A Dual Dielectric Approach for Performance Aware Gate Tunneling Reduction in Combinational Circuits

Valmiki Mukherjee
Computer Science and Engineering
University of North Texas
Denton, TX 76203.
Email: valmiki@unt.edu

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

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

***Abstract*— With continued and aggressive scaling, using ultra-low thickness SiO_2 for the transistor gates, tunneling current has emerged as the major component of leakage in CMOS circuits. In this paper, we propose a new approach called dual dielectrics of dual thicknesses (DKDT) for the reduction of both ON and OFF state gate tunneling currents. We claim that the simultaneous utilization of SiON and SiO_2 each with multiple thicknesses is a better approach for gate leakage reduction than the conventional one that uses a single gate dielectric, SiO_2, of multiple thicknesses. We develop an algorithm for the corresponding assignment of dual dielectric and dual thickness cells that minimizes the overall tunneling current for a circuit without compromising its performance. We performed extensive experiments on ISCAS'85 benchmarks using 45 nm technology which demonstrate that our approach can reduce the tunneling current by as much as 98.7% (on average 94.8%), without performance degradation.**

I. Introduction

There has been a phenomenal increase in the demand for low power and high performance digital VLSI circuits. The transistor-feature sizes have dramatically shrunk with technology scaling, resulting in a drastic change in the leakage components of the CMOS devices. While the dynamic power has remained almost unchanged, the leakage power has increased significantly to become a large portion of the total power. Therefore, there is a critical need for the reduction of leakage power, which continues to be dissipated even when a device is not performing any useful operations. The leakage current in short channel nanometer transistors has several forms, such as reverse biased diode leakage, subthreshold leakage, gate oxide tunneling current, hot carrier gate current, gate induced drain leakage, channel punch through current [1]. However, as we approach the low-end of nanotechnology, the leakage component that dominates is the gate oxide leakage current, more specifically the direct gate tunneling current.

According to the ITRS roadmap, high performance CMOS circuits will require very low gate oxide thicknesses [2]. Such ultra-thin oxide devices will be susceptible to new leakage mechanisms due to tunneling through gate oxide, which leads to gate oxide tunneling current (I_{gate}) [3]. Assuming that V_{dd} = supply voltage of the transistor and T_{gate} = gate silicon dioxide (SiO_2) thickness, the gate oxide leakage current can be expressed as follows [4], [5] : $I_{gate} \propto \left(\frac{V_{dd}}{T_{gate}}\right)^2 \exp\left(-\beta\frac{T_{gate}}{V_{dd}}\right)$, where β is an experimentally derived factor. This explains the fact that a small change in T_{gate} can have a tremendous impact on gate oxide current. This also gives the following possible options for reducing gate leakage power consumption: decreasing the supply voltage V_{dd} and/or increasing the gate SiO_2 thickness T_{gate}.

With the limits being reached in the efficacy of gate oxide thickness, it has now become desirable to find suitable alternatives to SiO_2 as the gate dielectric itself [6], [7], [8]. Recently, silicon oxynitride (SiON) has attracted attention as a replacement for SiO_2, as it has been used in silicon based processes before and is compatible with the established IC technology [7], [9], [8]. In this paper we focus on the use of two gate dielectrics SiO_2 and SiON each with two different thicknesses to optimize gate tunneling current. We develop an algorithm for the assignment of alternative logic gates comprised of transistors of different dielectrics to a CMOS circuit logic network. The algorithm ensures that the total gate oxide tunneling current of the circuit is minimized while preserving its performance.

Most of the available works in the literature have been addressing sub-threshold leakage. There are few works available that provide methodologies for gate tunneling current reduction at the logic or transistor level. Techniques like BGMOS using dual T_{gate} and dual V_{Th} (Threshold Voltage) have been proposed by Inukai, et. al. [10]. Rao, et. al. [11] have presented a sleep state assignment technique and applied this to MTCMOS circuits for reduction of both gate and subthreshold leakage. Lee, et. al. [12], [13] discuss the effect of gate tunneling with scaling and present pin-reordering as a solution for minimization of gate leakage and resolving state dependencies. Sultania, et. al. [3] have proposed a heuristic for dual T_{gate} assignment and the consequent tradeoff with the delay in the circuit. Also, Sultania, et. al. in [14], suggest an approach to minimize gate tunneling utilizing dual T_{gate} along with pin-reordering to reduce the total leakage current.

As can be seen from above, these works have focused on development of methods that use oxide of different thicknesses for gate tunneling reduction. We see that they do not address emerging dielectrics that will replace SiO_2 to reduce the tunneling current for low-end nanotechnology. They either consider ON state tunneling or OFF state tunneling, but do not account for both. The work proposed in this paper presents a new and unified performance-aware approach called DKDT for tunneling current reduction of CMOS circuits.

