Analytical Delay and Variations Modeling in the Subthreshold Region

Sungil Kim
Electrical and Computer Engineering
Auburn University
341 War Eagle Way
Auburn, Alabama 36849 USA

Faculty Advisor: Prof. Vishwani Agrawal

Abstract

Power constraint on modern electronics has been an increasingly significant part of VLSI design. Thus, subthreshold circuit design where the supply voltage is less than the device threshold voltage gained renewed attention to reduce the energy. However, such reduction comes with process, voltage, and temperature (PVT) variations. Furthermore, as previous analytical delay models (i.e. Alpha-power law MOSFET model) do not take account of these variations, their errors in subthreshold region are large. Therefore, this paper presents analytical delay model that considers PVT variations and models variations of analytical delay in the subthreshold region. The proposed model is verified through the simulation on PTM low power 16-nm technology (threshold voltage = 0.68V). Result shows that the proposed analytical delay model has a maximum error of 33.4% while Alpha-power law MOSFET model has a maximum error of 71.1% for 0 to 1 input transition over the range of 0.2V to 0.4V of supply voltage. In addition, three analytical delay variations are modeled and verified. In contrast to previous research, the proposed model considers the drain-induced barrier lowering (DIBL) effect. The accuracy is improved with a compensation factor and verified for the range of 0.15V to 0.4V of supply voltage along with threshold voltage variation up to 30mV, supply voltage variation up to 50mV, and temperature variation up to 20°C. By considering process variations, supply voltage, and temperature variations, we demonstrate the delay variations up to 4.31×. Result shows an average error of 1%, 14.6%, and 6.8% for PVT variations respectively. These results can further the applicability of subthreshold circuit for ultra-low-power electronics.

Keywords: Analytical Delay Model, Subthreshold Circuits, PVT Variations

1. Introduction

Concerns about power consumption and its management are important, particularly in modern ultra-low power electronics. In 1971, Meindl and Swanson [1] introduced subthreshold circuits—where the supply voltage is less than the device threshold—and provided the insight to the potential 10 to 1000 times improvement in power-speed product by using lower supply voltage. In the 1900s and 2000s, researchers proved the workability of subthreshold circuits. For instance, the feasibility of implementation of subthreshold CMOS circuits for low performance application has been proven, and the analysis of threshold voltage and temperature variations are carried [2]. The results from [3, 9] showed that the minimum energy operation occurs at the subthreshold region.

From the results of feasibility of subthreshold circuits, in the last decade, the ultra-low power systems attracted particular attention, such as circuits for biomedical sensing and monitoring [4]-[5], ultra-low power processors and memories [6]-[7], and energy harvesting and power management units [8]. For low- to moderate-speed systems whose primary interest is to reduce power consumption or that have a significant constraint on power consumption, operating at the subthreshold region can achieve the minimum energy operation [9]. While operating at subthreshold region significantly reduces the power consumption and can operate at the minimum energy, it introduces process, voltage, and temperature (PVT) variations [7] that directly affect circuit delay and performance. Therefore, analytical delay models that do not carefully characterize these variations or were previously applicable to superthreshold
circuits can no longer accurately model delay in the subthreshold region, and more accurate analytical models to characterize variations in the subthreshold region are needed.

For instance, the Alpha-power law MOSFET model [10] introduced a simple but practical MOSFET model and derived analytical expressions for the drain current, short-circuit power, logic threshold voltage, and propagation delay. In this paper, we focus on the analytical model for the propagation delay of CMOS inverter. While this model shows an average error of 0.55% from 0 to 1 input transition at \( V_{DD} = 1V \) on the predictive technology model (PTM) 45-nm bulk CMOS technology [11], the maximum error in the subthreshold region is 46.12%. This error prompted some of the previous efforts to better model analytical delay in the subthreshold region.

