOS including Ge fraction dependence of boron diffusivity and device simulation including Ge fraction dependence of the mobilities and effective masses are reported for the first time. The effect of Ge fraction on hole mobility was modeled using a combination of numerical, analytical and published results. A comparison of $Si_{1-x}Ge_x$ PMOS with standard baseline PMOS and also with a reference Si-epitaxy device shows significant transconductance enhancement.

2. Process integration and proposed device

The main feature of the baseline process is the non-selective $Si/Si_{1-x}Ge_x$ epitaxy, which is used to form the intrinsic base layer of the NPN and the poly-Si base electrode in a single step [16]. The epitaxy is carried out at 650 °C using $H_2/SiH_4/GeH_4/B_2H_6$ chemistry, so that the NPN base is heavily boron-doped $Si_{1-x}Ge_x$. The process integration is done with a *base-after-gate* scheme, in which the main CMOS process steps are performed before the bipolar steps. By doing so, the base is not subjected to CMOS thermal steps, such as gate oxidation, sidewall formation, S/D oxidation. This helps in maintaining a narrow base profile and ensuring that the SiGe base layer is strained. During bipolar processing, various films are deposited on the CMOS, which are then completely removed at the end of the HBT module.

Since the $Si_{1-x}Ge_x$ layer is p-type doped and the HBT emitter is n^+-polysilicon, the simplest PMOSFET can be made by using the emitter as the gate and p-$Si_{1-x}Ge_x$ as the channel. The dielectric can be obtained by not opening an emitter-cut before emitter poly-Si deposition, leading to

a buried channel device. The choice of this type of structure is further favored for some other reasons: (1) Using 1D Poisson simulations [14,17], it has been shown that for the same threshold voltage, hole confinement in the $Si_{1-x}Ge_x$ channel is significantly better in the n^+-gate design than in the p^+-gate. (2) Secondly, in a buried channel device, not only the holes flow in the high-mobility SiGe region, but there is also reduced surface roughness scattering. (3) Finally, to make a surface conduction PMOSFET, n-type dopant will have to be introduced during the SiGe epitaxy instead of boron. However, using AsH_3 or PH_3 in place of B_2H_6 that is currently used along with GeH_4, increases the process complexity and may lead to the contamination of the existing $Si_{1-x}Ge_x$ epitaxy reactor. The equilibrium band diagram of such a device is shown in Fig. 1b, showing that the hole confinement in the $Si_{1-x}Ge_x$ channel is further enhanced by the band bending in the top p-layer. This further enhances the hole mobility as compared to a conventional surface channel $Si_{1-x}Ge_x$ PMOSFET. Conventional buried channel devices have worse short-channel effects than the surface channel devices [18], but that should not be the case here as the total p-layer ($t_{Si} + t_{SiGe}$) is very thin.

For an enhancement-mode buried channel device, the top p-layer should be completely depleted of holes at zero gate bias. Process and device simulations showed that with the existing $Si_{1-x}Ge_x$ layer, it is not possible to have a depleted surface at zero bias because the boron concentration is too high, so that the device will be "on" at zero bias. Therefore, a separate epitaxy for the $Si_{1-x}Ge_x$ PMOS is needed, in which the boron layer thickness and doping concentration can be kept reasonably low.

Because of these integration constraints, we decided on a separate $Si_{1-x}Ge_x$ epitaxial layer for the PMOS. The gate oxide is formed by depositing PECVD oxide and not by thermal growth, so that both the old and the new $Si_{1-x}Ge_x$ layers remain strained. High quality PECVD oxides deposited on capped $Si_{1-x}Ge_x$ layers at 350 °C have been reported [8,14,19].

The proposed process flow with the $Si_{1-x}Ge_x$ PMOS integrated into the baseline process is shown in Fig. 2.

3. Simulation details

DIOS and DESSIS programs of ISE (now Synopsys) were used for process and device simulations, respectively. Strain in pseudomorphic $Si_{1-x}Ge_x$ layer affects dopant diffusivities also. Boron has higher solid solubility in $Si_{1-x}Ge_x$, retarding its diffusion. To account for this, the default model provided by ISE is used, which is based on Cowern's empirical law [20]. Drift-diffusion models for linear I_D–V_G simulations were used, as the channel simulated (0.35 μm) is long enough and does not need more advanced hydrodynamic models. Ge mole fraction dependent parameters for the band gap and for the hole and electron effective masses were used from literature [21–23].

Since the main benefit of strain is enhanced hole mobility, it is very important to include a model for mobility that accurately takes into account variation with doping concentration as well as with Ge fraction. Device simulations of $Si_{1-x}Ge_x$ devices with n^+-polysilicon reported in [24] included Ge-induced band gap narrowing, but did not include enhanced mobility due to the presence of Ge in their simulation. $Si_{1-x}Ge_x$ p-MODMOS simulations reported in [14] assumed a constant mobility in undoped $Si_{1-x}Ge_x$, independent of Ge fraction. While numerous experimental and analytical data exist in literature for hole mobility, it is not possible to use it in a device simulator. Therefore, a novel, semi-analytical approach for including strain-dependent mobility has been used for linear region I–V simulations, as described in the next section.

All the modified model parameters for process and device simulation except for mole fraction dependent mobility parameters have been used for modeling the HBT in the baseline technology, and HBT simulations showed very good agreement with experiment [25]. One main difference between $Si_{1-x}Ge_x$ PMOS simulations reported earlier [14,24] and in this paper is that ours start with process simulations, whereas others were device simulations.

