High-tilt implant and diffusion-less activation for lateral graded S/D engineering

**OVERVIEW**

As the industry continues to scale devices to meet the various system-on-a-chip applications in the future, several types of transistor designs from planar and vertical single-gate, double-gate, and multigate devices are emerging, as are several options in silicon starting material (bulk CZ, epi, blanket SOI, and selective/patterned SOI wafers) [1–7]. Although planar single-gate device scaling has been demonstrated down to a gate length ($L_g$) of 6nm, the end of the roadmap for planar single-gate CMOS seems to be drawing nearer as the industry increases research activities in double-gate and multigate CMOS devices [8]. Multiple device roadmaps that depend on specific applications are also emerging and both logic and memory devices will migrate from planar structures to vertical structures.

One of the critical areas in future transistor design scaling for improved device performance will be lateral source/drain (S/D) dopant engineering, including lateral abruptness, gate overlap control, and lateral dopant activation. Recent advances using high-tilt implantation and diffusion-less activation starting at the 65nm node and extendable to at least the 22nm node ($L_g = 13nm$) based on the roadmap (see the table) will be described [9].

**Single-gate structure**

There is no fundamental change in the design of the planar single-gate (SG) transistor to about 2015. However, changes for improved device performance are being made for its scaling and extendibility. The current oxynitride gate dielectric material will first evolve to a medium-$k$ dielectric of 12 (HiSiON, HfO$_2$/Al$_2$O$_3$, etc.) at the 65nm node and then to a high-$k$ of 25 (HfO) by the 45nm node. High-temperature diffusion-less activation (flash/RTA [rapid thermal anneal] or submelt laser annealing) can be used with medium-$k$ gate dielectrics, but once high-$k$ gate dielectrics are implemented, only low-temperature, diffusion-less activation, solid phase epitaxy (SPE) can be used to maintain their amorphous phase and high-$k$ integrity (Fig. 1) [9–11]. The boron solid solubility ($B_{ss}$) dopant activation values based on a box profile shape were calculated (atoms/cm$^2$) and identified (see the table and Fig. 1). Possible design evolutions of the transistor channel and S/D regions from the 90nm node through the 32nm node were collected from the top 10 logic companies worldwide [12; Fig. 2].

**300mm silicon wafer options**

For improved silicon material quality at 300mm, a majority of IC manufacturers have switched to epi wafers to maximize yield. Growing an epitaxial silicon layer on top of bulk Cz wafers eliminates grown-in defects such as crystal-originated precipitate [13, 14]. For the same reason, silicon-on-insulator (SOI) wafers are also using a top epitaxially grown silicon layer for both the SIMOX and wafer-bonding manufacturing methods in order to achieve the highest SOI silicon layer quality [15].

One of the main challenges for thin SOI wafer manufacturing is scaling the film thickness with each technology node while maintaining tight uniformities. At the 180nm node, the SOI film thickness for partially depleted (PD) SOI devices was 160–200nm. For the 130nm node, the SOI film thickness was reduced to a range between 79–120nm and for the 90nm node, the film thickness range is expected to be between 30–60nm. For this reason, PD/SOI film thickness is approximately equal to $L_g$ [2, 3, 5, 6]. With thinner SOI films, tighter control on thickness uniformity is critical since it directly defines the device channel region and variations will directly impact $V_{th}$ control and ring oscillator delay as reported by Shahidi [6]. For fully depleted (FD) SOI devices, the SOI film thickness is targeted to be about 0.2–0.3$L_g$, so at the 65nm node with $L_g = 35nm$, the SOI film is expected to be only 7–10nm thick [2, 3]. Very thin SOI film thickness uniformity and control for these fully depleted devices will be a major challenge, especially for bonded SOI wafer manufacturers in the future. Also, as the SOI top silicon film gets thinner, S/D parasitic resistance goes up and contacting silicide becomes a problem requiring the use of elevated S/D structures.

For embedded memory and other system-on-a-chip (SOC) applications, logic devices on SOI material and memory devices on bulk material may be preferred, creating a new classification of SOI wafers called selective/patterned SOI wafers. Two different approaches to realize logic devices in SOI and DRAM trench capacitor memory devices in bulk (non-SOI) have been recently reported. Researchers at IBM have used the SIMOX method with
a patterned mask for the oxygen implant, resulting in regions on a bulk wafer with patterned SOI structures [5]. At Toshiba, researchers used the bonded SOI wafer approach with patterned etch back through selected SOI regions down to the bulk wafer, followed by selective epitaxial growth in these non-SOI regions, forming selective SOI [16]. Another SOC embedded device application is the bipolar RF device, where the vertical SiGe bipolar device can be on bulk (non-SOI) or SOI material [6].

