Computation of Bus Current Variance for Reliability Estimation of VLSI Circuits*

Farid Najm†, Ibrahim Hajj

Coordinated Science Laboratory
University of Illinois at Urbana-Champaign
Urbana, Illinois 61801

Ping Yang

VLSI Design Laboratory
Texas Instruments Inc.
Dallas, Texas 75265

Abstract

This paper deals with the estimation of the median time-to-failure (MTF) due to electromigration in the power and ground busses of CMOS VLSI circuits. In our previous work [3, 4], we presented a novel technique for MTF estimation based on a *stochastic current waveform* model. In [6], we argued that including the *variance waveform* of the current, in addition to its *expected waveform* derived in [3, 4], would further improve the accuracy of the MTF estimate. In this paper, we present a novel technique for deriving the variance waveform for CMOS circuits. Using this technique, we establish the importance of the variance waveform by showing that its contribution to the MTF estimate can be in the range of 100% to 200% relative to that of the expected waveform. The technique has been built into the probabilistic simulator CREST [3, 4], and has shown good agreement with SPICE, as well as excellent speedup.

*This work was supported by Texas Instruments Inc. and the US Air Force Rome Air Development Center.

†Dr. Najm is now with the VLSI Design Lab., Texas Instruments Inc. MS-369, Dallas, TX. 75265.

IEEE International Conference on Computer-Aided Design, 1989.

1 Introduction

Reliability is a major concern in integrated circuit design. As higher levels of integration are used, metal line width and line separation will continue to decrease, thereby increasing a chip's susceptibility to failures resulting from line shorts or opens. Thus the importance of reliability can only increase in the future, and circuits should be designed with reliability in mind.

While the results to be presented below can be used to study a variety of reliability problems, we will illustrate their utility by focusing on the problem of Electromigration (EM). EM is a major reliability problem [1, 2] caused by the transport of atoms in a metal line due to the electron flow. Under persistent current stress, EM can cause deformations of the metal lines which may result in shorts or open circuits. The failure rate due to EM depends on the current density in the metal lines and is usually expressed as a median time-to-failure (MTF). There is a need for CAD tools that can predict the susceptibility of a given design to EM failures.

In [3, 4] we presented a novel technique for MTF estimation, based on a *stochastic current waveform* model. The implementation of this technique in the program CREST (CuRrent ESTimator) has proven to be very effective both in terms of accuracy and speed. We focus on the power and ground busses, and derive currents for them to be used for MTF estimation. In the interest of clarity, we will review some of the basic concepts behind this approach. The reader is referred to [7] for a more detailed description.

The argument presented in [7] is that the correct current waveform to be used for MTF estimation is one that combines the effects of all possible input waveforms. By considering the set of logical waveforms allowed at the circuit inputs as a *probability space* [5], the current in any branch of the bus becomes a *stochastic process*. CREST derives the *expected (*or *mean) waveform* (not a time-average) of this process, which we call an *expected current waveform*, $E[i(t)]$. This is a waveform whose value at a given time is the weighted average of all possible current values at that time. CREST uses *statistical* information about the inputs to directly derive the expected current waveform. The resulting methodology is what we call a *probabilistic simulation* of the circuit. In general, it can be slightly more time consuming than a single standard timing simulation run, but it needs to be applied only once, resulting in significant speedup.

In [6], we derived the exact relationship between the MTF and the statistics of the stochastic current and showed that a more accurate MTF estimate would be obtained if the *variance waveform*, $V[i(t)]$, of the stochastic current is included as well. We also discussed how the variances of the bus branch currents can be obtained from those of gate currents. $V[i(t)]$ is a waveform whose value at a given time is the variance of the current values at that time, $V[i(t)] = E\left[(i(t) - E[i(t)])^2\right]$. It is an indication of the spread of the current values at every time point around their expected value.

In this paper we present a novel technique for deriving the variances of the individual gate currents in CMOS circuits. As a result, the requirements for accurate MTF estimation are complete : we know both how to derive the expected and variance waveforms, and how to use them to accurately estimate the MTF. Using this technique, we establish the importance of including the variance waveform in addition to the expected waveform. This is done by presenting a number of examples where the contribution of the variance waveform to the MTF estimate is in the range of 100% to 200% relative to that of the expected waveform. This technique has been built into the probabilistic simulator CREST, and has shown good agreement with SPICE. The technique is dramatically faster than traditional approaches : we demonstrate a speedup of over 11000X on a 648-transistor CMOS parallel multiplier circuit.

