# Device and Circuit Performance Issues with Deeply Scaled High-K MOS Transistors

**V. Ramgopal Rao and Nihar R. Mohapatra**

***Abstract*— In this paper we look at the effect of Fringe-Enhanced-Barrier-lowering (FEBL) for high-K dielectric MOSFETs and the dependence of FEBL on various technological parameters (spacer dielectrics, overlap length, dielectric stack, S/D junction depth and dielectric thickness). We show that FEBL needs to be contained in order to maintain the performance advantage with scaled high-K dielectric MOSFETs. The degradation in high-K dielectric MOSFETs is also identified as due to the additional coupling between the drain-to-source that occurs through the gate insulator, when the gate dielectric constant is significantly higher than the silicon dielectric constant. The technology parameters required to minimize the coupling through the high-K dielectric are identified. It is also shown that gate dielectric stack with a low-K material as bottom layer (very thin $SiO_2$ or oxy-nitride) will be helpful in minimizing FEBL. The circuit performance issues with high-K MOS transistors are also analyzed in this paper. An optimum range of values for the dielectric constant has been identified from the delay and the energy dissipation point of view. The dependence of the optimum K for different technology generations has been discussed. Circuit models for the parasitic capacitances in high-K transistors, by incorporating the fringing effects, have been presented.**

***Index Terms*—MOSFET high-k dielectric, DIBL, Parasitic capacitances, CMOS**

## I. INTRODUCTION

Continuous CMOS miniaturization necessitates the use of high dielectric constant materials to replace the thermally grown silicon dioxide as gate insulator. Some of the high-K materials such as $Al_2O_3$ ($K_{gate}$~10), $HfO_2/ZrO_2$ ($K_{gate}$~25), $La_2O_3$ ($K_{gate}$~27), $TiO_2$ ($K_{gate}$~60-100) have been widely investigated [1]. The main idea is to achieve an equivalent oxide thickness of less that 15Å by combining the higher K with a larger physical thickness to prevent direct tunneling. Some of the problems currently being looked at by various research groups include interfacial layer formation during the thermal process, microcrystalline film formation during the deposition of gate insulator, the mobility degradation and instability due to charge trapping. These studies are essential to ensure good interface between the gate insulator and silicon and for successful integration of the high-K technologies into the mainstream CMOS.

However, with increase in $K_{gate}$, the ratio of the physical thickness of the gate dielectric to channel length increases. Therefore, the percentage of field lines originating from bottom of the gate electrode and terminating on the source/drain regions (fringing lines) increases compared to the field lines terminating on the channel, thereby weakening the gate control and aggravating the short-channel effects. This phenomenon is generally termed as Fringing Induced Barrier Lowering (FIBL) [2]-[7]. Over the years, several studies have been done to investigate the effect of FIBL on device and circuit performance of high-K MOS transistors. However, these studies do not adequately describe the physics behind the FIBL phenomenon, which is very important for designing the high-K transistors for future generation low power, high performance circuits. Further, because of the thicker

Manuscript received AUGUST 29, 2003; revised FEBRUARY 16, 2004.
Department of Electrical Engineering, Indian Institute of Technology, Bombay, 400 076, India
E-mail: rrao@ee.iitb.ac.in , nihar@ee.iitb.ac.in

gate stack in the high-K gate dielectric transistors and the corresponding reduction in the transistor gate dimensions, the parasitic capacitance components need to be remodeled, and their effect on the circuit performance must be understood.

The scope of this paper is therefore twofold. In the first part of the work, we show the device performance issues of deeply scaled high-K MOS transistors under different operating bias conditions. We investigate the impact of change in technological parameters (overlap length, junction depth, dielectric stack and effective dielectric thickness) on the device performance of high-K MOS transistors. The parasitic capacitances in ultra short channel transistors are then modeled, including the high-K. In the second part of the paper, we analyse the circuit performance issues with high-K transistors, and present a correlation between the device performance degradation and its consequences on the circuit design.

Also, in this work, a quantitative estimate of fringing on the device performance is provided by extracting various capacitance components using Monte-Carlo techniques. It is shown that the device performance of high-K MOS transistors degrade significantly during low gate bias ($V_G$) and high drain bias ($V_D$) operation. From detailed simulations, this observation is linked to the coupling of drain-to-channel through gate insulator, which *enhances* the DIBL. For this reason, in this paper, this effect is termed as Fringing Enhanced Barrier Lowering (FEBL). Detailed study is done to investigate the physics behind FEBL. Finally, we show that the device and circuit performance with high-K transistors can be improved by a proper optimization of the technological parameters.

## II. SIMULATION DETAILS

Device simulations are performed using two-dimensional (2-D) device simulator MEDICI [8]. The structure used for the simulation (schematic shown in figure 1) is based on the scaled dimensions outlined in the SIA roadmap [9]. A spacer technology (spacer width, $t_{sp}$ of 60nm) with heavily doped source/drain extensions is used. The depth of source/drain extension ($X_{ext}$) and deep source/drain junction ($X_j$) is 30nm and 50nm respectively. Unless otherwise specified, for each channel length, the channel doping concentration is optimized to achieve a threshold voltage ($V_T$) of 0.25V for transistors with $K_{gate}$=3.9 ($SiO_2$). During the simulation, the permittivities of the gate dielectric are varied from 3.9 to 100 (this higher limit for $K_{gate}$ is used sometime in order to bring out the underlying physical mechanism) keeping the effective gate dielectric thickness ($T_{ox,eff}$) constant at 1.5nm.