Channel and S/D engineering

Starting at the 90nm node, higher-dose HALOs (super HALO) will be used for short channel effect (SCE) control, reducing \( V_T \) roll-off [12; Fig. 2]. As \( L_g \) scales, however, the HALOs will become overlapped (Fig. 2), raising the channel-doping level into the mid- to upper \( \text{E}18/\text{cm}^2 \) levels [1]. Impurity scattering occurs as the channel-doping level becomes >2\( \text{E}18/\text{cm}^2 \), thereby degrading channel mobility and leading to the adoption of the strained-Si channel [17, 18]. The best strained-Si on SiGe epitaxial layer has about \( 10^5 \) defects/cm\(^2\), so compromises have been made to achieve strained-Si by combination of other techniques such as 1) localized Ge implantation of a few percent; 2) strain-induced CVD layer deposition; and 3) strain-induced STI formation. The use of strained Si technology can be avoided with the lateral graded S/D terraced (quad/multimode implant) or wedged (rotation-mode implant) structure improving lateral abruptness, gate overlap control, SCE, and \( R_{\text{ext}} \) thereby reducing the HALO dose [4].

Current transistor designs use separate shallow source/drain extension (SDE) junction and deep S/D contacting junction implants and masks. After efforts to maximize device drive current, Ghani et al. reported on the criticality of the optimization of SDE vertical junction depth \( (X) \) and lateral junction depth \( (Y) \) called gate overlap/under diffusion and lateral abruptness [17]. To control the SDE gate overlap precisely, off-set spacers were introduced to compensate for excessive lateral diffusion for both bulk and SOI devices becoming standard at the 90nm node [19-21]. The updated roadmap (see the table) has the new target values for \( X \), \( Y \), lateral abruptness, and box profile activated dopant concentration for the most advanced logic devices taken from several end-user inputs.

For high-performance (HP) logic devices using oxynitride gate dielectric with poly electrode, high-temperature dopant activation annealing can be used for the SDE, deep S/D, and poly-gate-doping activation and diffusion (>10nm). Starting at the 65nm node, however, the ultra-shallow junction (USJ) requirements for SDE necessitate diffusion-less (<3.0nm) dopant activation annealing at either high temperatures or low temperatures, while the deep S/D and poly-gate doping will still require some dopant diffusion (>10nm). Therefore, from a process integration point of view, several companies will adopt the disposable spacer process flow also called reverse S/D for SDE. This adoption will allow the use of traditional high-temperature spike/RTA annealing with >10nm dopant diffusion for the deep S/D and gate poly doping. Following this will be either high-temperature flash (<4nm dopant diffusion), submelt laser annealing (<1nm dopant diffusion), or low-temperature SPE (solid phase epitaxy) annealing (no dopant diffusion) for diffusion-less dopant activation to achieve shallow and abrupt SDE junctions with gate overlap control (Fig. 2).

For low-power logic devices using high-k gate dielectrics and metal electrodes starting at the 45nm node, low-temperature dopant activation techniques will be required to prevent crystallization of the amorphous high-k gate material. Fortunately, the acceptable levels of junction leakage current for HP and low-operating-power (LOP) logic devices are within the acceptable range observed with
Further improvements in device performance can also be achieved with single-S/D, the merging of shallow SDE junction and deep S/D contact junction into a single junction as reported by Ozterk et al. [22]. Single-S/D structures will require the use of elevated S/D contacting structures in order to achieve low-S/D parasitic resistance and good salicide contacting just as PD/SOI devices at 90nm and beyond with thin SOI <40nm thick now require undoped elevated S/D structures. For this reason, single-S/D structures will first be implemented in SOI technology, while bulk non-SOI devices will still require deep S/D to maintain low-S/D parasitic resistance and good salicide contacting. Some companies have proposed a third, deeper contacting junction for bulk CMOS devices [23, 24]. Once bulk CMOS devices switch to elevated S/D to maintain good junction contacting, as the contacting S/D junction scales below 40nm at the 45nm node, they too could merge the deep and shallow S/D junctions, forming single-S/D.

