# Reduction of Direct Tunneling Power Dissipation during Behavioral Synthesis of Nanometer CMOS Circuits

Saraju P. Mohanty, R. Velagapudi, & V. Mukherjee
Department of Computer Science and Engineering
University of North Texas, Denton, TX 76203.
Email : {smohanty,rv0063,vm0058}@unt.edu

Hao Li
Department of Computer Science and Engineering
University of South Florida, Tampa, FL 33620.
Email : hli5@csee.usf.edu

***Abstract*— Direct tunneling current is the major component of static power dissipation of a CMOS circuit for technology below $65nm$, where the gate dielectric (SiO$_2$) is very low. We intuitively believe that multiple oxide thickness may be useful to reduce the direct tunneling current dissipation. Since no foundry design rules are available for design and layout using technology below $90nm$ we provide analytical models to calculate the tunneling current and the propagation delay of behavioral level components. We then characterize those components for $45nm$ technology and provide an algorithm for scheduling of datapath operations such that the overall tunneling power dissipation of the circuit is minimal. We have carried out extensive experiments for various behavioral level benchmarks under various constraints and observed significant reductions.**

## I. Introduction

As per ITRS, high performance CMOS circuits will require gate oxide thickness of $0.7nm - 1.2nm$ in near future. Such ultra-thin oxide devices will be more susceptible to new leakage mechanisms due to tunneling through gate oxide leading to gate oxide current [1]. The probability of electron tunneling is a strong function of the barrier height and the barrier thickness. The oxide tunneling current is strongly dependent on gate oxide thickness. Increase in the gate SiO$_2$ thickness leads to the increase in propagation delay. Thus, we conclude that use of multiple oxide thickness may be able to reduce the leakage current while maintaining the performance.

To our knowledge there are no behavioral synthesis methods available at present to reduce the tunneling current of a datapath circuit. However, there are few behavioral synthesis works considering the reduction of subthreshold current. In [2], authors have proposed algorithms for subthreshold leakage power analysis and reduction during behavioral synthesis. The authors in [3] also use MTCMOS approach for reduction of subthreshold current during high-level synthesis.

In this paper we develop models for direct tunneling current and propagation delay calculation of functional units. Moreover, we assume that functional units of different oxide thickness are available as standard cell and introduce an algorithm for scheduling of the datapath operations such that overall tunneling current is minimal. We assume that all transistors used in a functional unit have oxide of equal thickness, but the thickness of different functional units may differ.

## II. Analytical Models

We use a three level approach to form the models for direct tunneling current and propagation delay calculation of functional units (FUs). We use transistor level equations and derive models for logic level followed by behavioral level components. Let us assume that there are total $n_{total}$ NAND gates in the network of NAND gates constituting a $n$ bit functional unit out of which $n_{cp}$ number of NAND gates are in the critical path.

The direct tunneling current is expressed by Eqn. 1 [4], [5].

$$I_{DT} \quad = \quad \frac{WL\ q^3 V_{ox}^2}{16\pi^2 \hbar \phi_B T_{ox}^2}\ exp\left[-\frac{4\sqrt{2m_{eff}}\ {\phi_B}^{1.5} T_{ox}}{3\hbar q V_{ox}} \left\{1 - \left(1 - \frac{V_{ox}}{\phi_B}\right)^{1.5}\right\}\right] \qquad (1)$$

The voltage across the MOSFET gate oxide $V_{ox}$ is expressed as follows [6], [5] : $V_{ox} = V_{gs} - V_{fb} - \psi_S - V_{poly}$. We calculate the tunneling current of a $n$ bit functional unit using the above NAND gates as building blocks in the following manner.

$$I_{DT}\text{FU} = \sum_{j=1}^{n_{total}} Pr_j \sum_{\text{MOS}_i\ \in\ \text{NAND}_j} Pr_i\ I_{DT\,i} \qquad (2)$$

Here, $Pr_j$ is the probability that input of the NAND gate is at logic “0”, and $Pr_i$ is the probability that inputs of the parallel connected transistors are at logic “0”.

