Analyzing Reconvergent Fanouts in Gate Delay Fault Simulation

Hillary Grimes and Vishwani D. Agrawal
Dept. of ECE, Auburn University
Auburn, AL 36849
grimehh@auburn.edu, vagrawal@eng.auburn.edu

Abstract - *To determine the quality of gate delay tests, Min/Max delay fault simulation must determine the detectable sizes of faults. In conventional Min/Max timing simulation, correlations at the inputs of reconvergent gates are ignored. This paper shows how correlation information can be used when fanouts reconverge to produce more accurate results.*

1. Introduction

Delay testing ensures that a manufactured VLSI device meets its timing constraints. Two types of delay fault models are common, the path delay fault model and the gate delay fault model [1, 2, 3]. In this work, we are using the gate delay fault model which assumes that a delay fault is lumped at a single gate. Due to process variations in today's manufactured devices, circuit delays need to be modeled as imprecise delays rather than a simple nominal delay value. One way to model imprecise gate delays is with Min/Max delay simulation, where a gate's delay is assumed to be somewhere between some minimum and maximum value [5, 6, 7, 8, 9, 10]. Given a set of vectors, Min/Max delay fault simulation is used to determine the quality the set of vectors provides for testing delay faults in the circuit [1, 3, 4].

In order to detect a gate delay fault, the test must both place a transition at the fault site and propagate its effect to an observation point. Therefore, testing a delay fault requires 2 vectors: the first to set the initial value of the transition at the fault site, and the second to both place the final value at the site and propagate its effect to an observation point. For a rising (falling) transition, the first vector places a logic 0 (logic 1) at the fault site, and the second vector is a stuck-at 0 (stuck-at 1) test for the same fault site [2, 3].

In order to determine the quality of a set of tests, gate delay fault simulation determines both how many faults are detected and their minimum size detectable. The minimum size detectable is the minimum faulty delay that must be present for the test to detect the fault [2, 3]. Given two tests that detect the same fault, the test that detects the fault with a smaller size is considered a higher quality test for that fault.

When reconvergent fanouts are present in a circuit, the signals at the inputs of reconvergent gates are correlated [1, 4, 9, 10]. In conventional Min/Max timing analysis [2, 3, 5, 8, 9, 10], these correlations are ignored. In this work, we are showing that ignoring these correlations can produce pessimistic results. Section 2 of this paper describes Min/Max delay simulation and signal correlations at the inputs of reconvergent gates. Section 3 describes the gate delay fault model used, and how correlations at the inputs to reconvergent gates can be used to improve accuracy in the fault-free timing analysis. Section 4 describes how these correlations can be used to calculate a more accurate detection threshold. Section 5 describes test quality and detection gaps. Section 6 shows experimental results for ISCAS85 benchmark circuits, and Section 7 concludes the paper.

2. Motivation

Figure 1 shows an example to illustrate bounded delay simulation and how correlations at inputs of reconvergent gates can be used to provide more accurate simulation results. Minimum and maximum gate delays are shown beside each gate and simulation waveforms are shown below the circuit. The output signal waveform at gate C is a logic 1 before time t=1, and logic 0 after time t=3. Due to process variations, the signal can

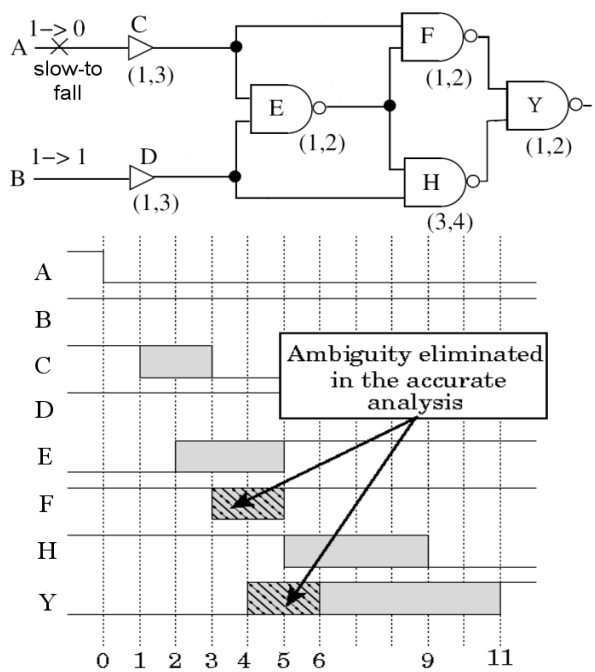

Figure 1: Min/Max Simulation Waveforms.