Several analytical delay models were proposed for the subthreshold region to characterize the increased variations and degraded device characteristics due to the reduced \( \frac{I_{on}}{I_{off}} \) [7] and the saturation current that exponentially depends on the gate and threshold voltages [12]. In particular, the model [13] present the analytical delay model for CMOS inverter, considering process variability and on-current transient variations, and simulate on the STM 45-nm technology. The model [14] develops PVT variations modeling and verify against CMOS 130-nm technology.

This paper derives three analytical delay variations modeling for the subthreshold region. In contrast to [14], the proposed model considers the drain-induced barrier lowering (DIBL) effect. While authors from [13] do not derive separate analytical variations model, we derive simple models to characterize PVT variations based on the analytical delay model developed from [13], improve the accuracy with a compensation factor, and verify for the range of \( 0.15 \leq V_{DD} \leq 0.4 \) along with \( V_{DD} \) variation up to 50mV, \( V_{th} \) variation up to 30mV, and temperature variation up to 20°C through the simulation on PTM low power 16-nm technology (\( V_{th} = 0.68191V \)) [11].

2. Analytical Delay Modeling

2.1. Alpha-Power Law MOSFET Model

The Alpha-Power law MOSFET model expresses drain current as follows:

\[
I_D = \begin{cases} 
0, & V_{GS} \leq V_{th} \text{ (cutoff region)} \\
\left( \frac{I_{D0}}{V_{D0}} \right)^{\alpha} V_{DS}, & V_{DS} < V_{DSAT} \text{ (linear region)} \\
\left( \frac{V_{D0}'}{V_{D0}} \right)^{\alpha/2} V_{DSAT}, & V_{DS} \geq V_{DSAT} \text{ (saturation region)}
\end{cases}
\]

where

\[
I_{D0}' = I_{D0} \left( \frac{V_{GS} - V_{th}}{V_{DD} - V_{th}} \right)^{\alpha} \\
V_{D0}' = V_{D0} \left( \frac{V_{GS} - V_{th}}{V_{DD} - V_{th}} \right)^{\alpha/2}
\]

In the above equation, it is shown that the \( \alpha \)-power law depends on \( V_{th} \) (threshold voltage), \( \alpha \) (velocity saturation index), \( V_{D0} \) (drain saturation voltage), and \( I_{D0} \) (drain current). Voltage and current characteristics are shown in Figure 1.

![Figure 1. α-power law MOS model [10]](image-url)
From the $\alpha$-power law, the simple yet useful analytic delay model for CMOS inverter is derived [10].

$$\text{Delay} \propto \frac{V_{DD}}{(V_{DD} - V_{th})^\alpha}$$

(2)

The authors [10] note that this model should not be used for VLSI operating in the near- and subthreshold region because of sensitivity to variations. Therefore, we verify errors of the $\alpha$-power law in those regions and propose an analytical delay model specifically for subthreshold circuits.

2.2. Proposed Analytical Delay Model

We consider the nominal inverter delay with the input transition from 0 to 1. The delay can be estimated by the time taken to charge and discharge the output node.

$$\text{Delay} = \frac{Q_{\text{output}}}{I_{\text{on}}} = \frac{K V_{DD}(C_L + C_{\text{out}})}{I_{\text{leakage}}}$$

(3)

$$I_{\text{leakage}} = I_0 \exp \left( \frac{JV_{DS}}{mV_T} \right) \left[ 1 - \exp \left( -\frac{V_{DS}}{V_T} \right) \right]$$

(4)

since $\exp \left( -\frac{V_{DS}}{V_T} \right) \approx 0$, $I_{\text{leakage}} = I_0 \exp \left( \frac{JV_{DS}}{mV_T} \right) = I_0 \exp \left( \frac{JV_{DD}}{mV_T} \right)$

(5)

$$\text{Delay} = \frac{Q_{\text{output}}}{I_{\text{on}}} = \frac{K V_{DD}(C_L + C_{\text{out}})}{I_0 \exp \left( \frac{JV_{DD}}{mV_T} \right)}$$

(6)

$C_L$ is an external capacitance, and $C_{\text{int}}$ is internal subthreshold capacitance. $\lambda$ is the DIBL coefficient in subthreshold region, $m$ is the subthreshold slope factor, $V_T$ is the thermal voltage ($V_T = \frac{kT}{q}$). Delay depends on the product of $V_{DD}$ (supply voltage) and total capacitance.