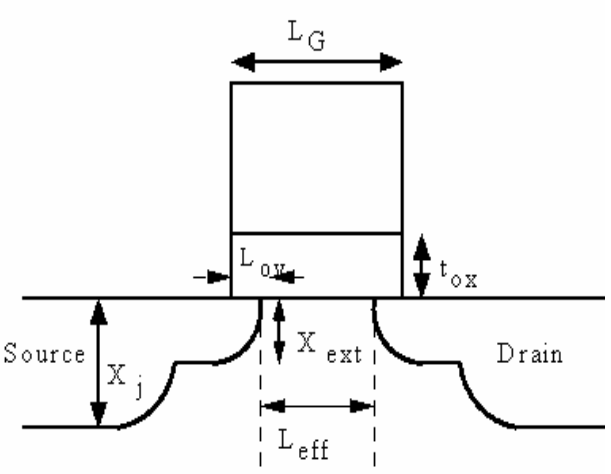


**Fig. 1.** Schematic of the MOS transistor structure used for device simulation.

The calibrated energy balance model is used along with drift-diffusion models to consider the spatial variation of carrier energy in order to account for the velocity overshoot and non-local transport phenomena. The Fermi-Dirac statistics are used to determine the active carrier density within the simulation structures. A combination of coupled and uncoupled solution techniques is used for the simulation of devices at room temperature. To isolate the degradation essentially arising out of fringing fields, the poly depletion and quantum mechanical effects are not taken into account. The $I_{ON}$ and $I_{OFF}$ of the devices used in the simulation are matched with the values reported in the SIA roadmap by a judicious choice of models and model parameters. Mixed-mode simulations are used as supported by MEDICI for understanding the circuit performance issues. For this purpose, the zero carrier solution as well as the doping information, mesh, material properties, electrode contact and model information of the NMOS and PMOS transistors are extracted using device simulation module of MEDICI. The above information is exported to the circuit simulation module for estimating

the circuit performance.

All the device capacitance extractions are extracted using CAPEM [10], a 3D interconnect capacitance extractor which can handle complex conductor and dielectric configurations. This capacitance extractor uses a floating random walk technique first proposed by by Lecoz and Iverson. More details of the Monte-Carlo simulator can be found in ref [10].

## III. EFFECT OF HIGH-K DIELECTRICS: DEVICE PERFORMANCE

MOS transistors are the basic unit of CMOS logic circuits. During logic operations it operates under two different bias conditions. These are, $V_G$ high-$V_{DS}$ low, and $V_G$ low-$V_{DS}$ high. For successful induction of high-K transistors into the mainstream technology it is therefore essential to study the operation of high-K transistors under the above bias conditions.

*(a) Operation under low $V_D$*

It is clear from previous discussions that high-K transistors need thicker gate dielectric for achieving equivalent gate capacitance and drive current as conventional $SiO_2$. The higher physical gate dielectric thickness results in higher fringing of field lines, which alters the device intrinsic capacitances as shown in figure 2.

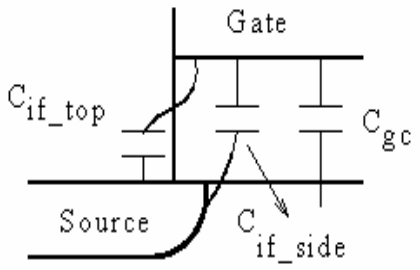


**Fig. 2.** Schematic of a MOS transistor showing various intrinsic capacitance components.

As shown in Fig. 2, $C_{if\_side}$ is the fringing capacitance associated with the electric field lines emerging from bottom of the gate electrode and terminating on the sides of source/drain regions. $C_{if\_top}$ is the fringing capacitance associated with the electric field lines emerging from bottom of the gate electrode, going through the gate dielectric and side-wall spacer and terminating on the top of source/drain regions. $C_{gc}$ is the gate to channel capacitance associated with the electric field lines originating from bottom of the gate electrode and terminating on the channel. To have a quantitative estimate of fringing effects, we have extracted the fringing capacitances using the Monte-Carlo based 3-D capacitance extractor, CAPEM [10]. Figure 3 shows the variation of $C_{if\_side}$ and $C_{if\_top}$ for different $K_{gate}$. As shown, more number of electric field lines are terminating on the inside edges of source/drain regions with increasing $K_{gate}$. These field lines induce an electric field from source-to-channel thereby reducing the source-to-channel barrier potential. Since the threshold voltage ($V_T$) of the device is controlled by the injection of electrons over the source-to-channel potential barrier, the $V_T$ of a MOSFET decreases with increase in $K_{gate}$ as shown in figure 4.

Figure 5 shows the variation of $C_{gc}$ for different $K_{gate}$. As expected, $C_{gc}$ decreases with increase in $K_{gate}$ thereby reducing the gate control over the channel. This directly affects the device sub-threshold swing (S), which is an important parameter for low power applications. For a device to have good turn-off characteristics, S should be as small as possible. Figure 5 shows the variation of S for different $K_{gate}$. As can be seen, the turn-off characteristics of the MOS transistor are degraded when the conventional $SiO_2$ is replaced by a high-K gate dielectric. This can further aggravate the already severe leakage issues for the scaled technologies when the high-K is introduced.

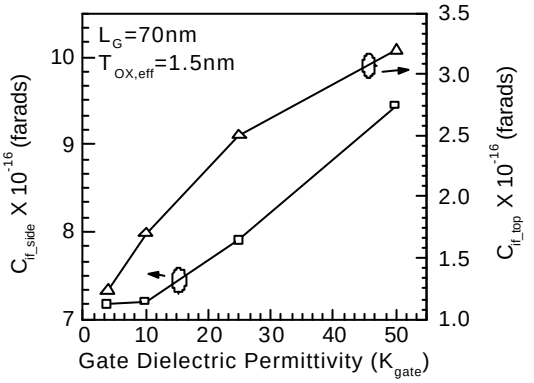


**Fig. 3.** Variation of $C_{if\_side}$ and $C_{if\_top}$ of a 70nm MOS transistor as a function of gate dielectric permittivity. The effective gate dielectric thickness is kept constant at 1.5nm during the simulation.

