

## Scaling of SOI FinFETs down to Fin Width of 4 nm for the 10nm technology node

J. B. Chang, M. Guillorn, P.M. Solomon, C.-H. Lin,  
S.U. Engelmann, A. Pyzyna, J. A. Ott, W.E Haensch

IBM Research, IBM T.J. Watson Research Center, Yorktown Heights, NY

### Abstract

Using a novel replacement gate SOI FinFET device structure, we have fabricated FinFETs with fin width ( $D_{Fin}$ ) of 4nm, fin pitch ( $FP$ ) of 40nm, and gate length ( $L_G$ ) of 20nm. With this structure, we have achieved arrays of thousands of fins for  $D_{Fin}$  down to 4nm with robust yield and structural integrity. We observe performance degradation, increased variability, and  $V_T$  shift as  $D_{Fin}$  is reduced. Capacitance measurements agree with quantum confinement behavior which has been predicted to pose a fundamental limit to scaling FinFETs below 10nm  $L_G$ .

### Introduction

Advanced non-planar multi-gate MOSFETs appear on the ITRS roadmap as a solution needed for scaling MOSFETs below  $L_G$  of 10nm [1]. Accordingly, FinFETs have attracted significant development effort [2,3]; as  $L_G$  in FinFETs is decreased to enable continued gate pitch scaling,  $D_{Fin}$  must be reduced to maintain proper electrostatics. The majority of FinFET work has focused on (110) surface, [110] directed channels. This offers the advantage of compatibility with conventional (100) Si substrates and, depending on  $D_{Fin}$  and channel strain, may also provide a performance advantage over other orientations [4]. However, for channels less than about 5 nm thick, quantum confinement is predicted to pose a challenge to electrostatic scaling [5-7] due to strong  $V_T$  increase and mobility degradation as  $D_{Fin}$  decreases. These effects are exacerbated for (110) FinFETs compared to devices with (100) surface channels.

### Device fabrication

$D_{Fin}$  scaling in conventional SOI FinFET structures is limited by erosion, undercut of the buried oxide (BOX), and damage to the fin during device processing. To circumvent these issues we have developed a novel replacement gate FinFET device architecture that allows formation of the fin late in the process flow (Fig. 1). With this process, robust yield was obtained for fins with thicknesses averaging down to 4nm as determined by TEM measurements (Fig. 2).

### Results

Gate inversion capacitance taken from capacitors (Fig. 3) show quantum effects which are consistent with the TEM-derived  $D_{Fin}$  of 4nm and 9nm. The observed  $V_T$  shift matches values that would be expected from quantum confinement, with PMOS showing a shift of ~50mV and NMOS a shift of ~80mV. The thinner fin NMOS capacitors also show steeper slope and higher capacitance than the thicker fins, as expected [8]. This result also shows that even the thinnest fins are not significantly eroded, as fin erosion would result in reduction of measured capacitance.

Transfer characteristics (Fig. 4), show much larger  $V_T$  shifts than the C-V curves: ~100mV for PFETs and ~250mV for NFETs. Though the capacitors are larger than the FETs (3080 fins vs. 20 fins), averaging alone cannot explain this difference. If roughness of the fin along the channel creates constrictions which act to limit current flow (Fig. 5), this could affect current flow along the entire fin, but would only affect inversion layer formation in a localized area, causing minimum  $D_{Fin}$  to dominate I-V characteristics, while C-V curves represent average  $D_{Fin}$ .

NFETs show both greater  $V_T$  shift and greater  $V_T$  variability (Fig. 6) than PFETs in I-V characteristics. The enhanced variability can be partially explained by that fact that, for (110) channel FinFETs, quantum capacitance is expected to affect

electrons more strongly than holes due to differences in effective mass. For a given amount of surface roughness and  $D_{Fin}$  distribution, therefore, NFETs would be expected to exhibit greater  $V_T$  variability due to quantum confinement.

