Threshold Voltage and Power-Supply Tolerance of CMOS Logic Design Families

Madhuban Kishor
Dept. of Electrical Engineering
Indian Institute of Technology
New Delhi, India

José Pineda de Gyvez
Philips Research
Prof. Holstlaan 4 – WAY 4.71
5656 AA Eindhoven, The Netherlands
Email: jose.pineda.de.gyvez@philips.com

Abstract
The advent of deep submicron technologies brings new challenges to digital circuit design. A reduced threshold voltage ($V_T$) and power supply ($V_{dd}$) in addition to process variabilities have a direct impact on circuit design. In a semiconductor environment it is conventionally thought that parametric yield is high and stable and that the main yield losses are functional. Although functional yield remains the main focus of attention, modern and future circuits may not have the presumed high parametric yield. We present a study that compares the tolerance to process variability of various design families for metrics including timing and power consumption under $V_T$-$V_{dd}$ scalability using a NAND gate as a test vehicle. Basically, the fundamental limitations to the scaling of the supply voltage due to the statistical variation of MOS $V_T$ are investigated and defined. The four logic families under study are: static CMOS, Differential Complementary Voltage Swing Logic (DCVSL), Domino and Pass Logic.

1. Introduction
The problem of threshold voltage variability is considered to be one of the most serious concerns in future Gigabit scale VLSI and ULSI designs[1-8]. However, until now there have been limited studies of performance fluctuations of digital circuits in terms of the statistical fluctuation of the MOSFET threshold voltage ($V_T$)[9]. With the non-scalability of $V_T$, due to increase in leakage power dissipation, circuit design in the deep sub-micron regime becomes a major challenge.

In a semiconductor environment it is conventionally thought that parametric yield is high and stable and that the main yield detractors are spot defects. Although functional yield remains the main focus of attention, modern and future circuits may not have this high parametric yield[10]. In fact, due to the use of submicron transistor sizes, modern circuits become quite sensitive to intra-die process variations. It may be possible that design strategies using a worst-case approach, which accounts primarily for inter-die variations, do not suffice anymore. Worst-case design styles assume that all transistors use the same worst-case $V_T$ whose average and standard deviation come from inter-die statistical variations. However, intra-die differences, such as random local $V_T$ variations of very short channel transistors are not considered and may pose a serious problem for designs based on low-voltage low-power premises, e.g. clock skews, excessive leakage current, out of spec critical-path delays, etc. Noise is also another important parameter in the proper design of mixed-mode analog-digital low-voltage and low-power circuits[11].

There is a natural tendency to shrink device dimensions to attain very large scale integration. Unfortunately $V_T$ does not scale linearly and as such the power supply needs to be adjusted for constant electrical field scaling. With this present trend in the scaling scenario there is an obvious trade-off between power supply and threshold voltage for
low-voltage and low-power applications[12-16]. Worth observing is that under these conditions a circuit begins to deviate from its normal behavior and shows significant variations on its performance parameters. Burnett[12] showed that for technologies of 0.5μm minimum dimensions and 3.3V nominal power supply, the 3σVT specification is about 10% or less of the nominal VT. Moreover, the impact of fundamental VT variability and transistor mismatch has significant implications for the performance of SRAMs and logic circuits [9] as well.

Based on these premises we have chosen to study the effect of process variability on digital circuit design using a NAND gate as a test vehicle. The gate is implemented in four different design styles representing in a broad sense today's logic design families. These styles are: 1) CMOS static logic, 2) Differential Complementary Voltage Swing Logic (DCVSL), 3) Pass transistor logic, and 4) Domino logic. The objective of this study is: i) to predict the performance of the four generic logic families in the voltage scaling scenario, under nominal conditions and under the condition of statistical VT variation. ii) to compare the logic families to bring out their important features and to determine which is the best under process variability conditions.

