Technology Options for 22nm and Beyond

Kelin J. Kuhn, Mark Y. Liu and Harold Kennel

Logic Technology Development, Intel Corporation, Hillsboro, OR, 97124, U.S.A.
Contact: kelin.ptd.kuhn@intel.com

Abstract

This paper explores the challenges facing the 22nm process generation and beyond. CMOS transistor architectures such as ultra-thin body, FinFET, and nanowire will be compared and contrasted. Mobility enhancements such as channel stress, alternative orientations, and exotic materials will be explored. Resistance challenges will be reviewed in relation to key process techniques such as silicidation, implantation and anneal. Capacitance challenges with traditional and new architectures will be discussed in light of new materials and processing techniques. The impact of new transistor architectures and enhanced channel materials on traditional junction engineering solutions will be summarized.

1. Introduction

For the past 40 years, relentless focus on Moore’s Law transistor scaling has provided ever-increasing transistor performance and density. For much of that time, Moore’s Law transistor scaling meant “classic” Dennard scaling [1] where oxide thickness \( (T_{ox}) \), transistor length \( (L_g) \) and transistor width \( (W) \) were scaled by a constant factor \( (1/k) \) in order to provide a delay improvement of \( 1/k \) at constant power density.

![Fig. 1. Performance improvement enabled by enhancers [2]](image)

In recent years, Dennard scaling has become less influential in Moore’s Law scaling. For generations after the 130nm node (90nm, 65nm, 45nm, 32nm, see Fig. 1) performance enhancers have been added to continue to drive the transistor roadmap forward (e-SiGe and strained SiN for strain in the 90nm and 65nm nodes [3,4], and high-k metal-gate (HiK-MG) in the 45nm and 32nm nodes [2,5-7]).

As we look forward, there are a number of challenges to be addressed. Increased off-state current \( (I_{off}) \) from degraded drain-induced barrier lowering (DIBL) and subthreshold slope (SS) caused by poorer short channel effects (SCE) represents a significant limitation for effective gate lengths \( (L_{eff}) \) shorter than approximately 15nm. Decreasing \( T_{ox} \) to provide better channel control comes with a penalty of increased gate leakage current \( (I_{gate}) \) and increased channel doping.

Increased channel doping decreases mobility (degrading performance due to impurity scattering) and increases random dopant fluctuations (RDF) degrading the minimum operating voltage \( (V_{min}) \). Decreasing gate pitch increases the parasitic capacitance contribution for both contact-to-gate and epi-to-gate thus increasing overall gate capacitance \( (C_{gs}) \). Decreasing source/drain opening size increases the source drain resistance \( (R_{sd}) \) thus decreasing drive current. Additionally, decreasing gate pitch decreases the stress enhancement for both NMOS (stress induced by overlayer films) and PMOS (stress induced by embedded-SiGe, e-SiGe) thus decreasing mobility and drive.

2. Future transistor architectures

Maintaining the scaling roadmap will require continual improvement in short channel properties. A variety of device architectures which improve electrostatic confinement are being investigated for advanced technology nodes. These architectures can be broadly categorized by the method of electrostatic confinement. There are architectures which provide additional electrostatic confinement with a planar architecture (ultra-thin body (UTB), fully-depleted SOI (FDSOI), etc.), those which use 1D electrostatic confinement (double gate, FinFET, etc.), those with more than 1D, but less than 2D (Trigate, Omega-FET, etc.) and those with full 2D confinement (gate-all-around (GAA), nanowire etc.) [8].
A. Additional electrostatic confinement in planar

The potential value of fully-depleted UTB SOI for planar electrostatic confinement (as well as the requirements for extremely Si thin layers to achieve well-designed fully-depleted devices) has been recognized since the mid-1980s [9,10]. There has been a steady reduction in the minimum demonstrated body thicknesses ($T_{Si}$) moving from $\sim$100nm in the 1980s and early 90s [11-13], down to the 15-20nm range in early 2000 [14-16], and more recently to values significantly below 10nm [17-19].

UTB SOI devices benefit from using similar manufacturing to planar SOI technology, but with improved SCE, potential for improved RDF (due to lower channel doping) and the possibility for body bias (with thin BOX).

Challenges of UTB SOI include thin $T_{Si}$ effects (external resistance, $R_{ext}$, scattering, and quantum confinement changes in $V_T$), difficulties in inducing strain and manufacturing challenges with the thin $T_{Si}$.

B. 1'D and 1'D+ confinement

There is a tradeoff between the electrostatic improvement of a GAA device and the fabrication complexity of making gates on all sides of a channel. A number of intermediate architectures (sometimes called multiple gate FET devices or MuGFETs) have been developed in an attempt to get the best SCE with the minimum process complexity [20-36].

Double-gate devices first appeared in the literature in the mid-1980s [20], and a variety of different geometries were explored in the next two decades [21-26]. Categories (see Fig. 2) include:

FinFET: Combines double-gate and vertical device concepts for a more manufacturable version of a double gate device [27].

Trigate: Differs from FinFETs in the absence of a gate-blocking layer on the top of the gate. Trigate devices have gates around three sides of the device, providing improved SCE with reduced vertical topography requirements [28].

