A Comparative Analysis of Gate Leakage and Performance of High-K Nanoscale CMOS Logic Gates

Saraju P. Mohanty
Computer Science & Engineering
University of North Texas
Denton, TX 76203.
Email: smohanty@cse.unt.edu

Elias Kougianos
Electrical Engineering Technology
University of North Texas
Denton, TX 76203.
Email: eliask@unt.edu

Rabi N. Mahapatra
Computer Science
Texas A&M University
College Station, TX 77843.
Email: rabi@cs.tamu.edu

Abstract

To support the growing market demand for ever smaller portable devices, there has been continuous scaling of device feature size and geometry. This has led to a completely new dimension of power dissipation in CMOS devices, e.g. gate leakage (direct tunneling current) has emerged as the major component of power dissipation due to the use of ever lower thickness of SiO$_2$ layer. Replacement of SiO$_2$ as gate dielectric with alternative high-K dielectric materials, such as SiON, Si$_3$N$_4$, Al$_2$O$_3$, etc. is considered as a method to contain the gate leakage current, thus constructing a type of non-classical nano-CMOS transistor. Intel has recently prototyped a processor called Penryn using such transistors of 45nm technology [1]. A proper characterization of the dielectric is necessary in order to give the designer a complete idea of the efficacy of the material as a gate dielectric and to evaluate its potential in replacing SiO$_2$. It further necessitates the development of an analytical method for on-the-fly calculation of electrical characteristics of logic level representations of circuits to facilitate optimization and automatic synthesis.

Figure 1. Nanoscale Planer CMOS Transistors

Few research exist for analysis and estimation of gate leakage in CMOS circuits. In [20], the authors presented mechanisms for estimation of leakage current. In [13], a heuristic method is presented for simultaneous analysis of $I_{gate}$. The authors in [5] have formulated “state-dependent” gate leakage data. The authors in [2] have developed several methods of modeling, estimation and analysis of total leakage and its components, especially gate leakage. In [16], gate leakage analysis is done for high-K transistor based logic gates for constant loading conditions.

The rest of the paper is organized as follows: Section 2 highlights contributions of this paper. In Section 3 we present the current state of the art of high-K dielectrics in CMOS technology. In Section 4, we present a complete transient study of a prototype 2-input NAND gate. We present the characterization of logic gates with respect to design and process parameters in Section 5. The concept of analytical modeling of high-K devices is presented in Section 6. Finally, our findings and directions for future work are summarized in Section 7.
2 Our Contributions

This paper presents a novel characterization of nanoscale high-\textit{K} CMOS logic gates and provides a complete analysis of the various design and process parameters needed to enhance the design task. We introduce a new approach for the characterization of high-\textit{K} dielectric CMOS transistors and corresponding logic gates. We give a complete view of the design and process parameter variations of the devices when high-\textit{K} gate dielectric materials are used. The effect of variable load on gate oxide leakage and propagation delay of logic gates is also studied. We present novel analytical functions to provide on-the-fly calculations of the device characteristics which would provide a useful mechanism for the design engineer for design space and technology exploration during the automatic synthesis of circuits.

3 High-K Dielectrics in CMOS Technology

The CMOS fabrication process technology using high-\textit{K} dielectrics as well as their compact modeling are at early stages. In this section, we provide a brief summary.

3.1 Fabrication Technology for High-\textit{K}

Several materials have been investigated for use in nanoscale CMOS technology, such as ZrO$_2$, TiO$_2$, BST, HfO$_2$, Al$_2$O$_3$, SiON, and Si$_3$N$_4$ [10, 15, 25]. It is a challenging task in itself to integrate these materials into the conventional CMOS process [6]. Progress has been made in the development of various technologies for high-\textit{K} gate dielectric deposition [8]. This includes the extension of chemical vapor deposition (CVD), rapid thermal chemical vapor deposition (RTCVD), rapid plasma-enhanced chemical vapor deposition (RPECVD), liquid source misted chemical vapor deposition (LSCVD), physical vapor deposition (PVD) [18], jet vapor deposition (JVD) [14], oxidation of metallic films [12], and molecular beam epitaxy [11].

In this new scenario it is necessary to study the effect of high-\textit{K} dielectrics on other device parameters, including the effective gate dielectric thickness, the threshold voltage, the gate capacitance, the carrier saturation velocity and mobility, the gate polysilicon depletion depth, gate leakage and delay to provide guidelines for design engineers.

