A Comparative Study of Fin-Last and Fin-First SOI FinFETs

Darsen Lu, Josephine Chang, Michael A. Guillorn, Chung-Hsun Lin, Jeffrey Johnson,
Philip Oldiges, Ken Rim, Mukesh Khare and Wilfried Haensch
IBM T. J. Watson Research Center, Yorktown Heights, NY 10598 (USA)
IBM Research at Albany Nanotech, Albany, NY 12203 (USA)
Tel: +1-914-945-1774, Fax: +1-914-945-4184, Email: ddlu@us.ibm.com

***Abstract*— Two FinFET fabrication processes are compared with simulation: the conventional fin-first process and the novel fin-last process. With the fin-last process, more longitudinal strain can be incorporated into the channel from source and drain SiGe stressor than fin-first. pFET mobility advantage is 15% at fully-strained condition and with silicon recess. Maintaining vertical junction uniformity is the main challenge for fin-last. However, its impact on parasitic resistance and capacitances are small. Vertical junction non-uniformity is improved with source and drain recess and doping optimization.**

***Keywords*— FinFET, SOI, Fin-last, Replacement Metal Gate Process.**

I. Introduction

A critical step of the FinFET [1] fabrication process is the etching of silicon to form the fin. Conventionally FinFETs are fabricated with the fin-first process [1], where fin etch is performed in the very beginning of the process sequence. Alternatively, in a replacement metal gate process, we have the option to delay fin etch until after the dummy gate is removed. This is known as the fin-last process [2].

In this simulation study, we compare the fin-first and the fin-last processes in terms of device DC and AC performance, as well as process integration benefits. For fin last, integration is simpler because all processing steps before the final fin etch is nearly identical to that of planar MOSFET. One of the key advantages of planar processing is the ability to maximize the effectiveness of source and drain stressors. In addition, lack of topology simplifies numerous process steps such as the etching steps to form the gate and the spacer. However, at the same time, the fin-last process introduces additional challenges to form a vertically uniform junction. For this reason, fin-first is slightly superior in terms of external resistance. We will discuss how this issue in fin-last is addressed with source and drain recess and additional implantation prior to the source and drain epitaxial growth.

II. The Fin-Last Process Flow

The fin-last process sequence is illustrated in Fig. 1. The fin pattern is defined on a hard mask by a high resolution lithography technique such as sidewall image transfer (Fig. 1(a)). However, unlike fin-first, the pattern is not transferred to the underlying silicon layer right away. Process continues with dummy gate definition and source/drain formation by implant or epitaxy (Fig 1(b)). The exposed hard mask is also

