

MODELING AND DESIGNING ENERGY-DELAY OPTIMIZED WIDE DOMINO CIRCUITS

Christine Kwong, Bhaskar Chatterjee, Manoj Sachdev

Department of Electrical and Computer Engineering
University of Waterloo, Waterloo, Ontario, Canada
ktckwong@engmail.uwaterloo.ca

ABSTRACT

In this paper, we present simple analytical models for energy (switching + short circuit) per transition and delay for wide-NOR domino logic gates. These gates are used to design register files (RFs) in high performance microprocessors and priority encoders for content addressable memory (CAMs). They contribute significantly to the overall switching energy and read delay, therefore require accurate modeling and optimized designing. Our results for a 130nm bulk CMOS technology indicate that the energy (delay) models track SPICE simulations to within 4% (7%) for a large range of load and delay conditions. The results show that, optimal energy-delay operation is a function of the number of pulldown paths. It is achieved when the wide-NOR gate equivalent fan-out is between 2.3-2.7.

1. INTRODUCTION

In the past 30 years there have been unprecedented improvements in digital IC performance, resulting primarily from aggressive CMOS technology scaling, innovative architectures, and improved circuit design techniques. One such circuit technique is the compound domino logic (CDL) that uses alternate stages of n-MOS domino and complementary static CMOS logic gates [1]. This method enhances circuit performance while maintaining acceptable robustness for deep submicron (DSM) technologies. Figure 1 shows a CDL based wide-NOR domino gate. Such gates are used as local and global bitlines (LBL/GBL) in high-end RFs [2] and priority encoders in CAMs [3]. These gates are characterized by high dynamic node capacitance and degraded evaluation times. In fact, the wide-NOR domino gates contribute significantly to the total read delay and switching energy of a functional unit block (FUB). Simulation results indicate that the fan-out achieving minimum energy-delay product (EDP) for conventional domino gates (without multiple parallel pulldowns) can be different from that of wide-NOR domino gates. In this paper we investigate the modeling and designing of energy-delay optimized wide-NOR gates for high performance circuits.

This paper is organized as follows: Section 2 presents the simulation results for different transistor parameters. In

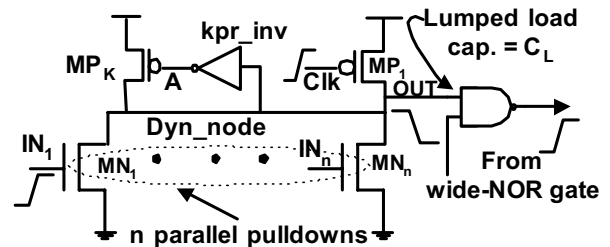


Figure 1: Typical wide-NOR domino gate

Section 3, the capacitance, delay and energy models are described while in Section 4 the simulation and model-predicted energy-delay results are presented. Section 5 offers some conclusions.

2. TRANSISTOR PARAMETER EXTRACTION

In this section, we present transistor level simulation results for the following transistor parameters: gate and diffusion capacitance, as well as saturation current under DC operating conditions. Figure 2 shows the normalized plots for gate-drain capacitance (C_{GD}), gate capacitance (C_G), and saturation current (I_{DSAT}) as functions of transistor width for a 130nm n-MOS transistor. As the transistor width increases, saturation current and capacitances undergo an approximately linear increase. This trend allows us to model the capacitances (C_{GD} , C_G) and saturation current as indicated in Eqs. (1(a-c)):

$$I_{DSAT} = K_I^N W_{MN} \quad (1a)$$

$$C_{GD} = K_{GD}^N W_{MN} \quad (1b)$$

$$C_G = K_G^N W_{MN} \quad (1c)$$

where K_I^N, K_{GD}^N, K_G^N are constants of proportionality and W_{MN} refers to the transistor width. The absolute values of these constants are technology dependent and calibration is required for each new generation. For the technology used, these constants are as follows: $K_I^N = 67\mu A/\mu m$, $K_{GD}^N = 0.84fF/\mu m$ and $K_G^N = 0.83fF/\mu m$. The above approach can be adopted to characterize p-MOS transistor parameters using different proportionality constants.

