Abstract

Device Improvement with strain engineering is considered a way to enhance the carrier mobility. Several stress-transfer techniques (such as etch-stop liner, stress transfer technique, e-SiGe) using extra integration process into an existing baseline process is demonstrated. In addition, new preparation techniques of strained-Si surface (e.g. biaxial tensile stress) and different substrate orientation to enhance mobility are introduced. The challenges and vitality of each method will be discussed and compared. In addition, we will highlight how the stress oriented from the layout geometry affects the device electrical behavior. The issues and improvement in the circuit level device modeling will be discussed.

Shrinking the Transistor

In the coming decade, advances in CMOS fabrication will lead to devices with gate lengths from 45 to below 10 nm. Shrinking size can pack more devices and improves the logic switching speed. The challenges in reducing device sizes are the feasibility of fabrication, maintaining performance and reliability, as well as gate oxide thinning. After more than two decades of continual and rapid progress in CMOS devices technology, especially aggressive scaling and significant pace of scaling in the last few years, the scaling will apparently approach the fundamental limits of scaling, the feasibility of fabrication and device reliability. Besides scaling, several innovative mobility enhancement techniques are suggested to maintain the MOS performance improvement trajectory laid out in the ITRS roadmap.

Strain helps carriers to travel faster

Mobility enhancement is an attractive option because it can potentially improve device performance beyond any of the benefits from device scaling. Compressive stresses are introduced by Shallow Trench Isolation (STI) towards channels longitudinally and laterally. Mechanical stress breaks crystal symmetry and removes the 2-fold and 6-fold degeneracy of the valence and conduction bands respectively. This will lead into changes of the band scattering rates and / or the carrier effective mass, which in turn affect carrier mobility [1, 2]. However, NMOS and PMOS carriers have different interaction on the channel strain in three directions, namely longitudinal, lateral and Si depth directions. Fig. 1 illustrates possible directions of uniaxial stresses and their effects on N and PMOS. Different mechanical stresses can be introduced through the wafer substrate and during the fabrication process, which will be discussed below. NMOS current is enhanced with tensile stress in channel, and PMOS one is with compressive stress. 40% difference may be resulted if the longitudinal stress is applied (Fig 2).

Strain for CMOS performance Improvement

'Victor chan, 'Ken Rim, 'Meikei Ieong, 'Sam Yang, 'Rajeev Malik, 'Young Way Teh, 'Min Yang, 'Qiqing (Christine) Ouyang

' IBM Systems & Technology Group, 'IBM Research Division, T. J. Watson Research Center and 'Chartered Semiconductor Mfg. Ltd, at IBM Semiconductor Research and Development Center (SRDC)

email: vchan@us.ibm.com

Shrinking the Transistor

In the coming decade, advances in CMOS fabrication will lead to devices with gate lengths from 45 to below 10 nm. Shrinking size can pack more devices and improves the logic switching speed. The challenges in reducing device sizes are the feasibility of fabrication, maintaining performance and reliability, as well as gate oxide thinning. After more than two decades of continual and rapid progress in CMOS devices technology, especially aggressive scaling and significant pace of scaling in the last few years, the scaling will apparently approach the fundamental limits of scaling, the feasibility of fabrication and device reliability. Besides scaling, several innovative mobility enhancement techniques are suggested to maintain the MOS performance improvement trajectory laid out in the ITRS roadmap.

Strain helps carriers to travel faster

Mobility enhancement is an attractive option because it can potentially improve device performance beyond any of the benefits from device scaling. Compressive stresses are introduced by Shallow Trench Isolation (STI) towards channels longitudinally and laterally. Mechanical stress breaks crystal symmetry and removes the 2-fold and 6-fold degeneracy of the valence and conduction bands respectively. This will lead into changes of the band scattering rates and / or the carrier effective mass, which in turn affect carrier mobility [1, 2]. However, NMOS and PMOS carriers have different interaction on the channel strain in three directions, namely longitudinal, lateral and Si depth directions. Fig. 1 illustrates possible directions of uniaxial stresses and their effects on N and PMOS. Different mechanical stresses can be introduced through the wafer substrate and during the fabrication process, which will be discussed below. NMOS current is enhanced with tensile stress in channel, and PMOS one is with compressive stress. 40% difference may be resulted if the longitudinal stress is applied (Fig 2).

