FinFET Performance Advantage at 22nm: An AC perspective

M. Guillorn, J. Chang, A. Bryant, N. Fuller, O. Dokumaci, X. Wang, J. Newbury, K. Babich, J. Ott, B. Haran, R. Yu, C. Lavoie, D. Klaus, Y. Zhang, E. Sikorski, W. Graham, B. To, M. Lofaro, J. Tornello, D. Koli, B. Yang*, A. Pyzyna, D. Neumeyer, M. Khater, A. Yagishita†, H. Kawasaki† and W. Haensch

IBM Research, *AMD, Inc., †Toshiba, Inc, IBM T.J. Watson Research Center, Yorktown Heights, NY

Abstract

At the 22 nm node, we estimate that superior electrostatics and reduced junction capacitance in FinFETs may provide a 13~23% reduction in delay relative to planar FETs. However, this benefit is offset by enhanced gate-to-source/drain capacitance (Cgs) in FinFETs. Here, we measure FinFET C_{gs} capacitance at 22nm-like dimensions and determine that, with optimization, the FinFET capacitance penalty can be limited to <6%, resulting in an overall advantage of up to 17% over a planar technology. (Keywords: FinFET, parasitic capacitance)

Introduction

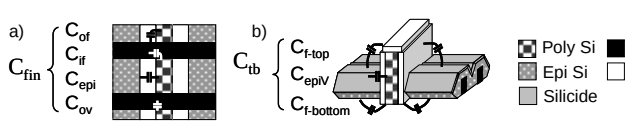
Continued gate pitch scaling demands gate length scaling, making FinFETs an attractive option for the 22nm technology node. At realistic 22nm node dimensions (Table 1), the heavy channel doping required to control off-state leakage in planar devices at 22nm will present a significant performance penalty [1]. Using TCAD and circuit simulations, we predict that improved short channel effects and reduced junction capacitance (C_j) in FinFETs can be exploited to reduce circuit delay by as much as 13 to 23% compared to a planar, silicon-on-insulator technology with SiON gate dielectric.

Table 1. Relevant dimensions for 22nm technology node

	planar	FinFET	This work
Vdd	.9V	.9V	
Gate Length (Lg)	25 nm	25 nm	≥25 nm
Contact-to-gate space	20 nm	20 nm	≥20 nm
Fin Width (Df)		12 nm	≥9 nm
Fin Height (Hf)		40 nm	≥25 nm
Fin Pitch (FP)		40 nm	≥50 nm

The FinFET advantage will be offset by the fundamental FinFET gate-to-source/drain capacitance (C_{gs}) penalty which arises from the fringing fields between the gate and the top and bottom of the source/drain ("C_{tb}", Fig 1b) [2]. While planar FETs have an analogous fringing capacitance that exists at width boundaries, the penalty from this decreases with larger device width. No such mitigating strategy is available in a FinFET technology with fixed fin height (H_f).

In this study, we use a simple analytical model (Fig 1c) to roughly separate C_{gs} into two components: a planar-like portion (C_{fin}) and the FinFET penalty (C_{tb}).



c) $C_{gs}=C_{tb}+C_{fin}*2*H_{fin}$
$C_{tb}=C_{f\text{-}top}+C_{f\text{-}bottom}+C_{epiV}$
$C_{fin}=C_{ov}+C_{of}+C_{if}+C_{epi}$

$C_{f\text{-}top/bottom}$: extracted from TCAD
C_{ov},C_{of},C_{if}: same as in planar
C_{epi},C_{epiV}: parallel plate capacitors

Fig 1. (a) C_{fin} roughly captures the planar-like components of C_{gs}, while (b) C_{tb} roughly captures the fundamental FinFET C_{gs} penalty. (c) Simple analytical model for C_{gs}.

Device Fabrication

FinFETs were fabricated with fin pitch (FP) down to 50nm and H_f from 25nm to 56nm using the flow in Fig 2. Two contact schemes were explored: via-contacted merged s/d and bar-contacted unmerged s/d regions. Devices have undoped channels and a poly-Si/SiON gate stack resulting in a negative threshold voltage. We measured C_{gs} on large arrays of several thousand fins biased with a V_{gs} of -0.5V.

The topology differences caused by varying H_f impact many subsequent unit processes, resulting in differences in fin thickness (D_{fin}), BOX undercut, and vertical epi growth (epiV). Many of the lithography, deposition, and etch processes are sensitive to pattern density, so changing FP also affects parameters such as D_{fin}, spacer thickness (tsp), and epiV. Structural variables were cataloged using extensive TEM inspection of cross-sections and used in subsequent simulation.