This research is supported in part by the National Science Foundation Grant CNS-0708962.

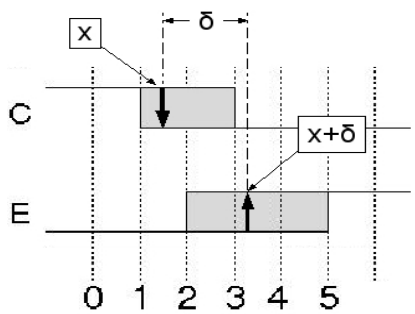

Figure 2: Correlated Waveforms at the Inputs to F.

change anytime between times 1 & 3. Similarly the signal at the output of gate E changes from 0 to 1 sometime between times 2 & 5.

The fanout at the output of gate C in Figure 1 reconverges at gate F. Figure 2 illustrates the correlation of both inputs to gate F (signals C & D). Suppose signal C changes at time x, which is somewhere between times 1 & 3. Since gate C's output is an input to gate E, the signal at E follows the signal at C in time, and δ cannot be smaller than the minimum delay of gate E, which is 1. Therefore, the output of E cannot change before time (x + min_delay(E)), or (x+1).

The inputs to gate F in Figure 1 are correlated and the input from C (top input) transitions to a dominating value before the input from E (bottom input) transitions away from a dominating value. Therefore, there is always a dominating value on at least one input to gate F. Conventional Min/Max timing analysis produces an ambiguity region at the output of F between times 3 & 5, which can never occur in an actual circuit. Propagating its effect to the output at gate Y would produce extra ambiguity between times 4 to 6. Removing, or suppressing, the erroneously produced hazard results in the signal at output Y transitioning from a 0 to a 1 at some time between 6 & 11, instead of times 4 & 11 from conventional timing analysis.

3. Gate Delay Fault Model

The gate delay fault model used in this work is based on the model used in [2, 3], which assumes delays are lumped at gates. Delays for each gate are specified by an upper (max) and lower (min) bound. A test consists of two vectors V1 and V2. Vector V1 is assumed to have stabilized before V2 is applied, and time t=0 is the time at which vector V2 is applied to the circuit's inputs [1, 2, 3].

Each gate has an initial and final value. The initial value (IV) is the logic value for that gate after vector V1 is applied, and the final value (FV) is the logic value after vector V2 is applied. Each gate also has two timing values, EA and LS. EA is the earliest arrival time for the output of the gate after V2 is applied, and LS is the latest stabilization time for the gate's output after V2 is applied. A gate's output is at its initial value before time EA, and at its final value after time LS. Between time EA and LS, the gate has an unknown (X) value [1, 2, 3]. Recall the discussion of Figure 1 in Section 2. Gate C has an initial value of 1, a final value of 0, an EA value of 1, and an LS value of 3, as shown in the waveform.

EA and LS provide timing information for the fault-free circuit. For gate F in Figure 1, conventional Min/Max timing simulation would calculate EA(F) and LS(F) to be 3 and 5. However, due to correlations at the inputs to gate F, the output is stable and the correct fault-free timing values should be:

$$EA(F) = \infty$$
$$LS(F) = -\infty$$

To propagate correlation information for reconvergent fanout analysis, we propagate hazard lists at each gate during fault-free timing simulation. Each element in the hazard list of a gate G contains a gate id of a fanout point effecting that gate, the minimum delay from that fanout point to G and a maximum delay from that fanout point to G. If inputs at a reconvergent gate are correlated due to a common fanout point, a hazard list element for that fanout point appears at all correlated inputs of G [1, 4].