The fact that  $V_T$  increases for both NFETs and PFETs simultaneously as  $D_{Fin}$  is reduced is consistent with a widening of the band gap due to quantum confinement. However, interface traps may also cause a  $V_T$  increase and mobility reduction for both NFETs and PFETs at the same time. Interface trap density ( $D_{it}$ ) derived from conductance is indeed higher in thinner fins (Fig. 7), though this may be due to increased series resistance. A passivation anneal reduces  $V_T$  for both N and PFETs (Fig. 8), but the  $V_T$  dependence on  $D_{Fin}$  remains, suggesting  $D_{it}$  alone cannot explain our observations.

Short channel effects such as  $V_T$  roll-off (Fig. 9) and drain induced barrier lowering (DIBL) (Fig. 10) improve as expected with reduced  $D_{Fin}$ . While  $D_{Fin}$  was not physically measured on every device, the universal relationship between electrostatic scale length ( $\lambda$ ) and DIBL (Fig. 11) provides a method of deriving  $D_{Fin}$  from measured DIBL for long channel devices, where the error in nominal  $L_G$  is small [9].  $V_T$  shift vs.  $D_{Fin}$  calculated in this way follows the expected trend for quantum confinement in FinFETs with (110) channels (Fig. 12) [8,10].

In addition to  $V_T$ , quantum confinement may also degrade mobility degradation due to enhanced sensitivity to surface roughness [10]. This, along with increased series resistance in thinner  $D_{Fin}$ s, results in the strong correlation between transconductance and  $V_T$  shift in thin fins (Fig. 13).

### Conclusion

Using a novel FinFET structure, we have fabricated for the first time arrays of thousands of fins down to 4nm  $D_{Fin}$  with robust structural integrity and yield, enabling quantum confinement effects to be observed unambiguously through capacitance measurements. In FET measurements, we observe strong fin scaling behavior including performance degradation, increased variability, and  $V_T$  increase in both NFETs and PFETs as  $D_{Fin}$  is reduced. The fin scaling behavior—and in particular the quantum confinement effects—experimentally observed in this work points to fundamental properties of thin body devices which may ultimately limit scaling of FinFETs to below 10nm  $L_G$ .

### Acknowledgements

The authors would like to thank D. Klaus, E. Kratschmer, L. Gignac, A. Bryant, M. Kobayashi, T. Ando, K. Choi, E. Carier, J. Tornello, I. Lauer, J. Sleight, and A. Majumdar for their contributions to this work. This work was completed at the IBM T.J. Watson Microelectronics Research Laboratory.

### References

- [1] <http://www.itrs.net/Links/2010ITRS/Home2010.htm>.
- [2] V. S. Basker *et al.*, VLSI 2010, p.19.
- [3] C. Yeh *et al.*, IEDM 2010, p.10-772.
- [4] G. Tsutsui *et al.*, IEEE Trans. Elec. Devl, 53, p. 2582 (2006).
- [5] J. Wang *et al.*, IEEE Trans. Elec. Dev., 51, p. 1366 (2004).
- [6] B. Yu *et al.*, IEEE Trans. Elec. Dev., 55, 2846 (2008).
- [7] M. Kobayashi *et al.*, J. Appl. Phys, vol 103, 053709, 2008.
- [8] G. Hu *et al.*, Jap. J. Appl. Phys, 49 (2010) 034001.
- [9] K. Suzuki *et al.*, IEEE Trans. Elec. Dev., vol. 40, 2326, 1993.
- [10] K. Uchida *et al.*, IEDM 2003, 33.5

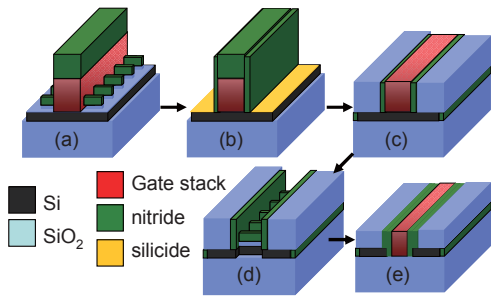


Fig. 1. FinFET fabrication process: (a) A hard mask is patterned on SOI mesas prior to the formation of a poly silicon dummy gate. (b,c) A conventional replacement gate fabrication flow is followed until the dummy gate is removed. (d) The exposed fin hard mask in the gate trench is used to define the fin using RIE. Source and drain regions are left intact. (e) An optional inner spacer is added, followed by a gate stack of  $\text{HfO}_2$ , TaN, and W gate fill.