The two main approaches being pursued are strain engineering (from both substrate- and process-induced) and orientation effects.

Substrate-induced strain. One of the most effective way to introduce high tensile strain to the channel is to epitaxially grow strained-silicon on a relaxed SiGe layer (Fig 3a). Because of the lattice mismatch between Si and SiGe, the lattice of the Si layer is stretched (strained) in the plane of the interface. This deformation breaks the symmetry of the energy band structure and results in band splitting (Fig. 4). The reduced inter-band/inter-valley scattering and effective masses result in enhanced carrier transport in the strained silicon layer that is used as the channel of the MOSFET[2]. Enhanced drive currents of 15 to 25% have been demonstrated on sub-100 nm bulk strained-silicon MOSFETs (Fig. 5) [2, 3].
The performance benefit of combining strained silicon with an SOI has also been demonstrated in a 60 nm gate-length, n-channel MOSFET with ultrathin thermally mixed strained silicon/SiGe on insulator substrate (strain-Si on insulator SSOI, SiGe on insulator SGOI) (Fig 3b) [4]. 20-25% device enhancement is demonstrated at short channel length.

Recently, Rim et al. [5] demonstrated transistors using ultrathin strained silicon directly on insulator (SSDOI) structures (Fig 3c) that eliminate the SiGe layer before transistor fabrication, thereby providing higher mobility while mitigating the SiGe-induced material and process integration problems. A SSOI structure is fabricated by a layer-transfer or “wafer-bonding” technique. First, an ultra-thin layer of silicon with an SOI has also been demonstrated in a 60 nm channel in any one of the three dimensions by different process techniques [6,7]. 65nm bulk and SOI CMOS technologies with 45nm channel length and 12A gate oxide are used for study. Different stress-transfer techniques, such as contact etch stop liner, Stress Memorization Technique, and embedded SiGe in the source/drain (S/D) will be used for study. Different stress-transfer techniques, such as PECVD (Plasma Enhanced Chemical Vapor Deposition) or “wafer-bonding” technique. First, an ultra-thin layer of silicon with an SOI has also been demonstrated in a 60 nm channel in any one of the three dimensions by different process techniques [6,7]. 65nm bulk and SOI CMOS technologies with 45nm channel length and 12A gate oxide are used for study. Different stress-transfer techniques, such as contact etch stop liner, Stress Memorization Technique, and embedded SiGe in the source/drain (S/D) will be used for study. Different stress-transfer techniques, such as PECVD (Plasma Enhanced Chemical Vapor Deposition)
Deposition) or RTCVD (Rapid Thermal CVD) nitride, is uniformly deposited over the wafer. The drive current dependence on stress is through the change in film thickness and material [8]. The remaining steps follow a conventional process flow, including interlayer dielectric and contact formation. Thicker nitride capping layer can increase the stress level. However, this will impact the conformity of gap fill itself, subsequent interlayer dielectric (ILD) gap fill, and contact opening process. 

Fig. 6 shows the drive current improvement for NMOS with tensile stress and PMOS with compressive stress liner [9]. Tensile liner improves NMOS current by 11% (and 17% after self-heating correction) and compressive liner improves PMOS current by 20% than that of the non-stressed process.

If one single liner is used, one drawback of this approach is that the device of the opposite type will be degraded. One of the possible solutions is to use high dose of Ge implantation to destroy the nitride film and relax its stress [10]. For example, a highly tensile stress nitride film is deposited over both N & PMOS, and Ge is selectively implanted into PMOS region to relax the nitride stress. It is advised that the subsequent processes to have low thermal budget to prevent the stress stability of the liner.