2. Threshold Voltage Variability

Two classes of VT variability can be identified: 1) local variability due to the randomness in the number of dopants in the depletion region of the MOSFET, and 2) global variability due to manufacturing fluctuations in the gate length, gate oxide thickness, implant impurity, etc. The statistical variation of VT affects identical transistors within the same chip, while the die-to-die and wafer-to-wafer variations in VT are mainly due to manufacturing fluctuations. The effects of fluctuations in dopant distribution on the MOSFET threshold voltage have also been investigated using device-level simulations[7]. The simulation results indicate that the microscopic fluctuation in dopant distribution not only induces threshold voltage deviations, but also lowers the mean threshold voltage value. For a MOSFET with 0.1μm effective channel length, σVT was found to be between 25-30mV[7]. Furthermore, three-dimensional simulations of 0.1μm MOSFETs[8] show that the correlation between VT variation and the deviation of the number of dopants in the depletion region from the mean value is between 0.6 and 0.77. Consequently it is predicted[8] that as MOSFET scaling approaches the 0.1μm regime, the number of dopants in the depletion region will be in the order of 10^2, and below 10^2 in the inversion layer, for minimum geometry devices. Thus, it is imperative to study in detail the consequences of having only a small number of dopants in the channel. From a circuit design point of view it is impractical to use device simulators and rather SPICE-like simulators are used. This implies that to capture the essence of dopant variability the threshold voltage is varied as a parameter.

Modern deep-submicron circuits are more prone to fail due to parameter mismatches. An extensive work on transistor matching has been done by Pelgrom et al.[5]. They discussed matching in detail and did experiments for the analysis and measurement of the mismatches in threshold voltage, current and substrate factors of the MOS transistor as a function of area, distance and orientation. Their experimental results led to Pelgrom's general parameter-variance model given in equation (1).
where $P$ is the matching parameter of interest, $A_p$ and $S_p$ are process-related fitting constants relating the parameter variance to the device area $WL$, and $D$ is the separation distance between devices. Their analysis and measurements led to the following conclusions: The variance of $V_T$, $\beta$ and the substrate factor are inversely proportional to the transistor area. The mismatch in $V_T$ dominates the transistor performance for normal gate-source potentials. Thinner gate-oxide decreases the $V_T$ and $\beta$ mismatch, while the relative current factor mismatch remains almost constant.

3. NAND Gate as Test Vehicle

Logic styles can be classified as clocked or non-clocked. Domino logic is the most popular and widely used in the clocked style. Static CMOS, DCVSL and Pass Logic are more common among the non-clocked styles.

The test-NAND gate is designed to operate at an input rate of 100 MHz for $V_{DD}$ of 1.8V and $V_{TO}$ of 0.4V. The frequency of one of the inputs is kept at 50 MHz, since both the low and the high states will act as separate inputs. The frequency of the other input is 25 MHz. The time taken by the input to swing from its high state to the low state and vice versa is 10% of the time period of the 50MHz signal. The output rise and fall times of the gate are designed to be almost equal to that of the input. The load capacitance is taken to be 100fF to ground, to represent the fanout capacitance of the next stages as well as the interconnect capacitance. The channel length of the transistors is taken to be 0.18μm for high speed.

Fig. 1 shows the NAND gate designed in the various logic style families. The static CMOS gate is designed to have identical rise and fall times. The pull-up PMOS transistors in the DCVSL gate shown in Fig. 1b are designed to be weak so that they do not interfere with the pull-down path. The corresponding NMOS transistors are designed to meet equal fall and rise times. The pass transistor logic is shown in Fig. 1c in which the inverter is used to achieve full logic swing.

