Modeling and Estimating Leakage Current in Series-Parallel CMOS Networks

Paulo F. Butzen¹, Andre I. Reis², Chris H. Kim³, Renato P. Ribas¹
¹Instituto de Informática - UFRGS
Porto Alegre, RS, Brazil
{pbutzen, rpribas}@inf.ufrgs.br
²Nangate Inc.
Menlo Park, CA, USA
are@nangate.com
³EECS - University of Minnesota
Minneapolis, MN, USA
chriskim@umn.edu

ABSTRACT
This paper reviews the modeling of subthreshold leakage current and proposes an improved model for general series-parallel CMOS networks. The presence of on-switches in off-networks, ignored by previous works, is considered in static current analysis. Both contributions present significant influence in the logic circuit leakage prediction when CMOS complex gates are extensively used. The proposed leakage model has been validated through electrical simulations, taking into account a 130nm CMOS technology, with good correlation of the results.

Categories and Subject Descriptors
B.8.2 [Performance and Reliability]: Performance Analysis and Design Aids

General Terms
Design, Performance

1. INTRODUCTION
Power consumption has become an important issue in recent years due to emergent mobile products. Moreover, the leakage currents responsible for power dissipation during idle mode are increasing significantly in advanced CMOS technologies, where the device threshold voltage and the gate oxide thickness tend to reduce. As a consequence, great effort has been concentrated in order to understand the leakage mechanisms, to model their behavior and to develop design techniques for static power saving [1]-[3], [5]-[9].

The main contributors to the total leakage dissipation in CMOS design are the subthreshold current and the gate oxide current. The subthreshold current, negligible in old processes when compared to the dynamic charge and short-circuit currents, start to be taken into account in 180nm technologies. Gate leakage, on the other hand, tends to become the main factor responsible for the static consumption from sub-100nm CMOS processes [3]. Design techniques for leakage reduction have been proposed considering for instance the input state dependency, multi-Vth devices, bulk biasing, as future high-K gate dielectric and metal gate electrodes are expected to solve the gate leakage troubles [1]. The most considered design strategies are based on the well known stack effect which, in fact, presents a distinct influence on different leakage mechanisms [3].

The modeling of the stack effect has been treated in the literature for subthreshold and gate leakages. However, only basic NOR and NAND gates, i.e. purely series and parallel arrangements of transistors have been actually addressed. It should be quite enough for standard cell libraries applied to the technology mapping performed by commercial EDA tools. However, in most recent EDA technologies, the library-free mapping technique is targeted [4]. In this case, CMOS complex gates, composed by a great variety of mixed series-parallel pull-up/pull-down logic networks, are extensively used. Moreover, on-transistors present in off-networks for subthreshold current analysis have also been largely ignored by previous works, as discussed in Section 2. In this paper, the subthreshold leakage modeling is improved to take into account both factors, as described in Section 3. Section 4 presents experimental results in order to validate this new model and to demonstrate its benefits. Finally, the conclusions are given in Section 5.

2. RELATED WORKS
Several subthreshold leakage models have been recently reported in the literature. R. Gu and M. Elmasry in [5] presented the subthreshold current model for purely series and parallel off-transistors arrangements (Inverter, NAND and NOR gates), and assuming equivalent standby current in both NMOS and PMOS devices. The two XOR topologies, presented by Gu [5], result, in fact, in single off-transistor configurations during the steady state analysis. Transistor stacks composed by single transistors in chain are also addressed in [6]-[8], while on-devices are considered as ideal short-circuits.

AOI and OAI gates are mentioned as examples for the analysis of series-parallel off-networks by D. Lee et al. in [9]. Multiple parallel transistor stacks are computed, taking into account separately each branch, and the values are then added to obtain the total leakage of the logic gate. However, in the case of parallel transistors within a stack, such transistors are initially collapsed and replaced by a single device with transistor size equal to the sum of their sizes. Such strategy works very well but cannot be applied to general series-parallel network, as illustrated in Fig.1. Again, on-transistor in the stack is neglected, being considered as ideal on-switch.