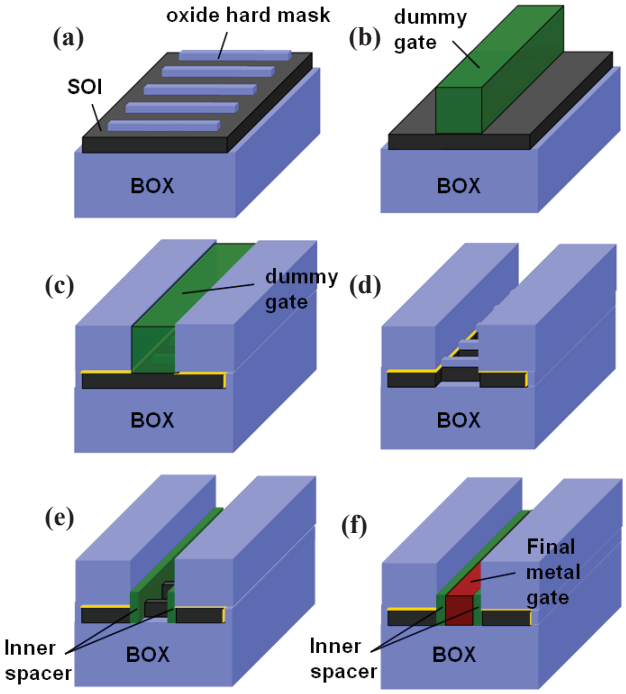


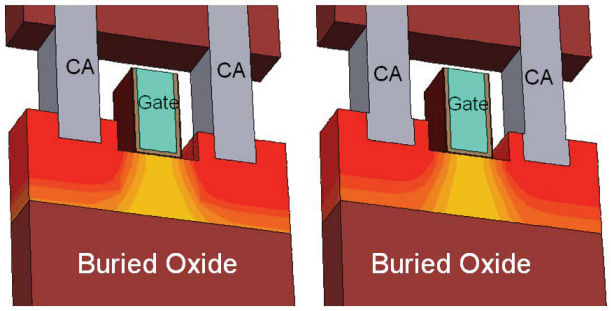
Fig. 1 FinFET formation with the fin-last process on SOI substrate. (a) The oxide hard mask is patterned without etching of silicon. (b) The dummy gate is patterned and the exposed hard mask oxide is removed. Source and drain are formed at this step. (c) The contact trench is filled with insulator and planarized. (d) The dummy gate is removed, revealing the patterned hard mask. The fin pattern is transferred to the underlying silicon layer by anisotropic etching. (e) The inner spacer is deposited on source and drain sidewalls to ensure sufficient separation between the gate and the flared source and drain. (f) The final gate stack is deposited into the gate trench.

etched away after gate etch, leaving behind only the remaining hard mask covered by the dummy gate. The dummy gate material can be an insulator such as silicon nitride. Subsequently, we perform planarization in preparation for removing the dummy gate (Fig. 1(c)). Once dummy gate is removed, the patterned hard mask is uncovered. An anisotropic silicon etch defines the fin based on hard mask pattern, as shown in Fig. 1(d). Prior to final gate stack formation, an inner spacer is created on source and drain sidewalls to ensure there is sufficient separation between the final metal gate and the flared source / drain (Fig. 1(e)). Without the inner spacer, the gate and source/drain would be

978-1-4673-5736-4/13/$31.00 ©2013 IEEE

Parameter	Nominal Value
Fin Height	25nm
Fin Thickness	8nm
Dummy Gate Length	18nm (fin-first) 31nm (fin-last)
Final Metal Gate Length	14nm
Final Metal Gate Height above the buried oxide	58nm
Spacer thickness (fin-first) Inner spacer thickness (fin-last)	6.5nm 6.5nm
Source/drain recess	16nm
Equivalent Oxide Thickness	1.05nm
Metal Gate Work-function	4.525eV
Gate pitch	60nm
RTA temperature	1000C

Table. 1 Nominal device geometry, process conditions, and gate work-function assumed in TCAD simulation.



(a) (b)

Fig. 2 Near-identical final device cross section for (a) fin-last and (b) fin-first processes. Process simulation is performed with Sprocess [3]. The nitride spacer and part of the interlayer dielectric are removed in these pictures to show the transistor structure. (CA: contact)

only separated by the gate dielectric, resulting in enormous parasitic capacitance. Finally, the final gate stack is formed inIf the gate trench (Fig. 1(f)).

III. FinFET Process Simulation Setup

The fin-first and fin-last processes are compared with TCAD simulations [3] with identical process and device models. The device electrical simulations are drift-diffusion based with accurate calibration to hardware.

Identical nominal device geometry for fin-first and fin-last are considered (Table 1). For fin-last, since the inner spacer deposited on the sidewall of the gate trench will reduce the final gate length, we start with longer dummy gate than fin-first to begin with. The final gate lengths for fin-first and fin-last become identical. The gate height above the top of the fin is assumed to be the same for simple comparison, even though in practice fin-first devices may require a taller gate than fin-last due to more topology. As illustrated in Fig. 2, with the assumptions in Table 1, the final device structures of fin-first and fin-last are identical. However, doping and mechanical