Due to the presence of the very thin $Si_{1-x}Ge_x$ layer (few nm), a mesh resolution in the SiGe layer down to 1 nm is required to obtain accurate doping profiles. A full 2-D

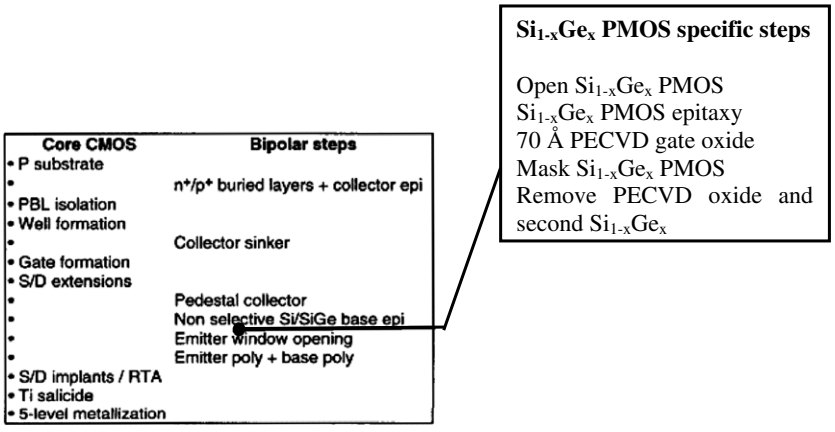

Fig. 2. Process integration of $Si_{1-x}Ge_x$ PMOS.

process simulation would need a very large number of mesh points, which makes it impractical due to very large computation time. Therefore, a full 2-D approach was not used. Instead, very accurate 1-D simulations were carried out initially, and were combined with appropriate lateral spreads to get the final 0.35 μm gate length device structure suitable for 2D device simulation.

To simulate threshold voltage, Q_{inv}–V_G device simulations were carried out, where Q_{inv} is the total inversion hole density. This was obtained through a 1-D process simulation for only the gate region, followed by a Poisson solution, and then integrating the hole density (for PMOS) from the Si–SiO_2 interface to a depth of few hundred angstroms, where the hole density becomes negligible. Saturation region characteristics were not simulated due to modeling limitations, as explained in the next section.

4. Optimization strategy and semi-analytical modeling

Optimization was done on the design parameters shown in Fig. 3: (a) Ge profile (Ge fraction, x and $Si_{1-x}Ge_x$ layer thickness, t_{SiGe}), (b) cap layer thickness (t_{cap}), (c) p-type doping in the cap layer (N_{cap}), and (d) doping in the Si_{1-x}-Ge_x layer (N_A). Device constraints including strain relaxation and oxide interface quality were also taken into account.

Although triangular profiles have been shown to be better than rectangular profiles for MODFETs [13], only rectangular profiles were considered in our simulations. Rectangular profiles make drain current modeling possible as described later, as well as it is easier to implement in the process. Since the growth rate is typically very high during $Si_{1-x}Ge_x$ epitaxy, and the growth time is very short, profiles other than rectangular increase the complexity significantly.

A typical plot of integrated hole densities in the Si cap (Q_{Si}) and in the $Si_{1-x}Ge_x$ layer (Q_{SiGe}) against gate voltage obtained by 1-D Poisson solution is shown in Fig. 4.

From this plot, we have, V_{co} = cross-over voltage; gate voltage where Q_{Si} exceeds Q_{SiGe}, V_{TO} = intercept of linear extrapolation of Q_{total} on V_G axis. This is equivalent to the threshold voltage. $V_{hc} = |V_{co} - V_{TO}|$. This is the gate voltage range of the linear region of the MOSFET over which holes flow predominantly in the high-mobility $Si_{1-x}Ge_x$ channel.

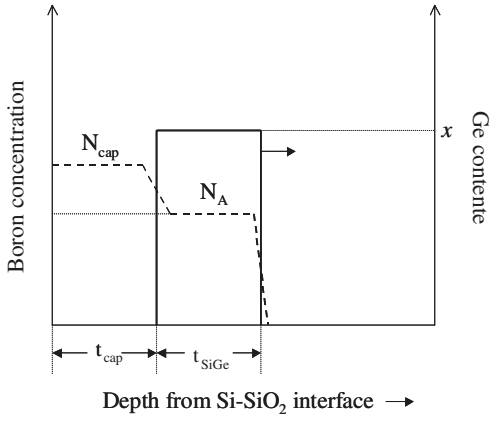

Fig. 3. Design parameters for $Si_{1-x}Ge_x$ epitaxy.

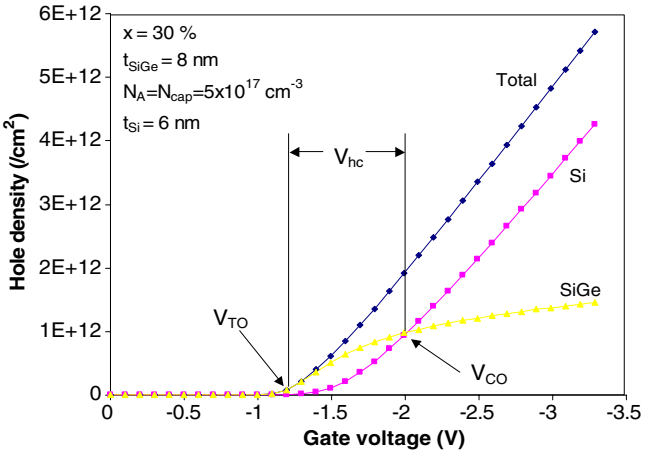

Fig. 4. Typical plot of integrated hole density in Si, in $Si_{1-x}Ge_x$ and in total.

For maximum hole confinement in $Si_{1-x}Ge_x$, V_{hc} should be maximized. This is the *first* optimization criterion. It is not enough to maximize only V_{co}, because the V_T's may be different for different designs, and V_T's can be adjusted easily using N_A. V_{hc} tells us that after the device has turned on, how long does it work better than a Si MOSFET. Ideally, we would like to have $V_{hc} = V_{dd}$, the maximum drain voltage, so that there are always more holes in the SiGe channel than in the parasitic Si channel.

Maximizing V_{hc} is also insufficient, as two structures may have same V_{hc}, but if the Ge content differs, the mobility enhancements will be different. Therefore 2-D I_D–V_G simulations that include Ge fraction-dependent mobility are required, so that the drive currents and transconductances can be compared. Thus, further optimization was done using 2-D device simulations with 0.35 μm gate length. These are actually quasi-2D simulations, in which doping profiles in the gate and in the source/drain regions are obtained using separate 1-D simulations, and then are appropriately combined to get a 2-D structure. We used the same spacer width and polysilicon etch profile that is used for the emitter polysilicon.