The above method cannot provide both NMOS and PMOS to their highest performance. In order to achieve ultimate CMOS performance, two types of stress liners should be applied to NMOS and PMOS accordingly. For example, a highly tensile nitride liner is deposited first. It is selectively etched on the PMOS regions. Next, a highly compressive nitride liner is deposited, and this film is also selectively etched on the NMOS region. This process can be applied in the reverse order. Fig. 7 shows the SEM cross-section taken from a SRAM cell with the Dual Stress Liner (DSL) process [9]. Compared to non-stressed process, DSL process results in a 7% improvement with DSL compared to single tensile liner. AMD Athlon™ 64 microprocessor [9] F_max vs. power improves 12 % with DSL compared to non-stressed liner.

b) Stress Memorization Technique.

Local strain can be applied to the channel through a Stress Memorization Technique. In the conventional fabrication process, Source / Drain Si area and poly gate are amorphized by S/D and extension implantation. In the stress memorization process, conventional dopant activation spike anneal is performed after the deposition of a tensile stressor capping layer, such as nitride (Fig 8) [11-13]. The nitride layer will subsequently be removed, and the next step is salicidation. The stress effect is transferred from the nitride film to the channel during annealing and memorized by the re-crystallization of the S/D and poly gate amorphized layers. After active area and gate re-crystallization, the stress inside the channel is memorized. Since the nitride film is disposable, a very thick capping layer can be used to increase the stress level without any process limitation. A 15% of NMOS improvement is demonstrated when this modular stressor element is added on top of a etch-stop liner, as shown in Fig. 9. The amount of drive current improvement is related to the amorphous layer, anneal condition and stress level of the nitride liner.

c) Embedded SiGe in S/D

Besides the introduction of nitride stressor, a more direct way to provide strain to channel is filling S/D with SiGe. One suggestion is, epitaxial-grown strained SiGe (e-SiGe) is embedded in the S/D regions [14-16,26] or extension location [17] (Fig. 10). When the lattice spacing of this SiGe material is larger than Si, uniaxial compressive strain is induced to the channel. Extra integration processes are inserted after spacer formation, such as selective Si recess
etch in the S/D regions, and selective and insitu-doped epitaxial SiGe growth. NMOS was protected by capping layer to prevent Si recess and SiGe epitaxial growth. This structure creates a large uniaxial compressive stress to the PMOS channel, thereby resulting in significant hole mobility improvement. An apparent drive current improvement from Intel published data is shown in Fig. 11.

The compressive stress is mainly dependent on the e-SiGe thickness in the S/D, both below and above (raise S/D) the Si surface, and % Ge content in the SiGe. Too high Ge content may cause defects and yield issue. Moreover, the improvement will be also degraded for small S/D overhang regions ((distance from poly gate to STI edge), Further optimization in process is necessary to reduce the adverse effect [15].

**Implications**

The electron and hole mobility has different dependence on the type of uni-axial stresses and this has been discussed in Fig. 2. Device drive current can be effectively modulated by mechanical built in stresses. Careful selection on process modular enhancer is necessary to prevent enhancement of one type and degradation of another type of carriers, and the expense of higher process complexity, cost and integration yield. A complete stress model on device parameters, especially Ion and Vt, is recommended for an accurate circuit simulation and verification.

Strained Si begins to be implemented in 65 and 90nm technologies. Fig. 12 shows the IonN / IonP ratio in the recent three generations. The enhancement of drive current from strained Si process will influence the rise and fall time of circuit logic stages and hence overall performance. Applying strain enhancement to both N and PMOS can adjust the ratio back to the appropriate levels, such as DSL (Fig. 7).

Moreover, Vt shift associated with shift in band edges due to channel strain is seen on both NMOS and PMOS...
devices (Fig 13). The uni-axial stress produced by the process transfers the stress through S/D overhang regions and poly gate to the channel resulting variation of stress along the channel. Hence, the Vt shift is dependent on both channel length and active area geometry. Appropriate Vt adjustment through channel or halo implant can compensate this.