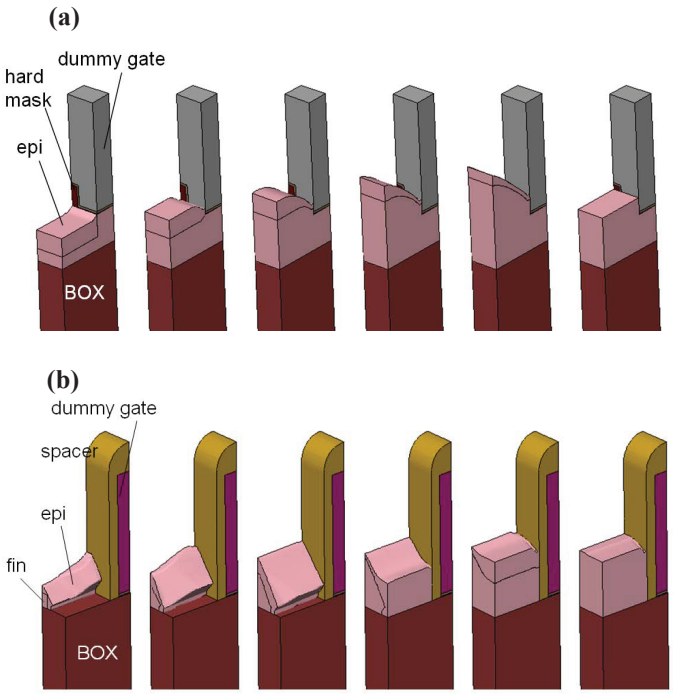


Fig. 3 Illustration of step-by-step SiGe epitaxial growth simulation. The same growth procedure is applied to the (a) fin-last and (b) fin-first cases. The entire growth process is divided into 5 steps to illustrate how the epitaxial front evolves, as predicted by simulation. In the last (6^{th}) step (right-most figures in (a) and (b)) the top of the epitaxial SiGe is recessed to the target height with a planar top.

stress profiles for the two processes are quite different.

IV. Effect of Source and Drain Stressors

Incorporation of uniaxial compressive strain with source and drain SiGe epitaxial stressors in pFETs is beneficial for device performance. However, stressing FinFETs is challenging due to its three-dimensional nature, and strain relaxation may occur during epitaxial growth. With the fin-last process, the silicon surface is planar prior to SiGe epitaxy. Stressor techniques for planar devices [4] can be utilized.

We simulate epitaxial growth for both fin-last and fin-first processes with TCAD. The amount of initial strain (before re-balancing calculations) in the SiGe source and drain stressor is calculated based on the lattice mismatch between silicon and SiGe with 35% germanium. Epitaxial SiGe is assumed to be fully strained without defects or dislocations. The shape of the epitaxial SiGe during growth is obtained via crystal-orientation dependent growth rates, as illustrated in Fig. 3. For fin-first, the HomNeumann (free) boundary condition [3] is considered at the boundary in between fins to model free growth in the fin-to-fin direction.

The average normal stress in the channel is simulated as function of silicon recess. Then, hole mobility enhancement is calculated based on these average stress values using a 3D stress model [5] (Fig. 4, 5). For fin-last, large increase in compressive stress in the channel length direction (S_{yy}) with increasing silicon recess is because of the significant volume increase of embedded SiGe compressing the channel. The

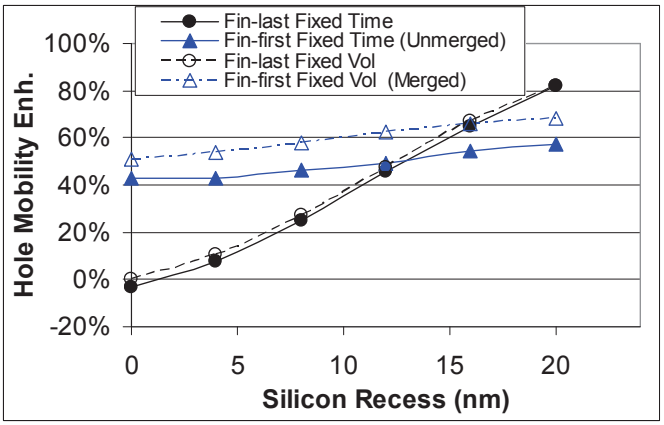


Fig. 4 Hole mobility enhancement due to stress in the channel (measured prior to silicide formation) as function of silicon recess for both the fin-last (black circles) and the fin-first (blue triangles) processes. Both fixed-time cases (cross sections shown in the 3rd figures in Fig. 3(a) and 3(b)) and fixed volume cases (6th figures in Fig. 3(a) and 3(b)) are plotted.

