Timing Modeling for Digital Sub-threshold Circuits

Niklas Lotze, Jacob Göppert, Yiannos Manoli
University of Freiburg - IMTEK, Department of Microsystems Engineering, Laboratory for Microelectronics
Email: {lotze, goeppert, manoli}@imtek.uni-freiburg.de

Abstract—Despite an increasing interest in digital sub-threshold circuits little research has been dedicated to timing modeling in this voltage domain so far. Especially high timing variabilities makes proper modeling necessary to allow for the prediction of timing behavior and timing yield on the path towards design automation.

This paper first deals with gate timing characterization at sub-threshold voltages and a characterization waveform well resembling the actual transistor-level waveforms in this voltage domain is proposed. The error made in this abstraction step is identified and shown to be typically below 3%.

Secondly, the modeling of timing variability is considered and the high correlation between gate delays due to slope propagation combined with strong non-linearities in the delay-slope dependencies are pointed out as modeling challenges. A path-based logic-level Monte-Carlo technique, already magnitudes faster than transistor-level simulation, is applied and shown to match transistor-level Monte-Carlo simulation results better than 3% in mean and 7% in standard deviation values.

Index Terms—sub-threshold circuits, timing modeling, timing analysis, SSTA, characterization

I. INTRODUCTION

For applications requiring ultra-low power consumption, digital sub-threshold circuits are a very attractive choice as they allow for a reduction of supply voltages to the operation point where energy per operation becomes minimal [1]. Even though several successful designs fully operational at sub-threshold supply voltages have been demonstrated [2], [3], the design of sub-threshold circuits remains challenging. This is especially due to the increasing impact of process variations with decreasing supply voltage, impacting both the reliability of output levels produced by logic gates [4] and the variability of gate delays. It has been shown in [5] that the delay variabilities present in sub-threshold circuits have fundamental impact on the design, e.g. shifting the supply voltage at which the minimum energy operation point is expected, which has been also verified by measurements [6].

Despite the recognition of the importance of circuit timing for sub-threshold designs, little research has been dedicated to the modeling of timing in this supply voltage range so far. The authors of [7] propose a methodology which allows identifying critical paths in a design to then characterize these paths using transistor-level Monte-Carlo simulation. Albeit yielding proper results, this approach is very time-consuming due to the incorporation of transistor-level simulations.

This paper analyzes the viability of modeling sub-threshold timing at logic level. Section II first considers the characterization of the gates used, as this step is essential to avoid systematic errors in the modeling process. The results are used in section III, where the modeling of delay variabilities is discussed, resulting in the proposal of a logic-level Monte-Carlo methodology, whose results are finally verified by comparison with transistor-level simulation results.

II. WAVEFORM MODELING

The transfer from transistor- to logic level implies that the complex waveforms present at gate in- and outputs are reduced to simple values like gate delay and input-output slew. To avoid the introduction of large errors, it is crucial that the input waveforms used for characterization resemble the actual transistor level signals. As matching errors in this step introduce systematic errors, these also cannot be expected to be averaged out due to the high variabilities for delay and slew present at sub-threshold voltages, again emphasizing the need for accurate modeling. Due to the considerably different transistor current characteristics at weak inversion, the modeling waveforms used for characterization at sub-threshold voltages should not simply be scaled. Therefore, a characterization waveform specific to sub-threshold circuits is proposed in the following. An ideal inverter step-response at sub-threshold supply, which can be derived in closed form, is used as base function and is generalized for parameter-fitting.

A. Characterization Waveform Derivation

The sub-threshold leakage in a CMOS transistor is expressed as [8]

\[ I_{\text{Leak}} = I_0 e^{\frac{V_{GS}-V_P-\lambda V_{DS}}{nVT}} (1 - e^{-\frac{V_{DS}}{nVT}}) \]  (1)

with \( I_0 \) a constant, \( V_T \) threshold voltage, \( \lambda \) DIBL coefficient, \( n \) sub-threshold swing coefficient and \( V_{th} \) the thermal voltage. The DIBL effect and currents through off-transistors are omitted as otherwise no closed-form solution can be identified for the characterization waveform, which though is highly desirable for this application. For a step input at the inverter, only one of the transistors is active in charging/discharging the output load \( C_{out} \), and solving the differential equation

\[ I(t) \, dt = dV_{out} \, C_{out} \]  (2)

with the boundary value \( V_{out}(0) = 0 \) (rising edge output) results in the desired ideal step response

\[ V_{out}(t) = V_{dd} - V_{th} \ln \left( 1 - \exp \left( -\frac{I_0 e^{\frac{-V_P}{TVT}}}{C_{out}VT} \right) \left( 1 - e^{\frac{V_{dd}}{nVT}} \right) \right) \]  (3)
with \( V_{dd} \) the supply voltage. (3) is generalized with \( \tau = C_{out} V_{th} \left( 1 - \exp \left( \frac{V_{dd}}{V_{th}} \right) \right) / \left( I_0 \exp \left( \frac{V_p}{nV_{th}} \right) \right) \) and \( V_{shape} = V_{th} \) to the proposed characterization waveform

\[
V_{char}(t) = V_{dd} - V_{shape} \ln \left( 1 - e^{-\frac{t}{\tau}} \left( 1 - e^{-\frac{V_{dd}}{V_{shape}}} \right) \right) \tag{4}
\]

