# Compact Models for Real Device Effects in FinFETs

## Quantum-Mechanical confinement and Double junctions in FinFETs

Sriramkumar Venugopalan, Muhammed A. Karim, Ali M. Niknejad and Chenming Hu
Dept. of Electrical Engineering and Computer Sciences
University of California, Berkeley
Berkeley, CA, USA
sriram@eecs.berkeley.edu

Darsen D. Lu
IBM Research Division
IBM T.J. Watson Research Center
Yorktown Heights, NY, USA

*Abstract*— **A novel geometrically scalable, phenomenological model for quantum mechanical carrier charge centroid in thin fins is presented. A model for capturing the capacitance characteristics of a graded double-junction arising out of punch-through stop implant in bulk-FinFETs is also proposed. Developed models have been included in BSIM-CMG multi-gate transistor compact model.**

***Keywords - FinFET; Quantum Mechanical Confinement; Double Junction Capacitance; BSIM-CMG; Compact Model***

### I. INTRODUCTION

FinFETs / Trigate MOSFETs are poised to take over bulk planar MOSFETs at sub-20nm technology nodes [1]. BSIM-CMG has recently been voted to become the first industry standard symmetric multi-gate compact SPICE model [2]. BSIM-CMG is physics based model with computation times that is a fraction of numerical device TCAD simulations. BSIM-CMG enables circuit designers to predict the performances of their circuits at these advance nodes with high accuracy leading to first time success of their designs. At the core BSIM-CMG supports FinFETs on bulk and SOI substrates as well as Gate-all-around (nanowire) transistors with circular and rectangular cross sections, Fig. 1 [3, 4]. Besides the core models, BSIM-CMG captures a variety of real device effects like mobility degradation, velocity saturation, drain-induced barrier lowering, parasitic resistance and capacitances, leakage currents and noise to capture real hardware data [2]. In this paper we report two new real device features introduced in the model – a novel phenomenological quantum mechanical (QM) confinement model and an enchanced junction capacitance model that accounts for the punch-through stop (PTS) implant in FinFETs on bulk substrates.

### II. QM CHARGE CENTROID MODEL

FinFETs with fin thicknesses below 20nm exhibit both structural and electrical quantum mechanical (QM) confinement of the carriers in the channel. This leads to a threshold voltage shift due to band-gap widening and a reduction in effective oxide capacitance due to the charge centroid being away from the oxide-channel interface, Fig. 2a [5, 6]. For this work TCAD device simulations (Schrodinger-Poisson-Continuity equations solved self consistently) were performed for an intrinsic silicon long (L=10μm) channel

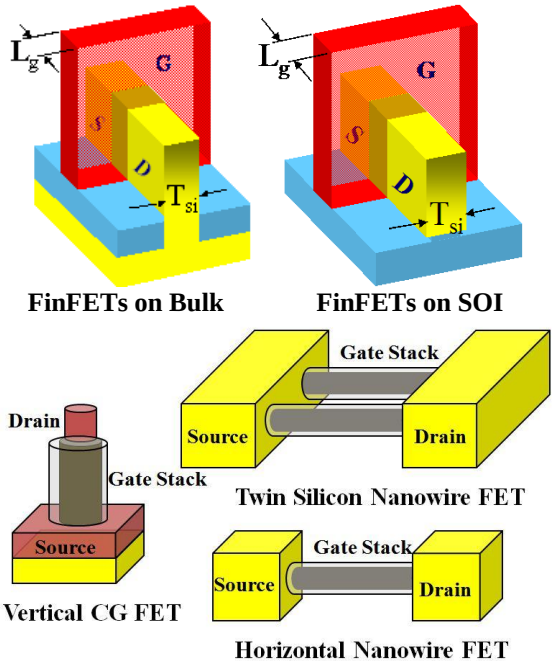


**Figure 1**: Various Multi-Gate transistor architectures supported in BSIM-CMG. All these structures allow for a common gate voltage to be applied to the multiple gates.

double-gate structure (DG FET) assuming constant mobility and with abrupt source-drain junctions [7]. A long channel length was used to avoid any drain coupling or short channel effects from affecting the charge distribution. The charge distribution within the channel was observed for different gate bias conditions at low drain terminal bias and the charge centroid was numerically extracted from the profile. We observe that the centroid position at low gate bias is a function of the channel thickness, $T_{FIN}$, Fig. 2b. We now determine analytically the asymptotic values of the charge centroid in sub-threshold regime of FET operation. For a very thin fin it is known from the solution of the Schrodinger's equation for a 1-D quantum well (assuming the carriers occupy just the first sub-band) that the distribution of charge in the channel follows a $\cos^2(\pi x/T_{FIN})$ dependence [8]. For this case, the position of