Pi-gates: Differs from Trigates in having the gate extend below the channel. This creates a virtual back gate which shields the back of the channel from electric field lines from the drain, providing improved SCE [29].

Omega-FETs: Differs from Trigates in that the gate not only wraps around three sides, but underlaps part of the fourth. This has an effect similar to Pi-gate in shielding the back of the channel from field lines, resulting in improved SCE [30].

These multiple-gate devices have similar DIBL and RDF advantages over planar as UTB SOI. In addition, the increased confinement in comparison with UTB devices relaxes the manufacturing constraints ($W_{Si} \sim 2T_{Si}$). Furthermore, tying the gates together provides nearly ideal sub-threshold slope. Note also that independent gate operation is possible in some of these architectures.

C. 2'D confinement

GAA devices were first reported in the late 1990s [37-38]. GAA devices differ from Omega-FETs in that the gate wraps entirely around the device. Note that both lateral [37,40,42], and vertical [38,41], devices are possible with this architecture. Both types provide full two dimensional confinement with the associated SCE benefits.

GAA devices offer the best potential solution to electrostatic confinement challenges. However, these devices face significant challenges. Not only do they have the strain, $R_{ext}$, vertical topography, tight pitch, and complex gate patterning challenges of the MuGFET devices, they also face new challenges with gate conformity and excess parasitic capacitance.

Nanowires are an extreme case of GAA devices, having height and width dimensions roughly the same (or even cylindrical) and small (<10nm) dimensions [Fig. 3, 43-47]. Nanowires add the challenges of phonon scattering [48], (along with possible benefits due to reduction in interface scattering [49, 50]).
3. Mobility enhancements

Maintaining the scaling roadmap will require continual improvement in channel mobility. Short term approaches include reorienting the surface or channel of the device, and implementing improved strain techniques. Long term solutions may include more exotic channel materials (Ge, III-V, etc.).

A. Wafer and channel orientation

There are two potential wafer orientations for advanced planar silicon devices, (100) surface and (110) surface (see Fig. 4). The best unstrained NMOS devices are fabricated on the (100) surface with a <110> channel direction, and the best unstrained PMOS devices are fabricated on the (110) surface with a <110> channel direction [51-54]. Significant research in the last five years has focused on the challenge obtaining the enhanced PMOS mobility on (110) <110> material, without degrading the NMOS (for example, the HOT process, which integrates both orientations on the same wafer [51]).

Notice that non-planar devices can significantly complicate optimizing wafer and channel orientation. For example, FinFET devices may be oriented in different directions to expose different orientations (placing an NMOS FinFET device at 45 deg. from a PMOS FinFET device on a 100 surface gives 110 <110> PMOS and a (100) <110> NMOS [52]). As a more extreme example, devices with higher dimensionality (for example, a round nanowire) will contain multiple orientations.

B. Strain

Strain has had tremendous impact in advancing the transistor scaling roadmap [Fig. 5, refs. 2-7, 51-54]. A large number of process-induced strain techniques are employed in today’s fabrication (e-SiGe, e-SiC, contact-etch-stop layer, stress-memorization technique, stressed gate metal, stressed contact metal, etc.). Future transistor architecture solutions (whether (100) or (110), planar or non-planar) must possess significant strain enhancement on both N and PMOS to continue to drive the scaling roadmap forward.

C. Strain and Orientation

While significant improvements can result from combining strain and orientation changes, the performance results may not be additive. An example is given by comparing (110)<110> and (100)<110> orientations for PMOS under uniaxial compressive strain. PMOS in the (100)<110> orientation is not strongly affected by vertical confinement in a MOS device. However it is strongly affected by compressive uniaxial strain. Thus, PMOS (100)<110> devices experience a large enhancement in mobility with increasing strain. In contrast, PMOS in the (110)<110> orientation is strongly affected by confinement (the origin of the (110) vs (100) improvement discussed earlier), but not significantly affected by strain. Thus, PMOS (110)<110> devices have larger mobility at lower strain, but improve less with increasing strain [54].

D. Exotic materials (Ge and III-V)

Advanced channel materials (such as Ge and III-V materials) offer potential long term solutions for mobility enhancement when integrated with silicon substrates [55]. Unfortunately, the most interesting advanced channel materials are lattice-mismatched to silicon, with associated fabrication challenges. In addition, the low band-gap materials (InAs, InSb, Ge, etc.), display significant band-to-band tunneling (which may limit them to low voltage operation, or UTB implementation). Finally, use of these materials requires developing new gate dielectric fabrication techniques (or moving away from MOS-style gate architectures).
4. Resistance and next generation transistors

Improving the traditional resistive elements (Figs. 6-7), such as the accumulation ($R_{acc}$), spreading, silicide and contact resistances, will become more challenging at the reduced dimensions of advanced technologies. Furthermore, resistance elements previously neglected (including interface and epi resistance) are becoming significant issues. Moreover, the various non-planar architectures will introduce new resistance components associated with small dimension fins and wires.