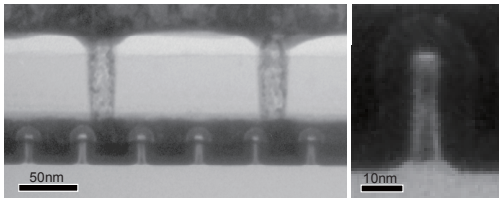


Fig. 2. TEM of nominal 4nm wide Si fins embedded within the gate.

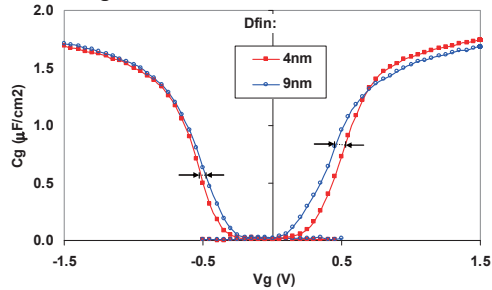


Fig. 3. Gate inversion capacitance for 3080 fins at 60nm  $FP$ , 250nm  $L_G$ , and  $D_{Fin}$  of 4 and 9nm. Arrows indicate the amount of  $V_T$  shift expected from simulation [8].  $V_T$  is defined by the point of maximal slope in the C-V curve.

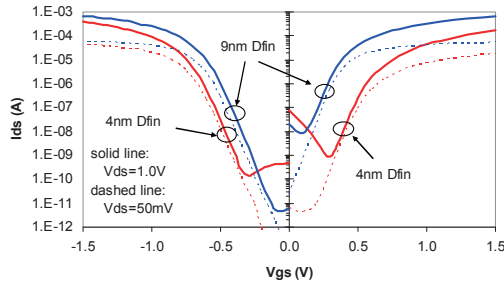


Fig. 4. Transfer characteristics of FETs with 60nm  $L_G$ , 80nm  $FP$ , 20 fins, and  $D_{Fin}$  of 4 and 9nm.

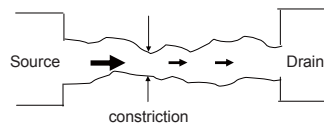


Fig. 5. A point of constriction along a fin would impede current flow (but not inversion layer formation) along the entire fin. Thus, I-V's would tend to reflect minimum  $D_{Fin}$ , while C-V's reflect average  $D_{Fin}$ .

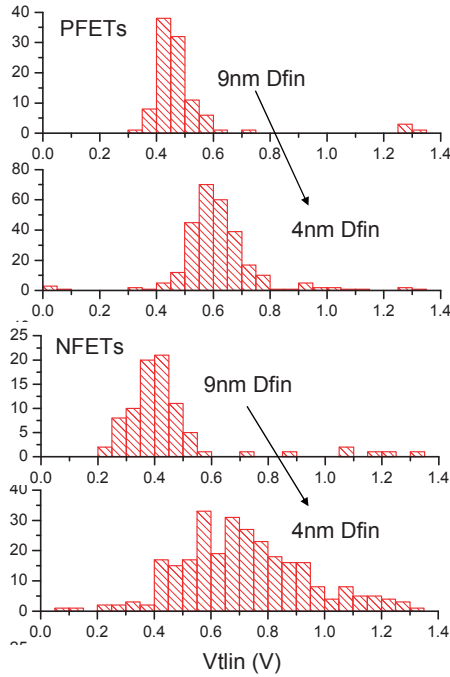


Fig. 6. Long-channel  $V_T$  distributions for 9nm and 4nm  $D_{Fin}$ . NFETs show a greater  $V_T$  shift from thick to thin fin and a significantly wider  $V_T$  distribution for the thinner fins.