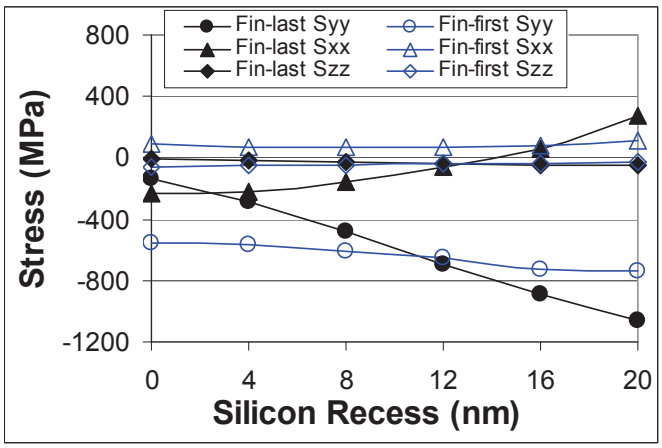


Fig. 5 Average channel stress in FinFETs with the fin-first and the fin-last processes. The stress values for fixed-time epitaxial growth are plotted. (positive indicates tensile; negative indicates compressive; S_{xx}: stress along fin height; S_{yy}: stress along channel length; S_{zz}: stress along fin width)

fixed time case (3rd picture in Fig. 3(a)) and fixed volume case (6th picture in Fig. 3(a)) show little difference because SiGe above the fin height is not very effective in stressing the channel. On the other hand, the fin-first case show weaker dependence on silicon recess since the epitaxial SiGe volume shape is a weak function of silicon recess. For fin-first, merged epitaxy results in larger stress compared to unmerged epitaxy simply due to larger volume of SiGe. With sufficient silicon recess, the fin-last process has strain and mobility advantages compared to fin-first.

As shown in Fig. 5, for fin last the hole mobility improvement is due not only to the compressive S_{yy}, but also to the tensile S_{xx} (strain in the fin height direction). This is because for fin-last, epitaxial SiGe is growing against the silicon sidewall of the recessed trench in a lattice matched fashion. On the other hand, for fin-first epitaxial SiGe is adjacent to the gate sidewall spacer which is amorphous and cannot be lattice matched to.

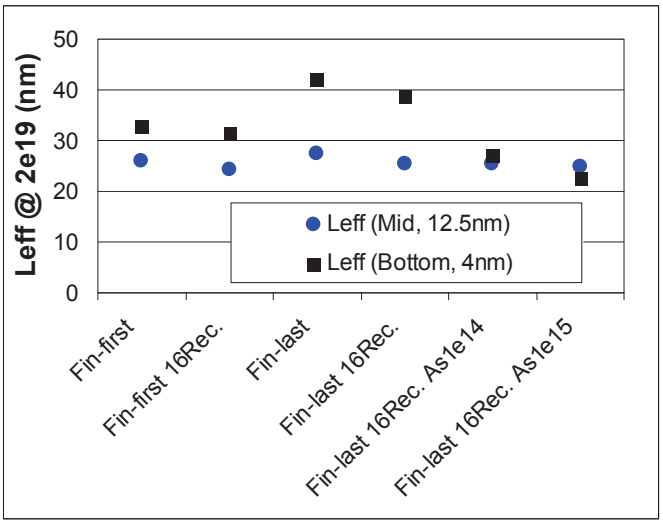


Fig. 6 Effective channel length (defined at the point of 2e19cm^{-3} source/drain doping) at various processing conditions at mid-point (12.5nm below fin top) and near bottom of the fin (4nm above the buried oxide). "16Rec" indicates 16nm silicon recess. "As1e14" indicates 4keV arsenic implantation at 10^{14}cm^{-3} dose after silicon recess and before epitaxial SiGe)