All these resistive components are influenced by activation and doping profile shapes. For example, $R_{epi}$ can be improved by reductions in the sheet resistance $R_s$ ($R_s$ is an integral of the active doping and mobility). $R_{acc}$ can be improved both with highly activated dopants and abrupt lateral junctions. $R_{silicide}$ (with moderate barrier heights) exponentially decreases with increasing activation.

A. Characterizing annealing techniques

One way to characterize annealing techniques is by their characteristic interaction time (see Fig 8). Conventional RTA techniques have characteristic interaction times on the order of seconds. Flash and non-melt laser techniques are on the order of milliseconds, and laser melt techniques on the order of nanoseconds [58-63].

<table>
<thead>
<tr>
<th>Cycle</th>
<th>Rampup Rate (C/S)</th>
<th>Typical peak time (S)</th>
<th>Rampdown Rate (C/S)</th>
<th>Effective Time (S)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Soak</td>
<td>75</td>
<td>5-30</td>
<td>40</td>
<td>~5 + thold</td>
</tr>
<tr>
<td>Flash</td>
<td>$1E5 - 1E6$</td>
<td>&lt; $1E-6$</td>
<td>~ $1E6$</td>
<td>0.1-1 mS</td>
</tr>
<tr>
<td>Scanning</td>
<td>$1E5 - 1E6$</td>
<td>&lt; $1E-6$</td>
<td>&gt; $1E6$</td>
<td>0.1-1 mS</td>
</tr>
<tr>
<td>Melt (laser)</td>
<td>$1E7 - 1E8$</td>
<td>&lt; $1E-6$</td>
<td>&gt; $1E7$</td>
<td>10-100nS</td>
</tr>
</tbody>
</table>

Annealing techniques can also be characterized by the physics of activation [64]. This is illustrated in Fig. 9, where the various physical mechanisms of activation are plotted against the characteristic interaction time. Also shown are boxes delineating the timescales of conventional RTA, flash, and melt lasers.

Consider first the slowest processes (to the right) which include B and As dopant diffusion, {311} dissolution, and boron-interstitial cluster (BIC) dissolution. While this group of processes is faster than RTA processes, they are slower than flash/laser anneals. Thus, flash/laser processes have the potential to “freeze” dopant profiles in place (Fig. 10).
Processes with characteristic times on the order of flash and laser anneals include solid phase epitaxial regrowth and evolution of interstitial (Int) clusters to \{311\} defects. Processes with very short characteristic times include Si Int diffusion and clustering, cascade recovery, and dopant substitutionality via substitutional-interstitial exchange reactions. Dopant solubility limits are controlled by the slower processes of clustering or precipitation reactions rather than the faster processes of substitutional-interstitial exchange. This permits non-equilibrium enhanced activation (superactivation).

## B. Superactivation

Solid-phase epitaxial regrowth (SPER) of a doped amorphous layer is one method to obtain superactivation. The regrowth process forces dopant to substitutional sites far above the normal occupancies expected from solubility relations. A solute trapping process controls the incorporation, since dopant is unable to diffuse away from the interface faster than the interface sweeps by.

Liquid phase epitaxial regrowth (LPER) after localized melting is another method to enable activation above normal solubility levels (typically higher than those achievable with SPER). The technique uses an amorphizing implant to define an aSi region, which will be selectively melted by a 10-200 ns pulsed laser exposure. The melt is selective because the melting point of aSi is several hundred degrees lower than that of cSi. Liquid phase diffusivities are rapid (~1E-4 cm²/s), leading to boxlike doping profiles over the melt depth (see Fig. 11).

## C. Other resistance improvements

Looking ahead to further resistance improvements, interface resistance improvement through modulation of the Schottky barrier height (SBH) also offers significant opportunities. Unfortunately, the key practical challenge with improving SBH is that most materials on silicon pin the SBH at mid-gap. One rich field of research to address this is implant modifications to traditional silicides, either on single metals or alloys [65-66].

## 5. Capacitance and next generation transistors

The traditional capacitance elements (Fig. 12), such as under-lap capacitance ($C_{XUD}$), channel capacitance, junction capacitances (both gated edge and area) and the inner and outer fringe capacitance; will become more challenging at the reduced dimensions of advanced technologies. Furthermore, in recent generations, gate and contact CD dimensions have been scaling slower than contacted gate pitch. This means that parasitic fringe capacitances (for example, contact-to-gate and epi-to-gate) are becoming significant issues.

![Fig. 12. Key capacitance elements [56]](image)

The key knob for parasitic capacitance improvement in the front end (for either planar or non-planar devices) is to reduce the $k$ of the spacer with airgaps or lower-$k$ materials [67-68].
While significant transistor challenges (SC, resistance, capacitance, mobility, etc.) exist for technologies past 22nm, numerous solutions are being explored to drive Moore’s Law forward. Advanced junction engineering will play a critical role in the transistor roadmap past 22nm.

References

[38] Li, M., et al., “Sub-10nm gate-all-around CMOS nanowire transistors on bulk Si substrate,” 2009 Symp. on VLSI Tech, pp.89-90, June 2009.