Algorithms for Estimating Number of Glitches and Dynamic Power in CMOS Circuits with Delay Variations∗

Jins D. Alexander and Vishwani D. Agrawal
Auburn University, Department of Electrical and Computer Engineering, Auburn, AL 36849, USA
alexaji@auburn.edu, vagrawal@eng.auburn.edu

Abstract

Dynamic power dissipation of a CMOS VLSI circuit depends on the signal activity at gate outputs. The activity includes the steady-state logic transitions as well as glitches. The latter are a function of gate delays, which, for modern VLSI circuits, have wide process-related variations. Both average and peak power dissipations are useful and are traditionally estimated by Monte Carlo simulation. This is expensive and the accuracy, especially for peak power, depends upon the number of circuit delay samples that are simulated. We present an alternative. We use zero-delay simulation of a vector pair to determine the steady-state logic activity. We derive linear-time algorithms that, using delay bounds for gates, determine the maximum, minimum and average number of transitions that each gate output can produce. From this information, we estimate the average and peak energy consumed by each vector pair in a given vector set. For a set of random vectors applied to c7552 circuit, our analysis determined the per-vector energy consumption as 82.2 picojoules average and 196.3 picojoules peak. In comparison, Monte Carlo simulation of 1,000 circuit samples gave 82.8 picojoules average and 146.1 picojoules peak. The discrepancy of the peak consumption will reduce if more samples were simulated in the Monte Carlo method. Even with 1,000 samples the CPU time of the Monte Carlo analysis was three orders of magnitude greater than the alternative method we offer in this paper.

1. Introduction

Delays in combinational logic produce transient behavior generally known as glitches. In synchronous systems, clocks prevent the glitches from affecting the system function. Glitches are, however, responsible for consuming significant amount of dynamic power. The transient behavior of a circuit is often verified through simulation, which must have delay information for gates and interconnects. Gate delays often change depending on signal states, device temperature, power supply fluctuations, and interconnect coupling noise. Also, in today’s nanoscale technologies, there are wider process variations. Thus, the delays of gates may be modeled as ranges usually specified as (min, max).

The cost and accuracy of simulation depend upon the available detail and the complexity of analysis. For verifying the correctness of a synchronous system, it is sufficient to determine the duration of transient intervals. For power analysis, we need the actual number of transitions within those intervals. This paper provides analytical algorithms for obtaining these. When gate delays are characterized as bounds, we determine the transient intervals as well as the minimum and maximum numbers of transitions that signals can make. These algorithms allow us to use very simple zero-delay logic simulation for power analysis as discussed in recent publications [2, 3]. Those results are reproduced in Section 4 to demonstrate the usefulness of the present analysis.

2. Background

Prevailing power estimation techniques simulate functional or random vectors or propagate signal probabilities [20]. Randomly generated vectors are simulated to estimate power until a desired confidence level is obtained [8]. Specific statistical and deterministic approaches [8, 25] can speed up the estimation over the traditional simulation-based methods. These include generation of test vectors that can cause high power consumption [26]. One may assign transitions to gate nodes based on fan-outs and then justify them through backtracing to primary input. Alternative techniques involve state diagrams with state to state transitions for estimating the energy consumption [22]. Genetic algorithms (GA) has been used to generate patterns that stimulate high power consumption based on the fitness property of current vector [12, 16]. This is similar to a weighted activity function approach to generate patterns that maximize power consumption [11].

Some methods use a zero-delay model and hence do not consider glitches. Others use fixed (nominal or worst-case) delays but do not consider process variation. Effect of process variation on power has been discussed [10, 19, 21] mainly for optimization and reduction of leakage power. Another way is to use a Monte Carlo approach [25]. Here we consider delays to be a random variables. We must simulate a large number of circuit samples to get reasonable estimates.
for average and peak power. The simulation must have sufficient detail to accurately generate transients. Indeed, this technique is quite time consuming. In other contexts, process variation has been considered in logic verification, critical path delay analysis, and delay testing. Both statistical [5, 14] and bounded delay models [6, 7, 9] have been used.

3. Delays, Ambiguity and Transitions

In the bounded delay [23] model, the delay of a gate is modeled as a range, typically expressed as min and max values. For the bounded delay model, we derive the transient intervals and the number of transitions that the outputs of gates make for each input vector pair. It is assumed that the steady-state signal values, which can be easily predetermined by zero-delay logic simulation, are known. We then determine the ambiguity intervals (transient periods) and the bounds on the number of signal transitions that gate outputs could make.

3.1 Ambiguity Interval

Events (signal changes) occur at gate outputs in a circuit after a new vector is applied at primary inputs. We determine the ambiguity (transient) interval for the signal at the output of a gate from (1) bounded delay specification of the gate, and (2) steady-state signal values termed as initial value (IV) and final value (FV). We will assume that the primary inputs change at deterministic times synchronized with a clock. In general, however, input change ambiguities can be specified and treated in a similar way.