A Domino logic gate has output states which are developed in only one direction, e.g. in a cascaded set of logic blocks each state evaluates and causes the next stage to evaluate. Thus, a single ‘device polarity’ (usually NMOS) is used
predominantly in the evaluate path. The feedback PMOS transistor of the domino logic (Fig. 1d) is designed to be approximately five times weaker than the NMOSs in the evaluate path. This is required so that the fall time at node N1 is small and the performance of the gate does not deteriorate. Since the gate is optimized for pull-up transitions, the output capacitance is dramatically reduced. Also, the drain current of the evaluation logic device is usually devoted 100% to switching logic states, rather than to sinking static short-circuit currents. Hence, Domino logic is the fastest of all the four logic families. There are many variations of Pass gate logic, among them complementary pass logic (CPL) which is very fast and is the one that has been used in the present work. This logic makes use of NMOS logic trees which couple logical inputs to inverting buffers driving outputs. However, the performance of the gate goes down at lower supply voltages. In Static CMOS gates, the logic is built redundantly. The PMOS transistors are of comparatively larger sizes and they contribute significantly to the output capacitance such that the speed of the gate is reduced. Nevertheless it compares well with the Pass gate logic at lower supply voltages and in fact, its performance is better than the Pass gate logic. DCVSL is the slowest of all the four logic families, because the pull-up PMOS transistors work against the pull-down path.

The Static CMOS logic has the lowest power dissipation of all the four styles because it works in a complementary fashion. It has a low short circuit power dissipation, which can be almost eliminated if the supply voltage is equal to the sum of the magnitude of the threshold voltages of the NMOS and PMOS. Domino and Pass gate logics also have a low power dissipation but they have a quite higher leakage power dissipation. A significant portion of the dissipated power by the Domino logic is associated to the clock. The Pass gate logic suffers from short circuit power dissipation because the logic high at the intermediate node is $V_{DD} - V_T$, and thus the PMOS of the output inverter is not fully cut-off. The DCVSL logic has the highest power dissipation because the current flows from the supply to ground till the two output nodes reach their final values. Also, since both nodes switch states, current is drawn from the power supply for every change in output.

4. Performance Analysis

Variability of threshold voltage induces greater variation on the circuit’s performance. An approximate relation for the delay is given as[15]

$$ T_{pd} \approx (V_{DD} - V_T)^{1.5} \tag{2} $$

Differentiating this equation with respect to the MOSFET $V_T$, a relationship between the circuit delay variation $\sigma_{Tpd}$ and the MOSFET $V_T$ variation can be established as

$$ \sigma_{Tpd} \approx \frac{\sigma_{V_T}}{(V_{DD} - V_T)^{2.5}} \tag{3} $$

Thus, for $V_{DD} = 1.8V$ and $V_T = 0.3V$, $\sigma_{Tpd} \approx 0.444 \times \sigma_{V_T}$, while for $V_{DD} = 1.6V$, $\sigma_{Tpd} \approx 0.5917 \times \sigma_{V_T}$. This means that there is an approximate 25% increase in $\sigma_{Tpd}$ as $V_{DD}$ is reduced from 1.8V to 1.6V. An increase in delay is also observed if $V_T$ increases to $V_T + \sigma_{V_T}$. Because the delay variability strongly depends on $V_T$ and $V_{DD}$ as
can be seen from $\sigma_{TPD}/\tau_{PD} \propto \sigma_{VT}/(V_{DD} - V_{T})$, we have that the control over $V_T$ variability has to significantly be improved to maintain the same $\sigma_{TPD}/\tau_{PD}$ ratio at lower $V_{DD}$. This obviously poses a challenge for the processing capabilities of deep submicron technologies. The fact is that $\sigma_{VT}$ is increasing with each future generation. This requires the designers at circuit level to make the $\sigma_{TPD}/\tau_{PD}$ ratio as small as possible. Moreover, $V_T$ cannot be reduced without considering the effects of subthreshold leakage on the operation of the circuit. As will be seen in this Section, the Domino logic fails when $V_T$ is reduced. Also, circuits suffering from body-effects will have a larger variation in their delay parameters. This will be seen for the case of the Pass gate which suffers from body-effects while it is passing a logic '1'.

Noise margin is a measure of noise tolerance for digital circuits. Two noise margins are defined: noise margin for low signal levels (NML) and noise margin for high signal levels (NMH) where $NML = V_{IL} - V_{OL}$ and $NMH = V_{OH} - V_{IH}$. The $V_{OL}$ and $V_{OH}$ voltages of CMOS circuits are 0V and $V_{DD}$, respectively. $V_{IL}$ is the maximum input voltage that can be accepted as logic low, and it corresponds to the unity gain point on the voltage-transfer characteristics, i.e. $dV_{out}/dV_{in} = -1$. $V_{IH}$ is the minimum input voltage that can be accepted as logic high, and corresponds to the second unity gain point on the voltage-transfer characteristics. $V_{IL}$ is called the lower critical voltage and $V_{IH}$ is called the higher critical voltage.