This paper is organized as follows. Section 2 reviews some of the specifics of the CREST approach to lay the groundwork for the remaining sections. Section 3 describes the derivation of the gate current variances. Section 4 discusses implementation issues and shows the results of several CREST runs, and section 5 summarizes the results and draws some conclusions.

2 Basic Concepts

As explained in [3, 7], the probabilistic simulation approach follows an event-driven simulation strategy. *Probability waveforms*, which represent a large number of possible input logical waveforms, are propagated through the circuit as a sequence of *probabilistic events*. A probabilistic event embodies a number of possible logical transitions at that time instant.

Whenever a gate is simulated, the events at its inputs are used to derive an event at its output, and an *expected current pulse*, to be added to the global expected current waveform. This pulse is considered to be a *triangular pulse*. It starts with a *peak* value at the time of transition and decays linearly to zero after a time interval called the *time span*. The variance waveform can be derived with little modification to the overall simulation strategy. Whenever an expected pulse is derived for a gate, a *variance pulse* will be derived as well. The next section will describe how this pulse is derived.

Fig. 1 shows a generic CMOS gate structure. The p-block or p-part (n-block or n-part) of a gate will be used to refer to the p (n)-channel transistor mesh between its output node and the power supply (ground). A gate will be assumed to have *independent* inputs. While this may be true at the primary circuit inputs, it may not be true in general. However, the general case can be properly handled using the concept of a supergate, as described in [4], with the independent-inputs-gate-solver used as a subroutine.

We will focus on the *charging* current component. The output node capacitance is split into two lumped capacitors C_p to V_{dd} and C_n to V_{ss}. Similarly, each internal gate node n_i has two capacitances C_{in} and C_{ip}. The values of these capacitances are derived from the circuit description and the transistor model parameters. On a low-to-high transition, the currents flowing through C_n and C_p at the output node are i_{p1} and i_{p2}, respectively, as shown in the figure. The corresponding i_{n1} and i_{n2} for a high-to-low transition are also shown. The currents i_{p2} and i_{n2} are discharging currents that redistribute locally, and we are interested in $i = i_{p1} + i_{n1}$. Of course these currents are associated with the output node only, and the total gate current i_{tot} will be larger than i. However, the output current plays a central role in the derivation of $V[i_{tot}(t)]$.

3 Derivation of the Variance

The variance waveforms for the gate total and output currents will be modeled by triangular pulses $V[i_{tot}(t)]$ and $V[i(t)]$, respectively, with peak values of $V[I_{tot}]$ and $V[I]$. If an event occurs at the gate input at time t, then we denote by t^- and t^+ the instances of time immediately before and after the event, respectively. Focusing for now on the output current pulse, its variance waveform starts with a peak of $V[I] = V[i(t^+)]$ at time t and decays linearly to zero at time $t + \tau$. Since $V[I] = E[I^2] - E[I]^2$ [5], and since CREST already derives the expected pulse peak ($E[I]$), then it will be enough to derive $E[I^2]$.

Let $i_p = i_{p1} + i_{p2}$ and $i_n = i_{n1} + i_{n2}$. It's easy to verify that $i_{p1} = i_p \times C_n/(C_p + C_n)$, and $i_{n1} = i_n \times C_p/(C_p + C_n)$. Therefore :

$$E[i^2(t)] = E[i_p^2(t)]\left(\frac{C_n}{C_p + C_n}\right)^2 + E[i_n^2(t)]\left(\frac{C_p}{C_p + C_n}\right)^2$$

The term containing $E[i_p(t) i_n(t)]$ is omitted (it is zero) since at least one of the charging currents is zero at any given time. In particular, the value at the peak is :

$$E[I^2(t)] = E[I_p^2(t)] \times \left(\frac{C_n}{C_p + C_n}\right)^2 + E[I_n^2(t)] \times \left(\frac{C_p}{C_p + C_n}\right)^2$$