This work was supported by Semiconductor Research Corporation (SRC), Task IDs : 1149 and 2055



ISBN 978-0-615-71756-2

the centroid from the oxide-channel interface can be calculated as,

$$\frac{T_{cen0}}{T_{FIN}} = 0.5 - \frac{\int_0^{0.5} |\cos(\pi x)|^2 \, x.dx}{\int_0^{0.5} |\cos(\pi x)|^2 \, .dx} = 0.351 \quad (1)$$

For a thick fin, all the sub-bands merge to a continuum and the carriers are uniformly distributed across the fin. For this case, the position can be calculated as below,

$$\frac{T_{cen0}}{T_{FIN}} = 0.5 - \frac{\int_0^{0.5} 1.x.dx}{\int_0^{0.5} 1.dx} = 0.25 \quad (2)$$

For fins with thicknesses in between, the carriers tend to occupy multiple sub-bands. We developed a predictive model to capture this geometric dependence.

$$\frac{T_{cen0}}{T_{FIN}} = 0.25 + (0.351 - 0.25).\exp\left(\frac{T_{FIN}}{T_0}\right) \quad (3)$$

where $T_0$ is a tuning parameter. The model in Eqn. (3) agrees well with TCAD device simulations (with QM effects considered) for DG FET, Fig. 2b. We can observe that both the model and the extracted TCAD data approach the above estimated asymptotic values for the centroid. It is worthwhile to also notice that these asymptotic values calculated are independent of the channel material used.

As the gate bias increases and the device moves into strong inversion, the centroid tends to move towards the interface increasing the effective gate capacitance. As shown in Fig. 2c, this tends to happen after a critical charge in the channel. This critical charge can be calculated from the fact that the Debye screening length of the charges at this concentration is of the order of the fin thickness. For a DG FET device we write,

$$T_{FIN} \approx \sqrt{\frac{\varepsilon_0 \varepsilon_{ch}.kT}{q^2.N_{ch}}} \quad (4)$$

where $\varepsilon_0$ is the permittivity of free space, $\varepsilon_{ch}$ is the dielectric constant of channel material, q is the electronic charge, kT/q is the thermal voltage and $N_{ch}$ is the carrier density in the channel (in /cm$^3$). An empirical form for the charge centroid can now be written as follows,

$$T_{cen} = \frac{T_{cen0}}{1 + \left(\frac{Q_{inv}}{Q_0}\right)^{\alpha}} = \frac{T_{cen0}}{1 + \left(\frac{T_{FIN}.Q_{inv}}{Q_1}\right)^{\alpha}} \quad (5)$$

where $T_{cen0}$ is the upper bound of centroid obtained in Eqn. (3) and $Q_{inv}$ (in C/cm$^2$) is the inversion charge in the channel. $Q_0$ (in C/cm$^2$) is related to $N_{ch}$ as, $Q_0 = qN_{ch}T_{FIN}$. Using (4) we observe that $Q_0$ is inversely proportional to $T_{FIN}$. $Q_{inv}$ is already available for use in a Verilog-A code for a charge based model like BSIM-CMG [2]. $T_0$, $Q_1$ and $\alpha$ will be used as tuning parameters to capture any inadequacies in the theory, to support holes as well as electrons as carriers, the usage of different channel materials in the transistor and process