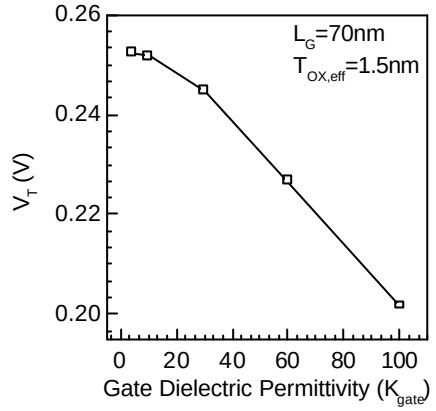


**Fig. 4.** Threshold voltage of a 70nm transistor for different gate dielectric permittivities.

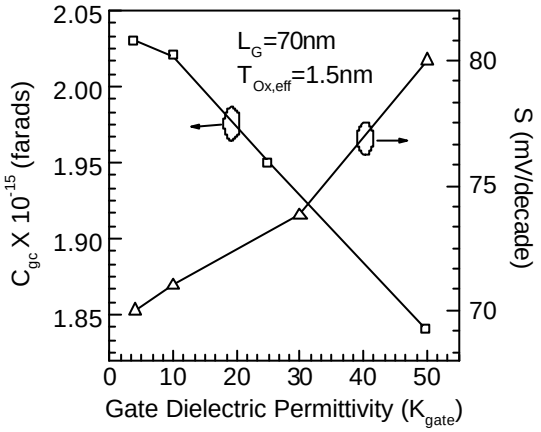


**Fig. 5.** Variation of $C_{gc}$ and the sub-threshold swing (S) of a 70nm MOS transistor as a function of gate dielectric permittivity.

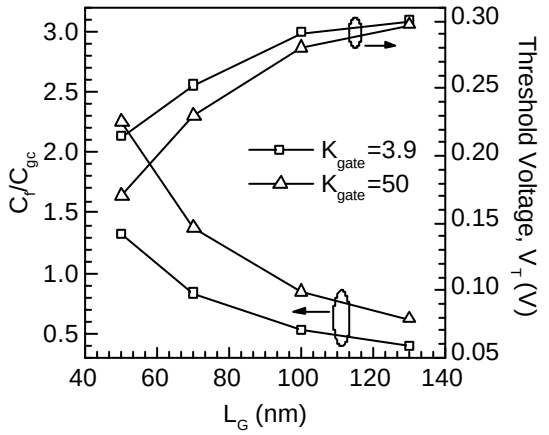


**Fig. 6.** Ratio of total fringing capacitance to gate-to-channel capacitance and threshold voltage as a function of channel length for a gate dielectric permittivity of 3.9 and 50. The effective gate dielectric thickness is kept constant at 1.5nm during the simulation.

We have also looked at the effect of $L_G$ on the device performance of high-K MOS transistors. For this purpose, we have extracted $C_{if\_side}$, $C_{if\_top}$ and $C_{gc}$ for different $L_G$ using the 3-D capacitance extractor. Note that, with the change in $L_G$, the electric field lines terminating on the channel decrease due to a decrease in the area of gate electrode without affecting the fringing field lines. Therefore, the ratio of fringing field lines from gate to source/drain regions and field lines from gate to channel is a good parameter for studying the effect of scaling on the high-K device performance. To accomplish this, we have defined the total fringing capacitance ($C_f$) as,

$$C_f = 2(C_{if\_side} + C_{if\_top}) \quad (1)$$

Figure 6 shows the ratio of $C_f$ and $C_{gc}$ as a function of $L_G$ for a $K_{gate}$ of 3.9 and 50. As can be seen, $C_f/C_{gc}$ increases for lower $L_G$ and this increase is higher for higher $K_{gate}$. In general, one can see that the fringing field lines become a significant portion of total field lines coming from the gate electrode for transistors with lower $L_G$ and higher $K_{gate}$, which leads to an increased degradation in device performance as indicated in figure 6 (right axis).

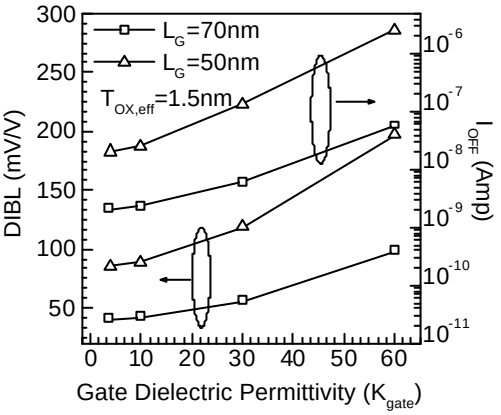


**Fig. 7.** DIBL and off current as a function of gate dielectric permittivity for 70nm and 50nm channel length NMOS transistors. During the simulation, the equivalent dielectric thickness is kept constant at 1.5nm.

*(b) Operation under high $V_D$*

Figure 7 shows the Drain Induced Barrier Lowering (DIBL) and $I_{OFF}$ as a function of $K_{gate}$ for 50nm and 70nm MOS transistors. $I_{OFF}$ is taken at a $V_G$ and $V_D$ of

0V and 1V respectively. It can be seen that at higher $V_D$ degradation in device performance is exponential with $K_{gate}$ and it is more severe for transistors with lower $L_G$.

We now explain the physical mechanism responsible for the above behavior. Figure 8 shows the physical distance ($L_{physical}$) for a point (point A) in the channel (near the source) from source, drain and gate electrodes of a MOS transistor through different paths. Note that, point A is electrically coupled to the electrodes through paths I, II and III. In our study, we have defined the equivalent electrical distance ($L_e$) of each path as the ratio of physical distance ($L_{physical}$) and the dielectric constant ($\varepsilon_r$) of the medium.

$$L_e = L_{physical} / \varepsilon_r \quad (2)$$

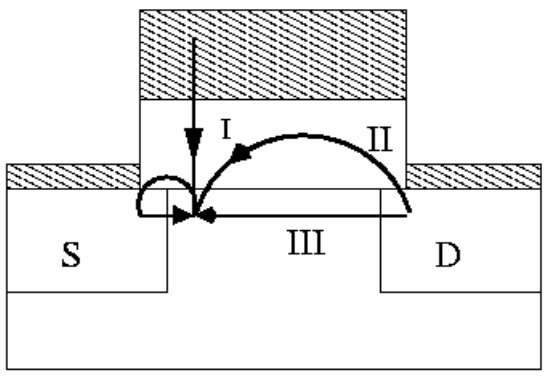


**Fig. 8.** Schematic of a MOS transistor showing the coupling of a point near source with gate, source and drain electrodes through different paths.