2.3. Simulation Result

The CMOS inverter was implemented and simulated using the predictive technology model (PTM) low-power 16-nm technology [11] $(V_{th} = 0.68V)$. For both NMOS and PMOS, the channel length of 50nm was used. For parameters $(K, V_{th}$, and $\alpha$) used in the $\alpha$-power law, HSPICE and MATLAB were used to extract values. After obtaining I-V characteristics through HSPICE, we ran the best-fit lines in MATLAB. Shown in Figure 2 and Table 1, the result shows the improvement in accuracy. Over the range of 0.2V to 0.4V of supply voltage, the average error of proposed model is 16.5% compared to 39.8% of the $\alpha$-power law model. The worst-case occurs when supply voltage is 0.2V. Note that it is rare for subthreshold circuits to operate at that low voltage because of the performance degradation.

Figure 2. Accuracy of proposed model compared to the $\alpha$-power law model
3. Variations Modeling

3.1. Analytical Delay Model for Variations Analysis

We consider the nominal inverter delay with the input transition from 0 to 1. The subthreshold leakage current is expressed in [15]:

\[ I_{\text{leakage}} = I_0 \exp \left( \frac{V_G - V_S - V_T + V_D + \lambda V_D}{mV_T} \right) \left[ 1 - \exp \left( -\frac{V_D}{V_T} \right) \right] \]

where \( I_0 = \mu C_{ox} \frac{W}{L} (m - 1) V_T^2 \exp \left( \frac{V_G - V_T}{mV_T} \right) \).

\( m \) is the subthreshold slope factor, \( V_T \) is the thermal voltage \( (V_T = \frac{kT}{q}) \), \( \mu \) is the carrier mobility, \( C_{ox} \) is the oxide capacitance for the unit area. We ignore other exponential parameters other than \( \frac{V_D}{mV_T} \) in leakage current and get the following:

\[ I_{\text{leakage}} = I_0 \exp \left( \frac{V_D}{mV_T} \right) \left[ 1 - \exp \left( -\frac{V_D}{V_T} \right) \right] \]

Previous research [14] derives the following propagation delay equation by applying Equation (2) to CMOS inverter with an external capacitance load \( C_L \) and internal subthreshold capacitance \( C_{int} \) assuming 0 to 1 input transition.

\[ \text{Delay} = \ln 2 \left( \frac{2mV_T}{I_{\text{th}}^2} \right) \frac{V_D}{mV_T} \frac{\exp \left( \frac{V_D}{2mV_T} \right)}{\exp \left( \frac{V_D}{mV_T} \right)} (C_L + C_{int}) \]

On the basis of Equation (9), we conduct analysis of the PVT variations modeling in the subthreshold region.

3.2. Proposed Variations Models

3.2.1. Effect of Supply Voltage \((V_{DD})\) Variation

Supply voltage variation impacts the delay of subthreshold circuits. Let \( V_{DD1} \) as the nominal supply voltage and \( V_{DD2} \) as the varied supply voltage. \( \Delta \text{Delay} \) indicates the variations in delay (i.e., 1 indicates no change in delay after certain variation and 0.5 indicates the delay is reduced 50% from the nominal delay). From Equation (9), the following analytical delay model has been derived.