II. High-K dielectrics for low-end nanometer CMOS Devices

SiO_2 has reached the limit in its role as the gate dielectric of choice due to the fact that decrease in its thickness is associated with a concomitant and significant increase in tunneling current [8]. This inevitable drawback and the impending increase in power dissipation has necessitated the investigation of alternative candidates for the replacement of SiO_2. A suitable candidate needs to have a higher dielectric constant (relative permittivity) than SiO_2 [7]. This would allow for scaling down the gate thickness as well as maintaining the effective dielectric barrier height to prevent gate tunneling current. There are various metrics that define the performance of a high-K gate dielectric such as interface trap density, low frequency CV hysteresis and frequency dispersion, fixed charge density, minimal dielectric charging and interface degradation or stress induced leakage current (SICL) which result from voltage stress and surface mobility.

Recently several materials have been investigated for use in the CMOS devices such as ZrO_2, TiO_2, BST, HfO_2, Al_2O_3, SiON and a host of Silicon Nitride (Si_3N_4) compounds [7], [15], [8]. It is a challenging task in itself to integrate these materials into the conventional process and is a topic of continued research [16]. There has been a lot of progress in the development of various technologies for high-K gate dielectric deposition [9]. This includes the extension of chemical vapor deposition (CVD), single wafer methodologies such as rapid thermal chemical vapor deposition (RTCVD), rapid plasma-enhanced chemical vapor deposition (RPECVD) and liquid source misted chemical vapor deposition (LSCVD). Other techniques include physical vapor deposition (PVD) [17], jet vapor deposition (JVD) [18], oxidation of metallic films [19], molecular beam epitaxy (MBE) [20], [21].

We believe that along with the efforts in introducing high-K gate dielectrics, future synthesis methodologies should be developed in order to incorporate them into the existing automated design flows. This leads us to propose the DKDT idea that uses logic cells of dual gate dielectrics along with dual thicknesses which promises to be an efficient and effective alternative to the conventional method of a single dielectric, predominantly composed of SiO_2.

III. Problem Definition and Contributions

In this section we summarize the state-of-the art and the needs of future technologies, as discussed above. Based on these considerations we formulate our problem definition and present our contribution to address these needs. As per our discussion in section II and the current works as cited in section I the need for alternative future methodologies using dual dielectrics can be summarized as follows:

- Various synthesis works available in current literature (section I) focus on subthreshold leakage only. This calls for a new synthesis approach considering tunneling current for low-end nanotechnology CMOS circuits.
- The synthesis works available at present focus on solutions along the traditional lines of using SiO_2 only. Also various existing techniques for tunneling current reduction that exist offer methods that are limited to methodologies without any consideration to the new material perspective.

Thus, we believe that the simultaneous use of logic gates of dual dielectrics of dual thickness will prove to be an effective technique aimed towards minimization of the gate oxide tunneling current of the logic circuits while keeping the performance degradation under control.

A combinational circuit can be modeled as a weighted directed acyclic graph $G(V, E)$. The nodes $v_i \epsilon V$ are comprised of the primary inputs (PIs), the primary outputs (POs), and the combinational active elements. The edges $e_{i,j} \epsilon E$ represent the interconnections between nodes v_i and v_j. A PI is a node that has no fanins (incoming edges) and a PO is a node that has no fanouts (outgoing edges). Using this interpretation we will introduce the formulation of the dual dielectric assignment problem. The weight on the nodes can be associated with the delay in the active elements. Also, the dual dielectric assignment occurs at the technology mapping phase and as the exact layout information is not available at this stage, the interconnecting edges can be modeled as a constant delay.

Given a weighted directed acyclic graph (WDAG) $G(V, E)$ it is required to find the best possible assignment of dielectric and thickness such that the total tunneling current is minimized and latency constraint (circuit performance) is satisfied.

This can be viewed as an optimization problem as follows: Let V be the set of all vertices and V_{CP} be the set of all vertices in the critical path from the PIs to POs. The power- and performance-driven two dimensional problem can thus be formulated as follows:

$$Minimize \quad \sum_{v_i \in V} I_{gate}(v_i) \tag{1}$$

Where, $I_{gate}(v_i)$ is the tunneling current consumed per sample node v_i of the DAG, such that the following latency constraint is satisfied :

$$\sum_{v_i \in V_P} D_i(v_i) \quad \leq \quad D_{CP} \; [\forall \; v_i \in V_P \; \text{(where, } V_P \subset V)] \tag{2}$$

The constraints in Eqn. (2) ensure that the summation of all delays $D_i(v_i)$ in a given path (P) is less than the critical path delay D_{CP}.