From the figure, it can be observed that the drain electrode is electrically coupled to the channel through paths II (through gate dielectric) and III (through silicon). The coupling through silicon is well understood and is a major contributor for enhanced short channel effects. This coupling through the channel region can be reduced by properly optimizing the junction depth ($X_J$) and using halos. On the other hand, the coupling through gate dielectric is a strong function of physical thickness of the gate dielectric as well as the drain-side channel overlap area. It may be noted that, this coupling is small (can be neglected) in conventional $SiO_2$ MOS transistors due to the smaller physical thickness of the gate oxide and its lower relative permittivity as compared to the silicon. In the case of MOS transistors with high-K gate oxides, the equivalent electrical distance ($L_e$) from gate electrode to the channel (path I) is kept constant by a scaling of the dielectric thickness in the ratio of its dielectric permittivity. However, the equivalent electrical distance between the drain and channel (path II) is not scaled leading to a decrease in $L_e$ in the lateral direction. Also, the strength of field lines originating in the drain region is higher in the higher permittivity dielectric, since $\varepsilon_r$ (high-K) > $\varepsilon_r$ (silicon). Therefore, in high-K gate dielectric MOS transistors the drain electrode is more tightly coupled to the channel through the gate insulator and electric field lines from the drain reach a longer distance into the channel. This electrically closure proximity of drain to source region forward biases the source to channel junction thereby degrading the device performance.

In order to understand this coupling through gate dielectric (path II) further in high-K transistors, we have extracted DIBL for different $X_J$. This simulation is done for two transistors having overlap length ($L_{ov}$) of 1nm and 12nm. To isolate the degradation essentially due to $L_{ov}$, which, to first order, is a measure of coupling through the high-K dielectric, the effective channel length is kept constant at 55nm for both transistors. The effective gate dielectric thickness is also kept constant at 1.5nm. The difference in DIBL ($\Delta$DIBL) of transistors (eq. 3 ) having $L_{ov}$=1nm and 12nm is calculated as a function of $X_J$ for different $K_{gate}$.

$$\Delta DIBL = DIBL\ (L_{ov} = 12nm) - DIBL\ (L_{ov} = 1nm) \quad (3)$$

Note that, DIBL occurs due to an electrical coupling between drain and source through path II and III (shown in figure 8). As mentioned earlier, the coupling through path III depends on $X_J$ while coupling through path II depends on $L_{ov}$. Therefore, $\Delta$DIBL at a fixed $X_J$ for a fixed $K_{gate}$ symbolizes the coupling between source and drain through the gate dielectric (since coupling through silicon remains almost identical and cancels out). Figure 9 shows the variation of $\Delta$DIBL as a function of $X_J$ for different $K_{gate}$. From the figure, it can be observed that for constant physical gate dielectric thickness (fixed $K_{gate}$), $\Delta$DIBL is almost independent of $X_J$. However, for fixed $X_J$, $\Delta$DIBL increases with an increase in the $K_{gate}$ due to higher physical gate dielectric thickness leading to a stronger coupling between source and drain through the gate insulator (path II). With scaling of channel

length ($L_G$), this coupling further increases due to a decrease in $L_{physical}$ resulting in a higher ΔDIBL.

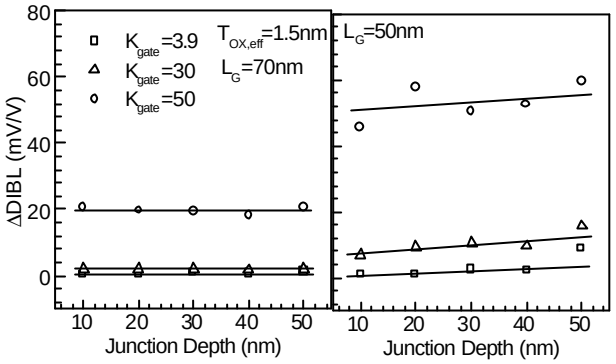


**Fig. 9.** The difference in DIBL between transistors having $L_{ov}$ of 1nm and 12nm as a function of junction depth for different gate dielectric permittivity. The effective channel length and effective dielectric thickness of the transistors is kept fixed during simulation.

*(c) Effect of technology parameters*

Note that the device performance of high-K transistors can be improved by minimizing the drain-to-source coupling through the gate insulator. In this section, we study the effect of technology parameters to accomplish this.

*Effect of spacer dielectrics*

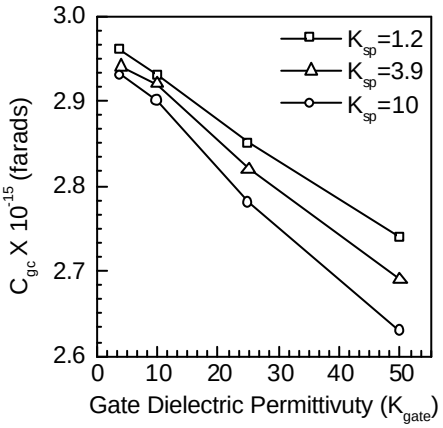


**Fig. 10.** Variation of $C_{gc}$ of a 70nm transistor as a function of gate dielectric permittivity for different spacer dielectric material. The effective gate dielectric thickness is kept constant at 1.5nm during the simulation.

Figure 10 shows $C_{gc}$ as a function of $K_{gate}$ for different spacer dielectric permittivities ($K_{sp}$). As shown, the use of low-K material as spacer dielectric can well confine the electric field lines within the channel region thereby increasing the gate control. Since, different low-K dielectrics such as porous inorganic sol-gel based material (K ~ 2-2.3), Hydrogensilsesquioxane (HSQ, K ~ 2.9), Nano-porous silica (K ~ 1.3-2.5), Aliphatic tetraflourinated poly-p-xylylene (Parylene AF4, K ~ 1.5), Poly-p-xylylene (K ~ 1.9) have been successfully evaluated as inter level dielectrics (ILD) in conventional CMOS technologies, it would be interesting to see what effect they would have on the device performance, when used as a spacer dielectric. One, of course, is aware of the technology issues and the process integration challenges of such low-K spacers. However, as can be seen from Fig. 10, the device performance improvement with low-K spacer does not look significant enough to warrant such an effort.