Both \( \tau \) and \( V_{shape} \) are used as fitting parameters: \( V_{shape} \) is constant over all gates and is adjusted for an optimum fit of “curve roundness”, deviating from the theoretical value \( V_{th} \) e.g. due to non-step inputs. Parameter \( \tau \) represents signal slew and is calculated by solving (4) for the desired slope between two extraction points.

Fig. 1 shows the waveform resulting from (4) for various values of \( V_{shape} \) along with an exemplary simulated nand3 gate output signal within a signal chain (thus i.e. with non-step input). It becomes obvious that the proposed characterization waveform fits the simulated signal very well after crossing approx. the 50% point, but does not model the onset rounding of the signal transition caused by a finite input transition time. Our studies though indicate that when used as a gate input signal, this difference has very little effect on the resulting output waveform due to negligible currents through the charging transistor at the onset of the signal transition, originating from the exponential dependence from input voltage. It is though important to carefully select extraction points for slope matching.

### B. Model Verification

To verify the applicability of the proposed characterization waveform, the ISCAS’85 benchmark suite has been synthesized using a gate library optimized for sub-threshold operation following the gate sizing guidelines given in [2] and implementing a limited set of gates (max. 4-input, limited complexity). Critical paths have been exported for transistor-level simulation, along with all gates connected to nodes on the critical path to also account for the voltage-dependent and Miller input capacitances of these gates. Wire capacitances are incorporated as determined by the wireload model. The complete paths are simulated on transistor level as timing reference.

The path is then traversed one gate at a time, applying the proposed characterization waveform as input and determining delay and output slew, which is then used to drive the next gate on the path. Wire and input capacitances are lumped into a single ideal capacitance here.

Wire resistances are not taken into consideration, because they become negligible compared to the effective transistor output resistances at sub-threshold supply voltages as already pointed regarding the supply lines in [6]. To estimate the worst-case error made by omitting wire resistance, delays for the strongest inverter in the considered library have been determined using the wire resistance/capacitance combinations given for different fanouts in the wireload. The results in Fig. 2 indicate that the assumption taken is generally valid at sub-threshold voltages (supply voltages below \( \sim 350mV \) in the technology used here).

### C. Results

Two major parameters are to be fitted in the modeling process: the value for \( V_{shape} \) and the extraction points (i.e. voltage levels) used to determine the parameter \( \tau \), having major impact on the modeling error achieved as discussed in Sect. II-A. Fig. 3 shows the resulting maximum errors at the end of all paths for varying upper and lower extraction points along with the error variabilities within the paths, which are a measure for the delay errors of the individual gates in the paths. As both accumulated and individual errors should be minimized, the extraction points 50%-10% for falling transitions have been selected to achieve low error combined with high modeling reliability. Accordant considerations yield 40%-80% extraction points for rising transitions.

An equivalent analysis was ushered for optimization of \( V_{shape} \), but observation of the resulting analog waveforms revealed that values resulting in minimum error did not exhibit
Table I
MODELING ERRORS FOR CRITICAL PATHS FROM ISCAS’ 85 BENCHMARK

<table>
<thead>
<tr>
<th>Design</th>
<th>transistor-level [ns]</th>
<th>modeling [ns]</th>
<th>error [%]</th>
<th>error variabilities [%]</th>
</tr>
</thead>
<tbody>
<tr>
<td>c17</td>
<td>404.8</td>
<td>411.8</td>
<td>1.73</td>
<td>0.09</td>
</tr>
<tr>
<td>c432</td>
<td>14010.2</td>
<td>14039.0</td>
<td>0.21</td>
<td>2.50</td>
</tr>
<tr>
<td>c499</td>
<td>3149.7</td>
<td>3209.6</td>
<td>1.90</td>
<td>1.61</td>
</tr>
<tr>
<td>c880</td>
<td>4984.4</td>
<td>5010.8</td>
<td>0.53</td>
<td>1.83</td>
</tr>
<tr>
<td>c1355</td>
<td>4627.2</td>
<td>4708.8</td>
<td>1.76</td>
<td>1.27</td>
</tr>
<tr>
<td>c1908</td>
<td>6871.3</td>
<td>6939.7</td>
<td>1.00</td>
<td>2.26</td>
</tr>
<tr>
<td>c2670</td>
<td>16635.3</td>
<td>16641.0</td>
<td>0.03</td>
<td>2.96</td>
</tr>
<tr>
<td>c3540</td>
<td>11698.4</td>
<td>11786.9</td>
<td>0.76</td>
<td>1.98</td>
</tr>
<tr>
<td>c5315</td>
<td>8920.0</td>
<td>8957.9</td>
<td>0.42</td>
<td>2.26</td>
</tr>
<tr>
<td>c6288</td>
<td>26630.5</td>
<td>27142.6</td>
<td>1.92</td>
<td>1.75</td>
</tr>
<tr>
<td>c7552</td>
<td>10356.4</td>
<td>10479.8</td>
<td>1.19</td>
<td>2.07</td>
</tr>
</tbody>
</table>