The contributions of this paper are listed below :

- It introduces a new approach of dual dielectric (SiO_2 and SiON) assignment for the tunneling current reduction.
- Logic level minimization of average-case gate oxide tunneling current dissipation of static CMOS circuits accounting for both ON and OFF states of the device.
- Introduction of a heuristic algorithm that assigns dual dielectric material in combination with dual thickness and achieves the objective of tunneling current reduction of CMOS circuits while maintaining performance.
- A characterization methodology is shown for logic gates using $45nm$ technology to calculate the average case tunneling. Also a dual gate dielectric and thickness component library that includes various logic gates like inverter, AND, OR, NAND, and NOR is developed.

IV. Modeling and Calculation of the Gate Tunneling Current

We performed complete transistor level characterization of a number of logic gates with respect to tunneling leakage and input-output delay using Cadence Design Systems' SPECTRE analog circuit simulator [22]. Even though we concentrated on leakage and timing in this analysis, the test benches are fully parameterized to account for load and power supply variations as well as the physical dimensions of the devices. One of the obstacles in performing such technology characterization, particularly in the sub-$65nm$ region, is the lack of commercially available processes for which data have been published.

We chose to use the Berkeley Predictive Technology Model (BPTM) [23] since it is well received. The BPTM BSIM 4.4 decks generated represent a hypothetical $45nm$ CMOS process with oxide thickness $T_{gate} = 1.4nm$, $V_{Th} = 0.22V$ 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_{gate} and channel length. The effect of varying oxide thickness was incorporated by varying the parameter $TOXE$ in the spice model deck directly while the effect of varying dielectric material was modeled by first calculating an equivalent oxide thickness (T^*_{gate}) according to the formula :

$$T^*_{gate} = \left(\frac{K_{gate}}{K_{\mathrm{SiO_2}}}\right) \times T_{gate} \quad (3)$$

Here, K_{gate} is the dielectric constant of the gate dielectric material other than SiO_2, while $K_{\mathrm{SiO_2}}$ is the dielectric constant of SiO_2. It may be noted that the length of the device is proportionately changed (maintaining a constant L/T_{gate} ratio) in order to minimize the impact of higher dielectric thickness on the device performance and to maintain the per width gate capacitance constant as per fabrication requirements [3], [24]. Moreover, length and width of the transistors are chosen to maintain a (W/L) ratio of $4:1$ for NMOS and $8:1$ for PMOS to ensure proper flow of current.

The first step in the characterization was the selection of an appropriate capacitive load. A value of 10 times the total gate capacitance C_{gg} of the PMOS device was used [25]. This value depends strongly on the condition of the channel and has been calculated from the BPTM model for each case and operating condition. The effect of the switching pulse rise time (t_r) was initially examined on the delay characteristics of the various gates. Following standard approaches [26] we define the delay as the time difference between the 50% level of the input and the output waveforms. For worst-case scenarios in the development of the algorithm, we chose the maximum delay time regardless of whether this was due to a Low-to-High or a High-to-Low transition. In order to eliminate an explicit dependence of the algorithm results on t_r, we chose a value that is realistic yet does not affect the delay significantly. For $t_r = 10ps$ the dependence of the delay on t_r is minimal for all gates.

After holding t_r fixed at the selected value, we characterized first the maximum delay time for each gate and subsequently the gate direct tunneling current by evaluating all tunneling components for each PMOS and NMOS device in the logic gate. There are also several components of the gate tunneling current within each device, such as I_{gs} and I_{gd} (components due to the overlap of gate and diffusions), I_{gcs} and I_{gcd} (components due to tunneling from the gate to the diffusions via the channel) and I_{gb}, the component due to tunneling from the gate to the bulk via the channel. The total gate tunneling current for each device was then calculated by summing all components; of course, their values depend on state (ON or OFF) and type (NMOS or PMOS) of a device.