When EA and LS are evaluated at a gate G, the hazard list at G is evaluated. Hazard lists at the inputs to G are used to determine the hazard list at the output. A hazard list element is added to G for every fanout element found at the inputs (i) to G, and the min and max delays of hazard list elements are updated:

$$d(G,f) = min\{d(i,f)\}+min\_delay(G)$$
$$D(G,f) = max\{D(i,f)\}+max\_delay(G)$$

The quantity d(G,f) is the minimum delay from fanout point f to gate G, and D(G,f) is the maximum delay from fanout f to gate G. If a fanout point, f, appears in the hazard list of more than one input to G, that fanout is reconverging at gate G, and those inputs are correlated. If the correlated inputs to G are both transitioning to a dominating value and transitioning away from a dominating value for gate G, the following quantities are evaluated:

$$min_{DV}(f) = max\{d(M,f)\}$$
$$max_{SV}(f) = max\{D(m,f)\}$$

where M is an input to G that is transitioning away from a dominating value and m is an input to G that is transitioning to a dominating value. If $min_{DV}(f) \geq max_{SV}(f)$, then the correlated inputs to G are such that there is always a dominating value at some input to G. Hazard suppression occurs in which the hazard list at the output of G is set to NULL, and the following fault-free

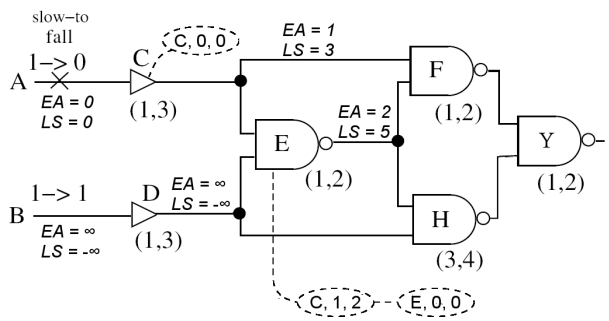

Figure 3: EA and LS Calculations.

signal	EA	LS	FPV	ρ(s)	RTa	RTb
A	0	0	1	0	-∞	0
B	∞	-∞	1	0	-∞	∞
C	1	3	1	0	-∞	1
D	∞	-∞	1	0	-∞	∞
E	2	5	0	0	-∞	2
F	3	5	1	0	-∞	3
H	5	9	1	0	-∞	5
Y	4	11	0	0	-∞	4

Figure 4: Detection Threshold Calculations.

timing values are used:

$$EA(G) = \infty$$
$$LS(G) = -\infty$$

If no fanout point, f, results in $min_{DV}(f) \geq max_{SV}(f)$, hazard suppression does not occur, and EA and LS are evaluated as in [2].

Figure 3 shows hazard lists and fault-free timing values, EA and LS, for the circuit of Figure 1, when gate F is about to be evaluated. Since an element for fanout point C appears in both hazard lists at the inputs to F, $min_{DV}(C)$ and $max_{SV}(C)$ are evaluated as:

$$min_{DV}(C) = 1$$
$$max_{SV}(C) = 0$$

Since $min_{DV}(C) \geq max_{SV}(C)$, the hazard is suppressed and the output of gate F is stable. The hazard list at F is then set to NULL, and EA and LS are set to ∞ and -∞.

If a delay fault is present, the output of the faulty gate is delayed by a time, δ. This faulty gate's output is "stuck-at" its initial value for δ time units after time EA(G). Therefore, the fault propagating value (FPV) at a fault site, gate G, is the initial value of that gate:

$$FPV(G) = IV(G)$$

At a gate, G, outside the cone of influence of the faulty gate, the waveforms are unaffected by the fault, so:

$$FPV(G) = FV(G)$$

At a gate G inside the cone of influence, the FPV is evaluated using boolean logic on the FPVs of its inputs [1, 2, 3]. Figure 4 shows the values for the circuit from Figure 1. The FPVs are shown in the table for a slow to fall delay fault on input A. Propagating the FPVs propagates the fault effects through the circuit, and can be used to determine whether or not a fault is detected by the test. If at an output, the FPV differs from the output's final value, the fault is considered detected by that test, but we are not only interested in whether or not the fault is detected. We also need to determine what size fault is detected [2, 3]. For this, we propagate three additional reference quantities, shown in the last three columns of Figure 4, to determine the detection threshold.

4. Detection Threshold

The detection threshold for a gate is the minimum fault size detectable by a given test. A delay fault with a size greater than the detection threshold is always detectable by the test [2, 3]. Suppose the output in Figure 1 is sampled at time Tc = 12. In order for a fault to be detected, its size needs to be large enough to shift the output waveform to the right such that an incorrect value is sampled at time 12.