Yang et al. [3], in turn, presents a detailed analysis to estimate the total leakage, including and interacting subthreshold and gate leakage currents, for NAND and NOR gates. Different AOI and OAI gates are also discussed emphasizing the strong influence of the electrical topology in the leakage value. Although not

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

GLSVLSI’07, March 11–13, 2007, Stresa-Lago Maggiore, Italy.
Copyright 2007 ACM 978-1-59593-605-9/07/0003...$5.00.
mentioned, probably the same strategy of transistor width equivalence is applied to treat parallel devices, as proposed in [9].

The major contributions of this paper are: (a) the extension of subthreshold current modeling for general mixed series-parallel off-networks, and (b) the analysis and modeling of subthreshold leakage considering the presence of on-devices in off-networks. These two factors have no influence in the gate leakage estimation already proposed by Yang et al. [3], improving this method in the total leakage calculation.

3. SUBTHRESHOLD LEAKAGE MODEL

Standard CMOS logic gates are composed of series-parallel transistor networks. As mentioned previously, the total leakage dissipation results from the sum of the current in each branch. To present the proposed method the off-network illustrated in Fig. 1 can be considered as the entire NMOS pull-down arrangement, or a branch from a more complex CMOS gate. The same analysis is applicable to a PMOS pull-up tree.

![Figure 1 - NMOS series-parallel network.](image)

From the BSIM MOS transistor model [10], the subthreshold current for a MOSFET device can be modeled as

\[
I_s = I_0 W e^{-\frac{V_{gs}-V_{th}}{nV_T}} \left[1 - e^{-\frac{V_{th}}{V_T}}\right] \tag{1}
\]

where \(I_0 = \mu_0 C_{ox} V_T^2 e^{V_{gs}/kT}\) and \(V_T = kT/q\). \(V_{gs}, V_{th}, V_{oi}\) and \(V_{so}\) are the gate, drain and bulk voltage of the transistor respectively. \(V_{th}\) is the zero bias threshold voltage. \(W\) and \(L\) are the effective transistor width and length, respectively. \(\gamma\) is the body effect coefficient and \(\eta\) is the DIBL coefficient. \(C_{ox}\) is the gate oxide capacitance, \(\mu_0\) is the mobility and \(n\) is the subthreshold swing coefficient.

In Fig. 1, the currents passing through the transistors are given by

\[
I_{s1} = I_0 W_1 e^{-\frac{-(V_1+V_2+V_5)-(V_{th_1}-\eta(V_{th_2}-V_1-V_5)+\gamma(V_1+V_5))}{nV_T}} \tag{2}
\]

\[
I_{s2} = I_0 W_2 e^{-\frac{-(V_1+V_2+V_6)-(V_{th_1}-\eta(V_{th_2}-V_1-V_6)+\gamma(V_1+V_6))}{nV_T}} \tag{3}
\]

\[
I_{s3} = I_0 W_3 e^{-\frac{-(V_1+V_3)-(V_{th_3}-\eta(V_{th_2}-V_1-V_3)+\gamma(V_1+V_3))}{nV_T}} \tag{4}
\]

\[
I_{s4} = I_0 W_4 e^{-\frac{-(V_1+V_3)-(V_{th_4}-\eta(V_{th_3}-V_1-V_3)+\gamma(V_1+V_3))}{nV_T}} \tag{5}
\]

\[
I_{s5} = I_0 W_5 e^{-\frac{-(V_3)-(V_{th_5}-\eta(V_{th_3}-V_3)+\gamma(V_1+V_3))}{nV_T}} \tag{6}
\]

\[
I_{s6} = I_0 W_6 e^{-\frac{-(V_6)+\gamma(V_3)}{nV_T}} \tag{7}
\]

The derivation assumes that \(V_{dd} \gg V_i, i = 1,2,3; V_i \gg V_T\) and \(V_j \gg V_T\) which will be confirmed in the results through the Hspice simulation. Thus, the term \([1-e^{-(V_{th_3}/V_T)}]\) in equations (2), (3), (4), (5) and (6) has been ignored.