The values of $E[I_p^2]$ and $E[I_n^2]$ are derived as follows. For $E[I_p^2]$, consider the p-part of the gate, and let every transistor T_k be represented by a switch of on-conductance $g_{on,k}$ [7]. Based on this switch-network model of the p-block, let $G_p(t)$ be the random conductance between the output node and V_{dd}. G_p is a function of the individual transistor random conductances g_k, where g_k is 0 if the transistor is off and $g_{on,k}$ if it is on. If an event occurs at the gate at time t, then the value of $G_p(t^+)$ and the previous state of the output node, $V_o(t^-)$, will determine I_p. Formally, we have $E[I_p^2] = E[(V_{dd} - V_o(t^-))^2 \times G_p^2(t^+)]$, which becomes :

$$E[I_p^2] = V_{dd}^2 \times E[G_p^2(t^+) \mid G_p(t^-) = 0] \times P(G_p(t^-) = 0)$$

where $P(A)$ is the probability of the event A, and $E[A \mid B]$ denotes the *conditional expected value* [5] of A given B. The formula is correct because if $G_p(t^-) = 0$ ($\neq 0$) then $V_o(t^-) = 0$ (V_{dd}). Similarly for the n-part of the gate, we get :

$$E[I_n^2] = V_{dd}^2 \times E[G_n^2(t^+) \mid G_n(t^-) = 0] \times P(G_n(t^-) = 0)$$

To derive the conditional expectations, consider a graph representation of the p-block (or n-block), where every edge in the graph is labeled with $E[g_k^2(t^+) \mid G_p(t^-) = 0]$, $E[g_k(t^+) \mid G_p(t^-) = 0]$, and the gate node probabilities of its corresponding transistors. The details of how these quantities can be derived for every transistor can be found in [7]. Then perform a *graph reduction* operation [7], which, simply stated, involves a number of series/parallel combinations and node eliminations that reduce the graph to a single edge, whose labels are the required statistics $E[G_p^2(t^+) \mid G_p(t^-) = 0]$ and $E[G_p(t^+) \mid G_p(t^-) = 0]$. Similarly for the n-block.

Having found the peak $V[I] = E[I^2] - E[I]^2$ for the output current, the time span τ will be found by first solving for the area under the $V[i(t)]$ pulse. Notice that, if $i(t)$ is a triangular pulse of height I and area q, then :

$$\int_0^\infty i^2(t)dt = \frac{2}{3}Iq$$

In this case, q is equal to the charge delivered to (or from) the output node capacitors. From this it follows that :

$$\int_0^\infty E[i^2(t)]dt = \frac{2}{3}E[Iq], \text{ and } \int_0^\infty E[i(t)]^2 dt = \frac{2}{3}E[I]E[q].$$

The second equation follows since $E[i(t)]$ is a triangular pulse of height $E[I]$ and area $E[q]$. Therefore, the variance pulse has an area :

$$\frac{V[I] \times \tau}{2} = \int_0^\infty V[i(t)]dt = \frac{2}{3}\Big(E[Iq] - E[I]E[q]\Big).$$

The value of $E[q]$ is available from equation (5) of [3], and the value of $E[Iq]$ can be written as :

$$E[Iq] = E[(I_{p1} + I_{n1}) \times (q_{p1} + q_{n1})].$$

where I_{p1} (I_{n1}) is the peak of $i_{p1}(t)$ ($i_{n1}(t)$), and q_{p1} (q_{n1}) is the charge delivered by $i_{p1}(t)$ ($i_{n1}(t)$). Since q_{p1} (q_{n1}) is equal to $V_{dd}C_n$ ($V_{dd}C_p$) if $i_{p1}(t)$ ($i_{n1}(t)$) is non-zero, and is otherwise zero, then :

$$E[Iq] = \frac{V_{dd}C_n^2}{C_p + C_n}E[I_p] + \frac{V_{dd}C_p^2}{C_p + C_n}E[I_n].$$