Based on the $\alpha$-power law and physical-$\alpha$-power model the propagation delay of a MOS $T_{pd}$ is computed [7], [8] :

$$T_{pd} \quad = \quad \frac{0.5 C_L V_{dd}}{I_{D\,Sat0}} + T_T \left\{\frac{0.5 - \left(\frac{V_{dd} - V_{Th}}{V_{dd}}\right)}{\alpha + 1}\right\} \qquad (3)$$

We calculate the critical path delay of a $n$ bit FU as :

$$T_{pd}\text{FU} = \sum_{i=1}^{n_{cp}} 0.5 \left(n_{fan-in} T_{pd}\text{NMOS} + T_{pd}\text{PMOS}\right) \qquad (4)$$

The $n_{fan-in}$ is the effective fan-in factor and is calculated for short channel devices with velocity saturation and strong inversion as shown below [9].

$$n_{fan-in} \quad = \quad 1 + \left\{\frac{\left(2-\sqrt{2}\right)(n_{series}-1) V_{ds\,Sat}}{V_{dd} + V_{Th} - 0.5 V_{ds\,Sat}}\right\} \left(1 + \frac{T_{ox}}{\epsilon_{ox}} \sqrt{\frac{q N_{channel} \epsilon_{Si}}{2\psi_S}}\right) \qquad (5)$$

Here, $n_{series}$ is the number of series connected MOSFET and fermi-level is assumed to be half of $\psi_S$ for strong inversion.

TABLE I

SELECTED RESOURCE CONSTRAINTS (RC) USED IN OUR EXPERIMENTS

| Number of FUs of Different Oxide Thickness $T_{ox}$ | | | | No. |
|---|---|---|---|---|
| Multiplier | | Adder-Subtractor | | |
| $1.7nm$ | $1.4nm$ | $1.7nm$ | $1.4nm$ | |
| 1 | 1 | 2 | 0 | 1 |
| 2 | 1 | 1 | 1 | 2 |
| 2 | 0 | 0 | 2 | 3 |
| 3 | 0 | 1 | 1 | 4 |

## III. BEHAVIORAL SCHEDULER

Let us assume that the datapath is specified as a sequencing data flow graph (DFG). The combined reduction of tunneling power dissipation and execution time translates to reduction of the current-delay-product (CDP). Thus, the objective of the algorithm is to minimize the CDP while assigning a schedule for the DFG. The inputs to the algorithm are an unscheduled DFG, and the resource constraints including FUs of different oxide thickness. The algorithm generates various outputs, such as scheduled DFG with appropriate functional unit assignment to a datapath operation, and estimates of current and delay. Fig. 1 presents the heuristic algorithm proposed to assign functional units of multiple oxide thickness.

```
Find total number of FUs of all available thickness.
Get resource constrained ASAP & ALAP schedules.
Find the vertices in critical and off-critical path.
Assume above ASAP schedule as current schedule.
For critical path vertices assign largest thickness
    multipliers and smallest thickness adder-subtractor.
for each off-critical vertex of the current schedule
    if vertex is a multiplication then assign the
            multiplier of largest available thickness.
    else assign the adder-subtractor of
            smallest available thickness.
    Calculate CDP of the current schedule.
    for each off-critical vertex
      for each allowable control steps
        Assign multipliers of next higher thickness
          or adder-subtractor of next lower thickness.
        Find CDP of the DFG at each case.
      end for
      Fix time stamp of the vertex with the FU
              assignment for which CDP is minimum.
    end for
end for
```

Fig. 1. Behavioral Scheduler Heuristic

## IV. EXPERIMENTAL RESULTS AND CONCLUSIONS

We characterized functional units of 16-bit size. While the adder-subtractor unit is a ripple carry unit, the multiplier is an array multiplier. It is assumed that the probability of logic "1" and logic "0" is same. While changing the oxide thickness the channel length of the transistor is changed proportionately [1]. The algorithm was implemented in C and tested with several behavioral level benchmark circuits for several constraints. A selected set of resource constraint is given in Table I.