A simple way to transition from separate shallow SDE and deep S/D junction structures to a single-S/D structure is by using multiple, high-tilt, quad-mode implantation through the off-set spacer to form lateral graded SDE (LG-SDE) or single S/D (LG-SS/D) called PoGI (post-gate implant) [12; Fig. 3]. Diffusion-less activation simulation results for sub-65nm node devices (Lg = 25nm, height = 65nm, and Tfax = 0.8nm) are shown in Figs. 3 and 4 for a Taurus Monte Carlo implant model developed by Synopsys. An arsenic 3keV/1E15/cm² implant was used for the nSDE. Results for the traditional zero tilt nSDE showed no gate overlap, while a 30°, high-tilt, PoGI quad-mode implant (0, 90, 180, and 270° wafer twist) for an arsenic 3keV/1E15/cm² implant is shown in Fig. 3, realizing 8.5nm of gate overlap (encroachment). Geometrically, sin 30° = 0.5, so Yg = 0.5Xj and this is what the simulations also show. Also, the ratio of Yg/Xj is approximately 0.5 for the 130nm, 90nm, and 65nm nodes (see the table) and increases to 0.6 (sin 35°) for the 45nm and 32nm nodes. So a tilt of 30–35° seems optimum with diffusion-less activation for gate overlap control.

Both zero and 30° high-tilt implantation results satisfy the 65nm node for nSDE Xj = 10–17nm depth and lateral abruptness of <2.8nm/decade. However, the Yg gate overlap target of 8nm could not be achieved for zero tilt (0nm overlap), while for the 30° high-tilt case Yg = 8.5nm using a 5nm offset spacing results in a net gate overlap/encroachment under the gate edge of 3.5nm, thus enabling the use of a thinner offset spacer. With the traditional zero-tilt implant, the peak arsenic dopant concentration at the gate edge is zero, and under the sidewall spacer it is 5E20/cm³. With the LG-SDE terraced structure, the peak arsenic dopant concentration at the gate edge is 5E20/cm³, and under the sidewall spacer it is 1.3E21/cm³, resulting in a lower value of Rn and the length of SDE is now controlled by the gate stack height shadowing effects (sin X) and not the sidewall spacer thickness variations.

As the industry moves toward diffusion-less activation (65nm node and beyond) for shallow junction formation, high-tilt directional implantation in the lateral dimension will become desirable to precisely control gate overlap (encroachment); this is now on the roadmap of several companies. All diffusion-less activation processes (high-temperature flash/RTA and submelt laser or low-temperature SPE annealing) that use some form of pre-amorphizing implant (PAI) for the SDE gate overlap precision control and optimization of lateral dopant activation, will also require high tilt for the PAI implantation step.

A simulation comparison between a 10keV/5E14/cm² Ge-PAI 0° tilt and 30° tilt for a B11 = 500keV/1E15/cm², 30°-tilt LG-SDE was performed (Fig. 4). Outlined in black is the simulated amorphous/crystalline silicon interface — vertical and lateral. No lateral placement under the gate sidewall for zero tilt was observed, while a 7.1nm (Yg/Xj = 0.5) lateral placement under the gate with the 30° tilt was achieved as expected. High-tilt dopant and PAI implantation will therefore allow precise optimization and engineering of lateral dopant activation for gate overlap control, especially for diffusion-less activation (high- or low-temperature) processing starting at the 65nm node. P. Chidambaram et al. recently reported on improved 50nm pMOS devices using tilted PAI implantation [25]. Reduction in overlap capacitance of 8%, and a 3% improvement in drive current were reported. Additionally, improvements were found in lateral abruptness, SCE, Vt roll-off, as well as decreased channel doping.

Various options for the LG-SS/D structure when using PD/SOI devices with elevated S/D were performed (Fig. 5). Rotating the

Figure 2. HALO overlapping with Lg scaling [1].

2003 (90nm)

2005 (65nm)

2007 (45nm)

2009 (32nm)

High temp.

Spike/RTA

(Bn = 8E19/cm³)

Flash/substrmelt

(Bn = 1.2E20/cm³)

Low temp.

SPE

(No diffusion)

(Bn = 2.5E20/cm³)

High-k (25) gate

(low-temp. activation)

Medium-k (12) gate

Diffusion-less activation (disposable spacer)

Figure 1. Diffusion-less activation roadmap. Bn is based on box profile boron solid solubility limit.
Figure 3. Taurus Monte Carlo implant simulation for 30° quad-mode arsenic 3keV/1E15 implant through off-set spacer.

Figure 4. High-tilt 30° vs. zero-tilt Ge-PAI implantation for pSDE gate overlap control and optimization of lateral dopant activation when using diffusion-less activation.

Figure 5. Lg SS/D terraced structure dependence on PD/SOI wafer twist during quad-mode implant with elevated S/D.