When the technology reaches deep submicron regime and the active area becomes small, the drive current can be modulated by the mechanical stress from the STI. The source is mainly from STI corners. Longitudinal stress can be modulated by different (e.g. irregular or asymmetrical) S/D overhang regions, whereas different lateral stress can be obtained by different channel widths and adjacent active area. Figs. 14-15 investigate the effect of STI proximity to the gate edge in device performance [1,18-20]. PMOS exhibits higher Ion at constant Ioff with smaller S/D overhang regions i.e. higher compressive stress from STI in longitudinal direction, whereas NMOS Ion is decreased. When the channel width is reduced, NMOS is less sensitive to the stress effect; on the contrary, PMOS becomes more sensitive to it. The stress varies along the channel and hence different channel lengths received different stress effects. STI proximity has minimal effect on long channel devices (Fig. 14). As the gate length reduces, the STI stress effect becomes significant, but reduces when the gate length is nominal and short. This suggests that the drive current of the nominal devices is controlled by the halo dose and carrier scattering.

From Figs. 14-15, devices always have STI stress proximity effects, especially in small S/D overhang regions. The effects can be reduced by process material but cannot be fully eliminated. In general, > 50% devices with 2x minimum distance of gate to STI edge in a product [1]. Therefore, understanding the phenomenon and fundamental behavior can help to build and develop an appropriate circuit level device model, rather than using isolated test structure as baseline model.

From the discussion above, stress engineering induced from process integration will affect the Vt and drive current. This will also affect the SRAM design, such as the cell stability (access disturb margin, ADM) and writability (write margin, WRM). ADM is bigger with weak NMOS and strong PMOS, whereas bigger WRM comes from strong NMOS and weak PMOS. SRAM operation margin (ADM & WRM) vary inversely with device current tolerance. If the drive current tolerance increases, the operation margins will be degraded. Besides stability margin and write margin, read and write performance are also important. Gamma Ratio of \( \frac{I_{\text{passgate}}}{I_{\text{pull-up}}} \) and Beta Ratio of \( \frac{I_{\text{pull-down}}}{I_{\text{passgate}}} \) must be adjusted to meet the operation requirements. If NMOS current becomes higher, Beta Ratio remains the same and Gamma Ratio increases. The WRM is better but ADM will be degraded. On the other hand, if PMOS current becomes higher, Gamma ratio is reduced. The cell will be written more slowly. ADM will be better and more stable. Thus, Vt and current have to be adjusted when stress engineering is implemented. In addition to conventional Vt-adjustment and halo implantation correction to modulate the drive and standby current, stress-transfer modular process is important. For example, the stress level in the nitride film during SMT process and etch stop liner in the DSL process can affect both drive current and Vt.

The annealing temperature is also critical in the stress transfer process. It will also affect the device reliability, such as hot carrier concern. Moreover, different methods of stressor deposition, such as nitride film, may provide different...
conformity and uniformity; hence affect the stress level in the device channel. Furthermore, many stress-induced processes involve lithography and etching. The new processes may violate the minimum design ground-rule, such as lithography overlay margin and etch bias in the cell layout design of the SRAM cells.

As discussed above, stress-induced process can impact device variances. This can translate into impact on circuit performance and yield. Circuit and layout designers should understand the impact of any new stressed-transfer processes and the chance of under-performance devices. Layout of devices in a critical path should be drawn in a less stress-sensitive way, for example, larger S/D overhang regions (Figs. 14-15).

The fabrication line should implement appropriate methods to monitor the process, such as in-line nitride deposition thickness and film stress, to control the uniformity and repeatability of stress transfer process on large amount of transistors. Devices with different stresor process (such as tensile liner and neutral liner on two different devices correspondingly) are built on same location of a wafer. From the electrical data, stresses under within chip variation, within wafer variation, within lot variation and lot to lot variation can be monitored. Furthermore, devices with different active area (and poly gate) geometries, sizes, and density should be monitored to check the micro-loading effect from film deposition and etching.