The time reference here is the start of the clock period. The delay of a gate is specified by minimum and maximum bounds, min and max, respectively. We assume that signal changes at primary inputs occur precisely at the beginnings of clock periods. However, as signals pass through gates, the information about the exact timing of transitions is lost. Thus, for a signal we define an earliest arrival time (EA) before which the signal would remain steady, and a latest stabilization time (LS) after which it is again guaranteed to be steady. The time (EA, LS) is called the ambiguity interval.

For applications like simulation [6] it is sufficient to evaluate the ambiguity interval and final value for all signals following the primary input changes. However, an estimation of dynamic power requires information on how many transitions could occur during ambiguity intervals.

Consider the example of Figure 1. We define [6] EAdv as the earliest arrival time of a signal that causes the input of the gate to change from controlling value (e.g., 0 for AND gate) to non-controlling value (1 for AND gate). Similarly, LSdv is the latest stabilization time of an input signal changing from controlling value to non-controlling value. Thus, EAsv and LSsv are the earliest arrival time and the latest stabilization time, respectively, for an input signal changing from non-controlling to controlling value. Note that for EA = −∞ and LS = −∞, the output is defined as having no ambiguity interval or is in steady state condition. The following theorem determines the output ambiguity interval [2].

**Theorem 1:** The ambiguity interval (EA, LS) for the output signal of a logic gate is determined from the ambiguity intervals of input signals, their pre-transition and post-transition steady-state values, and minimum and maximum gate delays, as follows:

For all inputs i of a gate,

\[ E1 = \max \{EAdv(i)\} \]  
\[ E2 = \min \{EAsv(i)\} \]  
\[ L1 = \min \{LSsv(i)\} \]  
\[ L2 = \max \{LSdv(i)\} \]  
\[ EA' = \max \{E1, E2\} \]  
\[ LS' = \min \{L1, L2\} \]

Then

\[
(\text{EA, LS}) = \begin{cases} 
\{EA' + \text{mindel}, LS' + \text{maxdel}\} & \text{if } (LS' - EA') \geq \text{maxdel} \\
\{\infty, -\infty\} & \text{if } (LS' - EA') < \text{mindel}
\end{cases}
\]

where the inertial delay of the gate is bounded as \(\text{mindel, maxdel}\).

The output of a gate can have multiple ambiguity intervals separated by deterministic signal values as we will demonstrate in the next subsection. In that case, each ambiguity interval as well as each deterministic interval will be affected by the inertial filtering caused by mindel. For simplicity, Theorem 1 takes a pessimistic view by combining all possible ambiguity intervals into a single continuous region.

Next, we use the steady-state values to derive more details within the ambiguity intervals. For the example of Figure 1, at the output of the AND gate, \(EA = EAdv\) and \(LS = LSsv\). Also notice when IV takes controlling (non-controlling) value, \(EAsv (EAdv) = -\infty\). Similarly, when FV takes controlling (non-controlling) value, \(LSdv (LSsv) = \infty\).
3.2 Maximum Number of Transitions

For glitch-free primary inputs, the number of transitions can only be 0 or 1 for a vector-pair. Assume that the available data for a gate consists of ambiguity intervals and the minimum and maximum numbers of transitions, mintran and maxtran, respectively, for fan-ins. We will estimate the value of maxtran at the output. We should consider: (1) cause - an output transition must be caused by an input transition, and (2) filtering - gate inertia can filter out transitions that are closer to each other than the gate delay. In the absence of detailed information, we assume that the transitions at a fan-in are evenly spaced within its ambiguity interval.

**Theorem 2:** The maximum number of transitions is the minimum of two upper bounds [2]:

\[
\text{maxtran} = \min(\text{Nd}, N)
\]

where \(\text{Nd}\) is the maximum number of transitions permitted by the gate inertial delay and \(N\) is the sum of all transitions present at all inputs of the gate.

**Proof:** We derive the two upper bounds for maximum transition and take the lower of those as a tighter upper bound. This analysis is an improvement over a previously reported result [1].

**First upper bound (Nd):** We calculate the maximum number of transitions that can be accommodated in the ambiguity interval given by the gate delay bounds and the \((IV, FV)\) output values.

We consider the filtering of glitches by gate inertia. Note that largest number of transitions will be accommodated if they were evenly spaced over the output ambiguity interval with a spacing equal to the inertial delay. Consider the following cases:

1. If the output has a static hazard, then we allow an even number of transitions determined by \(t_{pd} \times (2n - 1) \leq \text{LS} - \text{EA}\), and the number of transitions is given by \(2n\), where \(t_{pd}\) is the gate delay given by minimum delay bound, \(n\) is the number of hazards that can possibly be accommodated in the ambiguity interval, \(\text{LS}\) is the latest stabilization time and \(\text{EA}\) is the earliest arrival time for the output signal, as given by Theorem 1.