Results

C_{tb} and C_{fin} dependence on parameters such as FP, H_f, and spacer thickness (t_{sp}) was extracted from 3-D TCAD simulations of merged s/d FinFET structures (Fig 3a). The resulting model shows good agreement to sensitivities observed in experimental data (Fig 3b). Note that due to a discrepancy between TCAD predictions and experimental data, a constant 10aF/fin adder has been added to all experimental data presented here to aid comparison to model predictions.

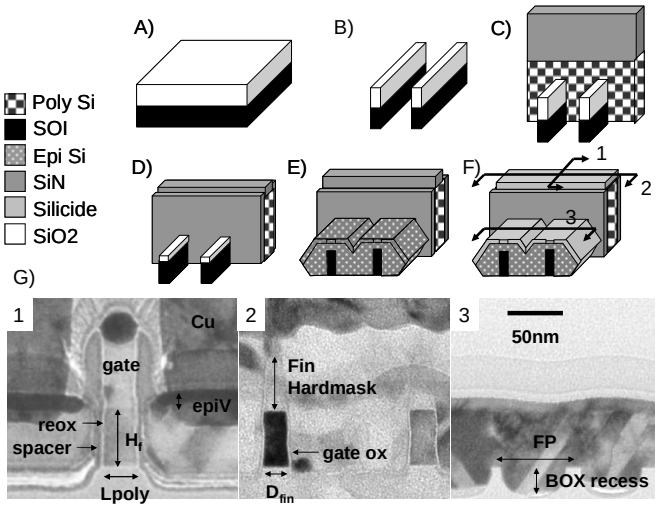


Fig. 2 Process flow: A) SOI thinning and fin oxide hardmask (HM) growth; B) Fin definition: electron beam lithography and RIE; C) Gate stack deposition (1.1nm SiON + polysilicon), gate electron beam lithography and RIE; D) Spacer formation followed by shallow tilted As implant; E) Epitaxial Si growth; F) Deep source/drain As implant, RTA and NiPt silicide formation. Devices are contacted using copper vias and wiring. (G) TEM of device cross-sections.

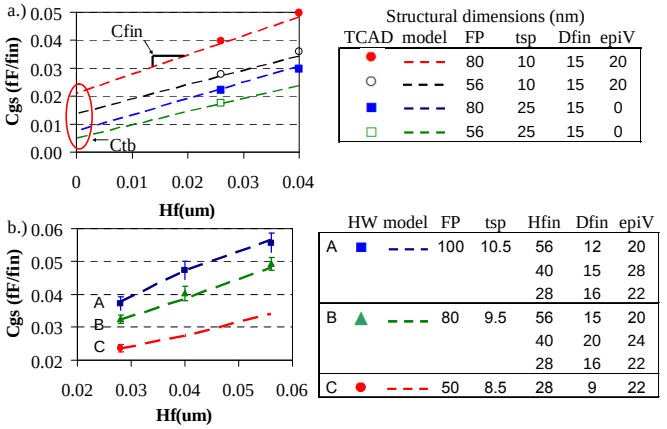


Fig. 3. (a) Extraction of model parameters from TCAD simulation (b) Comparison of model predictions to experimental data.

For 22nm node merged s/d FinFET technology, we predict FP to be the most effective means for C_{gs} reduction. C_{tb} and C_{fin} are both reduced by minimizing FP (Fig 4), predominantly due to a decrease in direct & fringing capacitance to the epi (C_{epi}, C_{epiV}, and $C_{f\text{-}top/bottom}$). Even at a fixed FP/H_f ratio, we find that reducing FP is beneficial; for example, going from a FP/H_f of 80/40nm to 40/20nm results in a 0.2fF/µm C_{gs} reduction. Using a CV/I-based model, we estimate this reduction to correspond to a ~10% inverter delay reduction (Fig 5).

Fortunately, minimizing FP is consonant with the need for granularity in device width. Assuming a minimum designed pitch for the 22nm node of 80nm, a FP of 40nm is possible using a pitch split technique (e.g. sidewall image transfer). For a fixed FP, C_{gs} may be further reduced by increasing H_{fin} (Fig 5). For a FP/H_f of 40/40nm, the delay penalty from C_{tb} is estimated to be ~9%. If the spacer is converted from nitride to oxide, the delay penalty falls to ~6%. Assuming a ~13 to 23% delay reduction from superior electrostatics and reduced C_j, FinFETs may therefore provide up to a 17% performance advantage over planar CMOS.

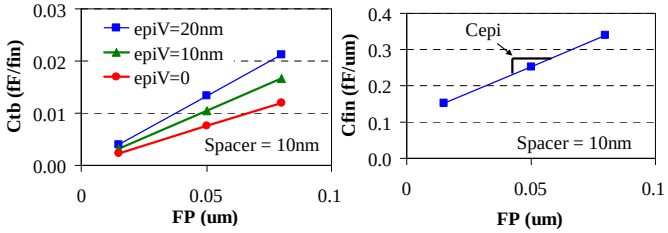


Fig. 4. Model predictions for C_{tb} and C_{fin} sensitivity to FP and epi.