$$I_{gate}[i] = I_{gs}[i] + I_{gd}[i] + I_{gcs}[i] + I_{gcd}[i] + I_{gb}[i], \quad (4)$$

where the index i identifies the device within a gate. A total gate tunneling current for the logic gate (I_{gate}) was then calculated by summing the absolute gate currents over all the MOS devices in the logic gate (both positive and negative gate current contributes to leakage and their absolute sum account both ON and OFF states of the MOS devices):

$$I_{gate} = \textstyle\sum_i |I_{gate}[i]|. \quad (5)$$

During its various states of operation, each gate presents different dominant leakage paths, depending on the combination of inputs. For the 2-input gates we considered in this work, the characterization is straightforward as all states can be simulated, thus resulting in a complete characterization. For each of the four possible states (00, 01, 10 and 11), the overall gate tunneling current ($I_{00}, I_{01}, I_{10},$ and I_{11}, respectively) is calculated from eqs. 4 and 5. Assuming that all states are to occur with equal probability, an average gate tunneling current ($\overline{I_{gate}}$) is calculated as :

$$\overline{I_{gate}} = \left(\tfrac{I_{00}+I_{01}+I_{10}+I_{11}}{4}\right). \quad (6)$$

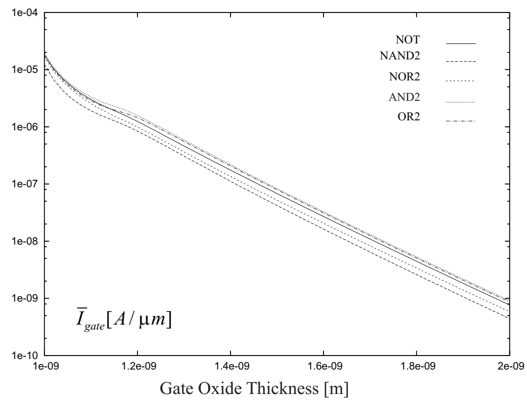
The only exception is the NOT gate which has only 2 possible inputs (0 or 1) and the average gate tunneling current was then calculated by :

$$\overline{I_{gate}} = \left(\tfrac{I_0+I_1}{2}\right). \quad (7)$$

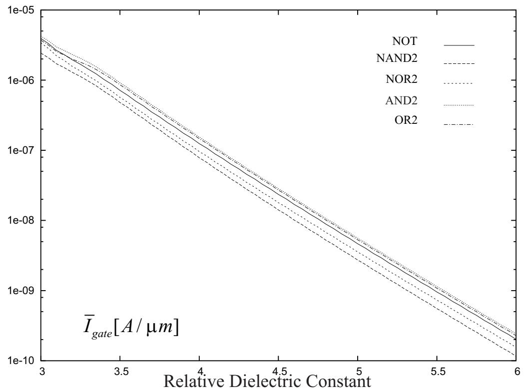
Table I shows the values of oxide tunneling current and maximum delay for various gates characterized for the experiment. A comparison of the results is shown in Fig. 1 for all logic gates under consideration. Fig. 1(a) shows the tunneling current variation with dielectric thickness when the dielectric is SiO_2. When the gate dielectric constant K_{gate} is varied the tunneling current also changes as shown in Fig. 1(b) assuming a fixed dielectric thickness T_{gate}, which is assumed as the default minimal thickness from the BSIM 4.4 model. Similar results for the propagation delay is presented in Fig. 1(c) and Fig. 1(d).

V. Dual Dielectric Assignment Algorithm

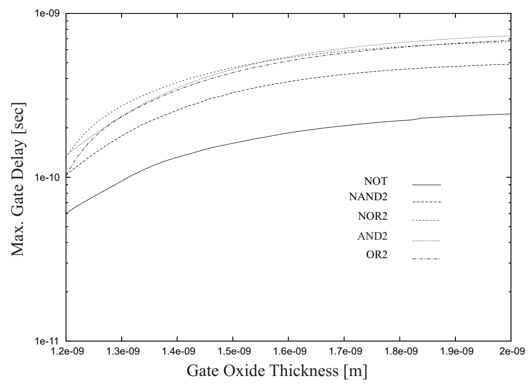
The dual dielectric assignment, described in the following paragraphs, plays a principal role in attaining the dual goal of optimizing tunneling leakage as well as integrating high-K gate technology into the synthesis flow. In the new design and synthesis flow, the dual dielectric assignment and



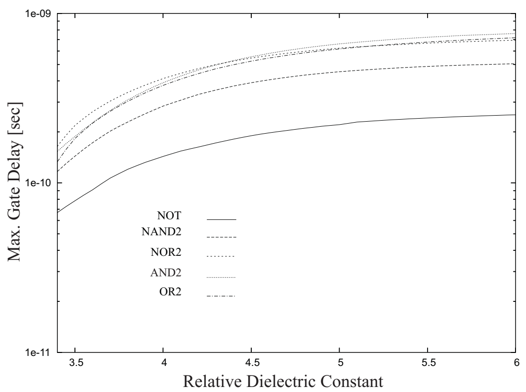
(a) Average Gate Tunneling Current ($\overline{I_{gate}}$) vs. Gate Dielectric Electrical Thickness T_{gate}. The order of the curves (top to bottom) is: AND2, OR2, NOT, NOR2 and NAND2.