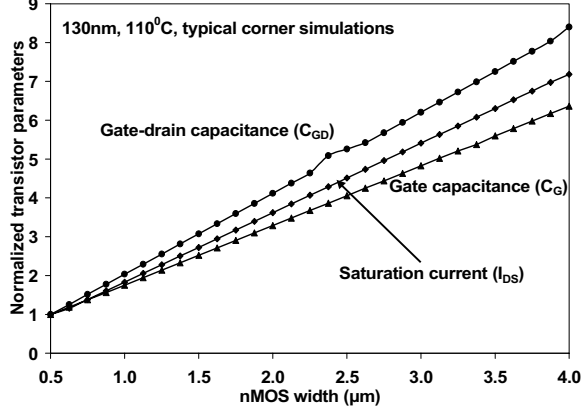


Figure 2: Normalized transistor current and capacitance

3. WIDE DOMINO GATE ANALYSIS AND MODELS

Wide-NOR domino gates may have several pulldown paths enabled at a given time. However, the analysis presented below follows the operation of high-performance MUX-es with only one pulldown path active at a given time, as it corresponds to the worst-case situation. The total dynamic node capacitance is dominated by the gate-drain capacitance when pulldown transistors are upsized beyond a certain width. In addition, increasing the width of the pulldown network only offers limited energy-delay improvements. Therefore, the overall capacitance increases more than proportionately than the transistor drive capability. This is unlike the case of conventional domino logic gates where the output load is typically dominated by the gate capacitance and significant energy-delay improvements can be achieved by increasing the pulldown n-MOS transistor sizes. Consequently, the fan-out which achieves minimum energy-delay product (EDP) for conventional domino designs can be different from that of wide-NOR gates. In this paper we develop our models in the context of an 8-wide 2-stack n-MOS pulldown domino gate (LBL circuit). This does not result in any loss of generality, the same methodology can be used for gates with different number of parallel paths and/or single transistor stacks as in GBL circuits.

3.1. Dynamic Node Capacitance Model

The dynamic gate energy and delay are directly related to the total capacitance associated with the domino node. For CDL, the wide-NOR domino gate output typically drives static logic gates (such as 2-input NAND or NOR), which make up the effective load capacitance (C_L). In addition, the parallel pulldown paths contribute to the total capacitance through the C_{GD} component and have a self-loading effect. The wide-NOR domino node (dyn_node) effective capacitance (C_{DYN}) can be approximately modeled as follows:

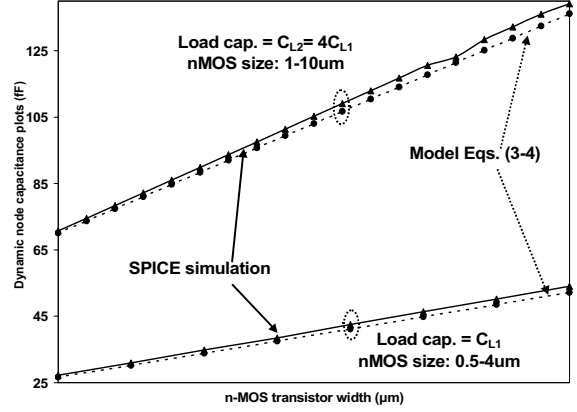


Figure 3: Wide-NOR dynamic node capacitance plots

$$C_{DYN} = C_{GD}^{TOT} + C_G^{TOT} \quad (2a)$$

$$C_{GD}^{TOT} = (C_{GD}^{MP1} + C_{GD}^{MPK}) + \sum_{i=1}^{i=n} C_{GD}^{MN_i} \quad (2b)$$

$$C_G^{TOT} = C_G^{kpr_inv} + C_L \quad (2c)$$

where the subscripts GD and G correspond to the gate-drain and gate capacitance components, “ n ” represents the total number of parallel paths, and the superscripts correspond to the individual transistors according to Figure 1. It is now possible to combine Eqs. (1-2) to obtain a simple expression for the total capacitance in terms of transistor widths as indicated in Eq. (3):

$$C_{DYN} = [K_{GD}^P (W_{MP1} + W_{MPK}) + nK_{GD}^N W_{MN}] + [K_G^P W_{MP}^{kpr_inv} + K_G^N W_{MN}^{kpr_inv}] + C_L \quad (3)$$