2. Similarly, for an output signal with a dynamic hazard we would get an odd number of transitions determined by \(t_{pd} \times 2n \leq \text{LS} - \text{EA}\), and the number of transitions is given by \(2n - 1\).

In our analysis, we have taken \(\text{maxdel}\) in our estimation of \(\text{Nd}\).

**Second upper bound (N):** We modify the sum \((\text{Nsum})\) of all input transitions as:

\[
N = \text{Nsum} - k
\]

where \(k = 0, 1, \text{or } 2\) for a 2-input gate and is determined by the ambiguity intervals and \((IV, FV)\) values of inputs. The procedure is explained by the example of Figure 2. In (a), for the given input transitions \(n1 = 6\) and \(n2 = 4\), the output cannot have the total sum of input transitions. This is because, when we consider the sum, in the case of a two signals going from a controlling value to a non-controlling value, only one of the two transitions should be counted. Thus, we see that a correction factor \(k\) is required, which would give us a \(\text{maxtran} = 8\) for the example in (a). The example in (b) is a possible deterministic signal representation.

Consider the example of Figure 3. In (a) the gate is assumed to have zero delay. Thus from the above principles we get the maximum number of transitions from the first upper bound, \(\text{Nd} = \infty\) and the second upper bound gives us \(N = 8\). The minimum of the two gives us \(\text{maxtran} = 8\). In (b), we consider gate delay bounds \((3, 5)\) and find that 8 transitions that must now be at least 3 time units apart can never occur within the output ambiguity interval \((6, 23)\) whose width is only 17 units. From Theorem 2, we get \(\text{Nd} = 6\), \(N = 8\) and \(\text{maxtran} = 6\). This agrees with the output in Figure 3(b) that assumes gate delay of 3 to produce maximum transitions.

3.3 Minimum Number of Transitions

A pessimistic lower bound on the minimum number of transitions, mintran, is easily found from the steady-state values at the gate output. This bound is 0 or 1 depending upon whether the output values before and
after transients, i.e., $IV$ and $FV$, are same or different. This has been used as the condition for minimum glitch power design [1]. When there are split ambiguity intervals, we can obtain a tighter lower bound.

**Theorem 3:** The minimum number of transitions is higher of the two lower bounds [2]:

$$mintran = \text{maximum}(Ns, Ndet) \quad (9)$$

where $Ns$ is the number of transitions required by the steady-state signal values and $Ndet$ is that needed to produce deterministic signal values separating any non-overlapping ambiguity intervals.

**Proof:** First lower bound ($Ns$): This is obtained from the steady-state values, without considering any details of the ambiguity interval. Logic changes 0→0, 1→1 need not have any transition and 0→1, 1→0 must have at least one transition. Provably, $Ns = 1$ if $IV \neq FV$, and $Ns = 0$ if $IV = FV$ [1].

Second lower bound ($Ndet$): Definite transitions can occur in the output ambiguity interval is the number of deterministic signal changes that occur within the ambiguity interval and such that signal changes are spaced at time intervals greater than or equal to the inertial delay of the gate.

The effect of the second lower bound can be seen in the example of Figure 4. There are at least two essential signal changes that must occur within the output ambiguity interval. Thus, there will always be a hazard in the output as long as:

$$(EAsv - LSdv) \geq \text{maxdel} \quad (10)$$

where $\text{maxdel}$ is the maximum delay of the gate producing the transient.

In this case $mintran$ is not 0, as given by the steady state first lower bound, but is 2. Detailed analysis of split ambiguity intervals is possible with the help of transient output functions [17] or timed Boolean functions [15, 18, 24]. In general, the value of $Ndet$ can be higher than 2 depending on how many ambiguity and deterministic intervals are produced and their widths with respect to the gate delay $d$.

4. Results

These results are taken from a recent implementation of a power analysis tool [2, 3] based on the algorithms discussed in the preceding section. The data consists of power analysis of ISCAS85 benchmark circuits for 1000 random vectors. The circuits were implemented using TSMC025 2.5V CMOS library. Process variation can be modeled by assuming 15% intra-die and 5% inter-die variation [13]. For illustrative purposes, standard size gate delay of 10ps and wire-load delay model was used to determine the nominal gate delays from which bounds (mindel, maxdel) were obtained by assuming a ±20% variation. It should be noted that any kind of variation modeled is suitable for our method. Node capacitances for the circuits were obtained from the Spice modeling files.