\[ \Delta \text{Delay}_{V_{DD}} = \frac{\tau_{\text{step}V_{DD}}}{\tau_{\text{step}V_{DD1}}} = \frac{V_{DD1}^{\frac{V_{DD1}+mV_T}{mV_T}} - V_{DD2}^{\frac{V_{DD2}+mV_T}{mV_T}}}{V_{DD2}^{\frac{V_{DD2}+mV_T}{mV_T}} - V_{DD1}^{\frac{V_{DD1}+mV_T}{mV_T}}} \]

\[ \approx \exp \left( \frac{V_{DD1}^{\frac{V_{DD2}}{mV_T}}}{mV_T} - \frac{V_{DD2}^{\frac{V_{DD2}}{mV_T}}}{mV_T} \right), \quad \text{because} \quad \exp \left( \frac{AV_{DD}}{mV_T} \right) \approx \exp \left( \frac{AV_{DD}}{mV_T} \right) \]

The above equation can be further reduced to the following:

\[ \Delta \text{Delay}_{V_{DD}} = \exp \left( \frac{V_{DD2}^{\frac{V_{DD1}^{\frac{V_{DD2}}{mV_T}}}{mV_T}} - V_{DD1}^{\frac{V_{DD2}}{mV_T}}}{mV_T} \right) \]

Because \( \exp \left( \frac{AV_{DD}}{mV_T} \right) \approx \exp \left( \frac{AV_{DD}}{mV_T} \right) \) is verified by simulating both equations, and the percent difference between the simplified equation and not simplified equation was negligible. The DIBL coefficient \( \lambda \) is given by the device. The subthreshold slope factor \( m \) can be calculated by Equation (6).

\[ m = \ln(10) V_T \left( 1 + \frac{C}{C_{ox}} \right) \]

where \( C_d \) is the depletion layer capacitance and \( C_{ox} \) is the gate-oxide capacitance.

Equation (11) shows that the delay variation exponentially depends on the supply voltage variation. It tends to underestimate the delay, and thus we consider the compensation factor suggested by [13]. After multiplying by \( \frac{V_{DD2}}{V_{DD1}} \), we get Equation (13).

### Table I. Average and worst cases for proposed model and the \( \alpha \)-power law model

<table>
<thead>
<tr>
<th></th>
<th>( \alpha )-power law</th>
<th>Proposed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average</td>
<td>39.8%</td>
<td>16.5%</td>
</tr>
<tr>
<td>Worst</td>
<td>71.1%</td>
<td>33.4%</td>
</tr>
</tbody>
</table>
\[ \Delta \text{Delay}_{\Delta V_{DD}} = \frac{V_{DD2}}{V_{DD1}} \exp \left[ \frac{2(V_{DD1} - V_{DD2})}{2mV_T} \right] \]  

(13)

### 3.2.2. Effect of Temperature Variation

The delay variations due to temperature variation can be written as Equation (7).

\[ \Delta \text{Delay}_{\Delta T} = \frac{t_{\text{step}T_2}}{t_{\text{step}T_1}} = \frac{V_{T1}f_{oT1}}{V_{T2}f_{oT2}} \left( \frac{\exp \left[ \frac{AV_{DD}+mV_{T1}}{V_{T1}^2} \right]}{\exp \left[ \frac{AV_{DD}+mV_{T2}}{V_{T2}^2} \right]} - \frac{\exp \left[ \frac{AV_{DD}+mV_{T2}}{V_{T2}^2} \right]}{\exp \left[ \frac{AV_{DD}+mV_{T1}}{V_{T1}^2} \right]} \right) \]  

(14)

The variation in temperature directly affects \( I_o \) due to the changes in \( \mu \) and \( V_T \). The change in temperature also directly influences \( V_{th} \). Therefore, the impact of the change on \( \mu \) and \( V_{th} \) is modeled using the mobility temperature exponent \( (UTE) \) and the temperature coefficient for \( V_{th} \) \( (K_{th}) \) for BSIM4 level-54 model. Previous research by Lin et al. [13] shows the temperature effects on \( \mu \) and \( V_T \) as follows:

\[ \mu_2 = \mu_1 \left( \frac{T_2}{T_1} \right)^{UTE} \]  