Figure 4. Illustration of non-linearities in input slew rate vs. gate delay dependency for various C_{out} values.

an optimum matching of the analog waveforms. This effect is caused by a conservative modeling of the (non-linear) gate input capacitances, using the maximum possible capacitance as modeling value, and the waveform deviation partly compensating this effect. This though is not desirable as the modeling strategy is purposely conservative to not underestimate delays, therefore values for V_{th_{appc}} have been hand-tuned for optimum waveform matching and shown optimum results at approx. 26mV for falling and 54mV for rising transitions.

Overall errors for the benchmark circuits using the optimized parameter set are shown in Tab. I. A modeling error generally below 3% could be achieved.

III. MODELING OF TIMING VARIABILITY

The massive variabilities at sub-threshold voltages would make timing predictions based on corner-based techniques extremely pessimistic, thus rendering statistical modeling necessary. This section outlines challenges in modeling sub-threshold timing variability and presents a path-based logic-level Monte-Carlo approach for variability analysis which is magnitudes faster than transistor-level approaches.

Timing variability at sub-threshold voltages is largely dominated by the impact of random dopant fluctuations (RDF) and the corresponding threshold voltage variabilities [5]. The exponential impact of V_{th} on on-currents (compare (1)) combined with an approximately Gaussian distribution of V_{th} causes a log-normal distribution of gate delays.

A. Library Characterization and Model Non-Linearity

The gate library described in sec. II-B is used as basis for the modeling approach. The process technology used for our experiments (a 130nm commercial process provided by UMC) provides variability data only via transistor-level Monte-Carlo models. Therefore these were applied for gate variability characterization, which is viable due to the RDF dominated nature of variabilities and also is in accordance with former approaches as [7]. Mean and standard deviation for signal delay and output slew, fully characterizing the distributions under the assumption of log-normally distributed results, are determined by Monte-Carlo simulation for a set of input slews, output capacitances and where applicable all relevant input value combinations (as in the non-linear delay model approach) using the characterization approach outlined in sec. II.

The characterization results display the expected high delay variabilities with delay- and slew-3-σ-corners deviating more than 300% from the mean values. The dependencies between capacitances and output slews/delays show relatively weak non-linearities, whereas strong non-linear behavior is found especially when considering input slew vs. 50-50% delays, as shown in Fig. 4. The relatively weak gates, which are usually preferred at sub-threshold voltages due to better energy efficiency [4], cause high possible input slew rates where output slew may increase slower than input slew, thus causing the decreasing 50-50% delay with increasing input slope. This effect may be mitigated by choice of other reference values for delay, but a significant non-linearity remains.

B. Modeling Approach

An initial analysis using transistor-level Monte-Carlo simulation data from paths as described in sec. II-B has been conducted to study the impact of slew rate variability propagation between gates. It reveals that negligence of this effect results in delay standard deviation errors in the range of 40%. Further analysis shows that slew propagation over one and two gates reduces the standard deviation error to approx. 10% and 5%, respectively, further propagation reducing it only marginally. The high absolute value of variability thus also causes that this variability is significantly propagated, resulting in a strong slew-based correlation between adjacent gates.

Sub-threshold variability modeling thus requires slew propagation combined with non-linear slew-delay dependencies and incorporation of log-normal delay distributions. As most existing statistical static timing analysis (SSTA) approaches either do not consider delay correlation caused by slew propagation (e.g. [9], [10]) or only allow for linear propagation (e.g. [11], [12]), with the latter supporting log-normal distributions only approximately, we at this point decided to implement a logic-level Monte-Carlo based methodology. The implementation uses a path-based approach for good verifiability using the methods from sec. II-B, but could be easily extended for block-based analysis.