(b) Average Gate Tunneling Current ($\overline{I_{gate}}$) vs. Gate Material Relative Dielectric Constant K_{gate}. The order of the curves (top to bottom) is: AND2, OR2, NOT, NOR2 and NAND2.



(c) Max. Propagation Delay vs. Gate Dielectric Electrical Thickness T_{gate}. NAND2 is the best performing of all 2-input gates.



(d) Max. Propagation Delay Versus Gate Material Relative Dielectric Constant K_{gate}. NAND2 is the best performing of all 2-input gates.

Fig. 1. Tunneling Current and Propagation Delay Variation with Gate Oxide Thickness and Dielectric for $45nm$ Technology

TABLE I

CHARACTERIZATION OF GATE LEAKAGE CURRENT AND DELAY FOR VARIOUS LOGIC GATES

	Dual Dielectric and Thickness Combinations																			
	K_1T_1					K_1T_2					K_2T_1					K_2T_2				
	I_{gate} in $nA/\mu m$				T_{pd} $in ps$	I_{gate} in $nA/\mu m$				T_{pd} $in ps$	I_{gate} in $nA/\mu m$				T_{pd} $in ps$	I_{gate} in $nA/\mu m$				T_{pd} $in ps$
	00	01	10	11		00	01	10	11		00	01	10	11		00	01	10	11	
INV	100.4	252.0	-	-	129.6	6.0	15.9	-	-	210.2	0.262	0.770	-	-	241.3	0.005	0.018	-	-	258.4
NAND2	55.8	172.0	35.8	247.6	256.9	3.1	10.7	1.0	15.6	423.7	0.134	0.506	0.029	0.754	495.3	0.0028	0.0121	0.0004	0.0185	519.2
NOR2	102.1	128.5	121.3n	246.6	378.2	6.0	8.0	7.7	15.6	586.2	0.260	0.382	0.375	0.755	680.3	0.0055	0.0094	0.0092	0.0186	724.0
AND2	179.6	295.7	160.0	298.5	350.0	11.0	18.6	8.9	18.7	611.3	0.513	0.885	0.409	0.887	741.3	0.012	0.021	0.010	0.021	802.8
OR2	225.4	179.6	171.8	297.7n	340.3	13.9	11.0	11.0	19.0	573.0	0.640	0.513	0.501	0.885	697.0	0.015	0.012	0.012	0.021	772.0

optimization are performed before placement and routing to obtain the tunneling leakage optimized netlist. This netlist is subsequently processed for the placement legalization and ECO (Engineering Change Orders) routing before generating the final layout [27], [28].

The DKDT assignment algorithm is performance aware which aims at minimizing the gate tunneling current without compromising the desired performance. In this technique, we aim at assigning from a combination of two dielectrics and two thicknesses that are assumed to be available to us as characterized cells. Let us assume that K_1 and K_2 are the relative permittivity of two gate dielectrics, where $K_1 < K_2$, and thickness $T_1 < T_2$. We assume that there are four different types of transistors available, such as K_1T_1, K_1T_2, K_2T_1, and K_2T_2. In other words, a transistor can use dielectric of relative permittivity K_1 or K_2 and of thickness T_1 or T_2. Assuming that all the transistors of a logic gate are made of same K_{gate} and equal T_{gate}, we consequently have four different types of logic gates. It is evident from the cell characterization explained in Section IV that the tunneling leakage current of logic gates increases and the propagation delay decreases in the order K_2T_2, K_2T_1, K_1T_2, and K_1T_1. This has served as the basis of the heuristic algorithm proposed in this section, where a logic gate under consideration is assigned a higher order K and T to reduce leakage whenever corresponding increase in delay of the path does not violate the target delay.