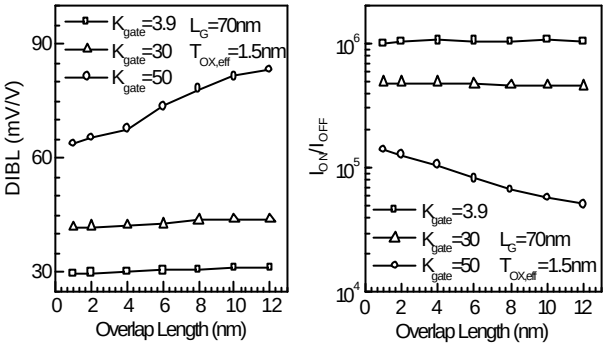


**Fig. 11.** DIBL and $I_{ON}/I_{OFF}$ of a 70nm NMOS transistor as a function of overlap length. The effective channel length and effective dielectric thickness of the transistors is kept fixed during simulation.

*Effect of overlap length*

Figure 11 shows DIBL and $I_{ON}/I_{OFF}$ as a function of gate-to-source/drain overlap length for different $K_{gate}$. As can be seen from the figure, the device performance is improved with decrease in the overlap length. This is mainly due to a suppression of the drain coupling to channel (path II as shown in Fig. 8) through the high-K dielectric, which reduces with decreasing overlap length. Also, reducing the overlap length decreases the accumulation and spread resistance by a significant amount, thereby increasing the drive current [11]. Thus, the overlap length is an important parameter for optimizing the DC performance of short channel high-K MOS transistors, which would mean tighter thermal budgets with high-K integration.

*Effect of dielectric stack*

Interfacial layers between high-K gate dielectric and the silicon are formed either intentionally or unintentionally. Intentional layers like thermal oxides ($K_{gate}$ ~ 3.9), silicon nitrides ($K_{gate}$ ~ 7) or oxy-nitrides ($K_{gate}$ ~ 5) are introduced to achieve better interfacial properties. Inadvertently formed interfacial layers, however, occur during dielectric deposition stages or during post deposition annealing in oxygen ambient. This interfacial layer may be composed of low quality nitrided oxides/oxy-nitrides and metal silicides depending on the high-K dielectric chemistry and growth conditions [12]-[13].

In this section, we study the effect of low-K interfacial layer on the device performance of high-K MOS transistors. To accomplish this the simulation is done by keeping the bottom dielectric fixed as $SiO_2$ and varying the permittivity of top dielectric ($K_{TOP}$) from 3.9 to 50. The effective dielectric thickness ($T_{eff}$) of the stack is kept constant at 1.5nm.

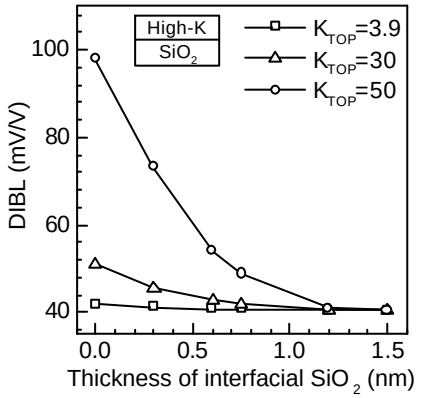


**Fig. 12.** DIBL of a NMOS transistor as a function of physical thickness of interfacial oxide. The effective dielectric thickness is kept fixed at 1.5nm during the simulation.

The above simulations are done for different physical thickness of interfacial (bottom) oxide.

$$T_{eff} = T_{OX} + (K_{TOP}/3.9) \times T_{TOP}$$

The results are shown in figure 12. As can be seen, the device performance is improved when $SiO_2$ (or a lower K material) is used as an interfacial layer in the dielectric stack. The reason for this can be attributed to the reduced coupling through the bottom dielectric, due to a suppression of path II in Fig. 8. Therefore, it can be concluded that a relatively lower K dielectric forming an interface with silicon is beneficial in controlling the EFBL.

*Effect of gate dielectric thickness*

From earlier discussions it is evident that the performance degradation in high-K transistors is caused due to drain coupling through gate insulator. This enhanced degradation of high-K transistors can be reduced by scaling the vertical dimensions (gate dielectric thickness) more aggressively. Figure 13 shows DIBL as a function of effective gate dielectric thickness ($T_{eff}$) for different $K_{gate}$. With decrease in $T_{eff}$, the gate capacitance (gate control) increases and the coupling through gate insulator (path II) reduces thereby improving the device performance. From figure 13(a), we have extracted $T_{eff}$ for different $K_{gate}$ to achieve a nominal DIBL of 50mV/V. This is shown in figure 13(b). It can be noted that, with an increase in $K_{gate}$ from 3.9 to 30, $T_{eff}$ has to be decreased by ~30% to match the $SiO_2$ performance (at $T_{eff}$ of 1.5nm). This would mean additional mobility degradation in the high-K gate dielectrics from the higher gate capacitance, even if the dielectric-silicon interface is made as good as conventional $SiO_2$. Furthermore, the decrease in $T_{eff}$ reduces $V_T$ (increases $I_{OFF}$), which must be adjusted by increasing the channel doping ($N_B$), which would mean an additional degradation in the mobility and ON currents. Higher $N_B$ will also increase the junction leakage.

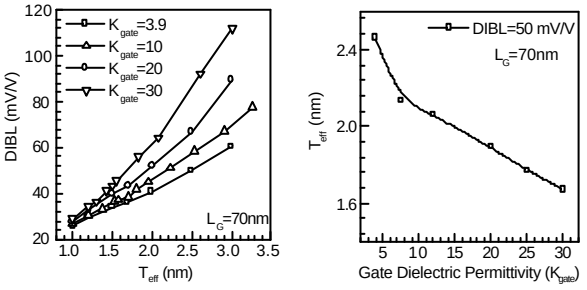


**Fig. 13.** DIBL as a function of effective gate dielectric thickness for different $K_{gate}$.

Hence, by considering all the above optimization factors, one can summarize that; better device performance of high-K MOS transistors can be achieved

through the following steps,

*Step-1:* Lower $T_{eff}$ to achieve required DIBL/off currents.