(15)

\[ V_{th2} = V_{th1} + K_{th} \left( \frac{T_2}{T_1} - 1 \right) \]  

(16)

Since \( \exp \left[ \frac{V_{GS-V_{th1}}}{mV_{T1}} - \frac{V_{GS-V_{th2}}}{mV_{T2}} \right] \equiv 0 \),

\[ \frac{i_{oT1}}{i_{oT2}} = \left( \frac{V_{T1}}{V_{T2}} \right)^{UTE} \]  

(17)

Using Equations (15) and (17), Equation (14) can be written as follows:

\[ \Delta \text{Delay}_{\Delta T} = \frac{V_{T1}}{V_{T2}} \left( \frac{T_1}{T_2} \right)^{UTE} \left[ \frac{\exp \left[ \frac{AV_{DD}+mV_{T1}}{V_{T1}^2} \right]}{\exp \left[ \frac{AV_{DD}+mV_{T2}}{V_{T2}^2} \right]} - \frac{\exp \left[ \frac{AV_{DD}+mV_{T2}}{V_{T2}^2} \right]}{\exp \left[ \frac{AV_{DD}+mV_{T1}}{V_{T1}^2} \right]} \right] \]  

(18)

After adding compensation factor of \( \left( \frac{T_2}{T_1} \right)^{UTE} \), we have

\[ \Delta \text{Delay}_{\Delta T} = \frac{V_{T1}}{V_{T2}} \left( \frac{T_1}{T_2} \right)^{UTE+3} \left[ \frac{\exp \left[ \frac{AV_{DD}+mV_{T1}}{V_{T1}^2} \right]}{\exp \left[ \frac{AV_{DD}+mV_{T2}}{V_{T2}^2} \right]} - \frac{\exp \left[ \frac{AV_{DD}+mV_{T2}}{V_{T2}^2} \right]}{\exp \left[ \frac{AV_{DD}+mV_{T1}}{V_{T1}^2} \right]} \right] \]  

(19)

The value of \( UTE \) for the chosen technology is -1.5. The result in 4 shows estimated delays from both Equation (18) and (19) and demonstrates that compensation factor significantly increases the accuracy of the delay variation model.

### 3.2.3. Effect of Process Variation

Process variations include variations in threshold voltage, effective channel length, doping concentration, and more parameters. Since threshold voltage variation is a dominant factor affecting the delay in the subthreshold region [16], the following model considers only the variations in \( V_{th} \).

\[ \Delta \text{Delay}_{\Delta V_{th}} = \frac{t_{\text{step}V_{th1}}}{t_{\text{step}V_{th2}}} \frac{i_{oV_{th1}}}{i_{oV_{th2}}} = \exp \left[ \frac{V_{th2} - V_{th1}}{mV_T} \right] \]  

(20)

Note that the Equation (20) is unchanged from the previous model developed by [13].

### 4. Simulation Results

The CMOS inverter was implemented and simulated using the predictive technology model (PTM) low-power 16-nm technology [11] \( (\lambda = 0.1, m = 0.05916 \text{ at } T = 25 \text{ °C}) \). For both NMOS and PMOS, the channel length of 50nm was used.

With the load capacitance of 1aF and the 0 to 1 step input, the inverter delay has been simulated @ \( T = 25 \text{ °C} \). Figure 3 shows the simulated and estimated inverter delay variations by Equation (6) due to the supply voltage variation. The maximum delay variation of up to \( 4.31x \) is found when \( V_{DD} = 0.38V, \Delta V_{DD} = 50mV \). The minimum error of 0.1% and maximum error of 79.1% are found at \( V_{DD} = 0.17V, \Delta V_{DD} = 50mV \) and \( V_{DD} = 0.225V, \Delta V_{DD} = 30mV \) respectively. Note that in this technology, the percent error is greater than less scaled technology (i.e. 130-nm).
Therefore, we also compare this result with the previously developed model [13]. For the worst-case of 4.31 × delay variations, the proposed model is ~100% more accurate as the previous model [13] does not consider DIBL effect.