To determine the detection threshold, we propagate three reference quantities along with fault propagation values as in [2], which propagate timing information about the faulty waveforms. The logic value at a gate G for a given delay fault of size δ is at FPV(G) between times RTa(G) and RTb(G)+δ, provided δ > ρ(G). RTa(G) and RTb(G) are two reference times, and ρ(G) is the reference size [1, 2, 3]. If the size of the fault is less than the reference size, the fault has no effect at the output of G. Hazard lists are also propagated along with these three reference quantities, but unlike those used in the fault-free timing analysis, these lists are only propagated in the downcone of the fault site.

For a gate G, at the fault site, FPV(G) is at IV(G) for the time between -∞ and EA(G):

$$\rho(G) = 0$$
$$RTa(G) = -\infty$$
$$RTb(G) = EA(G)$$

If the fault site is also a fanout stem, then a hazard list is created at G.

For a gate G, outside the cone of influence of the fault, FPV(G) is at FV(G) between the times LS(G) to ∞:

$$\rho(G) = 0$$
$$RTa(G) = LS(G)$$
$$RTb(G) = \infty$$

Reference quantities for a gate, G, inside the cone of influence of the fault site are evaluated as in [2], along with hazard lists. The hazard lists at the inputs of G are used to determine the hazard list at the output. Since these lists contain information about signal correlations, they are used to modify the reference quantities whenever a hazard is suppressed at a gate [1, 4].

For a gate G, inside the cone of influence of the fault site, if all inputs (i) have a sensitizing FPV, RTa(G) and RTb(G) are calculated as follows:

$$RTa(G) = max\{RTa(i)\}+max\_delay(G)$$
$$RTb(G) = min\{RTb(i)\}+min\_delay(G)$$

To calculate the reference size, ρ(G), we first calculate the following value, ω:

$$\omega = max\{0, max\{RTa(i)\}+max\_delay(G)-min\{RTb(i)\}\}$$

The reference size for gate G is then:

$$\rho(G) = max\{max\{\rho(i)\}, \omega\}$$

The hazard list at the output of G is updated by adding an element for every fanout element found at the inputs (i) of G, and updating min and max delays of hazard list elements, where d(G,f) is the min delay from fanout point f to gate G, and D(G,f) is the max delay from fanout point f to gate G:

$$d(G,f) = min\{d(i,f)\}+min\_delay(G)$$
$$D(G,f) = max\{D(i,f)\}+max\_delay(G)$$

For a gate, G, inside the cone of influence of the fault site, if inputs to G have both sensitizing and dominating FPVs, the following quantities are evaluated for all fanout points (f) of the hazard lists at inputs to G:

$$min_{DV}(f) = max\{d(M,f)\}$$
$$max_{SV}(f) = max\{D(m,f)\}$$

where M is an input to G with a dominating FPV and m is an input to G with a sensitizing FPV. Notice that these quantities are similar to those calculated during fault-free timing simulation discussed in Section 3. The main difference here is that the input FPVs are used instead of initial and final values. If no fanout point, f, results in $min_{DV}(f) \geq max_{SV}(f)$, the reference quantities at the output of G are calculated using an input, i, with a dominating FPV:

$$RTa(G) = RTa(i)+max\_delay(G)$$
$$RTb(G) = RTb(i)+min\_delay(G)$$
$$\rho(G) = max\{\rho(i)\}, Rta(i)+max\_delay(G)-RTb(i)\}$$

The hazard list at G is updated similar to before, adding an element to G's hazard list for every fanout point appearing at G's inputs:

$$d(G,f) = min\{d(i,f)\}+min\_delay(G)$$
$$D(G,f) = max\{D(i,f)\}+max\_delay(G)$$

If, however, a fanout point at the inputs to G results in $min_{DV}(f) \geq max_{SV}(f)$, then no hazard can occur at the output of G. Hazard suppression occurs in which the hazard list at the output of G is set to NULL, and the following reference values are used [1]:

$$\rho(G) = 0$$
$$RTa(G) = -\infty$$
$$RTb(G) = \infty$$

The detection threshold for a delay fault f (DT(f)) at output z is:

$$DT(f) = max\{\rho(z), Tc(z)-RTb(z)\}$$

where Tc is the sample time of the output. The minimum size fault detectable is then the minimum detection threshold for all outputs [2, 3].