*Step-2:* Reduced gate-drain overlap length.

*Step-3:* Optimized channel doping to achieve required $I_{ON}/I_{OFF}$.

## IV. EFFECT OF HIGH-K DIELECTRICS: CIRCUIT PERFORMANCE

In this section, we study the circuit performance issues with high-K MOS transistors. Various circuit performance factors like delay, noise margin, power dissipation and power-delay product are investigated for different technology generations [14]-[15]. The observations from the above studies are analyzed using the parasitic capacitance model recently reported [14]. Optimization methods for achieving improved circuit performance are also proposed.

Figure 14 shows the different parasitic capacitances associated with a MOS transistor. As shown, the parasitic capacitance referred to source or drain ($C_{P,D}$ or $C_{P,S}$) is the parallel combination of $C_{top}$, $C_{pp}$ and $C_{side}$.

$$C_{P,D} = C_{P,S} = C_{top} + C_{pp} + C_{side} \quad (4)$$

Note that, $C_{P,D}$ and $C_{P,S}$ are equal due to the symmetrical device structure. We have extracted $C_{P,D}$ and $C_{P,S}$ from 2-D device simulations using MEDICI. Figure 15 shows the variation of $C_{P,D}$ and $C_{P,S}$ of a 70nm MOS transistor for different $K_{gate}$ and $K_{sp}$. As shown, the parasitic capacitance decreases with increase in $K_{gate}$. This can be attributed to the higher physical gate dielectric thickness of high-K MOS transistors, as studied in Section III. Furthermore, in case of low-K material as spacer dielectric, there is a further reduction in fringing, leading to a lower $C_{P,D}$ and $C_{P,S}$.

We have also recently developed [14] circuit models for accurate extraction of parasitic components for scaled MOSFETs, including the effect of fringing in high-K MOSFETs. The final model equations (Eqn. 5) are given below. For further details on the approach and significance of the different constants, readers are referred to reference [14].

$$C_{side} = \frac{\beta_1 W K_{sp} \epsilon_0}{\pi} \ln\left[\left(M^2-1\right)\left(\frac{M^2}{M^2-1}\right)^{M^2}\right], M = \frac{t_{sp}}{t_g}$$

$$C_{top} = \frac{2\beta_2 K_{gp} \epsilon_0 n_c W_c}{\pi} \ln\left[1 + \frac{L_g}{2t_{sp}}\right]$$

$$C_{pp} = 1.5 K_{sp} \epsilon_0 n_c W_c \left(\frac{t_g + t_p}{t_{sp}} - \beta_3\right) \quad (5)$$

Where,

$$t_g = t_{ox,eff} \frac{K_{gate}}{K_{ox}}$$

The total parasitic capacitance is therefore given by,

$$C_p = 2\,(C_{side} + C_{top} + C_{pp}) \quad (6)$$

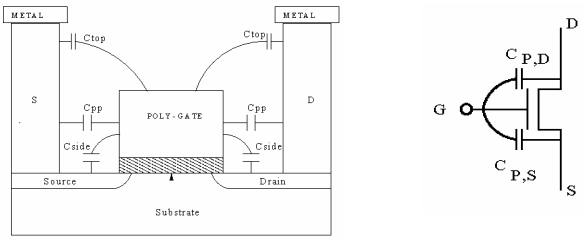


**Fig. 14.** Schematic of a MOS transistor with different parasitic capacitance components.

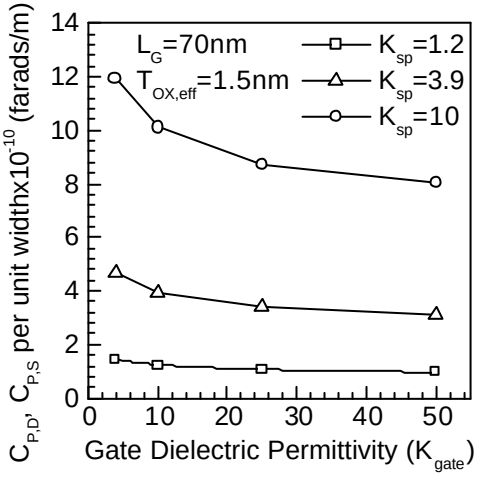


**Fig. 15.** Parasitic capacitance of a 70nm MOS transistor as a function of gate and spacer dielectric permittivity. The effective gate dielectric thickness is kept constant at 1.5nm during capacitance extraction.

Using mixed-mode simulations, we have analyzed the total energy dissipation (switching and leakage energy) of the static CMOS inverter for different $K_{gate}$ values and supply voltages. The charge transferred to the second stage during switching ($\Delta Q$) is calculated by integrating the switching current ($I_{SW}$). This integration is performed by the circuit consisting of CCCS and capacitor C1. The switching energy ($E_{switching}$) is calculated by multiplying $\Delta Q$ with the voltage swing ($V_{swing}$) at the output node of the first stage.

$$E_{switching} = \Delta Q V_{swing} \quad (7)$$

On the other hand, leakage energy is calculated using the following formula,

$$E_{leakage} = I_{OFF} V_{DD} T_C \quad (8)$$

Where, $I_{OFF}$ is the off current, $V_{DD}$ is the power supply voltage and $T_C$ is the clock width. In our study we have used a $T_C$ of 400ps. Note that, $I_{OFF}$ has several components. Some of these are PN junction reverse bias current ($I_J$), gate induced drain leakage current ($I_{GIDL}$), punch-through current ($I_{PT}$), gate oxide tunneling current ($I_{TUN}$), sub-threshold leakage current ($I_{leak}$) and current due to DIBL ($I_{DIBL}$). Among them $I_J$, $I_{GIDL}$, $I_{PT}$ are very small and can be neglected. Also, $I_{TUN}$ is expected to be small for high-K transistors, compared to $I_{leak}$ and $I_{DIBL}$. Thus,