Elevated S/D and contact salicide

Undoped elevated S/D structures for improved salicide contacting will first be used for thin (<40nm thick) PD/SOI devices at the 90nm node to maintain good salicide S/D contacting. For sub-90nm PD/SOI, the top silicon layer thickness (24-30nm) is approaching the depth of the SDE (15-24nm), and the SDE and deep S/D junctions are merging quickly into single-S/D, making LG-SS/D structures ideal for lower S/D tip resistance for thin PD/SOI devices beginning at the sub-65nm node. As the deep S/D junction on bulk CMOS devices scales below 40nm junction depth, the industry will also adopt elevated S/D at Lg between 20-40nm. The selective silicon (epi or poly) deposited structure does not have to be doped, provided the silicide contact metallization process step completely consumes the planar facet-free, elevated silicon [28-30].

The switch from CoSi2 to NiSi2 salicide for lower-S/D contact resistance and poly-gate resistance will start at the 90nm node for compatibility with SiGe material used in strained Si channel technology, lower contact, and gate/poly resistance and lower-temperature processing. To achieve good contacting, a certain amount of silicon must be consumed. This becomes problematic with SOI devices, where the S/D parasitic resistance increases with thinner SOI layers. The thin junction layer under the silicide and small interface contact area results in higher resistance and, therefore, the need for elevated S/D [31].

Rext. If elevated S/D structures are used, then a selective poly deposition process is preferred over a selective epi deposition process to avoid issues with sidewall-faceting control and interface surface-cleaning sensitivity. This can be achieved using controlled interface cleaning with selective HF vapor etching, for example [26, 27].

Solid State Technology - www.solid-state.com

continued on page 58
Double/multigate CMOS

With expected difficulties in the scaling of SG FD/SOI wafers and devices, the industry is rapidly looking at double-gate (DG) and multigate (MG) FD/SOI devices for possible introduction at the 32nm node for $L_g=10-20nm$. This would ease SOI wafer manufacturing, since the SOI thickness will be $0.7-1xL_g$ compared to $0.2-0.3xL_g$ for SG FD/SOI. Other benefits include an undoped-to-lightly-doped channel (thus avoiding mobility degradation and easing concern over the random dopant fluctuation effects in the channel region), better SCE control, and higher drive current by plasma implantation (Fig. 6) [34].

[2, 3]. Also, lateral graded S/D dopant engineering and elevated S/D contact formation technique will be used for both DG and MG novel devices as described earlier for the SG device architectures.

Planar (top/bottom) DG CMOS

One of the most difficult challenges, and therefore critical requirements, in the planar DG structure is the alignment of the top gate to the bottom gate [12; Fig. 1]. The simplest method is using wafer-bonding techniques, although gate alignment is most difficult. Another approach is to use selective epitaxy with lateral growth through a tunnel to form the source, and then the channel (tunnel), and last, the drain. For this structure, no SOI starting wafer is needed, but this is a complicated structure to achieve.

FIN-FET DG and MG CMOS

The FIN-FET DG structure is the easiest to fabricate and currently the most popular DG structure found in the literature. For the FIN-FET design, the SOI silicon film thickness determines the minimum gate width ($W_g$) dimension, so thicker SOI layers are required. With this structure, high-tilt implantation up to 45° for SDE and S/D doping prior to elevated S/D for improved salicide contacting will be needed as reported by Kedzierski et al. [32]. Additional flexibility for designing asymmetrical n+/p+ poly gates by high-tilt poly-gate implantation is another option in transistor design for undoped channels. If $W_g$ is scaled down to equal $L_g$, then a triple-/multiple-gate structure can easily be realized [33].

Again, multiple high-tilt implantation for LG-SDE and LG-SS/D structures can be used with the various FIN-FET device structures (Fig. 6) [34]. To create devices with various gate widths, multi-FIN structures will be required [34]. This 3-D poly-gate structure formed around multiple FIN structures may require conformal doping by plasma implantation (Fig. 6).

Vertical pillar DG CMOS

For the pillar DG design, also called vertical gate replacement (VGR) transistor, an SOI wafer is also not necessary and the S/D, SDE, and channel regions are selectively epitaxially grown and doped. If the gate structure goes completely around the pillar or planar DG structure, the result is a quadruple-gate, called gate-all-around or surround-gate [12; Fig. 1].

Conclusion

Multiple types of transistor designs and silicon wafers will be used as the industry continues to scale devices to meet the SOC applications of the future, ranging from planar SG and vertical SG structures, to vertical DG and MG options. Various silicon wafers will also be used, including selective/patterned SOI wafers. One of the key areas for extending SG and advancing DG transistor performance will be lateral graded S/D engineering, realized with high-tilt implantation and diffusion-less activation starting at the 65nm node.

References

22. M. Ozterk et al., The Electrochemical Society, May 2002 meeting, to be published.

For further information, contact John O. Borland, Varian Semiconductor Equipment Associates, 4 Stanley Tucker Dr., Newburyport, MA 01950; phone 978/463-5070, fax 978/462-0210, john.borland@vsea.com.