Fig. 2 presents noise margin metrics for the four families under a voltage scaling scenario. The noise margin of a Static CMOS gate depends on the relative strengths of the pull-up and the pull-down paths. If the paths are equally strong, then its NML is close to $V_{DD}/2$. The NML of a Domino gate is close to $V_{ta}$ as seen in Fig. 2, but it can be increased by making the feedback PMOS stronger. However, this increases the delay of the gate. The noise margin of a Pass gate logic depends on the ratio of the strength of the pull-up path and the pull-down path of the output inverter. Since the pull-down path of the output inverter is made stronger than the pull-up path to make the rise and fall time equal, the NML is small and close to $V_{ta}$ as can be seen in Fig. 2, but NMH is large. The DCVSL gate has good NML while its higher noise margin is poor as the $\beta_{ps}/\beta_{nk}$ value is kept low, otherwise the PMOS will act against the pull-down path. The Domino's logic NMH is good as can be seen from Fig. 2b.

![Figure 2. Noise Margin. (a)NML, (b)NMH](image-url)
Fig. 3 shows the functional output of the NAND gate subjected to statistical $V_T$ variations. One can see that variations in $V_T$ lead to timing problems in the static gate (Fig. 3a) but it essentially remains functional. Fig. 3b shows the results of the DCVS logic. As in the case of the static gate of Fig. 3a, functionality remains unaltered but in this case the timing problems are more accentuated. Fig. 3c shows the output of the pass transistor logic. We can see that the fall time is severely affected by $V_T$ variability and also that for many of the trials the gate cannot be considered functional. Essentially, the transistor that passes the logic '1' suffers from body-effects in addition to its $V_T$ dependency and this is reflected at the output of the inverter. The output of the domino-logic NAND gate is illustrated in Fig. 3d. This gate also fails to produce a correct logic output for some of the simulation trials. We have essentially that the relative strength of the PMOS transistor in the feedback path and the corresponding evaluate NMOS transistors need proper matching. What happens is that when the feedback PMOS drives more current, node $N_1$ in Fig. 1c cannot be discharged and so we see that the output fails to rise to the logic high level of $V_{DD}$. But when the evaluate NMOS's become stronger, or $V_T$ becomes smaller, then node $N_1$ discharges and the output goes to a logic high level. One can also observe glitches that in the normal operation do not appear.

![Figure 3. Functional behavior due to statistical $V_T$ variations. (a) Static logic, (b) DCVS logic, (c) Pass-transistor logic (d) Domino logic](image)

Fig. 4 shows plots of rise, fall and delay time as well as of power dissipation performance based on a sigma-to-mean ratio as a function of the supply voltage. The simulations are done for a supply voltage of 1.8V, 1.5V, 1.2V and 1.0V. Fig. 4a shows results for the rise time. The variation in rise-time is the lowest for Domino logic and does not degrade much with supply voltage, because the PMOS in the output inverter is designed to optimize the rise-time. The variation in DCVS logic also does not degrade much with supply voltage, because it is independent of the pull-down path. The effect on the Static CMOS and Pass gate logics is almost the same as shown in Fig. 4a. The fall-time for all the logic styles is almost the same, see Fig. 4b, but for the Pass gate logic it degrades severely at lower supply voltages due to body-effect and $V_T$ loss. Since the Static CMOS gate consists of only one stage its delay is the lowest; this is illustrated in Fig. 4c. For the Pass gate logic, it degrades severely at lower supply voltages due to reasons already discussed. The sigma/mean value of the power dissipation is comparatively smaller than other performance parameters, because the main component of power dissipation is dynamic power dissipation and it is relatively independent of variation of process parameters.