3.2 Compact Modeling for High-\textit{K}

While the materials research is in progress, there is no research addressing automatic design and synthesis of systems using high-\textit{K} MOS transistors. For compact modeling based study of high-\textit{K} non-classical transistors using BSIM4/5, two possible options can be considered: (i) varying the model parameter in the model card that denotes relative permittivity (EPSROX) and/or (ii) finding the equivalent oxide thickness (EOT) for a dielectric under consideration. Approach (i) may not be sufficient to model the behavior of non-classic nano-CMOS with non-SiO$_2$ dielectrics as it does not correctly account for the barrier height of non-SiO$_2$ dielectrics. Using method (ii) the EOT will be calculated so as to keep the ratio of relative permittivity over dielectric thickness constant.

Both of these approaches ignore several aspects of the physics behind non-SiO$_2$ dielectrics, particularly in the Si/dielectric interface. However, in the absence of published device data, the methodology presented will provide meaningful information of the various materials under consideration to match EDA development with materials.

We also believe that along with the efforts in introducing high-\textit{K} gate dielectrics, future physical-aware low power synthesis methodologies should be developed in order to incorporate them into existing automatic design or synthesis flows. This leads us to perform extensive modeling, analysis and characterization of a number of high-\textit{K} dielectric CMOS logic gates to facilitate design space exploration.

4 Analysis of Logic Gates

In our analysis we use the PTM [4] since it is well established and is able to predict the general trend of device attributes. In the absence of published data and device models, the PTM provides the means for timely and effective analysis. Since the PTM is physics based the simulations results are highly accurate and the calculated data are of comparable accuracy to TCAD simulations which are typically time and computation intensive.

We performed a complete transistor level characterization of several logic gates (NOT, NAND, NOR, AND and OR) with respect to gate leakage and delay using Spectre analog simulator [9], but present only the results for NAND for brevity. A capacitive load of value of 10 times the total gate capacitance \(C_{99}\) of the PMOS device has been used [7]. In order to obtain a fair comparison of the performance of the various dielectrics, this value was calculated for a SiO$_2$ device and kept fixed throughout the simulations.

4.1 Average Current and Delay Analysis

Following standard approaches we define the delay as the time difference between the 50\% level of the input and output waveforms [3]. For worst-case scenario, we chose the maximum delay time regardless of whether this was due to a low-to-high or a high-to-low transition.

The gate leakage of a logic gate can be calculated from all direct tunneling components for each PMOS and NMOS
device in the logic gate. There are also several components of the gate direct tunneling current within each device, such as \( I_{gs}, I_{gd}, I_{gcs}, I_{gcd}, \) and \( I_{gb} \) [16, 21]. We calculate the total gate tunneling current for each device by summing all components as:

\[
I_{\text{gate}[i]} = I_{gs}[i] + I_{gd}[i] + I_{gcs}[i] + I_{gcd}[i] + I_{gb}[i],
\]

where the index \( i \) identifies the device within a gate. A total gate leakage for the logic gate \( (I_{\text{gate}}) \) is then calculated by summing the gate currents over all the devices in the gate:

\[
I_{\text{gate}} = \sum_i I_{\text{gate}[i]}.
\]

During its various states of operation, a logic gate presents different dominant leakage paths, depending on the combination of inputs. For a 2-input logic gate, for each of the four possible states, the gate leakage currents \( I_{00}, I_{01}, I_{10}, \) and \( I_{11} \) are calculated from Eqn. 1 and 2. Assuming that all states are to occur with equal probability, an “average gate leakage (direct tunneling current)” \((I_{\text{gate}})\) is calculated as:

\[
\overline{I_{\text{gate}}} = \left( \frac{I_{00} + I_{01} + I_{10} + I_{11}}{4} \right).
\]

While characterizing the gate leakage current we present its average value over various switching states of the device. This ensures that we eliminate the effect of stacking. However, a state dependent look-up-table based gate leakage current approach can always be used to facilitate its estimation in larger CMOS circuits.