(01) Represent the network as a directed acyclic graph $G(V, E)$;
(02) Initialize each vertex $v \in G(V, E)$ with the values of leakage and delay corresponding to K_1T_1 assignment;
(03) Find the set of all paths $P\{\Pi_{in}\}\forall v \in \Pi_{in}$, the set of primary inputs, leading to primary outputs Π_{out};
(04) Compute the delay D_P for each path $p \in P\{\Pi_{in}\}$;
(05) Find the critical path delay D_{CP} for K_1T_1 assignment;
(06) Mark the critical path(s) P_{CP}, where $P_{CP} \subset P\{\Pi_{in}\}$;
(07) Assign target delay $D_T = D_{CP}$;
(08) FOR each vertex $v \in G(V,E)$ chosen in random order {
(09) Determine all paths P_v to which node v belongs;
(10) Assign K_2T_2 to v;
(11) Apply dynamic programmed LFO-NTF-DRF;
(12) Determine timing closure and insert buffers in the appropriate path;
(13) Calculate new critical delay D_{CP};
(14) Calculate slack in delay as $\Delta_D = D_T - D_{CP}$;
(15) IF ($\Delta_D < 0$) {
(16) Assign K_2T_1 to v;
(17) Apply dynamic programmed LFO-NTF-DRF;
(18) Determine timing closure and insert buffers in the appropriate path;
(19) Calculate D_{CP}; Calculate Δ_D;
(20) IF ($\Delta_D < 0$) {
(21) Assign K_1T_2 to v;
(22) Apply dynamic programmed LFO-NTF-DRF;
(23) Determine timing closure and insert buffers in the appropriate path;
(24) Calculate D_{CP}; Calculate Δ_D;
(25) IF ($\Delta_D < 0$) then reassign K_1T_1 to v;
(26) } // end IF
(27) } // end IF
(28) } // end FOR

Fig. 2. DKDT Assignment Algorithm for Performance Aware Tunneling Current reduction.

In Fig. 2 we outline the proposed heuristic algorithm. The network is represented as a weighted direct acyclic graph G(V,E). The algorithm performs assignment of dual dielectric of dual thicknesses and minimizing critical delay in the framework of a load independent delay model. It then applies an extension of the algorithms proposed in [29], [30] for computing the critical delay of the circuit. The algorithm traverses the graph in a bottom-up fashion starting from PIs to POs to identify the critical path of the circuit. Subsequently, the critical delay D_{CP} for the current assignment is calculated, which is compared against a target delay D_T (the critical delay with nodes originally assigned with K_1T_1), for any further assignment. This ensures that there is no compromise with the desired performance of the network. This also allows for a performance aware direct tunneling leakage reduction.

To begin with the assignment, the algorithm greedily considers each node as a candidate for assigning K_2T_2 which is the most desired case. In each iteration a node is considered and a determination is made whether the current assignment complies with the target delay limit D_T. Whenever an assignment is done, local fanout optimization is carried out based on a fixed node topology, as the logical structure of the network is not altered, and its effect on the overall critical delay is considered. For resolving as well as optimizing local fanouts we extend the inherent SIS [32] implementation of local fanout optimization-network topology different-rise-fall (LFO-NTF-DRF) [30] with same-rise-fall time to max-rise-fall time. We used a dynamic programming approach as suggested in [31] for resolving the LFO-NTF-DRF problem in polynomial time. At this point buffer insertion is done to implement local fanout optimization. This ensures proper timing closure at each node after assignment. For each assignment the slack Δ_D is calculated, which is defined as the difference between target delay D_T and critical delay D_{CP}. If $\Delta_D < 0$ the K_2T_2 assignment violates the target delay; then the next best assignment i. e. K_2T_1 is considered and again the value of Δ_D verified for meeting delay constraint. If this assignment still violates the target delay, K_1T_2 is assigned to the gate. If none of the above assignments meet the target delay requirement then K_1T_1 is reassigned to the node under consideration.

VI. Experimental Results

The performance aware dual dielectric assignment algorithm proposed in Sec. V was implemented in the framework of the SIS logic synthesis tool [32]. A dual dielectric and thickness library was characterized for $45nm$ technology using the SPECTRE simulator as described in section IV. The library included various logic gates like inverter, AND, OR, NAND, and NOR with a combination of dual dielectric and dual thickness as presented in Section V. We used $K_1 = 3.9$ (for SiO_2), $K_2 = 5.7$ (for SiON [6]), $T_1 = 1.4nm$, and $T_2 = 1.7nm$ to perform our experiments. The value of T_1 is chosen as the default value from the BSIM4.4 model card and the value of T_2 is intuitively chosen based on the characterization process in the previous section. The dual dielectric approach was tested on all major ISCAS'85 logic benchmarks.

The experimental results are presented in Table II. The tunneling currents reported correspond to the average case with contributions from both ON and OFF devices. It shows the values of tunneling current for K_1T_1 assignment (the base case), the tunneling current after dual dielectric assignment is done using the proposed algorithm, and percentage reduction. The results prove a considerable decrease in the gate tunneling leakage for all benchmark circuits under consideration without any tradeoff in delay.