The time span of the gate output current variance pulse is, therefore :

$$\tau = \frac{4}{3}\left(\frac{E[Iq] - E[I]E[q]}{V[I]}\right).$$

If $i_{tot}(t)$ is the total gate current, then :

$$\int_0^\infty V[i_{tot}]dt = \frac{2}{3}\Big(E[I_{tot}q_{tot}] - E[I_{tot}]E[q_{tot}]\Big).$$

where $E[q_{tot}]$ is available as equation (7) in [3]. Unfortunately, $E[I_{tot}q_{tot}]$ does not have a simple expression as was found for $E[Iq]$ above. We have chosen to use a conservative estimate based on the following assumption : whenever a node in the p-block (n-block) is charged to V_{dd} (V_{ss}), then every other node in the p-block (n-block) is also charged to V_{dd} (V_{ss}). This assumption is true for simple gates, and overestimates the current-charge product in more complex cases. Based on this assumption, one can show [7] that :

$$E[I_{tot}q_{tot}] \approx \frac{E[q_{p,tot}]}{E[q_p]}\frac{Q_pC_nE[I_p]}{C_p+C_n} + \frac{E[q_{n,tot}]}{E[q_n]}\frac{Q_nC_pE[I_n]}{C_p+C_n},$$

where $E[q_p]$ and $E[q_n]$ are available as equations (9) & (10) in [3], $E[q_{p,tot}]$ and $E[q_{n,tot}]$ are, respectively, the first and second summations in equation (7) in [3], and

$$Q_p = \sum_{i \in P\ block} V_{dd}C_{in}, \quad Q_n = \sum_{i \in N\ block} V_{dd}C_{ip}.$$

As was assumed for the expected current pulse [3, 7], we let the time span of the gate total current variance pulse be equal to that derived for the gate output current, therefore :

$$V[I_{tot}] = \left(\frac{E[I_{tot}q_{tot}] - E[I_{tot}]E[q_{tot}]}{E[Iq] - E[I]E[q]}\right) \times V[I].$$

4 Implementation and Results

The variance calculation technique outlined above has been implemented in CREST. We present below the results of CREST runs on a variety of circuits, showing both waveform comparisons and timing performance.

We start out with a simple example, a 2-input CMOS NAND gate. The variance waveform comparison between CREST and SPICE is shown in Fig. 2. The SPICE waveform is derived by running SPICE on the NAND gate for all possible logical transitions at its inputs, deriving the expected current waveform by doing a time-point averaging of the results, and then using that to find the variance as the time-point average of $(i - E[i])^2$. Since the object of this research is to handle very large chips, and since electromigration models for ac waveforms are still controversial, it makes little sense to shoot for perfect accuracy in the current waveforms. It is more important to be able to derive in a very *short time* a waveform that matches the peak and general shape of the SPICE waveform.

Another single-gate comparison is shown in Fig. 3 for a CMOS complex gate. The comparisons for two larger circuits are shown in Fig. 4 (for an XOR circuit) and Fig. 5 (for a 54-MOSFET 2-bit ripple adder circuit).

Our final example is a much bigger and more complex circuit; it is a 648-MOSFET 4-bit parallel multiplier. This circuit is too big to make the 2^8 required SPICE simulations. We will, therefore, show two different CREST runs to demonstrate that the variance computation works well even when the heuristics introduced for handling large circuits in [4] are used. In Fig. 6 we compare a full accuracy CREST run and a heuristic CREST run in which all internal nodes of the multiplier were assumed independent. The excellent agreement reaffirms the conclusion made in [4] that as a circuit becomes larger, the correlation between its internal nodes may be safely neglected.

We next examine the importance of the variance waveform for MTF estimation. As shown in [6], the expected and variance waveforms combine to provide a J_{eff} dc current density value to represent the ac current waveform for MTF estimation. We have measured this variance contribution as the increase in J_{eff} due to the variance waveform, divided by J_{eff} using only the expected waveform. This has been tabulated in Table 1 as a percentage for all the examples presented above. It clearly establishes the importance of the variance waveform in addition to the expected waveform.