### 4.2 Parameter Variation for Simulation

The BSIM 4.4 decks generated represent a hypothetical 45nm CMOS process with oxide thickness \( T_{ox} = 1.4nm \), threshold voltage \( V_{Th} = 0.22V \) for the NMOS and \( V_{Th} = -0.22V \) for the PMOS. The nominal power supply is \( V_{DD} = 0.7V \). These decks are also scalable with respect to \( T_{ox} \) and channel length. The effect of varying oxide thickness \( (T_{ox}) \) was incorporated by varying \( TOX \) in the spice model deck directly. The effect of varying dielectric material was modeled by calculating an equivalent oxide thickness \( (T_{ox}^*) \) according to the formula:

\[
T_{ox}^* = \left( \frac{K_{SiO_2}}{K_{gate}} \right) \times T_{gate},
\]

where \( K_{SiO_2} \) and \( K_{gate} \) are the dielectric constants of the oxide and gate material, respectively.
Figure 3. Low-to-high and high-to-low propagation delay ($T_{pd}$) for a 2-input NAND logic gate for different input switching states (00, 01, 10, 11) when $K$, $T_{gate}$, and $V_{DD}$ are varied.

Here, $K_{gate}$ is the relative permittivity and $T_{gate}$ is the thickness of the gate dielectric material other than SiO$_2$, while $K_{SiO_2}$ is the dielectric constant of SiO$_2$ (=3.9). In this section we perform a parametric study for $K$; however all real values of $K$ may not translate to a specific physical dielectric in nano-CMOS technology. The length of the device is proportionately changed to minimize the impact of higher dielectric thickness on device performance and to maintain the per width gate capacitance constant as per CMOS fabrication requirements [22, 24]. Thus, the scaling ratio of channel length to the gate thickness is maintained constant. Moreover, the length and width of the transistors are chosen to maintain a ($\frac{W}{L}$) ratio of 4 : 1 for NMOS and 8 : 1 for PMOS to ensure equal flow of current through the devices. The variation of $V_{DD}$ is achieved by running a parameter sweep in the simulator.

4.2.1 Effect of Variation on Gate Leakage

As can be observed from Fig. 2(a) all the components of the gate tunneling leakage current show a uniform trend when $K$ is varied assuming fixed $T_{gate}$ and $V_{DD}$ at their nominal values. It can be seen that with an increase in the value of $K$ there is a steady decrease in the value of tunneling leakage, $I_{gate}$. This trend continues up to a knee region of $K = 12$, after which all the components eventually become almost constant and there is not much of a decrease in the value of $I_{gate}$ with an increase in $K$. This indicates the presence of a saturation zone in the gate leakage versus $K$ characteristics which implies that there is not much benefit in deploying gate dielectrics of very high-$K$ (i.e. $K > 12$) for the gate tunneling leakage reduction.

The effect of a variation on the process parameter gate thickness, $T_{gate}$, is shown in Fig. 2(b). As can be seen there is a uniform decrease among all the components of the gate tunneling current with an increase in the gate thickness. However, this effect can be utilized in some design styles as in dual-thickness approaches as an effective means of minimizing the impact of gate tunneling leakage [17].

Fig. 2(c) shows the variation of $V_{DD}$ and its corresponding effect on $I_{gate}$. As can be seen an increase in the supply voltage leads to a corresponding increase in the level of the gate tunneling leakage. This trend supports the scaling down of the supply voltage along with scaling of the device.
4.2.2 Effect of Variation on Propagation Delay

The variation of $T_{pd}$ is shown in Fig. 3(a), from which we can see that there is a sharp increase in the value of $T_{pd}$ with an increase in the gate dielectric strength; this increase continues until a value of around $K = 6$ after which the slope is much lower. The increase in propagation delay can be attributed to the increase in capacitance per unit area ($C'_{gate}$) of gate oxide with dielectric constant i.e.

$$C'_{gate} = \left( \frac{K_{gate}}{T_{gate}} \right).$$

(5)

The effect of variation of propagation delay with respect to the process parameter gate thickness shown in Fig. 3(b) indicates that there is an increase in the propagation delay of the gate as the gate thickness increases. This happens due to the increase in gate capacitance ($C_{gate}$) with oxide thickness $T_{ox}$ for a particular dielectric, say SiO$_2$ with $K$ or $\epsilon_{ox} = 3.9$ as evident from the following discussion. The gate capacitance ($C_{gate}$) is given by [24]:

$$C_{gate} = \epsilon_{ox} \left( \frac{L}{T_{ox}} \right) W = \epsilon_{ox} \left( \frac{W}{T_{ox}} \right) L. \quad (6)$$

Thus, as per the suggestion in [22, 24], with increase in $T_{ox}$ when we maintain $\left( \frac{L}{T_{ox}} \right)$ and $\left( \frac{W}{T_{ox}} \right)$ constant by increasing $L$, $C_{gate}$ increases, and hence the propagation delay. This result is consistent with the results presented in [22].