For the Ge fraction-dependent mobility, a linear interpolation model is provided in the simulator, in which the mobility in $Si_{1-x}Ge_x$ varies linearly with the Ge content (x) from Si to Ge values. This may not be accurate for strained $Si_{1-x}Ge_x$. As a result, we developed a simple simulation approach based on the mobility model proposed by Carns et al. [26]. They proposed a modified Arora model for mobility, in which some parameters are molefraction-dependent, and fitted it to the experimental data. The mobility is given as:

$$\mu_p = \mu_{min} + \frac{\mu_0}{1 + \left(\frac{P}{P_{ref}}\right)^{\alpha}} \quad (1)$$

where $\alpha = 0.90$, $P_{ref} = 2.35 \times 10^{17}\ cm^{-3}$, and μ_{min} and μ_p are

$$\mu_{min} = 44 - 20x + 850x^2$$
$$\mu_0 = 400 + 29x + 4737x^2$$

where x is the Ge content and P the doping concentration. They fitted this relation to the experimental data only up to a Ge content of 22%. We used a Ge content of up to 30%. For $x = 0.3$ and $P = 10^{18}\ cm^{-3}$, the above equation gives the hole mobility in $Si_{0.7}Ge_{0.3}$ 2.3 times higher than in Si at the same doping concentration. For a doping concentration of $10^{17}\ cm^{-3}$, the enhancement factor becomes 2.2. These enhancement values are very close to those obtained by Manku et al. [27]. Thus, it has been assumed that the above model remains valid up to $x = 0.3$ also.

This model cannot be directly integrated into the simulator for a user. We first simulate the I_D–V_G curve at $V_{DS} = -0.1$ V for the $Si_{1-x}Ge_x$ PMOS using silicon values of mobility, i.e., using the above equation with $x = 0$. This gives us the conventional Arora model, which can be implemented in the simulator. Note that only for mobility, silicon values are used. For band gap and effective masses, Ge content dependency has been taken into account, as during 1-D simulations. Then, we modify this simulated current, by first partitioning the channel charge between Si and $Si_{1-x}Ge_x$ and then appropriately weighting it with corresponding mobilities, as follows:

The simulated current with Si values of mobility is:

$$I_{sim} = (Q_{Si} \cdot \mu_{Si,cap} + Q_{SiGe} \cdot \mu_{Si,bur}) \cdot \varepsilon \quad (2)$$

where Q_{Si} and Q_{SiGe} are the integrated hole charges in Si cap and buried SiGe, respectively, ε is the lateral electric field in the channel, $\mu_{Si,cap}$ the hole mobility in the cap layer, and the simulator takes into account the effect of perpendicular electric field and surface roughness scattering while calculating this mobility. $\mu_{Si,bur}$ is the hole mobility in the SiGe channel obtained using the Arora model in Eq. (1) with $x = 0$. It is assumed that this mobility is not affected by surface roughness scattering, and thus can be calculated using Eq. (1) with $x = 0$.

Now, the new current, using the appropriate mobility in the $Si_{1-x}Ge_x$ channel is:

$$I_{new} = (Q_{Si} \cdot \mu_{Si,cap} + Q_{SiGe} \cdot \mu_{SiGe}) \cdot \varepsilon \quad (3)$$

where μ_{SiGe} is the mobility in the $Si_{1-x}Ge_x$ channel obtained using Eq. (1) with appropriate value of x.

Subtracting (2) from (1) and re-arranging, we get

$$I_{new} = I_{sim} + Q_{SiGe} \cdot (\mu_{SiGe} - \mu_{Si,bur}) \cdot \varepsilon$$

Q_{SiGe} can be obtained by integrating the hole density in the $Si_{1-x}Ge_x$ region at each value of V_G. The lateral field, $\varepsilon = V_{DS}/L_{eff}$ is calculated by assuming that the drain voltage of −0.1 V drops across 0.3 μm. The modified I_D–V_G curve after this mobility correction along with the simulated curve are shown in Fig. 5. The V_T changes slightly because of higher current.

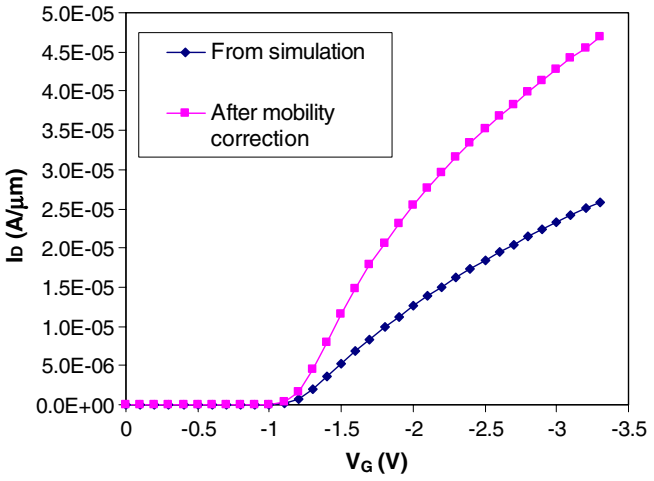

Fig. 5. Effect of mobility correction (same inputs used as in Fig. 4).