![Figure 3. Simulated vs. estimated inverter delay with compensation factor $V_{DD}$ variations @ $T = 25$ °C](image)

Figure 4 shows the simulated and estimated inverter delay variations by Equation (12) due to temperature variation. The CMOS inverter is simulated at different temperature from $\Delta T = 5 - 20$ °C. From $V_{DD} = 0.15 - 0.4V$, the maximum delay variation of up to 1.88 × is found when $V_{DD} = 0.225V, \Delta T = 20$ °C. The maximum error of 37.9% is found at $V_{DD} = 0.175V, \Delta T = 5$ °C. For greater temperature variations such as $\Delta T \geq 40$ °C, the average error exceeds 31% with a worst-case of 41%.

![Figure 4. Simulated vs. estimated inverter delay with compensation factor with temperature variation](image)
Figure 5 compares the simulated and estimated inverter delay variations by Equation (13) due to the process variation. We assume a normal distribution of $V_{th}$ and run Monte Carlo simulations (1000 runs) for the given $\Delta V_{th} = 10\text{mV}$ to $30\text{mV}$. The proposed model estimates with an average error of 1%, minimum error of 0%, and maximum error of 6%. The worst case occurs at $V_{DD} = 0.38V, \Delta V_{th} = 30\text{mV}$. We observed that for $\Delta V_{th} \geq 30\text{mV}$, the proposed model tends to underestimate the delay. It is partly because the DIBL coefficient, $\lambda$, changes near the threshold voltage (i.e. from 0.1 to 0.001 for the chosen technology).

Figure 5. Simulated vs. estimated inverter delay with $V_{th}$ variations

Table 2 shows the effect of compensation factor for two models that correspond to supply voltage and temperature variations. The result shows that the compensation factor greatly enhances the accuracy of estimation of delay and variations.

Table 2. Average, best, and worst cases for models with and without the compensation factor

<table>
<thead>
<tr>
<th>Error</th>
<th>Eq. (5)</th>
<th>Eq. (6) with compensation factor</th>
<th>Eq. (11)</th>
<th>Eq. (12) with compensation factor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average</td>
<td>27.4%</td>
<td>14.6%</td>
<td>62.1%</td>
<td>6.8%</td>
</tr>
<tr>
<td>Best Case</td>
<td>0.4%</td>
<td>0.1%</td>
<td>0.1%</td>
<td></td>
</tr>
<tr>
<td>Worst Case</td>
<td>131.7%</td>
<td>79.1%</td>
<td>87.5%</td>
<td>37.9%</td>
</tr>
</tbody>
</table>

5. Conclusion

Accurately measuring analytical delay is important for efficient circuits, and it still remains challenging for subthreshold CMOS circuits due to PVT (process, voltage, and temperature) variations.

In this paper, we derive and propose analytical delay model and three PVT variations models for subthreshold CMOS circuits. The result shows that proposed delay model has an average error of 16.5% compared to an average error of 39.8% of the alpha-power law model. We also found the delay variations up to $4.31 \times$, $1.29 \times$, and $1.88 \times$ for the chosen range of $V_{DD}, V_{th}$, and temperature. We verify our proposed model on simulation using PTM low power 16-nm technology [11] and demonstrated that proposed delay variations models estimate with the maximum errors of 79.1%, 37.9%, and 6% for supply voltage, threshold voltage, and temperature variations respectively. The proposed delay variations models can be used to predict circuits’ behaviors and functions in sensitive and extreme environments. Compared to the previous model [13], this model considers the DIBL effect, and compared to [14], this...
paper presents three separate, compact models to characterize $V_{DD}$, $V_{Tn}$, and temperature variations and considers adding a compensation factor.

6. References