The implementation uses two random vectors representing delay \( \bar{d}_i \) and slew \( \bar{s}_i \) for a node \( i \). For propagation through
a gate, a delay mean vector and standard deviation value are calculated as

\[
\begin{align*}
\mu_{\text{delay},i} &= t_{\text{gate},\mu}(\text{inp}, \text{state}, C_{\text{out}}, s_{\text{i}}) \\
\sigma_{\text{delay},i} &= t_{\text{gate},\sigma}(\text{inp}, \text{state}, C_{\text{out}}, s_{\text{i}})
\end{align*}
\]

(5)

wherein the functions \(t_{\text{gate},\mu}(\text{inp}, \text{state}, C_{\text{out}}, s_{\text{i}})\) and \(t_{\text{gate},\sigma}(\text{inp}, \text{state}, C_{\text{out}}, s_{\text{i}})\) are first-order interpolation functions for the mean and standard deviation library values for a certain input \(\text{inp}\), state of the other inputs \(\text{state}\) (where applicable), output load \(C_{\text{out}}\) and input slew rate \(s_{\text{i}}\). Input slew variability is thus propagated by regarding a mean vector \(\mu_{\text{delay},ij}\) instead of a scalar element. These values then undergo element-wise transformation to corresponding parameters for log-normal distributions

\[
\begin{align*}
\mu_{L\text{delay},ij} &= \log(\mu_{\text{delay},ij}) - \frac{1}{2} \log(1 + \frac{\mu_{\text{delay},ij}}{\sigma_{\text{delay},i}}) \\
\sigma_{L\text{delay},ij} &= \sqrt{\log(1 + \frac{\mu_{\text{delay},ij}}{\sigma_{\text{delay},i}})}
\end{align*}
\]

(6)

which are then used to incorporate the log-normally distributed random component of the gate itself by calculating the resulting distribution for \(t_{d_{i+1}}\) as

\[
t_{d_{(i+1)}} = t_{d_j} + \exp(\mu_{L\text{delay},ij} + \sigma_{L\text{delay},ij} N(0, 1))
\]

(7)

The vector \(N(0, 1)\) is a normally distributed random vector with zero mean and unity standard deviation. The slew random vectors \(s_{\text{i}}\) are propagated accordingly, with the random component \(N(0, 1)\) being equal to the one used for delay propagation to account for the correlation between gate delay and output slew values.

C. Results

The extracted critical paths from the ISCAS’85 benchmark as described in sec. II-B are again used for model verification. Transistor-level Monte-Carlo simulations of the complete path (500 runs) are used as reference and compared to the results from the approach outlined in the last section. The results for the individual circuits are shown in Tab. II, showing good matching for the modeling with errors below 2% for mean values and below 7% for standard deviation. Exemplary distributions for output delay in c7552 are shown in Fig. 5 and demonstrate the good matching in distribution shape between transistor- and logic-level modeling. The modeling speedup compared to transistor-level monte-carlo simulations is in the order of 2300x.

IV. Conclusions

It could be shown that modeling of digital sub-threshold circuits at logic level is possible with acceptable errors. Proper gate characterization is crucial to avoid systematic errors in the abstraction from transistor to logic level; using the proposed characterization waveform results in low deviations for resulting delay and slew values compared to transistor-level simulations. It also exhibits good results when used in modeling approaches considering timing variability. The proposed logic-level approach for timing variability characterization

| Table II: Statistical modeling results for ISCAS’85 benchmark |
|-----------------|------|-------|------|-------|
| Design          | transistor level [ns] | modeling [ns] | error [%] |
|                 | \(\mu\) | \(\sigma\) | \(\mu\) | \(\sigma\) |
| c17             | 509.4 | 169.9 | 506.2 | 171.8 | 0.63 | 1.08 |
| c432            | 16010.0 | 3111.4 | 15791.0 | 2971.0 | 1.37 | 4.51 |
| c499            | 3341.7 | 380.6 | 3390.9 | 400.9 | 1.47 | 5.33 |
| c880            | 708.5 | 879.6 | 701.2 | 888.1 | 0.13 | 0.97 |
| c1355           | 5274.1 | 678.9 | 5320.1 | 705.6 | 0.87 | 3.94 |
| c1908           | 8035.5 | 1445.3 | 7918.7 | 1373.9 | 1.43 | 4.94 |
| c2670           | 17977.0 | 3630.3 | 17901.0 | 3540.9 | 0.42 | 2.46 |
| c3540           | 12955.0 | 2329.1 | 12959.0 | 2334.2 | 0.03 | 0.22 |
| c5315           | 10042.0 | 1774.5 | 10030.0 | 1698.2 | 0.39 | 3.80 |
| c6288           | 30910.0 | 1356.0 | 30915.0 | 1422.3 | 0.52 | 4.89 |
| c7552           | 9930.2 | 1984.5 | 9912.9 | 1842.0 | 0.17 | 7.18 |

Figure 5. Comparison between transistor-level and logic level modeling delay distribution for c7552.

could be shown to yield results comparable to transistor-level simulations at a fraction of necessary simulation time. Future research will be directed towards the development of viable approaches for statistical static timing analysis considering the outlined modeling challenges.

REFERENCES