This semi-analytical (because we start with a numerically obtained current) modeling of the linear region I_D–V_G is based on the following assumptions and limitations:

(1) Drain current flows by drift only, which is reasonable for a MOSFET operating in linear region. We also compared the drain current obtained by the drift-diffusion model with that obtained using the hydrodynamic model, and found no difference between the two, because the channel length is ~0.35 μm, which is relatively large for dynamic transport effects to begin; and secondly the field in the channel is very low when $V_{DS} = -0.1$ V. Because current is assumed to flow by drift only, this modeling is not valid in the subthreshold region.
(2) Current flow (or charge density) in the channel is uniform from source to drain. In other words, we are using the gradual channel approximation.
(3) There is no surface roughness scattering for holes flowing in the buried $Si_{1-x}Ge_x$ layer, so that the hole mobility in the $Si_{1-x}Ge_x$ is determined only by doping, Ge content, and is given by the above equation.
(4) This method works only because there is a constant B and Ge concentrations in the $Si_{1-x}Ge_x$ layer. This will not work if the Ge profile is not rectangular, because the hole mobility in $Si_{1-x}Ge_x$ will vary with depth.

The second optimization criterion was then to maximize I_{new}.

5. Results and discussion

With the above described approach, the impact of each design parameter was evaluated:

5.1. Ge fraction and Si_{1-x} Ge_x layer thickness

Higher Ge content increases the hole mobility and improves the hole confinement through a deeper well in

Table 1
Effect of Ge content on hole confinement

Ge fraction (%)	SiGe thickness (nm)	V_{co} (V)	V_{TO} (V)	V_{hc} (V)
15	20	−1.88	−1.12	0.76
20	15	−2.02	−1.21	0.81
30	8	−2.02	−1.35	0.67

the valence band. However, to avoid strain relaxation, the thickness of the $Si_{1-x}Ge_x$ layer should not exceed the critical layer thickness which reduces with Ge fraction. Three different combinations of Ge fraction and $Si_{1-x}Ge_x$ layer thickness were simulated as shows in Table 1. All of them satisfy the Matthew–Blakeslee criterion strictly [28]. Higher Ge fractions were not used because of the unavailability of band structure model parameters for Ge fraction higher than 30%. Also, with the existing GeH_4 source that is available, the maximum Ge content that can be incorporated is 30–35%. Strained layers with thicknesses greater than the M–B criterion can be grown. However, we wanted to strictly adhere to the criterion because (1) the thermal budget at RTA is very severe (several seconds at >1000 °C), (2) MOSFETs made with 10 nm thick layers of 40% Ge (which does not satisfy M–B criterion), showed a severe performance degradation in [19]. In each of the cases shown in Table 1, the parameters $t_{Si} = 6$ nm, $N_A = N_{cap} = 5 \times 10^{17}$ cm^{-3} were held constant.

The variation in V_{hc} is modest, and it is not possible to make it −3.3 V. As the Ge content increases, the well in the valence band becomes deeper and is able to confine holes better. Therefore, V_{hc} increases when we go from 15% Ge to 20% Ge. For 30% Ge, the $Si_{1-x}Ge_x$ layer is so thin (8 nm) that it is not able to hold much of the inversion charge. In other words, even though the well is deeper, it is not wide enough, and some charge spills into the cap layer. Therefore, V_{hc} drops at $x = 0.3$. The $Si_{1-x}Ge_x$ thickness has to be high enough to contain most of the inversion charge, and low enough to allow a high Ge fraction.

These results suggest that we should use a 15 nm thick $Si_{0.8}Ge_{0.2}$ layer, because V_{hc} slightly higher than for other designs. However, with 2-D simulations, we find that there is more mobility enhancement for 30% Ge than for 20% Ge, so that it overcomes the lower V_{hc}. The I_D versus V_G–V_T curves for 20% and 30% Ge designs are shown in Fig. 6, with and without mobility correction. With mobility correction, the 30% Ge shows a higher drain current, indicating that even though holes are confined in $Si_{1-x}Ge_x$ for a lower gate voltage range for the 30%-8 nm layer than for the 20%-15 nm layer, they drift much faster. Thus, we select $x = 0.3$ and $t_{SiGe} = 8$ nm.

5.2. Cap layer thickness

A thinner cap layer results in a better hole confinement and wider operating region with transconductance remaining at its peak [15,17,29]. A thinner cap layer should also increase the gate-to-channel capacitance. A thicker cap layer, on the other hand, should give a higher off-state leakage current (I_{off}) because of reduced gate control in the subthreshold regime. This suggests using the thinnest possible cap layer.

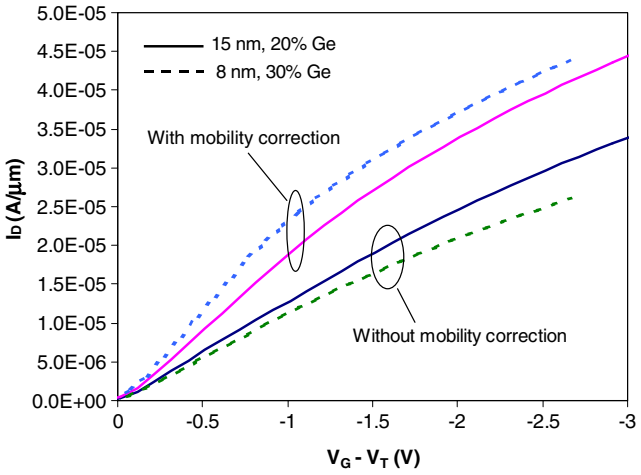

Fig. 6. I_D comparison for two Ge profiles.

We simulated hole confinement for five different cap layer thicknesses: 2, 4, 6, 8 and 10 nm, and found that the cross-over voltage and V_{hc} reduces as the cap thickness is increased, as shown in Fig. 7. Curves for 2 and 4 nm are not shown. For the 2 nm thick cap, at the end of the process, the Ge tail was found to have reached the oxide interface, leaving no Si cap. For the 4 nm thick cap, the Si–$Si_{1-x}Ge_x$ interface was found to be extremely close to the Si–SiO_2 interface, so that numerically it was not possible to integrate the hole density in the Si cap.