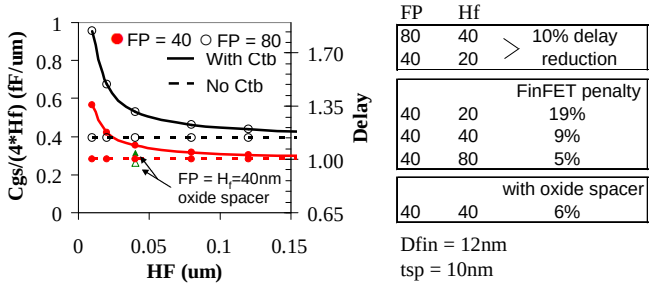


Fig. 5. Estimation of FinFET penalty (With C_{tb}) as compared to planar devices (idealized as "No C_{tb}"). The relative delay normalized to the case of "FP=40nm, no C_{tb}" is given by the 2nd y-axis.

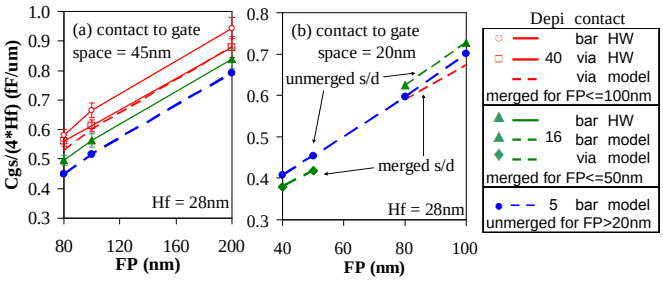


Fig. 6. Source/drain structure tradeoffs (a) at the dimensions of fabricated devices and (b) extrapolated to 22nm-like dimensions. (a) includes experimental data for bar-contacted devices with Depi = 40nm and 16nm as well as extrapolation for D_{epi} = 5nm assuming that C_{gs} varies linearly with D_{epi}. For (b), via and bar contact capacitances are estimated from TCAD simulation.

The preceding analysis assumes a merged s/d structure. Devices built in this manner may be contacted with vias identical to those used in planar technology. However, switching to an unmerged s/d structure by reducing epi thickness (D_{epi}) will reduce C_{gs} since the sensitivity of C_{tb} to FP depends strongly on vertical epi growth. Moreover, the slope of C_{fin} vs. FP is due almost entirely to C_{epi} (Fig 4).

Unmerged s/d regions require either the use of bar contacts or contact vias with a pitch equal to FP. The latter becomes less feasible as FP is reduced. Consequently, we focus on the tradeoff between via-contacted, merged FETs and bar-contacted, unmerged FETs. Experimental data, measured on structures with a contact-to-gate distance of 45nm, reveals lower C_{gs} in unmerged (D_{epi}= 16nm), bar-contacted devices than merged (D_{epi}=40nm), via-contacted devices (Fig 6a). However, when this tradeoff is extrapolated to a contact-to-gate distance of 20nm and FP = 40nm, the merged structure has lower C_{gs}.

Conclusion

We have fabricated FinFET devices at dimensions relevant for the 22nm technology node and extracted capacitance Cgs inherent to this 3D device structure. This capacitance component is a crucial parameter for evaluating FinFETs as an alternative to planar CMOS at the 22 nm technology node. The sensitivity of parasitic C_{gs} in FinFETs to structural variables such as H_f, FP, t_{sp}, and epiV is captured in a simple analytical model calibrated to both TCAD simulations and experimental measurements. Based on this model, a strategy for minimizing FinFET C_{gs} dictates minimizing FP, maximizing H_f, and using an oxide spacer and a via-contacted, merged s/d structure. This optimized structure should limit the FinFET Cgs penalty to ~6%, allowing the retention of up to a 17% FinFET performance advantage over a 22nm planar technology.

Acknowledgments

This work was performed by the Research Alliance Teams at various IBM Research and Development Facilities.

References

[1] W. Haensch, E. J. Nowak, R. H. Dennard, P. M. Solomon, A. Bryant, O. H. Dokumaci, A. Kumar, X. Wang, J. B. Johnson, and M. V. Fischetti, "Silicon CMOS devices beyond scaling," IBM J. Research & Development, vol. 50, pp. 339, 2006.

[2] W. Wu, M. Chan, "Analysis of Geometry-Dependent Parasitics in Multifin Double-Gate FinFETs," IEEE Trans. Electron Devices, vol. 54, pp. 692-698, April, 2007.