Figure 4 shows the reference quantity calculations for the example of Figure 1 for a slow to fall fault on input A. If the output Y were sampled at Tc=12, then the detection threshold would be:

$$DT(A) = Tc(Y)-RTb(Y)$$
$$DT(A) = 12-4 = 8$$

if signal correlations were ignored. Figure 5 shows the corrected detection threshold calculations and hazard list entries for the same example. Here, RTb(Y) is correctly evaluated as 6 instead of 4. The result is a less pessimistic detection threshold calculation:

$$DT(A) = 12-6 = 6$$

signal	FPV	ρ	RTa	RTb	Hazard List
A	1	0	-∞	0	Ø
B	1	0	-∞	∞	Ø
C	1	0	-∞	1	(C,0,0)
D	1	0	-∞	∞	Ø
E	0	0	-∞	2	(C,1,2),(E,0,0)
F	1	0	-∞	∞	Ø
H	1	0	-∞	5	(C,4,6),(E,3,4)
Y	0	0	-∞	6	(C,5,8),(E,4,6)

Figure 5: Corrected Detection Threshold Calculations.

5. Test Quality and Detection Gap

Given a set of vectors, Min/Max delay fault simulation is used to determine the quality the set of vectors provides for testing delay faults in the circuit [1,

Table 1: Largest Output EA and LS Values for 10,000 Random Vectors.

	Without Reconvergent Fanout Analysis		With Reconvergent Fanout Analysis		
	Largest EA	Largest LS	Largest EA	Largest LS	EA Difference
c3540	96.0	204.0	121.6	196.8	25.6
c5315	76.8	204.0	91.2	194.4	14.4
c6288	158.4	576.0	236.8	504.0	78.4
c7552	91.2	204.0	104.0	201.6	12.8

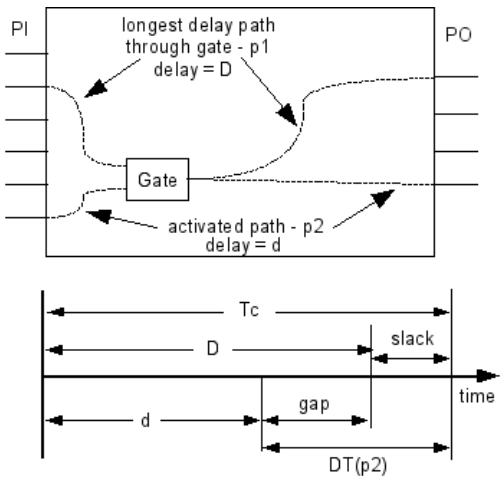

Figure 6: Detection Gap for a Gate Tested Through a Non-critical Path.

3, 4]. In order to determine the quality of a set of tests, gate delay fault simulation determines both how many faults are detected and their detection threshold [2, 3]. We can relate the detection threshold to the slack at the faulty gate by calculating a detection gap. The detection gap for a detected gate delay fault (gap(G)) is defined as [2]:

$$gap(G) = DT(G) - slack(G)$$

where slack(G) is the sum of all minimum gate delays along the longest path through gate G. If a vector pair detects a fault at G such that gap(G) = 0, the smallest possible delay fault has been detected. If gap(G) > 0, there is a possibility that there exists a better test to detect a gate delay fault at G with a smaller detection threshold [2].

The smaller the detection gaps are for a set of vectors, the better quality that set provides. Suppose a gate delay fault is detected at a gate by activating a path p2, which is shorter than the longest path (p1) through that gate, as shown in Figure 6. Path delays D and d in Figure 6 are the sums of all minimum delays of gates along paths p1 and p2 that pass through the gate, and Tc ≥ D ≥ d. The slack is Tc-D, where Tc is the clock period and D is the delay of the longest path, p1. If the fault is detected through the shorter path, p2, then the detection gap is DT(p2) – slack, which is larger than 0 by the amount D – d. Ideally we would like to detect the smallest gate delay fault possible, so a better test would detect the delay fault through path p1. If detected through p1, the detection gap would be DT(p1) – slack, which is 0 because DT(p1) = slack.