On the other hand, for thinner caps, the interface state density at the Si–SiO_2 interface increases [30]. It is not possible to include this effect in device simulation. These interface states reduce the subthreshold slope and degrade the performance. No D_{it} was assumed in simulations shown in Fig. 7. Secondly, as the buried channel moves closer to

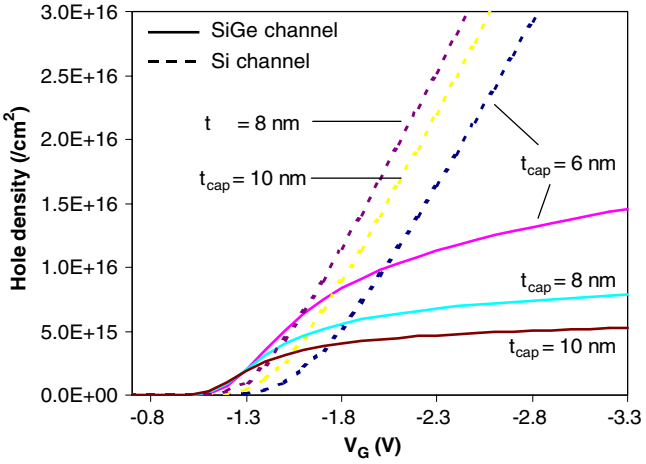

Fig. 7. Hole densities in the Si and $Si_{1-x}Ge_x$ channels for various cap thicknesses.

the interface, surface roughness scattering for holes in $Si_{1-x}Ge_x$ channel increases, and the benefit of mobility enhancement due to strain may diminish. It has been shown that the peak mobility in the buried $Si_{1-x}Ge_x$ channel increases as the cap thickness is increased [31]. Interface roughness scattering is less in devices with thicker cap layers, not only because of increased distance from the interface, but also because the lower amplitude of the Ge segregation tail at the Si–SiO_2 boundary leads to a smoother interface. Thus, there is a trade-off with the cap layer thickness.

A cap thickness down to 6–8 nm gives low values of D_{it} [30], and thus is a good compromise between hole confinement and mobility. In one study, a 2.3 nm thick cap layer was found give minimal parasitic conduction in the Si cap layer [32]. They used an exact Si etch methodology and rapid thermal oxidation to achieve this. The maximum thermal step was 120 s at 850 °C. This thin cap layer is not suitable for integration here, because of higher RTA thermal budget, which causes some Ge diffusion, and may bring some Ge tail to the Si–SiO_2 interface. Significant transconductance enhancement was found by device simulation in [33], for a 3 nm thick cap layer. However, such a thin cap layer is not practical because of high interface states and surface scattering.

Thus, for further optimization, we selected a cap layer thickness of 6 nm.

5.3. Cap layer doping concentration

We investigated the effect of cap layer doping concentration on the hole confinement. Six different values of cap layer doping density were chosen. The cross-over voltage did not show a strong dependence, as shown in Table 2. However, as the doping density in the cap layer increases, the threshold voltage reduces, and hence V_{hc} increases. Threshold voltage can be adjusted by reducing the doping density in the $Si_{1-x}Ge_x$ layer, as the threshold voltage depends upon the total charge in the p-layer [14]. In fact, we can keep the $Si_{1-x}Ge_x$ layer completely undoped and the cap layer doped. Then the doping density in the cap layer can be used to adjust the threshold voltage. This way we could have a MODMOSFET with the delta-doped layer above the $Si_{1-x}Ge_x$ channel.

However, we believe that switching from zero boron concentration in the SiGe layer, to very high boron concentration ($>1\times10^{18}$ cm^{-3}) in the cap layer will be very difficult, especially due to high growth rates and small thicknesses involved (8 nm of $Si_{1-x}Ge_x$ and 6 nm of Si cap). Therefore, we decided to keep the boron doping concentration in the cap layer same as in the $Si_{1-x}Ge_x$ layer, requiring no abrupt switching of diborane flow.

Table 2
Effect of cap layer doping density on hole confinement

Cap doping (cm^3)	V_{co} (V)	V_{TO} (V)	V_{hc} (V)
0	−2.17	−1.47	0.70
5.00E+16	−2.17	−1.46	0.71
1.00E+17	−2.16	−1.45	0.72
5.00E+17	−2.15	−1.36	0.78
7.50E+17	−2.16	−1.32	0.84
1.00E+18	−2.17	−1.27	0.91

5.4. Doping density in $Si_{1-x}Ge_x$ layer

We simulated hole confinement for four different boron concentrations in the $Si_{1-x}Ge_x$ and Si cap layers. Fig. 8 shows the hole densities in the Si and $Si_{1-x}Ge_x$ for two doping densities. V_{co} and V_{TO} both change slightly.

For a 0.35 μm CMOS technology, the threshold voltage of the PMOS is around −600 mV. Therefore, the doping concentration in the $Si_{1-x}Ge_x$ layer should be selected accordingly. More accurate V_T simulation requires 2-D device simulation, which showed that the V_T varies linearly with the doping concentration. This linear behavior is expected because for a buried channel PMOS, V_T depends linearly on the doping concentration in the top p-layer [34]. No mobility correction was applied here, as it does not change V_T significantly.

A doping concentration of 1.4×10^{18} cm^{-3} was selected to give a V_T of −620 mV, without mobility correction. Thus, the optimized values of design parameters for the SiGe PMOS are: $x = 0.3$, $t_{SiGe} = 8$ nm, $t_{cap} = 6$ nm, $N_{cap} = N_A = 1.4\times10^{18}$ cm^{-3}.

6. C–V curves

It is possible to plot total charge density ($Q = p + N_D - n - N_A$) versus V_G, as total space charge density is also one of the outputs of the device simulation. A derivative of these curves with respect to V_G gives the low frequency C–V curve of the MOS capacitor. One such curve is shown in Fig. 9.