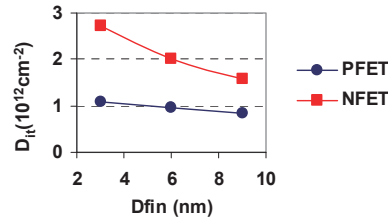


Fig. 7.  $D_{fb}$ , calculated from conductance at 1 MHz before passivation anneal, is higher for thinner fins.

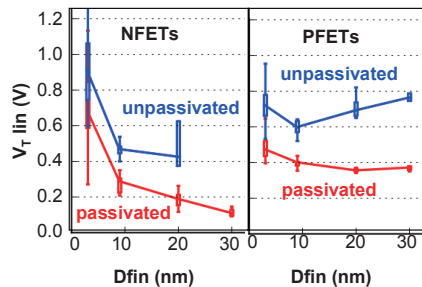


Fig. 8. Passivating interface traps reduces long-channel  $V_T$  for all  $D_{Fin}$ , but the  $V_T$  shift with thinner  $D_{Fin}$  remains.

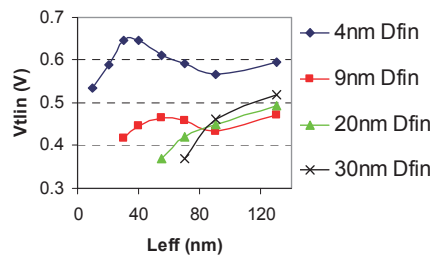


Fig. 9.  $V_T$  roll-off scales as expected with thinner  $D_{Fin}$ .

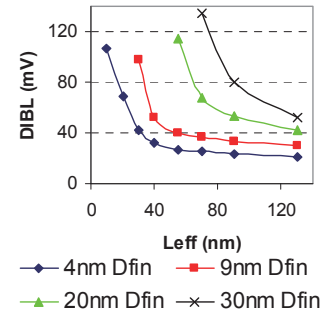


Fig. 10.  $DIBL$  (calculated from  $V_T$  shift between  $V_{ds} = 50\text{mV}$  and 1V.) improves with  $D_{Fin}$  reduction.

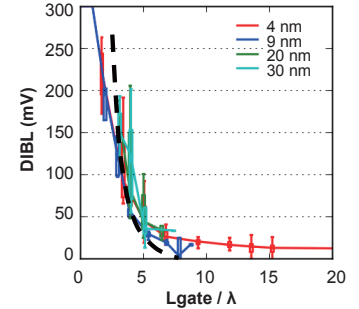


Fig. 11. Universal scaling of  $DIBL$  with electrostatic scale length ( $\lambda$ ). Dashed line indicates expected behavior [9].

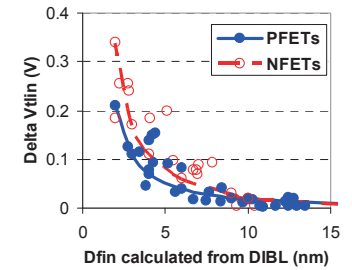


Fig. 12. Correlation of  $V_T$  shift ( $V_T$  of thick fin subtracted from  $V_T$  of thinner fins) to  $D_{Fin}$  calculated from  $DIBL$ , assuming nominal  $L_G$  with 20nm overlap. Lines show theoretical  $V_T$  shift expected from quantum confinement in (110) channel FinFETs [8].

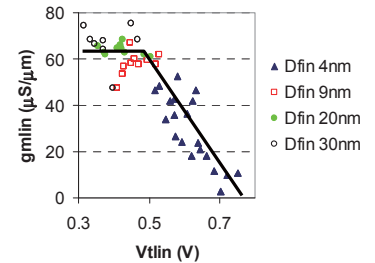


Fig. 13. The correlation of maximum transconductance to  $V_T$  is due to  $D_{Fin}$ : as  $D_{Fin}$  decreases,  $V_T$  increases, mobility decreases, and series resistance increases.