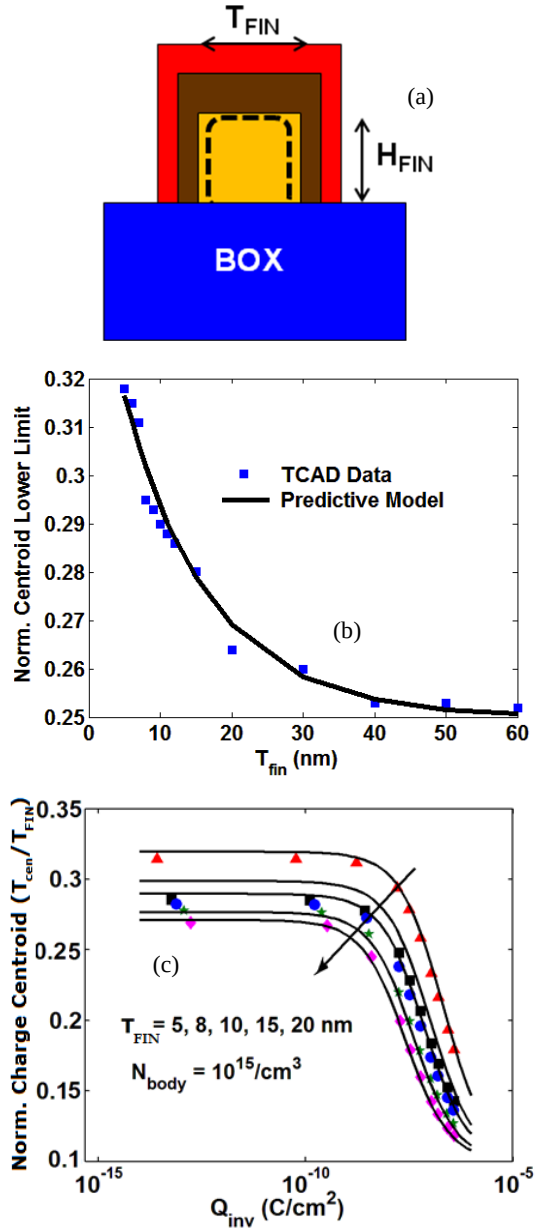


**Figure 2**: (a) Cross-section of a FinFET on SOI illustrating the charge centroid (dotted lines) being away from the oxide-channel interface. (b) Model vs. TCAD showing the geometry dependent component of the charge centroid (vs. fin thickness in sub-threshold region). (c) Model vs. TCAD overlay for the bias dependence of charge centroid. The centroid moves towards the interface with increasing gate bias. Zero on vertical-axis corresponds to the position of interface.

induced variations in the shape of the fin. Fig. 2c shows an overlay of the new model and charge centroid extracted from TCAD simulations for a silicon long channel DG FET structure. The model matches the data well for both gate bias and varying fin thicknesses. This model is then used to calculate the bias dependent effective oxide capacitance of a FinFET as follows,

$$\frac{1}{C_{oxeff}} = \frac{T_{ox}}{\varepsilon_0 \varepsilon_{ox}} + \frac{T_{cen}}{\varepsilon_0 \varepsilon_{ch}} \quad (3)$$

where $\varepsilon_{ox}$ is the dielectric constant and $T_{ox}$ is the physical thickness of the gate-oxide, $\varepsilon_{ch}$ is the dielectric constant of the channel material. In a compact model parameter extraction flow, the parameters ($T_0$, $Q_1$ and $\alpha$) introduced can be extracted from either $C_{gg}$ vs. $V_{gs}$ curve for the linear region or from a split-CV measurement of a large device.

## III. Double Junction Capacitance Model

As the channel length scales down, direct source-drain coupling leads to increased amount of leakage current. For FinFETs on bulk substrate a punch-through stop (PTS) implant is employed to prevent this coupling [9]. This implant present just below the intrinsic fin region would laterally diffuse under the source/drain junction region. The use of a high dose implant would increase the doping near the junction which leads to an increase in junction tunneling current leakage component. The magnitude of the doping at the junction would have to be optimized to balance punch-through prevention vs. high junction leakage and high junction capacitance. A graded PTS implant with lower doping near the junction and higher doping slightly below is a possible solution. An abrupt grading would create a double junction. In Figs. 3a&b the cross-section of a p-FinFET is illustrated showing the creation of a double junction with two different n-type doping – $p^+|n_{pts1}|n_{pts2}$ as a general case. When the reverse bias applied to this junction is increased (for ex: through increased positive drain voltage), the depletion region edge will traverse through the $n_{pts1}$ region and could enter the $n_{pts2}$ region. This leads to a deviation in the behavior of junction capacitance from that of an ideal uniformly doped $p^+|n$ kind of step junction diode observed in planar bulk MOSFETs. Fig. 3c shows that $1/C_{jn}^2$ vs. $V_{jn}$ curve for FinFET S/D junctions with graded PTS implant deviates from the linear behavior shown by a device without PTS implant (ideal step junction). The slope of this curve is inversely proportional to the n-type doping at the edge of the depletion region. FinFET S/D junctions in study tend to show two different slopes when a graded PTS implant is used. A new junction capacitance model has been developed to capture this process induced subtlety in the bulk FinFET junction region.