A plateau is clearly visible in this curve. This is because as the gate voltage becomes negative, initially inversion occurs in the buried SiGe channel, and the capacitance is

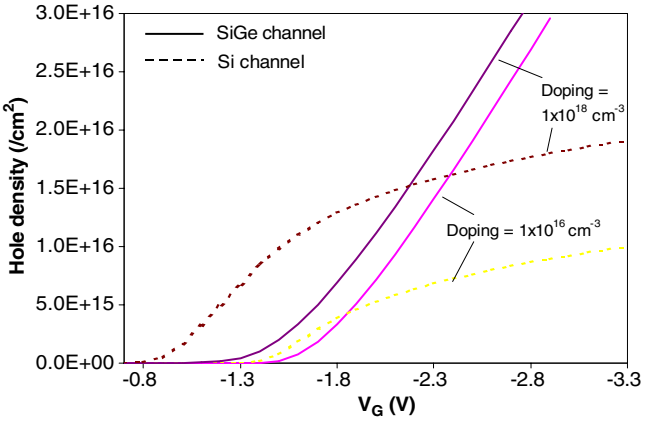

Fig. 8. Effect of doping density in $Si_{1-x}Ge_x$ on hole confinement.

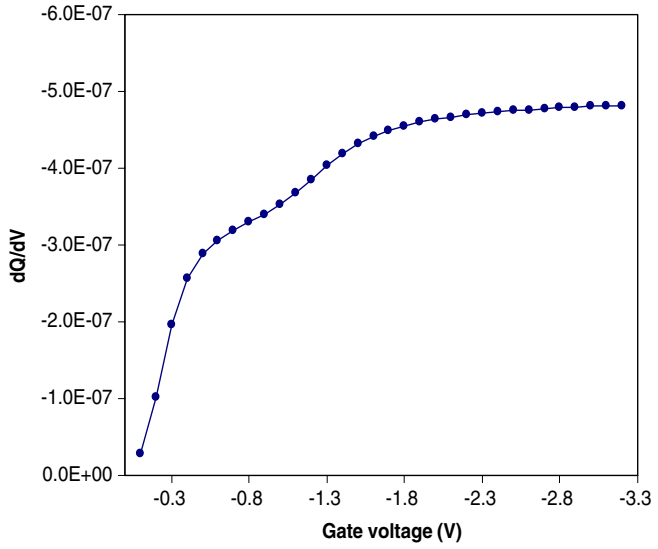

Fig. 9. Simulated low-frequency C–V curve. $N_A = N_{cap} = 1.4 \times 10^{18}$ cm^{-3}, $t_{SiGe} = 15$ nm, $x = 0.2$.

a series combination of C_{ox} and C_{cap}, and is thus lower than C_{ox}. This capacitance corresponds to the plateau. For more negative gate voltage, the hole density in the cap layer becomes dominates. For any increase in gate voltage, more holes are added in the Si cap layer. Therefore, the capacitance corresponds to C_{ox} at higher gate voltages, as in a conventional MOS device under inversion. This plateau has been observed experimentally [10,30,35]. Here it indicates that the band offsets have been simulated correctly in device simulation.

7. Comparison with standard PMOS

2-D simulations were also done for the standard PMOSFET available in the baseline process using the same set of process and device simulation models that were used for the $Si_{1-x}Ge_x$ PMOSFET. It is a conventional surface channel device with a p-type polysilicon gate. The simulated threshold voltage for the 0.35 μm standard PMOSFET is −477 mV.

Since the V_T's are different, the drain current is plotted against (V_{GS}–V_T) in Fig. 10a. I_D versus V_{GS}–V_T curves are used by others also [11,19]. Fig. 10b shows the corresponding transconductance plots for the $Si_{1-x}Ge_x$ PMOS and for the standard PMOS. Clearly, the $Si_{1-x}Ge_x$ PMOS has a >2× higher drain current and hence higher speed. Mobility correction has been taken into account for the $Si_{1-x}Ge_x$ PMOS.

Since the $Si_{1-x}Ge_x$ PMOS is a buried channel device (for $V_G < V_{co}$), one concern was that the subthreshold swing, S_t, might be higher than for the standard PMOS device, because of the reduced gate capacitance. The simulated S_t values for the standard PMOS and for the $Si_{1-x}Ge_x$ PMOS were found to be 86 mV/decade and 101 mV/decade, respectively. Thus, the $Si_{1-x}Ge_x$ PMOS turns on

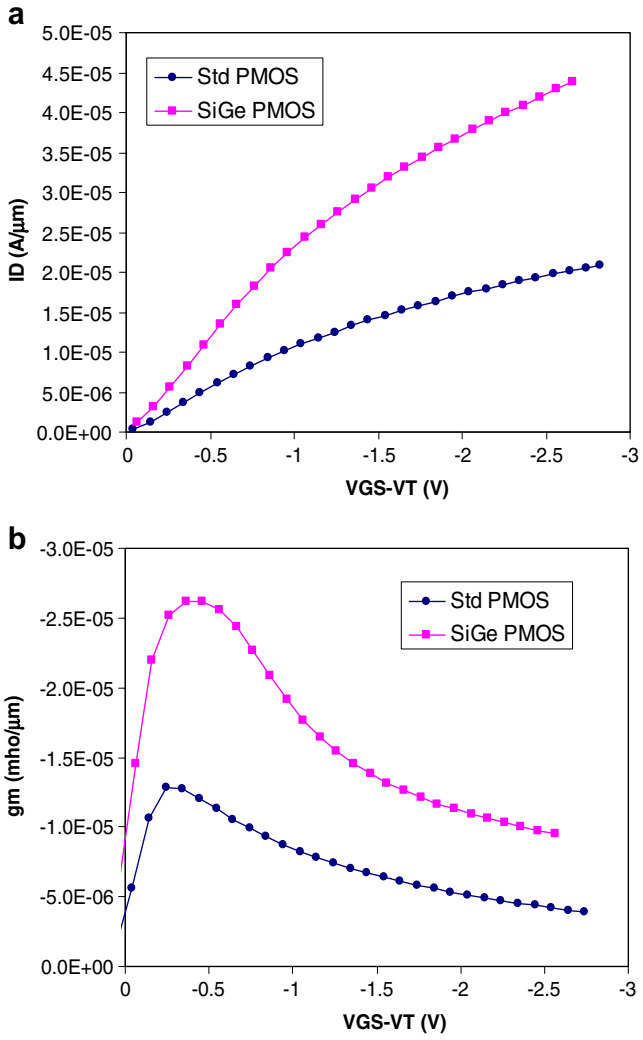

Fig. 10. Comparison of (a) drain current and (b) transconductance between standard PMOS and $Si_{1-x}Ge_x$ PMOS.

slower than the standard PMOS. However, no mobility correction was applied to the $Si_{1-x}Ge_x$ PMOS, because the model is not valid in the subthreshold region as mentioned earlier. Therefore, we believe that the actual difference in the two S_t values may be smaller than this.