First of all, the currents across the first, the second and the fourth transistors are equalized. By solving the equation \(I_{s1} + I_{s2} = I_{s4}\), then \(V_1\) is given by

\[
\eta V_{dd} + nV_T \ln \left(\frac{W_1 + W_2}{W_4}\right) = \frac{V_1}{1 + 2\eta + \gamma} \tag{8}
\]

In the next step, the \(V_2\) value is obtained from the equation \(I_{s3} + I_{s4} = I_{s5}\), as following:

\[
nV_T \ln \left(\frac{W_1 e^{\eta V_{dd}} + W_2 e^{\eta V_T}}{W_5}\right) = \frac{nV_T}{1 + \eta + \gamma} \tag{9}
\]

It is also assumed \(V_3 < V_T\). As a consequence, the term \(e^{-(V_3/V_T)}\) in (7) can be expressed as \((1 - V_3/V_T)\). Introducing this assumption and making \(I_{s5} = I_{s6}\), \(V_3\) is then expressed by the equation (10), which is accurately solved after some iteration.

\[
\frac{1 + \eta + \gamma}{n} \left(V_3/V_T\right) + \ln \left(V_3/V_T\right) = \eta V_2/nV_T + \ln \left(W_5/W_6\right) \tag{10}
\]

3.1 General subthreshold leakage model

Based on the previous calculation, the model can be generalized as follows. The subthreshold current through the top devices, i.e. transistors connected to \(V_{dd}\) can be expressed by equation (11). This equation considers the variable \(V_{th}\) as the voltage across every transistor placed below the top transistor in the stack.

\[
I_{s1} = I_0 W_1 e^{-\frac{-(V_1+V_2+V_5)-\sum V_{th_1}-(V_{th_2}-\sum V_j)+\gamma\sum V_j}{nV_T}} \tag{11}
\]
The subthreshold current through the other transistors in the network is expressed by equation (12). The differences between both equations are observed in the $\eta$ expression (DBL effect) and in the last term, which can be eliminated when $V_i \gg V_T$. Again, $V_i$ represents the voltage across every transistor below the node in the stack.

$$ I_{St} = I_0 W e^{-\sum \left[ V_{i} - \eta V_i' + \gamma \sum V_j \right] \left( 1 - e^{-V_i / V_T} \right) / n V_T} $$

The voltage across the transistors can be evaluated in three different situations, exemplified in the previous example. The subsequent analysis assumes that $V_{dd} \gg V_i$ which drop out all the $V_i$ terms. It also considers the fact that $V_i \gg V_T$, so that the $\left[ 1 - e^{-V_i / V_T} \right]$ term can be ignored.

The first situation is represented by the voltage $V_i$ in Fig. 1. In this case, it is possible to associate every transistor connected in that node by series-parallel association. The terms $W_{above}$ and $W_{below}$ in the equation (13), represent the width of the transistor above and below the node $V_n$, respectively. For this condition, $V_i$ is given by

$$ V_i = \frac{\eta V_{dd} + n V_T \ln \left( \frac{W_{above}}{W_{below}} \right)}{1 + 2\eta + \gamma} $$

The second situation, in turn, is presented by the voltage $V_2$ in Fig. 1. In this condition, it is not possible to make series-parallel associations between the transistors connected at $i$-index node. The term $V_{above}$ in the following equation represents the voltage of the transistors above the node $V_i$. For this state, the voltage $V_i$ is given by

$$ V_i = \frac{n V_T \ln \left( \sum W_{above} e^{\eta V_{above} / n V_T} \right)}{W_{below} (1 + \eta + \gamma)} $$

Finally, the third situation is represented by the voltage $V_3$ in previous example. This case only happens at the bottom transistors and the analysis cannot assume $V_i \gg V_T$, so that the term $\left[ 1 - e^{-V_i / V_T} \right]$ in (12) should not be ignored. To simplify the mathematic calculation, the term $e^{-V_i / V_T}$ can be expressed by $(1-V_i / V_T)$. Then, $V_i$ is obtained by equation (15), where $C = I + \eta + \gamma$

$$ \frac{C \left( \frac{V_i}{V_T} \right)}{n \left( \frac{V_i}{V_T} \right)} + \ln \left( \frac{V_i}{V_T} \right) = \eta V_{above} / n V_T + \ln \left( \frac{W_{above}}{W_{below}} \right) + \ln \left( \frac{V_{above}}{V_T} \right) $$

### 3.2 Influence of on-transistors in off-Networks

The previous analysis considers only off-networks composed exclusively by transistors that are turned off. Usually, in the most cases, the transistors that are turned on could be treated as ideal short-circuits, because the drop voltage across such devices is some orders of magnitude smaller than the drop voltage across the off-transistors.