It can be seen from Fig. 3(c) that the propagation delay shows a decreasing trend with an increase in the value of the supply voltage when $K_{gate}$ and $T_{gate}$ are kept fixed. This is due to the increase in the drive current resulting from the increase in supply voltage. However, a better insight of the situation can be obtained from the following. For a technology parameter $\alpha$, propagation delay is given by [19],

$$T_{pd} = \frac{\alpha C_{load} V_{DD}}{V_{TH}} \left( V_{DSAT} \right), \quad \text{where} \quad V_{TH} = V_{TH} + 0.5 V_{DSAT},$$

$$= \frac{\alpha C_{load} \left( \frac{1}{V_{TH}-V_{DSAT}} \right)}{V_{DD}} \left( \frac{V_{TH} + 0.5 V_{DSAT}}{V_{DD}} \right), \quad \text{(dividing by} V_{DD}). \quad (7)$$

Since, $-1 < \frac{V_{TH}}{V_{DD}} < 1$, using McLaurin series we get:

$$T_{pd} = \alpha C_{load} \left( 1 + \frac{V_{TH}}{V_{DD}} + \left( \frac{V_{TH}}{V_{DD}} \right)^2 + ... \right) \approx \alpha C_{load} \left( 1 + \frac{V_{TH}}{V_{DD}} \right). \quad (8)$$

This clearly suggest that for fixed load and threshold voltage, as $V_{DD}$ increases, $T_{pd}$ decreases. However, as scaling continues, the trend is to scale down the supply along with other device features and that is also compatible with the objective of decreasing the gate leakage current in the nanometer regime.

5 Logic Gates with Specific High-K CMOS

In this section we characterize a NAND gate for the effect of $T_{gate}$ and $V_{DD}$ for a complete picture of the gate leakage and propagation delay for various high-$K$ materials in their role as CMOS transistor gate dielectrics.

A number of dielectric materials are being investigated as potential alternatives to silicon. Among them some of the more prominent ones are SiON, Si$_3$N$_4$, Al$_2$O$_3$, ZrSiO$_4$, HfSiO$_4$ and HfO$_2$ [10, 15, 25]. Figs. 4(a) and 4(b) show how these alternative dielectrics behave with respect to variations in the design and process parameters.

5.1 Variation of Process Parameters

We studied the effect of process variation on the gate tunneling current and propagation delay for various dielectrics. The combined effect of a change in the gate thickness (process parameter) with different dielectrics can be seen in Fig. 4(a). Among the dielectrics considered, HfO$_2$ has the highest dielectric constant while SiO$_2$ has the lowest. It can be clearly seen that in all cases, the gate tunneling current for SiO$_2$, having the least $K$, is the highest across all thicknesses. As the value of $K$ increases the gate tunneling current reaches a constant level for all values of thickness. In other words, the tunneling current can be obviated to a great extent with the use of high-$K$ dielectrics.

Fig. 4(b) shows the behavior of different dielectrics for a change in the process parameter $T_{gate}$ with respect to the propagation delay. In this case the dielectric with the least value of $K$, i.e. SiO$_2$ has the least propagation delay across all thicknesses. However, as the thickness increases the propagation delay increases too. In the case of dielectrics with a medium value of $K$ the propagation delay stays constant over the range of gate thickness considered.

5.2 Variation of Design Parameters

The effect of variation of design parameters on a number of gate dielectrics was studied. As can be seen from Fig. 4(c), the gate tunneling current increases with an increase in the value of the supply voltage for all dielectrics. In the case of SiO$_2$, the gate tunneling current is maximum among all the other dielectrics across the range of supply voltages from 0.5$V$ to 1.0$V$. On the other hand, for HfO$_2$, having the highest value of $K$, the gate tunneling current is the smallest throughout the range of supply voltages. For any given supply voltage, it can be seen that the gate tunneling current is lower for dielectrics with higher $K$.