8. Comparison with a reference device

A reference Si-epitaxy device was also simulated to see the effect of purely the presence of strained $Si_{1-x}Ge_x$. The reference device has same structure and same gate and doping densities as the proposed $Si_{1-x}Ge_x$ PMOSFET except that it has no Ge. Thus, Si-epitaxy was done in place of $Si_{1-x}Ge_x$ epitaxy in the simulations.

The V_T of reference device was found to be −0.819 V, which is about 180 mV higher than that for the $Si_{1-x}Ge_x$ PMOS. The $Si_{1-x}Ge_x$ PMOS has lower V_T because of the valence band offset that increases the hole density. The I_D–(V_G–V_T) curves for the two devices are shown in Fig. 11a and the corresponding g_m–(V_G–V_T) in Fig. 11b.

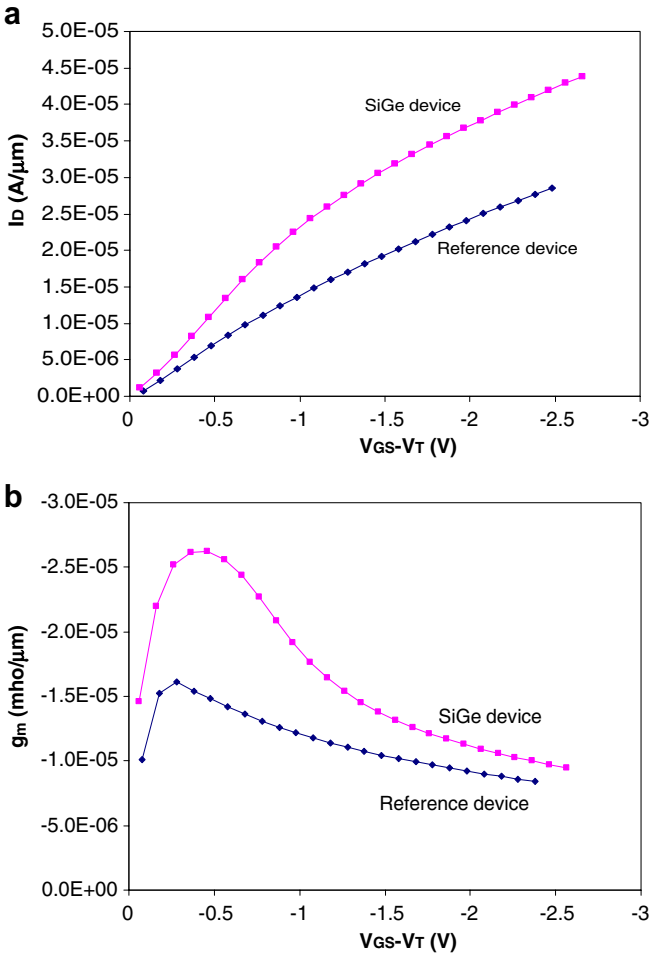

Fig. 11. Comparison of (a) drain current and (b) transconductance between reference PMOS and $Si_{1-x}Ge_x$ PMOS.

This shows that the performance enhancement in $Si_{1-x}Ge_x$ PMOS comes from the presence of the strained $Si_{1-x}Ge_x$, and not because of other device features such as a buried channel or n^+-gate.

9. Conclusion

We have proposed a new, truly buried channel strained $Si_{1-x}Ge_x$ PMOSFET that can be integrated into an existing BiCMOS process. The device structure was optimized using extensive process and device simulations. Accurate mobility modeling was done by combining published mobility data for strained $Si_{1-x}Ge_x$ with numerical simulations. Hole confinement and mobility enhancement were both found to be important in device design. An 8 nm thick $Si_{0.7}Ge_{0.3}$ layer with 6 nm thick Si cap layer and 1.4×10^{18} cm^{-3} boron doping density was found to be the best. The new device was shown to have a better performance than the standard Si device available in the baseline process, with 2× higher drive current and 2× increase in linear region transconductance. Modifications needed in the process flow have also been proposed.

Acknowledgements

The authors thank Dr. Thierry Schwartzmann and Dr. Herve Jaouen of TCAD Modeling Group in Central R&D, STMicroelectronics, Crolles, France for providing calibrated model parameters for TCAD simulations.