However, in the case of NMOS transistors switched on and connected to $V_{dd}$ power supply, the drop voltage across them should be taken into account. In the leakage current analysis, this voltage drop is somewhat important when the transistor stack presents only one off-device at the bottom of the stack. In stacks with more than one off-transistor in series configuration the on-devices could be considered as zero drop voltage short-circuit without impact in the result accuracy, as illustrated in Fig. 2.

Similar analysis is valid for PMOS transistors in off-networks when they are connected to the ground reference.

In the proposed model, the drop voltage across the transistors that are turned on is called $V_{drop}$ and the term $V_{dd} - \sum V_j$ in equation (11) must be replaced by $V_{dd} - V_{drop} - \sum V_j$.

![Figure 2 – Influence of on-transistor in off-stack leakage current.](image)

### 4. RESULTS

In order to validate this work, the results obtained from the proposed model were compared to Hspice simulation results, considering commercial 130nm CMOS process parameters and operating temperature at 100°C. It is known that down to 100nm processes, the subthreshold currents represent the main leakage mechanism, while for sub-100nm technologies, the model presented herein should be combined with gate leakage estimation, already proposed by other authors [3][9]. Table I presents the parameters used in the analytical modeling. In the first moment, transistors with equal sizing were applied to simplify the analysis, although the device size is a parameter in the model.

The leakage current was calculated and correlated with Hspice results for several pull-down NMOS off-networks, depicted in Fig 3. The results presented in Table II show a good agreement between the analytical model and the simulation data, showing an absolute average error less than 10%. It is interesting to note that the static current in both networks (h) and (i) from Fig. 3, not
treated by previous models, are accurately predicted. The main
difference is observed when three off-transistors are placed in
series arrangement in the network. This difference appears when
the model assumes $V_i < V_T$ and the term $e^{(-V_i/V_T)}$ in equation
12 is changed by $(1 - V_i/V_T)$.

Table I. Parameters used in the analytical model.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>$V_{dd}$ (V)</td>
<td>1.2</td>
</tr>
<tr>
<td>$V_{drop}$ (V)</td>
<td>0.14</td>
</tr>
<tr>
<td>$I_0$ (mA)</td>
<td>20.56</td>
</tr>
<tr>
<td>$W$ (µA)</td>
<td>0.4</td>
</tr>
<tr>
<td>$\eta$</td>
<td>0.078</td>
</tr>
<tr>
<td>$\gamma$</td>
<td>0.17</td>
</tr>
<tr>
<td>$n$</td>
<td>1.45</td>
</tr>
</tbody>
</table>

Table II. Subthreshold leakage current (nA) related to the off-
networks depicted in Fig. 3.

<table>
<thead>
<tr>
<th>Network</th>
<th>HSPICE</th>
<th>Model</th>
<th>Diff. (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a)</td>
<td>1.26</td>
<td>1.26</td>
<td>-</td>
</tr>
<tr>
<td>(b)</td>
<td>6.58</td>
<td>6.60</td>
<td>0.30</td>
</tr>
<tr>
<td>(c)</td>
<td>8.34</td>
<td>8.34</td>
<td>-</td>
</tr>
<tr>
<td>(d)</td>
<td>0.69</td>
<td>0.75</td>
<td>8.70</td>
</tr>
<tr>
<td>(e)</td>
<td>1.23</td>
<td>1.24</td>
<td>0.81</td>
</tr>
<tr>
<td>(f)</td>
<td>0.68</td>
<td>0.74</td>
<td>8.82</td>
</tr>
<tr>
<td>(g)</td>
<td>0.72</td>
<td>0.77</td>
<td>6.94</td>
</tr>
<tr>
<td>(h)</td>
<td>1.29</td>
<td>1.28</td>
<td>0.78</td>
</tr>
<tr>
<td>(i)</td>
<td>1.29</td>
<td>1.28</td>
<td>0.78</td>
</tr>
</tbody>
</table>