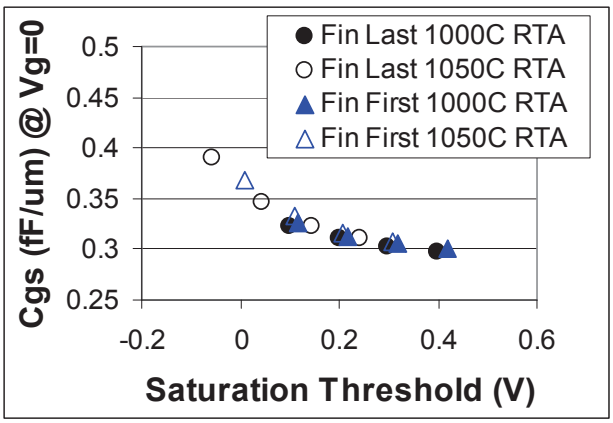


Fig. 7 Parasitic capacitance versus saturation threshold voltage for fin-first versus fin-last with different source and drain anneal. Gate work-function ranges from 4.325eV to 4.625eV.

V. Vertical Junction Uniformity

Despite strain advantages, one of the key challenges for the fin-last process is to maintain a vertically uniform doping profile along the height of the fin. In the absence of silicon recess, source and drain dopants, implanted prior to replacement gate formation, are introduced from the top silicon surface, making it difficult to form a vertically uniform junction. In Fig. 6 we plotted effective channel length as function of silicon recess to highlight the problem. The effective channel length near the bottom of the fin is significantly larger than that near the middle due to junction non-uniformity, much more so in fin-last than fin-first.

This issue can be addressed with ion implantation right after silicon recess. As shown in Fig. 6, with additional implantation the effective channel lengths at the top and near

the bottom of the fin become nearly identical. However, the electrical impact due to non-uniform doping may not be completely addressed, as we will discuss in the next section.

VI. Parasitic Capacitances and Resistances

In the fin-last process, an inner spacer is needed for gate-to-source and drain fringe capacitance reduction. We use TCAD to compare the fringe capacitance in fin-first and fin-last FinFETs with similar device cross sections (Fig. 2). The inner spacer thickness of fin-last is kept identical to the normal spacer thickness of fin-first. Fig. 7 shows the device parasitic capacitance as function of device threshold voltage ($V_{t,sat}$) at two different post-implant annealing conditions (1000C and 1050C spike anneals) and various gate work-functions. Although C_{gs} varies with overlap, for a given $V_{t,sat}$ which corresponds to a given overlap capacitance, total capacitance all lie on the same trend line. Therefore, fringe capacitance components are the same for fin-first and fin-last. This is because the shapes of metal gates and the flared source and drains from the two processes are the same.

As discussed in the previous section, with the fin-last process the junction is less vertically uniform. The impact on device characteristics is assessed with TCAD. Fig. 8(a) shows the on-resistance versus parasitic capacitance trade-off with varying spacer thickness, indicating very little difference between fin-first and fin-last from a parasitic resistance and capacitance point of view. However, For the same electrostatics (same *DIBL*), on resistance is lower in fin-first by about 25-35 Ohm-micron (Fig. 8(b)).

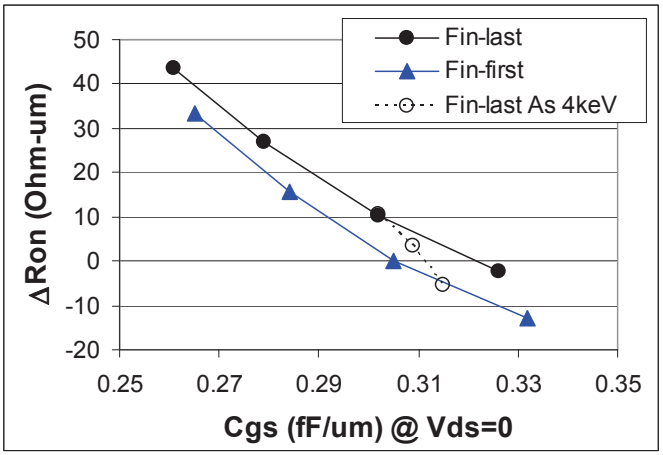
Although post silicon recess implant shortens L_{eff} near fin bottom (Fig. 6), simple implantation is insufficient to address the resistance and short channel effects trade-off (Fig. 8(b)).