The plot in Fig. 4(d) shows the behavior of propagation delay for various dielectrics when the design parameter supply voltage is changed. When the supply voltage is kept
fixed at 0.5V, the propagation delay is larger for gate dielectrics with higher values of $K$. As the supply voltage increases it can be seen from the same figure that there is a corresponding decrease in the propagation delay.

5.3 Analysis with Variable Load

We also performed the analysis of above gates in variable load conditions. The results are presented in Fig. 5(a) and 5(b) for the gate leakage and propagation delay of a NAND gate. It is observed that the gate leakage of the logic gates is almost independent of load variations. On the other hand the propagation delay increases as the load increases. This is observed for all dielectrics as evident from Fig. 5.

6 Logic Level Analytical Modeling

We perform curve-fitting of the data for each attribute viz. $K$, $T_{\text{gate}}$, and $V_{\text{DD}}$ for their effect on gate tunneling current ($I_{\text{gate}}$) and propagation delay ($T_{\text{pd}}$). The results of this modeling can be used in back-end tools of design and automatic synthesis frameworks for on-the-fly calculation.

We present a set of functions in Table 1 which show the best-fit for the effect of variation of various design and process parameters on the gate tunneling leakage and propagation delay for a 2-input NAND logic gate. Derivations for other gates are done in a similar manner, but we do not present the results for brevity. For the 2-input NAND gate, the gate tunneling current decreases exponentially with an increase in the dielectric constant. The propagation delay varies in a more complex fashion as can be seen from the table: it is represented as a combination of two functions over the range of dielectric strength considered.

Across the range of variation of $T_{\text{gate}}$ the tunneling shows a trend of exponential decrease with increase in the value of gate thickness. In the case of propagation delay, the fitting corresponds to a Langmuir function and shows an excellent fit over the range in consideration with correlation factor of 0.99913. This shows that the modeling functions are sufficiently reliable to be used for design.
The functions in Table 1 indicate an almost perfect fit of the metrics and an identical exponential fitting factor. It may be pointed out that the fitting factors for a given metric are different for the $T_{ox}$ and $V_{DD}$ fits. The reason for this is that we look at two separate curve fitting procedures: one keeping $V_{DD}$ constant and the other keeping $T_{ox}$ constant. We believe that this is more flexible than performing a simultaneous $T_{ox}$, $V_{DD}$ fit as we can isolate the effects of each factor separately. The advantages of our approach over experimentally derived relations (however, no published data are available as of now) is the predictiveness for process and design space exploration for a design engineer.

In the case of the variation of $V_{DD}$, both $I_{gate}$ and $T_{pd}$ behave in a similar fashion for all gates. While the gate leakage shows a uniform exponential increase with an increase in the value of the supply, the propagation delay shows an exponential decrease and the nature remains the same for all logic gates. All the functions yield a perfect fit and have a correlation factor of the order of 0.999 in all cases.

The fitting equations in all cases show a perfect fit and provide a very useful means of analysis and characterization of the various design and process parameters of the logic gates. This can be extended to other gates or compound gates and thus can result in a faster design process with the analytical functions replacing the look-up tables at the back end of design and automatic synthesis software.

7 Conclusions and Future Works

We presented a comprehensive analysis of the transient behavior of a CMOS NAND gate, for a realistic 45nm BSIM4 model. We used this information for the characteri-
characteristic  and modeling equations provides  an ex-
delay of the gates having various high-\( K \) gate dielectrics. It was observed that the NAND gate had minimal gate leakage and propagation delay among all the logic gates under consideration for the same experimental conditions.

We performed a further analysis of the inter-dependence of these parameters and used the data to derive curve-fitting parameters. The resulting fitting equations provide an excellent approach towards modeling the various process and design parameters of advanced nanoscale CMOS devices with high-\( K \) gate dielectrics. We propose that the data from the characterization and modeling equations provides an excellent means for the design of novel and complex low-leakage CMOS devices involving high-\( K \) dielectrics and provides valuable information to estimate the effect of gate leakage and propagation delay which can then be used to characterize entire cells and libraries.

Our ongoing research is on full swing in this field. As a first phase, we experimented based on compact models (BSIM4) and SPICE simulations. However, we are in the process of adopting two different approaches to thoroughly validate the data. Initially, a first principle physics based approach will be used, followed by TCAD based simulations.

References