TABLE II

EXPERIMENTAL RESULTS OF DIFFERENT BENCHMARK CIRCUITS

| Bench-marks | RCs | Tunneling Current in $\mu A$ | | | Critial Path Delay in $ns$ | | |
|---|---|---|---|---|---|---|---|
| | | $IDT_S$ | $IDT_M$ | $\%Red$ | $T_{pd_S}$ | $T_{pd_M}$ | $\%Penalty$ |
| ARF | 1 | 1360.53 | 647.79 | 52.38 | 34.86 | 49.72 | 42.62 |
| | 2 | 1360.53 | 409.23 | 69.92 | 34.92 | 43.80 | 33.04 |
| | 3 | 1360.53 | 218.58 | 83.93 | 34.86 | 45.74 | 31.21 |
| | 4 | 1360.53 | 195.12 | 85.65 | 32.93 | 43.80 | 33.00 |
| | Average Reduction | | | 72.97 | Average Penalty | | 34.96 |
| BPF | 1 | 1073.06 | 402.36 | 62.50 | 30.99 | 44.48 | 43.53 |
| | 2 | 1073.06 | 312.41 | 70.88 | 29.05 | 41.87 | 44.13 |
| | 3 | 1073.06 | 216.60 | 79.81 | 30.94 | 40.51 | 30.71 |
| | 4 | 1073.06 | 169.67 | 84.18 | 29.05 | 41.87 | 44.13 |
| | Average Reduction | | | 74.34 | Average Penalty | | 40.62 |
| DCT | 1 | 1232.15 | 205.58 | 83.31 | 52.29 | 70.65 | 35.11 |
| | 2 | 1232.15 | 222.19 | 81.96 | 52.29 | 69.97 | 33.81 |
| | 3 | 1232.15 | 304.32 | 75.30 | 52.29 | 68.61 | 31.21 |
| | 4 | 1232.15 | 222.19 | 81.96 | 52.29 | 69.97 | 33.81 |
| | Average Reduction | | | 80.63 | Average Penalty | | 33.48 |
| EWF | 1 | 811.92 | 88.43 | 89.10 | 44.55 | 62.80 | 40.96 |
| | 2 | 811.92 | 176.42 | 78.27 | 44.55 | 58.15 | 30.52 |
| | 3 | 811.92 | 240.95 | 70.32 | 44.55 | 54.07 | 21.36 |
| | 4 | 811.92 | 176.42 | 78.27 | 44.55 | 58.15 | 30.52 |
| | Average Reduction | | | 79.00 | Average Penalty | | 30.84 |
| FIR | 1 | 739.51 | 294.66 | 60.15 | 29.05 | 41.87 | 44.13 |
| | 2 | 739.51 | 145.07 | 80.38 | 29.05 | 35.17 | 21.06 |
| | 3 | 739.51 | 168.53 | 77.20 | 29.05 | 34.49 | 18.72 |
| | 4 | 739.51 | 145.07 | 80.38 | 29.05 | 35.17 | 21.06 |
| | Average Reduction | | | 74.52 | Average Penalty | | 22.24 |
| HAL | 1 | 513.49 | 198.67 | 61.30 | 13.55 | 20.93 | 54.46 |
| | 2 | 513.49 | 150.76 | 70.63 | 11.62 | 17.63 | 51.72 |
| | 3 | 513.49 | 85.26 | 83.39 | 13.55 | 17.63 | 30.11 |
| | 4 | 513.49 | 85.26 | 83.39 | 11.62 | 15.02 | 29.25 |
| | Average Reduction | | | 74.67 | Average Penalty | | 41.38 |

The experimental results are presented in Table II. The quantities with $S$ subscript represent the values for single oxide thickness and the multiple oxide thickness are shown with $M$ subscript. We assume the minimal oxide thickness case with $T_{ox}$ of $1.4nm$ as the base $S$ case. We anticipate that the critical path delay is going to increase due to the use multiple dielectric as delay increases with the increase in oxide thickness. We estimate the critical path delay of the circuit the sum of the delays of the vertices in the longest path of DFG.

In this paper we find that multiple oxide thickness is highly effective, however the use of multiple dielectrics alongwith multiple thickness will be explored in future. A heuristic based approach is presented here for functional unit assignment. We are anticipating that use of better optimization techniques may be further helpful. We also need to incorporate methods to accurately estimate the logic values for more accurate modeling of the tunneling current and propagation delay.

## REFERENCES

[1] 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.

[2] K. S. Khouri and N. K. Jha, "Leakage power analysis and reduction during behavioral synthesis," *IEEE Transactions on VLSI Systems*, vol. 10, no. 6, pp. 876–885, December 2002.

[3] C. Gopalakrishnan and S. Katkoori, "Knapbind: an area-efficient binding algorithm for low-leakage datapaths," in *Proceedings of 21st International Conference on Computer Design*, 2003, pp. 430–435.

[4] M. Depas, et. al., "Determination of Tunneling Parameters in Ultra-Thin Oxide Layer Poly-Si/SiO$_2$/Si Structures," *Elsevier Solid-State Electronics Journal*, vol. 38, no. 8, pp. 1465–1471, August 1995.

[5] 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.

[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] K. A. Bowman, L. Wang, X. Tang, and J. D. Meindl, "A Circuit-Level Perspective of the Optimum Gate Oxide Thickness," *IEEE Transactions on Electron Devices*, vol. 48, no. 8, pp. 1800–1810, August 2001.

[8] T. Sakurai and A. R. Newton, "Alpha-Power Law MOSFET Model and its Applications to CMOS Inverter Delay and Other Formulas," *IEEE Journal of Solid-State Circuits*, vol. 25, no. 2, pp. 584–594, April 1990.

[9] A. J. Bhavnagarwala, et. al., "A Minimum Total Power Methodology for Projecting Limits of CMOS GSI," *IEEE Transactions on VLSI Systems*, vol. 8, no. 3, pp. 235–251, June 2000.