where the superscripts (N, P) refer to n- and p-MOS transistors respectively. In general, the precharge transistor (MP_i) and inverter (kpr_inv) are significantly smaller when compared to the n-MOS pulldown network and can be represented by a constant lumped capacitive load. Furthermore, in order to maintain iso-robustness, the keeper bears a fixed relation (3% in this design) with the effective pulldown transistor width. Therefore, we can simplify the above capacitance expression as:

$$C_{DYN} \approx [0.03K_{GD}^P + K_{GD}^N] n W_{MN} + C_L' \quad (4)$$

where C_L' is the lumped equivalent capacitance representing the load (C_L), precharge (MP_i) and inverter (kpr_inv) transistors. Figure 3 compares the dynamic node capacitance obtained from SPICE simulations with the model discussed above for two load conditions as a function of the pulldown transistor size. Results indicate that our model tracks the total capacitance accurately to within 3% of SPICE.

3. 2. Wide-NOR Domino Energy Model

In this section, we model the total energy/transition for the wide-NOR domino gate accounting for the switching, short circuit and leakage currents. The switching energy/transition, E_{SW} , has three components: 1) enabled pulldown n-MOS transistor, 2) domino node and 3) keeper node A (as shown in Figure 1), and can be expressed as:

$$E_{SW} = C_{TOT} V_{DD}^2 = [C_G^{MN1} + C_{DYN} + C_{GD}^{INV} + C_G^{MPK}] V_{DD}^2 \quad (5)$$

where C_{TOT} represents the total switched capacitance. Thus, using Eqs. 1(a-c) and Eq. (5), we can now represent the switching energy as a function of transistor width.

For wide-NOR domino gates, the degraded fall-time causes a short circuit current to flow through the p-MOS keeper and the n-MOS pulldown transistors. Our results indicate that the short circuit energy, can make up $\sim 15\%$ of the total energy/transition. In this paper we model the short circuit energy, E_{SC} , according to [4, 5] as:

$$E_{SC} = \frac{I_{SC} t_{SC}}{2} V_{DD} \quad (6)$$

where I_{SC} is the peak short circuit current, and t_{SC} is the short circuit interval. Figure 4 shows the voltage and current waveforms for a representative data point during the switching transient. We approximate I_{SC} by the p-MOS keeper saturation current and express it as $K_I^p W_{MPK} = 0.03nK_I^p W_{MN}$. Our simulations indicate that t_{SC} interval starts when the input signal (IN_1) reaches V_{THN} and ends when node A reaches $V_{DD} - |V_{THP}|$. This allows us to model the short-circuit duration as follows:

$$t_{SC} = t_{50\%}^{kpr_inv} - t_{50\%}^{IN} + \frac{1.25}{V_{DD}} [t_r^{kpr_inv} V_{THN} - t_r^{IN} (V_{DD} - |V_{THP}|)] \quad (7)$$

where $t_{50\%}^{kpr_inv}$, $t_{50\%}^{IN}$ are the 50% V_{DD} crossover points of the inverter and input signals respectively, $t_r^{kpr_inv}$, t_r^{IN} are the inverter and input signal rise times. It should be noted that, for a given design, when the pulldown transistors and p-MOS keepers are upsized to maintain iso-robustness, the inverter (kpr_inv) is scaled up as well. This ensures the inverter delay remains approximately constant, and results in a fixed t_{SC} even for a wide range of load conditions and transistor sizes.

We now consider the impact of leakage energy associated with wide-NOR domino logic gates. At the start of the evaluation phase, the total leakage current is determined by the n-MOS transistor width and $I_{OFF}/\mu m$ of the ($n-1$) deselected pulldown paths and is expressed as:

$$I_{Leakage}^{TOT} = (n-1) K_I^N W_{MN} \left(\frac{I_{off}}{I_{on}} \right)^{110^\circ C} \quad (8)$$

where the worst case leakage at $110^\circ C$ is considered and the kpr_inv leakage currents are neglected. Our transistor