From Table II it is further evident that our technique gives a significant reduction in tunneling leakage values. The highest reduction is observed in the case of the benchmark C6288 which is 98.69% while the lowest reduction is 89.66% in the case of the benchmark C499. Overall, the dual dielectric approach achieves, on an average, a reduction of 94.8%. However, there is no increase in the critical path delay of the overall circuit, as the nodes belonging to the critical path are not assigned with K_2 or T_2. The results vary according to the number of nodes in the critical path and the network as a whole. The more parallelism exists in the network (with almost identical path lengths, all tending to the same number of critical nodes) the higher the probability that assignment of K_2T_2 will be sparse. However, for larger circuits with longer delays along the critical path, the chances of K_2T_2 assignment increase and so does the reduction in tunneling leakage as a consequence of this assignment.

VII. Conclusions

In this paper we proposed a new approach for tunneling current reduction considering both active and sleep states

TABLE II

EXPERIMENTAL RESULTS SHOWING REDUCTIONS IN TUNNELING CURRENT

Benchmark Circuits	Number of Logic Gates	Critical Path Delay (in ps)	Tunneling Current for SiO_2 with T_1 (in nA)	Tunneling Current with DKDT Assignment (in nA)	Percentage Reduction (%)
C432	160	3.848	3949.452	253.260	93.58
C499	202	2.054	5708.547	590.454	89.66
C880	383	6.162	6537.024	337.842	94.83
C1355	546	2.054	5708.547	274.644	95.19
C1908	880	6.675	9714.744	287.721	97.04
C2670	1193	24.643	17863.326	1560.672	91.27
C3540	1669	18.227	34637.148	2215.737	93.60
C5315	2406	23.103	28156.869	1098.801	96.10
C6288	2406	24.897	28474.641	372.564	98.69
C7552	3512	26.438	33899.463	625.842	98.15

using dual gate dielectric of dual thickness. A heuristic algorithm was developed that could carry out such assignment for benchmark circuits in a reasonable amount of time. The experiments yielded significant reductions in tunneling current without compromising the performance of the circuit. It may be noted that dual dielectric circuits may need to use more masks during the lithographic process of circuit fabrication. But, we believe that such costs would be compensated by the reduction of energy or power costs. However, the research on these materials by the material science and engineering as well as electrical engineering community is in full swing and we expect to see new process technologies in the future addressing these issues. We have focussed on heuristic based algorithms, but more optimal algorithms are under development.

REFERENCES

[1] K. Roy, S. Mukhopadhyay, and H. M. Meimand, "Leakage Current Mechanisms and Leakage Reduction Techniques in Deep-Submicrometer CMOS Circuits," *Proceedings of the IEEE*, vol. 91, no. 2, pp. 305–327, February 2003.

[2] "Semiconductor Industry Association, International Technology Roadmap for Semiconductors," `http://public.itrs.net`.

[3] A. K. Sultania, D. Sylvester, and S. S. Sapatnekar, "Tradeoffs Between Gate Oxide Leakage and Delay for Dual T_{ox} Circuits," in *Proceedings of Design Automation Conference*, 2004, pp. 761–766.

[4] N. S. Kim, et. al., "Leakage Current - Moore's Law Meets Static Power," *IEEE Computer*, pp. 68–75, December 2003.

[5] A. Chandrakasan, W. Bowhill, and F. Fox, *Design of High-Performance Microprocessor Circuits*, IEEE Press, 2001.

[6] E. M. Vogel, et. al., "Modeled Tunnel Currents for High Dielectric Constant Dielectrics," *IEEE Transactions on Electron Devices*, vol. 45, no. 6, pp. 1350–1355, June 1998.

[7] L. Manchanda, et. al., "High K gate Dielectrics for the Silicon Industry," in *Proc. of International Workshop on Gate Insulator*, 2001, pp. 56–60.

[8] A. Karamcheti, V.H.C. Watt, H.N. Al-Shareef, T.Y. Luo, G.A. Brown, M. D. Jackson, and H.R. Huff, "Silicon Oxynitride Films as Segue to the High-K Era," *Semiconductor Fabtech*, vol. 12, 2000.

[9] H. R. Huff, et. al., "Integration of high-k Gate Stack Systems into Planar CMOS Process Flows," in *Proceedings of the International Workshop on Gate Insulator*, 2001, pp. 2–11.

[10] T. Inukai, et. al., "Boosted gate MOS (BGMOS): Device / Circuit Cooperation Scheme to Achieve Leakage-Free Giga-Scale Integration," in *Proc. of Custom Integrated Circuits Conference*, 2000, pp. 409–412.

[11] R. M. Rao, J. L. Burns, and R. B. Brown, "Circuit Techniques for Gate and Sub-Threshold Leakage Minimization in Future CMOS Technologies," in *Proceedings of the European Solid-State Circuits Conference*, 2003, pp. 313–316.