Table III and IV correspond to the results related to both NMOS
trees in Fig. 3 (h) and (i), respectively. In these tables, the input
dependence leakage current is evaluated for all input
combinations. In some cases, different input vectors result in
equivalent off-device arrangements. For that, the Hspice values
are presented for minimum and maximum values obtained
applying the set of equivalent input states. Moreover, the previous
model presented in [6] was also calculated for such logic states.
Note that the first input vector in both cases, which represents the
entire network composed by off-switches, is not treated by the
model proposed in [6]. Basically, different values from both
methods are obtained when on-transistors are considered in the
off-networks, providing more correlation with the electrical
simulation results.

Table III. Input dependence leakage estimation (nA) in logic
network (h) from Fig. 3 (pull-down NMOS tree).

<table>
<thead>
<tr>
<th>Input-state (abcd)</th>
<th>HSPICE results*</th>
<th>Proposed model</th>
<th>Previous model [6]</th>
</tr>
</thead>
<tbody>
<tr>
<td>0000</td>
<td>1.29</td>
<td>1.28</td>
<td>-</td>
</tr>
<tr>
<td>0001</td>
<td>9.60</td>
<td>9.60</td>
<td>9.60</td>
</tr>
<tr>
<td>0010</td>
<td>6.30/6.70</td>
<td>6.60</td>
<td>8.34</td>
</tr>
<tr>
<td>0100</td>
<td>1.37</td>
<td>1.31</td>
<td>1.31</td>
</tr>
<tr>
<td>0101</td>
<td>16.67</td>
<td>16.69</td>
<td>16.69</td>
</tr>
<tr>
<td>1000</td>
<td>1.36</td>
<td>1.30</td>
<td>1.31</td>
</tr>
<tr>
<td>1001</td>
<td>14.91</td>
<td>14.94</td>
<td>16.69</td>
</tr>
</tbody>
</table>

* Equivalent vectors – 0110, 1010, 1100, 1110. The HSPICE value is
given for min./max. currents related to such equivalent vectors.

Table IV. Input dependence leakage estimation (nA) in logic
network (i) from Fig. 3 (pull-down NMOS tree).

<table>
<thead>
<tr>
<th>Input-state (abcde)</th>
<th>HSPICE results*</th>
<th>Proposed model</th>
<th>Previous model [6]</th>
</tr>
</thead>
<tbody>
<tr>
<td>00000</td>
<td>1.29</td>
<td>1.28</td>
<td>-</td>
</tr>
<tr>
<td>00001</td>
<td>9.71</td>
<td>9.65</td>
<td>9.65</td>
</tr>
<tr>
<td>00010</td>
<td>1.43</td>
<td>1.34</td>
<td>1.34</td>
</tr>
<tr>
<td>00011</td>
<td>25.00</td>
<td>25.02</td>
<td>25.02</td>
</tr>
<tr>
<td>00100</td>
<td>1.36/1.37</td>
<td>1.30</td>
<td>1.31</td>
</tr>
<tr>
<td>00101</td>
<td>14.91/15.14</td>
<td>14.94</td>
<td>16.69</td>
</tr>
<tr>
<td>00110</td>
<td>6.30/6.73</td>
<td>6.60</td>
<td>8.34</td>
</tr>
</tbody>
</table>

* The HSPICE value is given for min./max. currents related to such equivalent vectors.

Fig. 4 shows the subthreshold average leakage current related to
the NMOS networks illustrated in Fig. 3 (h) and (i). As discussed,
the previous model from [6] cannot estimate the subthreshold
current for the first input vector in both cases, and it is not
considered in the average static current calculation. Unlike the
previous model, the proposed one presents results close to Hspice
simulations. The main reason for that is the influence of on-transistors in off-networks, neglected in previous works.