Observe from equation (1) that the effect of $V_T$ variability is greater for devices with smaller dimensions. Consequently, if we increase the dimensions of the
transistor, the impact of process variations reduces. But increasing the sizes also increases the capacitance at the internal nodes. To overcome this problem the width of the transistor is increased to improve the current capability of the device. Hence, there is also an improvement in performance and it is also possible to operate the gates at higher input rates.

Let us investigate the effect of device resizing. The relative size of the transistors in the gate is kept the same, so that the functionality of the circuit is not affected. The nominal performance parameter and their one-sigma variation are measured and plotted as a function of channel area. The channel area is calculated as the sum of each transistor's area. The simulations are carried out at a supply voltage of 1.0 V so that the performance of the circuit at scaled voltages can be captured. The plots of Fig. 5 show the various parameters as a function of channel area. The vertical bars denote the one-sigma variation above and below the mean value.

It can be seen that the rise-time decreases for all logic styles as the device dimensions are increased (Fig. 5a). Also, there is a decrease in the one-sigma value of the rise-time. This is consistent with (1) and earlier works on matching which showed that the $V_T$ variation decreases with increase in device dimension. It is observed that the Static CMOS and Domino logic gate have the best improvement in one-sigma value with channel area, see Fig. 5a. One-sigma of fall-time also decreases with increase of device dimensions. From Fig 5b it can be seen that the improvement in one-sigma with increase of device dimensions is best for the Static
CMOS gate. The Pass gate logic suffers from body-effects when the NMOS pass gate is passing a logic '1'. This causes its one-sigma value to be quite high even at increased device dimensions. There is also a slight improvement in one-sigma value for the DCVSL gate but at the expense of a large increase in device dimensions. Fig. 5c shows that the Static CMOS gate has the smallest delay for a given channel area because it uses just two NMOSs and two PMOSs, while the Domino logic uses four NMOSs and three PMOSs. Also, the variation is much smaller for the Static CMOS gate. Power dissipation increases with increase in device dimensions for all the logic styles, see Fig. 5d. This is because of increase in short circuit power dissipation due to decrease of rise-time and fall-time at the output as compared to the input. It can be seen that the power dissipation as well as its one-sigma value is least for the Static CMOS gate.

![Graphs showing performance variability](image)

**Figure 5. Performance Variability as a function of the cell area. (a) rise time, (b) fall time, (c) delay, (d) power consumption**

5. Conclusions

Table 1 lists various parameters that are considered while designing a circuit, especially for future technologies. As a guideline for circuit design, the logic styles have an entry 1, 2, 3 and N.D. for each parameter. Entries 1, 2 and 3 are in decreasing order of preference, while N.D. means that the particular logic style is not desirable if the corresponding parameter is most important for circuit design.

From Table 1 we can conclude that the Static CMOS logic has characteristics which are suited in most applications, although it is not the best choice for high-
speed designs, but with mismatch in transistors and $V_T$ variation becoming important, it performs the best. Domino logic is a very good option for high-speed design, but if there is a surrounding noisy circuitry and $V_{DD}$ and $V_T$ need to be scaled down, then one must be extra careful when using this logic style. For implementing functions like multiplexers and full-adders, the Pass gate logic is a good choice as long as $V_{DD}$ is in between 3 and 5 times that of $V_T$ because as a rule of thumb this relationship of $V_{DD}$ and $V_T$ gives the optimum energy-delay product. DCVSL gives a good design when both true and complement functions are required and noise-immunity is a priority.

Table 1. Summary of performance characteristics

<table>
<thead>
<tr>
<th>Logic</th>
<th>Speed</th>
<th>Power</th>
<th>Area</th>
<th>$V_{DD}$ Scalability</th>
<th>$V_T$ Scalability</th>
<th>Noise Margin</th>
<th>$V_T$ variations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Static CMOS</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>DCVSL</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>N.D.</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Domino</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>N.D.</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Pass Gate</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>N.D.</td>
<td>2</td>
<td>3</td>
<td>3</td>
</tr>
</tbody>
</table>

References