VII. Process Compatibility and Scaling Advantages

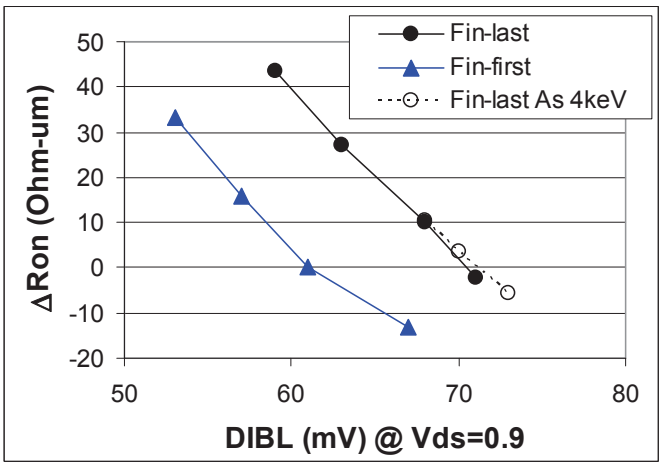
Since the fin-last process utilizes planar processing until the dummy gate is removed and the fin is etched, it is directly compatible with the embedded DRAM technology [6], which significantly enhances system performance. In addition, scaling of the FinFET technology to smaller gate lengths requires the scaling of fin thickness to control short channel effects. FinFET with a 4nm thick fin has been demonstrated experimentally with the fin-last process [2].

VIII. Conclusion

The novel fin-last FinFET process allows simpler planar processing by not etching the silicon to form the fin until the dummy gate is removed in a replacement gate process. This simplifies eDRAM integration and allows aggressive fin sacling. In addition, simulation shows 15% strain-induced mobility advantage. The real difference may be larger due to strain relaxation in the 3D epitaxial process in fin-first. Maintaining junction uniformity along the fin height is challenging for fin-last, and requires further optimization of silicon recess, source and drain epitaxy and doping schemes.



(a)



(b)

Fig. 8 (a) Device on-state resistance (R_{on}) versus total device parasitic capacitance trade-off and (b) R_{on} versus *DIBL* trade-off, for fin-last and fin-first. R_{on} is measured at 0.7V gate overdrive and V_{ds}=50mV. Spacer thickness is varied from 5.5nm to 8.5nm at 1nm increment to change L_{eff}. The 8.5nm spacer thickness case corresponds to the largest R_{on} (top-left corner of the figures). For the 6.5nm spacer case we also illustrated the effect of post silicon-recess 4keV arsenic implant at 10^{14} and 10^{15} dose.

References

[1] X. Huang et al., "Sub-50nm FinFET: PMOS," *International Electron Devices Meeting Technical Digest*, pp. 67-70, 1999.

[2] J. Chang et al., "Scaling of SOI FinFETs Down to Fin Width of 4nm for the 10nm Technology Node," *Symposium on VLSI Technology, Digest of Technical Papers*, pp. 12-13, 2011.

[3] Sentaurus Process User Guide, *Synopsys Inc.*, 2012.

[4] K. Cheng et al., "Extremely Thin SOI (ETSOI) CMOS with Record Low Variability for Low Power System-on-Chip Applications," *International Electron Devices Meeting Technical Digest*, pp. 49-52, 2009.

[5] A. Kumar et al., "A Simple, Unified 3D Stress Model for Device Design in Stress-Enhanced Mobility Technologies," in *Proceedings, SISPAD Conference*, pp. 300-303, 2012.

[6] S. Crowder et al., "Trade-offs in the Integration of High Performance Devices with Trench Capacitor DRAM," *International Electron Devices Meeting Technical Digest*, pp. 45-38, 1997.