$$I_{OFF} = I_{leak} + I_{DIBL} \quad (9)$$

Figure 16 shows total energy dissipation of a 70nm CMOS inverter as a function of $K_{gate}$. As shown, the total energy dissipation decreases with $K_{gate}$ and it shows a minimum at $K_{gate}$ of 25-30. Note that, with increase in $K_{gate}$ the parasitic capacitance as well as delay decreases. This reduces $E_{switching}$. However, $E_{leakage}$ increases with $K_{gate}$ due to an increase in the FEBL. For higher $K_{gate}$ MOS transistors, the increase in $E_{leakage}$ becomes higher compared to the decrease in $E_{switching}$. Therefore, the characteristics of total energy dissipation show a turn around at $K_{gate}$ of 25-30. Figure 17 shows the total energy dissipation as a function of delay (achieved by varying the power supply voltage) for different $K_{gate}$. As shown, for fixed energy consumption, higher performance can be achieved for circuits with high-K gate dielectrics. However, this performance improvement is limited to a narrow range of $K_{gate}$.

As discussed in section III, the higher channel doping required in high-K transistors (required to offset the lower $V_T$ caused by FEBL) and the lower $T_{eff}$ in high-K transistors (for a good control of the short-channel effects) mean an increase in the transverse electric field. This degrades the on-currents due to degraded mobility of the transistor, even if the dielectric-silicon interface is assumed to be as good as $SiO_2$. This, in turn, affects the circuit delay. Figure 19(a) shows delay of a 70nm CMOS static inverter as a function of $K_{gate}$ for a very conservatively specified $I_{OFF}$ (10, 1, 0.1nAmp/μm). As can be seen, the delay increases after a certain $K_{gate}$. The initial reduction in delay arises due to the reduced parasitic capacitances, while the subsequent increase arises due to the increased channel doping, reducing the mobility and hence the $I_{ON}$. From these characteristics, the existence of an optimum range for $K_{gate}$ for improved circuit performance can be clearly observed. Also, this optimum $K_{gate}$ is lower for stringent off-current requirements. Figure 19 (b) shows the delay for different technology generations. As the channel lengths are scaled, the optimum shifts towards the origin, which would rule out many high-K dielectrics for low power applications.

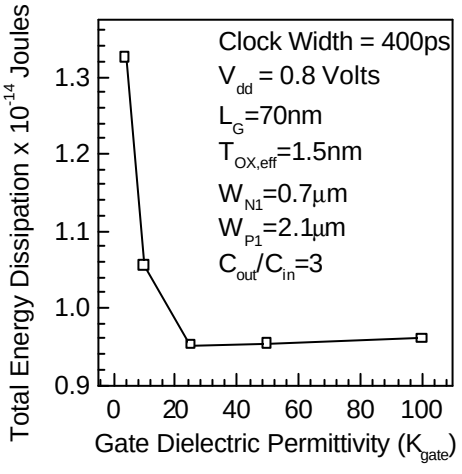


**Fig. 17.** Total energy dissipation of a 70nm CMOS static inverter as a function of gate dielectric permittivity. The effective gate dielectric thickness is kept constant at 1.5nm during simulation.

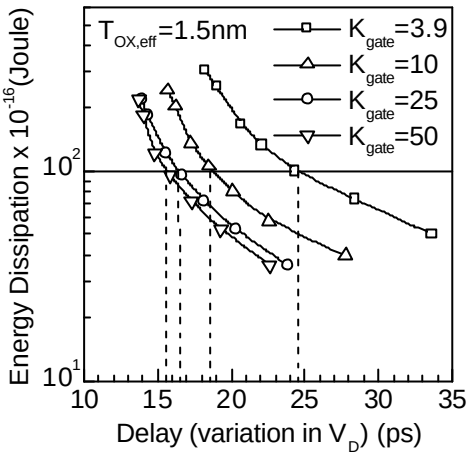


**Fig. 18.** Simulated total energy dissipation of a 70nm CMOS static inverter as a function of delay for different gate dielectric permittivity. The effective gate dielectric thickness is kept constant at 1.5nm during simulation.

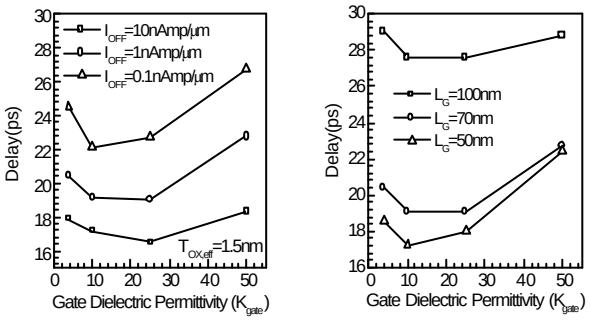


**Fig. 19.** Delay of a 70nm CMOS static inverter for different gate dielectric permittivity. The results are plotted for (a) Different $I_{OFF}$ (70nm technology) (b) Different technology ($I_{OFF}$=10nAmp/μm).

## IV. CONCLUSION

To summarize, in this paper we have studied the effect of FEBL and its dependence on different technological parameters (spacer dielectrics, overlap length, dielectric stack. S/D junction depth and dielectric thickness). It is shown that FEBL plays a significant role in determining the performance of deeply scaled high-K transistors. From the detailed simulations, we observe that the drain-to-channel coupling through the gate insulator is the main cause of FEBL. The technology parameters required to minimize the coupling through the high-K dielectric are identified, the most important one being the gate-to-source/drain overlap length. It is also shown that gate dielectric stack with a low-K material as bottom layer (very thin $SiO_2$ or oxy-nitride) will be helpful in minimizing FEBL. The circuit performance issues with high-K MOS transistors are analyzed. An optimum range of values for the dielectric constant has been identified from the delay and the energy dissipation point of view. The dependence of the optimum K for different technology generations has been discussed. Circuit models for the parasitic capacitances in high-K transistors, by incorporating the fringing effects, have been presented.