**Mobility improvement beyond stress engineering**

Besides substrate-induced and process-induced stress engineering, wafer substrate orientation and channel orientation can improve mobility. Different surface orientation and direction of applied field for different in-plane stress provide different interaction with carrier transports [21].

**a) Substrate Orientation**

Besides strained Si in the traditional (100) plane, it may be advantageous to change the crystal orientation to optimize CMOS circuit performance. The carrier mobility of inversion layers depends on surface orientation and current flow directions, due to asymmetry of the carrier effective masses in the Si crystal lattice [22]. The definition of substrate orientation and carrier direction is described in Fig. 16. For PMOS, hole mobility is 2.5 times as high on (110) surface orientation as on a standard (100) wafer surface, depending on the applied effective vertical field. However, electron mobility is degraded on the (110) substrate. Similar to devices fabricated on (100) substrate, those on (110) substrate are sensitive to longitudinal stress and have similar stress characteristics (Fig. 17).

To fully realize the advantage of the carrier mobility dependence on surface orientations, new design and integration have been developed for the Hybrid Orientation Technology (HOT). NMOS is fabricated on the Si substrate with (100) surface orientation and PMOS on (110) surface orientation [23, 24] (Fig. 18). Two types of hybrid substrates are shown in Fig. 19: Type A with PMOS on (110) SOI and NMOS on (100) Si epitaxial layer, and Type B with NMOS on (100) SOI and PMOS on (110) Si epitaxial layer. The hybrid substrate was formed by layer transfer technique through wafer bonding (Fig. 19). Layer transfer process, block-level trench etch and epitaxial re-growth were performed before the conventional CMOS device process (Fig. 20).

CMOS on the hybrid orientation substrates were fabricated with NMOS on (100) SOI and PMOS on (110) epi-Si (Fig. 18). The drive current of PMOS has 30-40% improvement compared to device on (100) substrate (Fig. 21). There is no NMOS degradation in HOT and the behavior is similar to those on (100) wafers. Other device characteristics such as Vt roll-off, junction capacitance and leakage and extension overlap region are very similar between (110) and (100) substrates. These indicates similar dopant diffusion rate in both substrates. With great PMOS current improvement, ring oscillator delay can be improved up to 21% [24, 25], dependent on channel length, fan-out and circuit components.

Further performance gain can be achieved by integrating HOT technology with strained Si, for example, e-SiGe [26], and DSL [27]. Fig. 22 shows our recent best PMOS performance with (110) and stress engineering. Drive current of 720µA/um at Ioff=100nA/um is observed.

**Fig. 17 Different longitudinal stress in channel provides different drive current in nominal devices in a) (100) and b) (110) substrate surfaces.**

**Fig. 18 Schematic cross-section and SEM of CMOS on a hybrid-substrate with PMOS on (110) surface orientation and NMOS on (100) surface orientation [24].**

![Fig 16 Devices are fabricated on two wafer orientations. Current flow direction relative to wafer flat is indicated.](Image)
b) Channel Orientation Design and Stress Control

On a (100) Si substrate wafer, carriers under the poly gate can flow in different directions (Fig. 23). Holes in the <100> channel has less sensitivity on mobility from channel stress than <110> channel, as shown in Fig. 24 [28, 29]. However, electrons mobility have the same sensitivity from the stress in both <100> and <110> channels. This advantage brings the technology using higher tensile stress (e.g. contact etch-stop liner) to improve NMOS mobility, and PMOS mobility is not degraded even without selective etch or implantation during stress-transfer process. This can reduce the fabrication complexity. No extra substrate preparation is required, except that appropriate wafer notch is considered.