Table 1 gives two sets of data. The first set (columns 2-5) is from a Monte Carlo event-driven simulation. These results are for 1000 circuit samples. Each gate in a sample circuit was assigned delay using a random number uniformly distributed in its (mindel, maxdel) range. From the simulation of 1000 random vectors, we have listed the energy consumption (pJ) for two vectors consuming the least and the most energy, and the average for all 1000 vectors. We observe that the bounded delay analysis always gives lower minimum energy and higher maximum energy. This is expected. As we simulate more vectors, the event-driven numbers drop on the minimum side and increase on the maximum side. Besides, to take the variability into account, the event-driven simulator will have to be used in a Monte Carlo experiment mode, which will take much more computing resources. The CPU times in Table 1 are for a Sun Sparc Ultra 10 with 4GB shared memory system.

The histogram in Figure 5 shows Monte Carlo simulation of circuit c880 for 1000 random vectors (999 vector-pairs) were simulated for 1000 circuit samples. Each vector-pair was simulated for 1000 random vector-pairs. The details of technology, process variation, and computing platform are discussed in Section 4. For a vector period of 1000 ns minimum power was 1.424 mW and maximum power was 11.598 mW. Monte Carlo simulation runs took 262.75 CPU s. Using the same variability, our analysis obtained a bound (1.35 mW, 11.89 mW) in just 0.3 CPU s. Considering that c880 is a small circuit and the impact of process variation on power continues to assume greater importance, this computational efficiency is a strong motivation for the
Table 1. Per vector energy consumption in picojoule in benchmark circuits for 1000 random vectors by Monte Carlo simulation of 1000 sample circuits and bounded delay analysis. (*Sun UltraSparc 10 with 4GB Shared Memory.)

<table>
<thead>
<tr>
<th>Circuit name</th>
<th>Monte Carlo simulation - picojoule per vector</th>
<th>Bounded delay analysis - picojoule per vector</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Minimum</td>
<td>Maximum</td>
</tr>
<tr>
<td>c880</td>
<td>1.086</td>
<td>10.847</td>
</tr>
<tr>
<td>c1355</td>
<td>3.606</td>
<td>13.577</td>
</tr>
<tr>
<td>c1908</td>
<td>4.870</td>
<td>29.470</td>
</tr>
<tr>
<td>c2670</td>
<td>6.036</td>
<td>66.660</td>
</tr>
<tr>
<td>c3540</td>
<td>29.810</td>
<td>91.100</td>
</tr>
<tr>
<td>c5315</td>
<td>45.360</td>
<td>194.860</td>
</tr>
<tr>
<td>c6288</td>
<td>6.036</td>
<td>66.660</td>
</tr>
<tr>
<td>c7552</td>
<td>35.050</td>
<td>146.120</td>
</tr>
</tbody>
</table>

present work.

The next result compares the accuracy of the bounded delay analysis with Monte Carlo analysis using event-driven simulation. As described before, 100 circuit samples were generated for c880. Each sample circuit was simulated by the event-driven simulator for the same set of 100 random vectors. For each vector-pair, we obtained the minimum, average and maximum power. The minimum and maximum numbers are shown in Figure 6 against the corresponding vector-pair results from the bounded delay analysis. $R^2$ shown on the regression graphs is the coefficient of determination from Microsoft Excel, whose ideal fit value is 1.0. We believe the vector-by-vector bounded delay analysis can give us information not just on minimum and maximum power but on the power distribution as well when we use statistical techniques [4]. This would be worth considering in the future.

Figure 7 shows the transition statistics for gate 1407 in c2670 [2]. This is a high activity gate of the circuit, which had up to 8 transitions on some vector-pairs. The left histogram (a) shows that the number of transitions on this gate for one vector-pair applied to 100 sample circuits. The delay bounds of the gate were (7ps, 12ps). So, in each sample, its delay was randomly selected from this range. The transitions on the gate range between 0 and 8. Bounded delay analysis gave $mintran = 0$ and $maxtran = 8$. Leaving all other gate delays as before, when the delay bounds of 1407 was changed to (11ps, 33ps), the analysis computed $mintran = 0$ and $maxtran = 8$. Histograms obtained by Monte Carlo simulation.

Figure 7. Bounded delay analysis of high-activity gate 1407 in c2670 for a random vector-pair: (a) delay bounds (7ps, 12ps), $mintran = 0$, $maxtran = 8$, (b) delay bounds (11ps, 33ps), $mintran = 0$, $maxtran = 8$. Histograms obtained by Monte Carlo simulation.
and maxtran = 4. The corresponding histogram from Monte Carlo simulation is shown in Figure 7(b).

5. Conclusion

We have used a bounded delay model for gates and developed new algorithms to determine bounds on transition intervals and the number transitions for signals. This analysis has a linear-time complexity in number of gates and it leads to an efficient alternative to the Monte Carlo analysis. The power estimation can be done using zero-delay simulation. Our expectation for the future is to consider process variation in leakage as well. Besides, node capacitances that are considered fixed here, can also have process-dependent variation.

References