In terms of combinatorial circuit static dissipation analysis, the technology mapping task divides the entire circuit in multiple logic gates. Thus, they can be treated separately for the leakage estimation, since the input state of each cell is known according to the primary input vector of the circuit. A complex CMOS logic gate, whose transistor sizes were determined by considering the Logical Effort method [11], is depicted in Fig. 5. Table V presents the comparison between electrical simulation data and the proposed model calculation.

Finally, the proposed model has been verified to the variation of power supply voltage and operating temperature, depicted in Figs. 4 and 5, respectively. The influence of temperature variation in the predicted current shows good agreement with Hspice results. On the other hand, the difference between the subthreshold currents obtained from the electrical simulator and the analytical model to voltage variation can be justified by eventual inaccuracy in the parameter extraction listed in Table I.

![Figure 5 – CMOS complex gate, with different transistor sizing, according to the Logic Effort [11].](image)

![Figure 4 – Subthreshold leakage average for Fig. 3 (h) and (i) pull-down networks.](image)

**Table V. Subthreshold leakage current related to the CMOS complex gate depicted in Fig 5.**

<table>
<thead>
<tr>
<th>Input state (abcd)</th>
<th>HSPICE results</th>
<th>Proposed model</th>
<th>Diff(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0000</td>
<td>4.01</td>
<td>4.13</td>
<td>3.0</td>
</tr>
<tr>
<td>0001</td>
<td>20.67</td>
<td>20.68</td>
<td>0.0</td>
</tr>
<tr>
<td>0010</td>
<td>19.93</td>
<td>19.99</td>
<td>0.3</td>
</tr>
<tr>
<td>0011</td>
<td>44.52</td>
<td>43.27</td>
<td>2.8</td>
</tr>
<tr>
<td>0100</td>
<td>4.44</td>
<td>4.29</td>
<td>3.4</td>
</tr>
<tr>
<td>0101</td>
<td>42.34</td>
<td>42.37</td>
<td>0.1</td>
</tr>
<tr>
<td>0110</td>
<td>19.93</td>
<td>19.99</td>
<td>0.3</td>
</tr>
<tr>
<td>0111</td>
<td>43.38</td>
<td>43.27</td>
<td>0.3</td>
</tr>
<tr>
<td>1000</td>
<td>4.40</td>
<td>4.26</td>
<td>3.2</td>
</tr>
<tr>
<td>1001</td>
<td>36.81</td>
<td>36.50</td>
<td>0.8</td>
</tr>
<tr>
<td>1010</td>
<td>19.93</td>
<td>19.99</td>
<td>0.3</td>
</tr>
<tr>
<td>1011</td>
<td>43.38</td>
<td>43.27</td>
<td>0.3</td>
</tr>
<tr>
<td>1100</td>
<td>19.50</td>
<td>19.99</td>
<td>2.5</td>
</tr>
<tr>
<td>1101</td>
<td>96.67</td>
<td>96.92</td>
<td>0.3</td>
</tr>
<tr>
<td>1110</td>
<td>20.43</td>
<td>19.99</td>
<td>2.2</td>
</tr>
<tr>
<td>1111</td>
<td>20.48</td>
<td>20.21</td>
<td>1.3</td>
</tr>
</tbody>
</table>

![Figure 6 – Variation of subthreshold leakage current in terms of power supply voltage variation.](image)

![Figure 7 – Variation of subthreshold leakage current according to the operating temperature variation.](image)
5. CONCLUSIONS
A new subthreshold leakage current model has been presented to be applied in general series-parallel off-networks, improving previous works not suitable to such a kind of static current prediction. The proposed model has been validated considering a 130nm CMOS technology, in which the subthreshold current is the most relevant leakage mechanism. In the case of sub-100nm processes where gate leakage becomes more significant, the present work should be combined with already published works which address the interaction between the subthreshold and the gate leakage currents, such as proposed by Yang et al. [3], in order to improve the accuracy of the total leakage estimation.

6. REFERENCES