The reverse bias depletion charge is modeled as follows,

$$Q_{jn,rev} = C_{j01}\phi_{b1}\frac{1-\left(1-\frac{V_{bs/d}}{\phi_{b1}}\right)^{1-m_1}}{1-m_1}, \quad V_{bc}>V_{bs/d}>0 \quad (5)$$

$$= C_{j01}\phi_{b1}\frac{1-\left(1-\frac{V_{bc}}{\phi_{b1}}\right)^{1-m_1}}{1-m_1} + C_{j02}\phi_{b2}\frac{1-\left(1-\frac{V_{bs/d}-V_{bc}}{\phi_{b2}}\right)^{1-m_2}}{1-m_2},$$

$V_{bs/d}>V_{bc}$

where $V_{bs/d}$ is the voltage across the junction, $C_{j01,2}$ are the capacitance coefficient values and $\varphi_{b1,2}$ is the barrier height of the $p^+|n_{pts1}$ and $p^+|n_{pts2}$ junctions. $m_{1,2}$ represent the gradient of the $p^+|n_{pts1}$ the $n_{pts1}|n_{pts2}$ junctions. We can observe that the first term in Eqn. (5) is the similar to that for a single junction diode (for example see implementation for a bulk planar MOSFET in [10]). Eqn. (5) has been written such that charge continuity is maintained at the cross-over voltage $V_{bs/d}=V_{bc}$. The continuity of the first and second derivatives of charge also need to be ascertained for accuracy in the prediction of up to third harmonic content in the output of a transistor in Analog/RF circuit simulations. The continuity of the first derivative of charge in Eqn. (5) (which is the junction capacitance) at $V_{bs/d}=V_{bc}$ yields,

$$C_{j01}\left(1-\frac{V_{bc}}{\phi_{b1}}\right)^{-m_1} = C_{j02} \quad (6)$$

Ensuring continuity of the second derivative of the charge (first derivative of capacitance), at $V_{bs/d}=V_{bc}$ gives rise to the below condition.

$$C_{j01}m_1\frac{\left(1-\frac{V_{bc}}{\phi_{b1}}\right)^{-1-m_1}}{\phi_{b1}} = \frac{C_{j02}m_2}{\phi_{b2}} \quad (7)$$

These conditions, Eqns. (6) and (7) are factored into the parameter extraction process. In the junction capacitance curve, $1/C_{jn}^2$-$V_{bs/d}$, Fig. 3c the first slope region corresponding to depletion edge traversing the PTS implant region, is used to extract the values for parameters $C_{j01}$, $\varphi_{b1}$ and $m_1$ in a similar way as that for a single junction diode. Among the remaining four parameters ($V_{bc}$, $C_{j02}$, $\varphi_{b2}$ and $m_2$) that correspond to the $n_{pts2}$ region (the second slope region), conditions Eqns. (6) and (7) allow us the flexibility to choose only two of them. We chose parameters $C_{j02}$ and $\varphi_{b2}$ that signify the depth of the PTS implant – $n_{pts2}$ region boundary and the $n_{pts2}$ region doping concentration. The parameters $V_{bc}$ and $m_2$ will now be determined by simultaneously solving Eqns. (6) and (7) using the values chosen for $C_{j02}$ and $\varphi_{b2}$. The junction capacitance is then given by the derivative $dQ_{jn,rev}/dV_{bs/d}$ as follows,

$$C_{jn,rev} = C_{j01}\left(1-\frac{V_{bs/d}}{\phi_{b1}}\right)^{-m_1}, V_{bc}>V_{bs/d}>0 \quad (8)$$

$$= C_{j02}\left(1 - \frac{V_{bs/d} - V_{bc}}{\phi_{b2}}\right)^{-m_2} \quad , V_{bs/d} > V_{bc}$$

To validate the model the fin based junction structure with graded PTS implant in Fig. 3b was simulated using TCAD [2]. This new model shows excellent agreement with the TCAD simulations for such a double junction, Fig. 3c. The model smoothly shifts from one region to another that exhibit two different slopes.

## IV. Conclusion

Enhancements to real device effects in BSIM-CMG multi-gate compact model were presented. A novel model to capture the quantum mechanical confinement effects of electrons/holes in thin fins of a FinFET was proposed. Developed model captures two key dependencies of the charge centroid which is then used to model the gate capacitance of a FinFET accurately – (a) the geometry dependence of the position of the centroid ranging from thin fins to thick fins and (b) the gate bias dependence of the centroid as it moves towards the gate-oxide channel interface. A capacitance model to capture the double S/D junction characteristics that arises in FinFETs on bulk substrate where a graded punch-through stop implant is employed is discussed. The new capacitance model accurately captures the dual slope nature exhibited by the double junction as witnessed in a $1/C_{jn}^2$-$V_{bs/d}$ curve. Parameter extraction strategies have also been outlined.