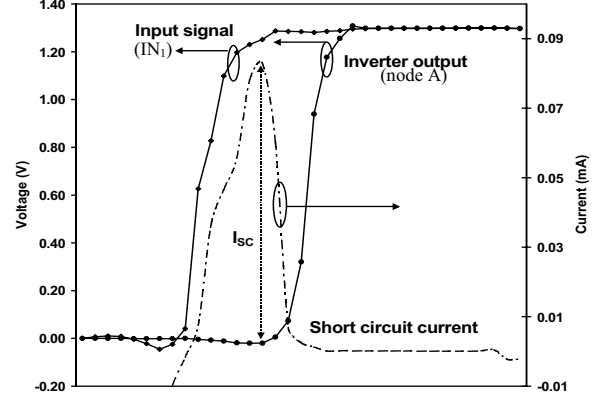


Figure 4: Wide-NOR domino voltage and current transients

level simulations for the 130nm-70nm technologies based on Berkeley Predictive Technology models [6] indicate that the worst case I_{ON}/I_{OFF} ratio is in the range of $\sim 10^4 - 10^3$. Hence, for our example of 8-wide domino gate, the worst-case combined leakage of the deselected 7 parallel paths is limited to within 0.07-0.7% of the ON-state current. However, in the case of wider domino logic gates, the leakage current contribution will have to be considered. Combining the above equations we can model the total energy/transition as:

$$E_{TOT} = C_{TOT} V_{DD}^2 + \frac{I_{SC} t_{SC}}{2} V_{DD} + I_{Leakage}^{TOT} V_{DD} t_{int} \quad (9)$$

where t_{int} is the integration interval and corresponds to the switching interval, and all other symbols are as previously defined in Eqs. (5-8).

3.3. Wide-NOR Domino Delay Model

In this section, we model the wide-NOR gate delay during evaluation using the alpha-power MOSFET model. For most practical designs, the input signals to the n-MOS pulldown network is timing critical and is driven by upsized drivers. This results in sharp $0 \rightarrow V_{DD}$ transition for IN_1 , while the domino node itself shows slow $V_{DD} \rightarrow 0$ transition. Thus, the delay can be modeled using equations pertaining to the “fast” input case discussed in [4] and can be approximated as:

$$t_{pHL} = \left(\frac{1}{2} - \frac{1 - V_{DD}/V_{THN}}{1 + \alpha} \right) t_r^{IN} + \frac{C_{DYN} V_{DD}}{2 I_{DSAT}} \quad (10)$$

where α is the velocity saturation coefficient, C_{DYN} is the dynamic node capacitance derived in Eq. (4-5), I_{DSAT} ($= K_I^N W_{MN}$) is the pulldown transistor saturation current, and t_r^{IN} is the input signal rise time. For our 130nm technology, V_{DD}/V_{THN} is 4 while the value of α equals 1.4. The procedure for parameter extraction is detailed in [4].

4. ENERGY-DELAY COMPARISONS

We now present the energy-delay results and compare the HSPICE simulations with model predicted data for different pulldowns sizes. Figure 5 shows the energy vs. delay plots of an 8-wide domino NOR gate for the 130nm generation at two loading conditions. The p-MOS keeper size is kept at 3% of the effective pulldown strength at all data points to maintain constant DC robustness. For a given design, the minimum energy-delay product is achieved under the following condition:

$$\frac{\partial E_{TOT}}{\partial t_{pHL}} = \left(\frac{\partial E_{TOT}}{\partial W_{MN}} \right) / \left(\frac{\partial t_{pHL}}{\partial W_{MN}} \right) = -E_{TOT} / t_{pHL} \quad (11)$$

By substituting the expressions for energy and delay and taking partial derivatives with respect to the pulldown transistor size, we obtain the following expressions:

$$\frac{\partial E_{TOT}}{\partial W_{MN}} = (nK_1 + K_2 + K_3)V_{DD}^2 + K_4 t_{SC} V_{DD} \quad (12a)$$

$$\frac{\partial t_{pHL}}{\partial W_{MN}} = \frac{-C_L}{K_5 W_{MN}^2} \quad (12b)$$

where K_1 - K_5 relates to the current and capacitance constants of proportionality. The pulldown transistor size for optimal EDP operation, W_N^{opt} , can be obtained using Eqs. (11-12). For Eq. (13), K_1 - K_5 are replaced by the corresponding proportionality constants, and the assumption that $C_{GD}^{TOT} + C_G^{kpr-inv} > C_L$ is made for simplification:

$$W_N^{opt} = \frac{C_L}{\left[\frac{2K_1^N}{V_{DD}} \left(\frac{1}{2} - \frac{1-V_{DD}/V_{THN}}{1+\alpha} \right) K_r^{IN} \right] + K_G^N \left(1 + \frac{\mu_p}{\mu_n} \right) + nK_{GD}^N} \quad (13)$$

where μ_p/μ_n is the p-MOS to n-MOS mobility ratio. Figure 5 indicates that the energy-delay models track the SPICE simulations over a wide range of energy (5x), delay (3x) and load (4x) conditions. The error related to energy has a range of 6% (RMS value of 4%) while the corresponding number for the delay model is 5% (7%). The points P and Q indicate the optimal EDP operation region for the wide-NOR domino gates and relate to fan-out of 2.3-2.7. Our simulations indicate that wide domino gates are in suboptimal operation if they are designed with the optimal EDP fan-out of $n=1$ domino gate. This can be explained with the help of Eq. 13. It suggests that the optimal pulldown transistor size is inversely proportional to "n". Using the above methodology, it is possible to design an energy-delay optimized wide-NOR domino gate with 10-20% reduction in EDP.

It should be noted that the models discussed in this paper for capacitance, energy and delay while being simple are approximate in nature. They do not incorporate the bias dependence of switching capacitances, short circuit

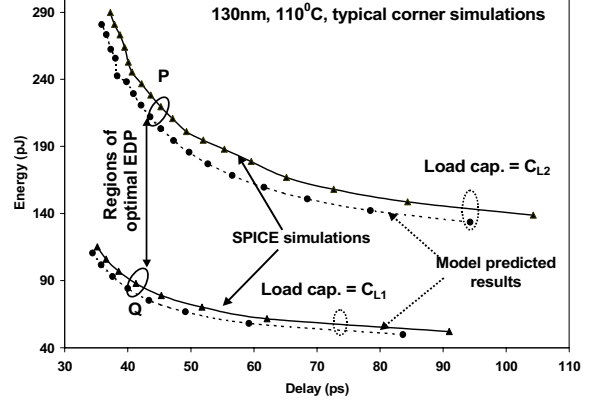


Figure 5: Energy-delay plots for wide-NOR domino gate

current though kpr_inv during evaluation, and stack effect when series transistor is present in the pulldown paths. In addition, the input and dynamic node signals are being approximated as linear ramps. These issues can be addressed in order to achieve better accuracy while making the overall models more complicated and remains the topic of future research.

5. CONCLUSION

In this paper, we discussed simple closed form analytical models for propagation delay and energy for wide-NOR domino gates. The models track simulation results to within 7% for different loading and sizing conditions. In addition, the points of optimal energy-delay operation for wide-NOR domino gates were identified and shown to be different from that of conventional domino gates. This is expected to improve energy-delay optimization for designing DSM datapath circuits.

ACKNOWLEDGEMENT

The authors would like to thank O. Semenov and S. Ardanian of University of Waterloo for discussions and their comments.

REFERENCES

- [1] K. Bernstein, K. Carrig, C. Durham, P. Hansen, D. Hogenmiller, E. Nowak, and N. Rohrer. *High Speed CMOS Design Styles*. Boston, MA: Kluwer Academic Publishers, 1999.
- [2] S. Hsu, B. Chatterjee, M. Sachdev, A. Alvandpour, R. Krishnamurthy, S. Borkar, "A 90nm 6.5GHz 256x64b Dual Supply Register File with Split Decoder Scheme," in *Proc. Symp. VLSI Circuits*, June 2003, pp. 237-238.
- [3] J. S. Wang, and C. H. Huang, "High Speed and Low Power CMOS Priority Encoders," *IEEE JSSC*, vol. 35, no. 10, pp. 1151-1154, Oct. 2000.
- [4] T. Sakurai, A.R. Newton, "Alpha-Power Law MOSFET Model and its Applications to CMOS Inverter Delay and Other Formulas," *IEEE JSSC*, vol.25, no. 2, pp. 584-594, Apr. 1990.
- [5] A. Bellaouar, M.I. Elmasry, *Low-Power Digital VLSI Design Circuits and System*. Boston: Kluwer Academic Publishers, 1995.
- [6] Berkeley PT M: <http://www-device.eecs.berkeley.edu>