Finally, we illustrate the speed performance of CREST with the variance estimation built in. Table 1 shows the speed comparisons between CREST and SPICE for all the examples presented above. The speedup becomes much better for larger circuits (1529X for the adder and 11595X for the multiplier). In fact, the speedup should grow exponentially, because an exponential number of deterministic simulation runs are replaced by a single probabilistic simulation run. We point out the case of the multiplier circuit (the largest circuit in the table) with the heuristic CREST run (last row in Table 1); considering the excellent waveform comparison in Fig. 6 along with the dramatic speedup of 11595X in Table 1, this establishes the feasibility of solving large VLSI chips.

5 Summary and Conclusions

We have discussed the problem of estimating the median time-to-failure (MTF) due to electromigration in the power and ground busses of CMOS VLSI circuits. In our previous work [3, 4, 6], we had presented a novel technique for MTF estimation based on a *stochastic current waveform* model. In this paper we have reviewed this technique, and verified that

Table 1. Execution time comparisons. Time is in CPU seconds on a VAX-11/780; size refers to the number of transistors.

Circuit	Size	SPICE	CREST	Variance Contribution	Speedup
Nand	4	41.75	0.90	110%	46X
Complex	6	244.15	1.09	185%	224X
Xor	16	456.00	3.13	236%	146X
Adder	54	32620.42	21.33	107%	1529X
Multiplier	648	697530.88*	1871.99	219%	373X
Multiplier	648	697530.88*	60.16†	200%	11595X

*Estimated (2^8 times the cost of a typical logical SPICE run).
†Heuristic CREST run, all others are full accuracy CREST.

including the *variance waveform* of the current, in addition to the *expected waveform* derived in [3, 4], further improves the accuracy of MTF estimation. This was done by showing that the variance contribution to the MTF estimate can be in the range of 100% to 200% relative to that of the expected current waveform. We have described a novel technique for deriving the variance waveform, and its implementation in the probabilistic simulator CREST. The results of several CREST runs have been presented, and they show good waveform agreement with SPICE, as well as excellent speedups over traditional approaches - a speedup of over 11000X was demonstrated on a 648-transistor circuit.

This work proves that the expected and variance waveforms of the stochastic current model are : (1) essential to derive an accurate MTF value, and (2) can be efficiently derived using the probabilistic simulation approach. The main advantage of this approach is the ability to handle large circuits by replacing an exponential number of deterministic simulation runs with a single probabilistic simulation run. Without such a technique, analyzing the reliability of large circuits would seem to be an impossibility. The reader is referred to [7] for a more detailed description of the probabilistic simulation approach.

References

[1] F. M. d'Heurle, "Electromigration and failure in electronics : an introduction," *Proceedings of the IEEE*, vol. 59, no. 10, October 1971.

[2] J. R. Black, "Electromigration failure modes in aluminum metallization for semiconductor devices," *Proceedings of the IEEE*, vol. 57, no. 9, September 1969.

[3] R. Burch, F. Najm, P. Yang, and D. Hocevar, "Pattern-independent current estimation for reliability analysis of CMOS circuits," *25th ACM/IEEE Design Automation Conference*, Anaheim, CA, pp. 294–299, June 12–15, 1988.

[4] F. Najm, R. Burch, P. Yang, and I. Hajj, "CREST - a current estimator for CMOS circuits," *IEEE International Conference on Computer-Aided Design*, Santa Clara, CA, pp. 204–207, November 7–10, 1988.

[5] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, 2nd Edition. New York, NY: McGraw-Hill Book Co., 1984.

[6] F. Najm, I. Hajj, and P. Yang, "Electromigration median time-to-failure based on a stochastic current waveform," *IEEE International Conference on Computer Design*, Cambridge, MA, October 2–4, 1989.

[7] F. Najm, "Probabilistic simulation for reliability analysis of VLSI circuits," Ph.D. thesis, Dept. of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign, June 1989.

Figure Captions

Figure 1: A generic CMOS gate structure.

Figure 2: CREST variance pulse result for a 2-input CMOS nand gate, compared to SPICE.

Figure 3: Variance results for a 3-input complex CMOS gate (inset).

Figure 4: Variance results for a 16-MOSFET exclusive-or (xor) CMOS circuit.

Figure 5: Variance results for a 54-MOSFET 2-bit ripple adder CMOS circuit.

Figure 6: Variance results for a 648-MOSFET 4-bit parallel multiplier CMOS circuit.