Summary and Trade-offs

In the nano-technology era, strained Si will be added to continue the Moore’s Law. Mobility enhancement technique is attractive and has been discussed above. Some options can be combined for even higher performance gains.
In fact, some of them have been integrated in a modular fashion into any existing baseline process for integrated circuit manufacturing. Extra process complexity, cost and yield should be carefully considered (Table 1). In addition, extra process steps may involve extra film deposition or etch. Extra plasma damage and low temperature process may lead to reliability issues.

**Biaxial tensile strained Si** is considered as a technique in the last decade. However, it has been difficult to implement because of misfit and threading dislocations, Ge out-diffusion, silicide formation difficulty, self-heating effect, higher As diffusion in the S/D extension region, and cost.

**DSL** (contact etch stop layer) provides enhancement in both NMOS and PMOS. It needs extra steps and ground rule in the liner boundary and contact opening may be a concern. This has been implemented in IBM PowerPC™ and AMD Athlon™.

**SMT** (Stress Memorization Technique) needs extra steps and is easy to be implemented. It does not have any major process or ground rule limitation. However, it can only improve NMOS performance.

**e-SiGe in S/D structure** provides highly compressive strain to the channel, and results in large hole mobility improvements. Extra complicated integration steps are necessary; nevertheless, it can provide very good PMOS performance.

**HOT** technology ((110) substrate) clearly shows an attractive option to improve device performance. However, the impact on circuit performance of mixing SOI and bulk devices on the same chip and circuit design will require more detailed analysis.

**Channel orientation** (<100> channel, 45 deg notch) is a highly manufacturable technique to reduce integration complexity. It allows NMOS improvement by other stress-transfer process, without any PMOS degradation. However, PMOS performance cannot be further improved by other stress transfer method.

Ultimate circuit level device model should include all the stress effects, including different channels and widths, S/D overhang regions, distance from adjacent active area in both longitudinal and transverse direction, and their sizes. Combination of different uni-axial stress from different directions will form bi-axial stress effects, leading the model to be more complicated.

The uniaxial channel stress is induced from the sides of the device. Devices with different channel lengths have different stress effects. Many mobility enhancement methods involve uni-axial stress through the active area, such as S/D area. The improvement may reduce from ideal isolated devices to devices inside a product.

**Acknowledgements**

This work is supported by the independent alliance programs for SOI Technology Development (Sony Electronics Inc., Toshiba America Electronic Components, Inc. and Advanced Micro Devices, Inc.), Bulk CMOS Technology Development (Chartered Semiconductor Mfg. Ltd, Infineon Technologies AG and Samsung Electronics Co. Ltd) and IBM Yorktown Research Division. The authors would like to thank Scott Crowder and Lisa Su for their support of this work.

**References**

[27] C. Sheraw et al., *VLSI Symposium*, June 2005, accepted

<table>
<thead>
<tr>
<th>NMOS</th>
<th>PMOS</th>
<th>Disadvantage &amp; limitation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bi-axial Tensile Strain</td>
<td>↑</td>
<td>Extra cost on substrate, difficulty in material, integration, &amp; device design.</td>
</tr>
<tr>
<td>Contact etch-stop liner (DSL)</td>
<td>↑</td>
<td>Extra steps in integration, ground rule consideration</td>
</tr>
<tr>
<td>SMT</td>
<td>↑</td>
<td>Extra steps in integration</td>
</tr>
<tr>
<td>e-SiGe</td>
<td>o</td>
<td>High process complexity, difficulty in epi-growth, yield</td>
</tr>
<tr>
<td>Substrate Orientation (HOT)</td>
<td>o</td>
<td>Extra Cost on hybrid substrate, extra steps eg. Epi to prepare isolation</td>
</tr>
<tr>
<td>Channel Orientation (&lt;100&gt;)</td>
<td>o</td>
<td>No further improvement in PMOS current</td>
</tr>
</tbody>
</table>

Table 1 Summary of all feasible technologies for mobility enhancement. Extra processes are considered, such as selective etch or implantation to prevent performance degradation.