## ACKNOWLEDGEMENTS

Authors wish to acknowledge the financial support received from the Intel Corporation for carrying out this work at IIT Bombay. Valuable feedback received from Siva.G.Narendra, Intel during the course of this project is gratefully acknowledged. Authors also wish to acknowledge the feedback from the members of Microelectronics group, IIT Bombay and in particular, the involvement of Prof. Madhav P. Desai, which greatly benefited this project.

## REFERENCES

[1] G. D. Wilk, R. M. Wallace and J. M. Anthony, "High-K gate dielectrics: current status and material properties considerations", *Journal of Applied Physics*, Vol. 89, p. 5243, May 2001.

[2] B. Cheng, M. Cao, V. Ramgopal Rao, A. Inani, P. V. Voorde, W. M. Greene, J. M. C. Stork, Z. Yu, P. M. Zaeitzoff and J. C. S Woo, "The impact of High-K gate dielectrics and metal gate electrodes on sub-100nm MOSFETs", *IEEE Transaction on Electron Devices*, Vol. 46, p. 1537, 1999.

[3] A.Inani, V.Ramgopal Rao, B.Cheng, and J.C.S. Woo, "Gate Stack Architecture Analysis and Channel Engineering in Deep Sub-Micron MOSFETs", Japanese Journal of Applied Physics, Part 1, April 1999, vol.38, (no.4B): p. 2266-71

[4] Nihar.R.Mohapatra, A.Dutta, G.Sridhar, Madhav.P.Desai and V.Ramgopal Rao "Sub 100 nm CMOS Circuit Performance with High-K Gate Dielectrics" Microelectronics Reliability, Vol. 41, p.1045-1048, 2001

[5] Nihar R. Mohapatra, Madhav P. Desai, Siva G. Narendra,

V. Ramgopal Rao, "The Effect of High-K Gate Dielectrics on Deep Sub-micrometer CMOS Device and Circuit Performance" IEEE Transactions on Electron Devices, vol.49, (no.5), May 2002, p.826- 831

[6] G. C. F. Yeap, "Fringing Induced Barrier Lowering in sub-100nm MOSFETs with High-K gate dielectrics", *Electronic Letters*, Vol. 34, p. 1150, 1997.

[7] D. L. Kencke, W. Chen, H. Wang, S. Mudanai, Q. Ouyang, A. Tasch and S. K. Banerjee, "Source-side barrier effects with High-K dielectrics in 50nm MOSFETs", *Device Research Conference Digest*, p. 22, 1999.

[8] User's manual for MEDICI 2-Dimensional Device Simulation, AVANT Corporation, USA, 2000.

[9] International Technology Roadmap for Semiconductors (SIA, San Jose, CA) http://public.itrs.net, 2001.

[10] Madhav P. Desai, "An efficient implementation of 3-D capacitance extractor, CAPEM", [Online], Available: http://www.ee.iitb.ac.in/~microel/download.

[11] T. Sugii, Y. Mamiyama and K. Goto, "Continued growth of in CMOS beyond 0.1 μm", *Solid State Electronics*, Vol. 46, p. 329, 2002.

[12] S. Suehele, E. M. Vogel, M. D. Edelstein, C. A. Richter, A. V. Nguyen and I. Levin," Challenges of high-K gate dielectrics for future MOS devices", *6th International Symposium on PPID*, p. 90, 2001.

[13] S. A. Campbell, D. C. Glimmer, X. C. Wang, M. T. Hsieh, H. S. Campbell, W. G. Glasdfelter and J. Yan, "MOS transistors fabricated with high permittivity $TiO_2$ gate dielectrics", *IEEE Transaction on Electron Devices*, Vol. 44, p. 104, 1997.

[14] Nihar R. Mohapatra, Madhav P. Desai, Siva.G.Narendra, V. Ramgopal Rao, "Modeling of Parasitic Capacitances in Deep Sub-micrometer Conventional and High-K dielectric MOS Transistors" IEEE Transactions on Electron Devices, vol. 50, No.4, pp. 959-966, April 2003

[15] Nihar. R. Mohapatra, Madhav. P. Desai, Siva G. Narendra, V. Ramgopal Rao, "The Impact of High-K Gate Dielectrics on Sub 100nm CMOS Circuit Performance", Proceedings of 31st European Solid-State Device Research Conference (ESSDERC), 11 - 13 September 2001, Nuremberg, Germany, September, 2001.

**V.Ramgopal Rao** V. Ramgopal Rao obtained his M.Tech from IIT Bombay in 1991 and Dr.Ingenieur degree from the Faculty of Electrical Engineering, Universitaet der Bundeswehr Munich, Germany in 1997. His doctoral thesis was on Planar-Doped-Barrier Sub 100 nm Channel Length MOSFETs. He was a DAAD fellow for three years during 1994-1996, and, during 1997-1998, and again in 2001, a visiting scholar with the EE Department, University of California, Los Angeles. He is currently an Associate Professor in the Department of Electrical Engineering, IIT Bombay. Dr.Rao's areas of interest include Physics, Technolology and Characterization of Silicon CMOS devices for logic and mixed-signal applications, Flash Memories, and Bio-MEMS. He has over 120 publications in these areas in refereed international journals and conference proceedings and holds two patents. Prof. Rao is an Editor for the IEEE Transactions on Electron Devices in the CMOS Devices and Technology area. He is Chairman, IEEE AP/ED Bombay Chapter, and was organizing committee chair for the 17th International Conference on VLSI Design. He is also a working group member setup by the Govt. of India on Nanotechnology. Prof. Rao is a Senior Member of IEEE and a Fellow of IETE.

**Nihar R. Mohapatra** Nihar R. Mohapatra received his B. Tech degree in Electrical Engineering from Sambalpur University, Orissa, India in 1998 and Ph. D degree from the Electrical Engineering department, Indian Institute of Technology (IIT), Bombay in 2004. He is currently a technical staff at the IHP, Frankfurt (Oder), Germany. His current research interests are in the areas of Simulation and Modeling of Semiconductor Devices, MOS Transistor Physics, Non-volatile Memories and their reliability. He has over 14 publications in these areas in refereed international journals and conference proceedings.