## References


[1] C.C. Wu et al.,"High performance 22/20nm FinFET CMOS devices with advanced high-K/metal gate scheme," *IEDM Technical Digest*, pp.27.1.1-27.1.4, Dec. 2010.

[2] "BSIM-CMG106.0.0 Technical Manual" http://www-device.eecs.berkeley.edu/bsim/?page=BSIMCMG.

[3] M.V. Dunga, PhD Dissertation Thesis, http://www.eecs.berkeley.edu/Pubs/TechRpts/2008/EECS-2008-20.html

[4] S. Venugopalan et al., "BSIM-CG: A compact model of cylindrical/surround gate MOSFET for circuit simulations", *Solid-State Electronics*, Volume 67, Issue 1, pp.79-89, Jan. 2012.

[5] Vishal P. Trivedi and Jerry G. Fossum, "Quantum-Mechanical Effects on the Threshold Voltage of Undoped Double-Gate MOSFETs" *IEEE Electron Device Letters*, Vol. 26, No. 8, Aug. 2005, pp. 579–582.

[6] Lixin Ge and Jerry G. Fossum, "Analytical Modeling of Quantization and Volume Inversion in Thin Si-Film DG MOSFETs" *IEEE Transaction on Electron Devices*, Vol. 49, No. 2, Feb. 2002, pp. 287–294.

[7] "Sentaurus", Synopsys Inc.

[8] D.J. Griffiths, "Introduction to Quantum Mechanics", Pearson Prentince Hall, 2005

[9] K. Okano et al., "Process integration technology and device characteristics of CMOS FinFET on bulk silicon substrate with sub-10 nm fin width and 20 nm gate length", IEDM Tech. Digest, Dec. 2005, pp. 721-724.

[10] "BSIM4.7.0 Technical Manual" http://www-device.eecs.berkeley.edu/bsim/?page=BSIM4


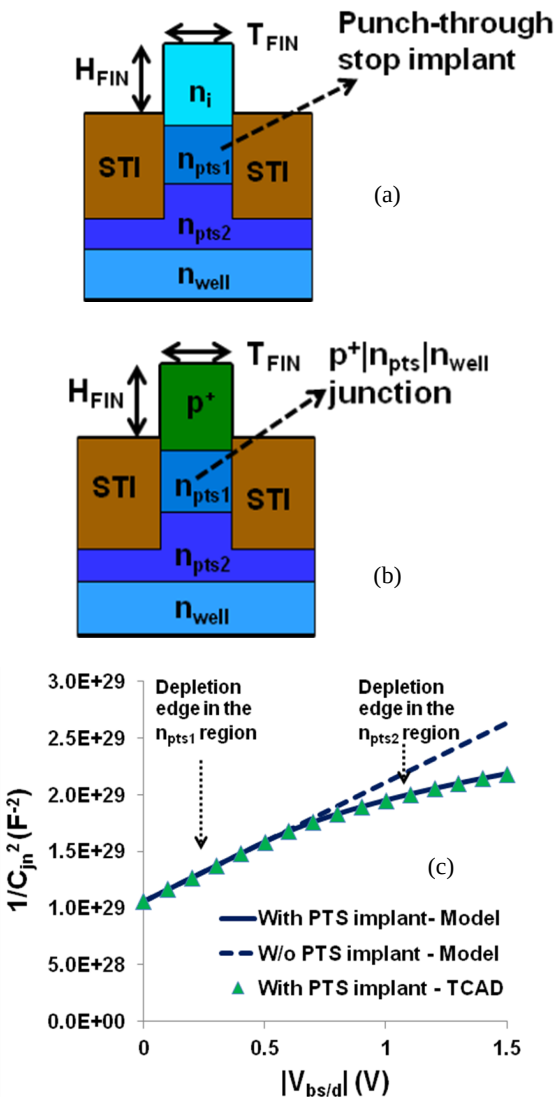


**Figure 3**: Cross-section of a bulk p-FinFET showing (a) graded punch-through stop implant below the fin (b) which laterally diffuses below the S/D junction region leading to $p^+|n_{pts1}|n_{pts2}$ type of junction. (c) This junction exhibits two slopes in a $1/C_{jn}^2$ vs $V_{jn}$ plot in comparison to single slope of the ideal step junction. Values of doping used were $p^+=3.10^{20}$, $n_{pts1}=1.10^{18}$, $n_{pts2}=3.10^{18}$ $n_{well}=5.10^{16}$ /cm$^3$ (Lines-Model, Symbols-TCAD)