References

[1] Jain S. Germanium–silicon strained layers and heterostructures. San Diego: Academic Press, Inc; 1994.
[2] Maiti C, Armstrong G. Applications of silicon–germanium heterostructure devices. London: Institute of Physics; 2001.
[3] Jain S, Decoutere S, Willander M, Maes H. "SiGe HBTs for applications in BiCMOS technology: I. Stability, reliability and material parameters. Semicond Sci Technol 2001;16:R51–65.
[4] Harame D, Ahlgren D, Coolbaugh D, Dunn J, Freeman G, Gillis J, et al. Current status and future trends of SiGe BiCMOS technology. IEEE Trans Electron Dev 2001;48:2575–94.
[5] Ghani T, Armstrong M, Auth C, Bost M, Charvat P, Glass G, et al. A 90 nm high volume manufacturing logic technology featuring novel 45 nm gate length strained silicon CMOS transistors. IEDM Tech Dig 2003:11.6.1–3.
[6] Hoyt J, Nayfeh H, Eguchi S, Aberg I, Xia G, Drake T, et al. Strained silicon MOSFET technology. IEDM Tech Dig 2002:23–6.
[7] Nayak D, Woo J, Park J, Wang K, MacWilliams K. Enhancement-mode quantum-well Ge_xSi_{1-x} pMOS. IEEE Electron Dev Lett 1991;12:154–6.
[8] Subbanna S, Kesan V, Tejwani M, Restle P, Mis D, Iyer S. Si/SiGe p-channel MOSFETs. Proc Symp VLSI Tech 1991:103–4.
[9] Kesan V, Subbanna S, Restle P, Tejwani M, Altken J, Iyer S, et al. High performance 0.25 μm p-MOSFETs with silicon–germanium channels for 300 K and 77 K operation. IEDM Tech Dig 1991: 25–8.
[10] Yeo Y, Lu Q, King T, Hu C, Kawashima T, Oishi M, et al. Enhanced performance in sub-100 nm CMOSFETs using strained epitaxial silicon–germanium. IEDM Tech Dig 2000:753–6.
[11] Collaert N, Verheyen P, De Meyer K, Loo R, Caymax M. High performance Si/SiGe pMOSFETs fabricated in a standard CMOS process technology. Solid State Electron 2003;47:1173–7.
[12] Verdonckt-Vandebroek S, Crabbé E, Meyerson B, Harame D, Restle P, Stork J, et al. High-mobility modulation-doped graded SiGe-channel p-MOSFET's. IEEE Electron Dev Lett 1991;12:447–9.
[13] Voinigescu S, Salama C, Noel J, Kamins T. Optimized Ge channel profiles for VLSI compatible Si/SiGe p-MOSFET's. IEDM Tech Dig 1994:369–72.
[14] Verdonckt-Vandebroek S, Crabbé E, Meyerson B, Harame D, Restle P, Stork J, et al. SiGe-channel heterojunction p-MOSFET's. IEEE Trans Electron Dev 1994;41:90–101.
[15] Niu G, Ruan G, Zhang D. Modeling of hole confinement gate voltage range for SiGe channel p-MOSFETs. Solid State Electron 1996;39:69–73.
[16] Monroy A, Laurens M, Marty M, Dutartre D, Gloria D, Carbonero J, et al. BiCMOS6G: a high performance 0.35 μm SiGe BiCMOS technology for wireless applications. Proceedings of the 1999 Bipolar/BiCMOS Circuits and Technology Meeting 1999:121–4.
[17] Niu G, Ruan G. Threshold voltage and inversion charge modeling of graded SiGe-channel modulation-doped p-MOSFET's. IEEE Trans Electron Dev 1995;42:2242–6.
[18] Wolf S. Silicon processing for the VLSI Era. The submicron MOSFET, vol. 3. USA: Lattice Press; 1995.
[19] Garone P, Venkataraman V, Sturm J. Hole mobility enhancement in MOS-gated Ge_xSi_{1-x}/Si heterostructure inversion layers. IEEE Electron Dev Lett 1992;13:56–8.
[20] Cowern N, Zalm P, Sluis P, Gravesteijn D, de Boer W. Phys Rev Lett 1994:2585.

[21] Dutartre D. Excitonic photoluminescence from Si-capped strained $Si_{(1-x)}Ge_x$ layers. Phys Rev B 1991;44:11525–7.
[22] People R. Indirect band gap of coherently strained $Ge_xSi_{(1-x)}$ bulk alloys on ⟨001⟩ silicon substrates. Phys Rev B 1985;32:1405–8.
[23] Jain S, Decoutere S, Willander M, Maes H. SiGe HBTs for applications in BiCMOS technology: I. Stability, reliability and material parameters. Semicond Sci Technol 2001;16:R51–65.
[24] Kuo J, Chen B, Chen H, Lu T, Sim J. A numerical simulation study of SiGe/Si-heterostructured pMOS and bipolar devices. VLSI Process Dev Model 1993:104–5.
[25] STMicroelectronics internal communication.
[26] Carns T, Chun S, Tanner M, Wang K, Kamins T, Turner J, et al. Hole mobility measurements in heavily doped $Si_{1-x}Ge_x$ strained layers. IEEE Trans Electron Dev 1994;41:1273–81.
[27] Manku T, McGregor J, Nathan A, Roulston D, Noel J, Houghton D. Drift hole mobility in strained and unstrained doped $Si_{1-x}Ge_x$ alloys. IEEE Trans Electron Dev 1993;40:1990–6.
[28] Matthews J, Blakeslee A. Defects in epitaxial multilayers – I. Misfit dislocations. J Cryst Growth 1974;27:118–25.
[29] Inieswki K, Voinigescu S, Atcha J, Salama C. Analytical modeling of threshold voltages in p-channel Si/SiGe/I MOS structures. Solid State Electron 1993;36:775–83.
[30] Iyer S, Solomon P, Kesan V, Bright A, Freeouf J, Nguyen T, et al. A gate-quality dielectric system for SiGe metal-oxide-semiconductor devices. IEEE Electron Dev Lett 1991;12:246–8.
[31] Palmer M, Braithwaite G, Grasby T, Phillips P, Prest M, Parker E, et al. Effective mobilities in pseudomorphic Si/SiGe/Si p-channel metal-oxide-semiconductor field-effect transistors with thin silicon capping layers. Appl Phys Lett 2001;78:1424–6.
[32] Sareen A, Wang Y, Södervall U, Lundgren P, Bengtsson S. Effect of Si cap layer on parasitic channel operation in Si/SiGe metal-oxide-semiconductor structures. J Appl Phys 2003;93:3545–52.
[33] Maiti C, Chakrabarti N, Ray S. Strained silicon heterostructures: materials and devices. London: The Institution of Electrical Engineers; 2001.
[34] Lee Y, Woo D, Lee J, Park B. Threshold voltage reduction model for buried channel pMOSFETs using quasi-2-D poisson equation. IEEE Trans Electron Dev 2000;47:2326–33.
[35] Garone P, Venkataraman V, Sturm J. Hole confinement in MOS-gated Ge_xSi_{1-x}/Si heterostructures. IEEE Electron Dev Lett 1991;12:230–2.