[12] D. Lee and D. Blaauw, "Static Leakage Reduction Through Simultaneous Threshold Voltage and State Assignment," in *Proceedings of the Design Automation Conference*, 2003, pp. 191–194.

[13] D. Lee, D. Blaauw, and D. Sylvester, "Gate Oxide Leakage Current Analysis and Reduction for VLSI Circuits," *IEEE Transactions on VLSI Systems*, vol. 12, no. 2, pp. 155–166, February 2004.

[14] A. K. Sultania, D. Sylvester, and S. S. Sapatnekar, "Transistor and Pin Reordering for Gate Oxide Leakage Reduction in Dual T_{ox} Circuits," in *Proceedings of ICCD*, 2004, pp. 228–233.

[15] M. Yang, et. al., "Performance Dependence of CMOS on Silicon Substrate Orientation for Ultrathin and HfO_2 Gate Dielectrics," *IEEE Electron Device Letters*, vol. 24, no. 5, pp. 339–341, May 2003.

[16] X. Guo and T. P. Ma, "Tuenneling Leakage Current in Oxynitride: Dependence on Oxygen/Nitrogen Content," *IEEE Electron Device Letters*, vol. 19, no. 6, pp. 207–209, June 1998.

[17] W.-J. Qi, et. al., "Ultrathin Zirconium Silicate Film With Good Thermal Stability for Alternative Gate Dielectric Application," *Applied Physics Letters*, vol. 77, pp. 1704–1706, 2000.

[18] T.P. Ma, "Making Silicon Nitride a Viable Gate Dielecrtric," *IEEE Transaction on Electron Devices*, vol. 45, pp. 680–690, 1999.

[19] B.H. Lee, L. Kang, R. Nieh, W-J Qi, and J.C. Lee, "Thermal Stability and Electrical Characteristics of Ultrathin Hafnium Oxide Gate Dielectric Reoxidized with Rapid Thermal Annealing," *Applied Physics Letters*, vol. 77, pp. 1926–1928, 2000.

[20] S. Guha and A. Cartier et.al., "Atomic Beam Decomposition of Lanthanum and Yttrium based Oxide Thin Films for Gate Dielectrics.," *Applied Physics Letters*, vol. 77, pp. 2710–2712, 2000.

[21] A.I. Kingon, J-P Maria, and S.K. Streifferr, "Alternative Dielectrics to Silicon Dioxide for memory and Logic Devices.," *Nature*, vol. 406, pp. 1021–1038, 2000.

[22] Cadence Design Systems, *Spectre Circuit Simulator User's Guide*, 2005.

[23] Y. Cao, et. al., "New Paradigm of Predictive MOSFET and Interconnect Modeling for Early Circuit Design," in *Proceedings of the IEEE Custom Integrated Circuits Conference*, 2000, pp. 201–204.

[24] N. H. E. Weste and D. Harris, *CMOS VLSI Design : A Circuit and Systems Perspective*, Addison Wesley, 2005.

[25] J. G. Hansen, "Design of CMOS Cell Libraries for Minimal Leakage Currents," M.S. Thesis, Dept. of Informatics and Mathematical Modelling, Technical University of Denmark, Fall, 2004.

[26] R. J. Baker, H. W. Li, and D. E. Boyce, *CMOS: Circuit Design, layout, and Simulation*, IEEE Press, 1998.

[27] U. Brenner and J. Vygen, "Legalizing a placement with minimum total movement," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 23, no. 12, pp. 1597–1613, December 2004.

[28] J. Cong, J. Fang, and K. Y. Khoo, "An implicit connection graph maze routing algorithm for ECO routing," in *Proceedings of the International Conference on Computer-Aided Design*, 1999, pp. 163–167.

[29] L. P. P. P. van Ginneken, "Buffer placement in Distributed RC-tree Networks for Minimal Elmore Delay," in *Proceedings of the IEEE International Sympoisum on Circuits and Systems*, 1990, p. 865868.

[30] R. Murgai, "Performance Optimization Under Rise and Fall Parameters," in *Proceedings of the IEEE/ACM International Conference on Computer-Aided Design*, 1999, pp. 185–190.

[31] R. Murgai, "On The Global Fanout Optimization Problem," in *Proceedings of the IEEE/ACM International Conference on Computer-Aided Design*, 1999, pp. 511–515.

[32] E. M. Sentovich, et. al., "SIS: A System for Sequential Circuit Synthesis," Tech. Rep. No. UCB/ERL M92/41, Dept of Electrical Engineering & Comp Science, University of California, Berkeley, 1992.