6. Results

Results on ISCAS85 combinational benchmark circuits for a set of 10,000 random vectors are shown in Table 1 and Table 2. A simple wireload delay model is used for Min/Max delays, where the delay of a gate is between {(nominal * #fanout) - tolerance} and {(nominal * #fanout) + tolerance}. For these results, the nominal value of 3.5 time units and a tolerance of 14% was used. The sample period Tc was chosen to be 1 + (longest path delay).

Results on fault-free timing analysis (EA and LS), discussed in Section 3, for a few larger combinational benchmark circuits for a set of 10,000 random vectors are shown in Table 1. These results show the difference seen at circuit outputs when reconvergent fanout analysis is used. The second and third columns show the largest EA and largest LS value at a circuit's output for all vectors without using reconvergent fanout analysis. Columns four and five show the same results when signal correlations are used, and hazards that cannot occur are suppressed. The last two columns show the difference in these calculations between those with reconvergent fanout analysis and those without reconvergent fanout analysis. The result of using information about correlated signals at reconvergent gates during simulation results in larger EA values and smaller LS values at outputs, and is more apparent for circuits that contain a large number of reconvergent fanout such as in the multiplier circuit c6288.

Table 2 shows the results when reconvergent fanout analysis is used during both fault-free timing analysis and detection threshold calculation. Using reconvergent fanout analysis on all gates of the fault-free circuit in addition to gates inside the cone of influence during fault simulation is the primary difference between this work and the work presented in [1]. The detection gap, which is defined in [2] and explained in Section 5, is used to display the data in Table 2. The smaller the detection gaps are for detected gate delay faults, the better quality the vector set provides for gate delay testing. Columns two and three of Table 2 show the detection gap results

Table 2: Detection Gap Results for 10,000 Random Vectors.

	Without Reconvergent Fanout Analysis		With Reconvergent Fanout Analysis	
ISCAS85	Average Detection Gap	Faults Detected with Gap ≤ 3.5	Average Detection Gap	Faults Detected with Gap ≤ 3.5
c432	110.4	7.35%	108.9	7.08%
c499	51.7	4.91%	44.0	12.85%
c880	16.4	48.41%	12.9	48.86%
c1355	50.8	4.80%	42.2	13.62%
c1908	55.2	21.70%	47.1	25.10%
c2670	41.8	34.25%	36.0	36.54%
c3540	50.4	32.60%	44.0	33.19%
c5315	21.7	55.72%	6.1	57.31%
c7552	39.4	13.43%	22.5	22.83%

without using reconvergent fanout analysis. Column two shows the average detection gap for all detected gate delay faults, and column three shows the fault coverage of detected gate delay faults with a detection gap that is less than or equal to the nominal gate delay used, which is 3.5. This is to allow faults to be counted as detected if they are either detected through the longest path through the gate, or if they are detected through a path which is less than the longest path by only one gate delay. Columns four and five of Table 2 show detection gap results when reconvergent fanout analysis was used during both fault-free timing analysis and detection threshold calculation. Column four shows the average detection threshold for all detected faults and Column 5 shows the fault coverage of detected faults with a detection gap less than or equal to 3.5. The data shown in Table 2 illustrates the pessimism when signal correlations are ignored in determining test quality. When reconvergent fanout analysis is used, the average detection gap is smaller and more faults are detected with smaller detection gaps.

7. Conclusion

The use of correlation information when simulating a reconvergent gate allows for a more accurate simulation when Min/Max delays are used to model imprecise circuit delays. When calculating the detection thresholds during gate delay fault simulation, signal correlations due to reconvergent fanouts should be used during fault-free timing calculations. This is because the reference quantities propagated in the cone of influence during fault simulation are initialized using EA and LS from the fault-free timing analysis. Results in Section 5 show that the use of signal correlations in both fault-free timing calculation and detection threshold calculation (which is used to determine the detection gap) reduce the pessimism of both conventional Min/Max timing analysis and gate delay fault simulation.

Acknowledgment:

Technical contributions of Dr. Soumitra Bose to this work are gratefully acknowledged.

8. References

[1] S. Bose, H. Grimes, and V. D. Agrawal, "Delay Fault Simulation with Bounded Gate Delay Model," *Proc. IEEE International Test Conference*, paper 26.3, 2007.

[2] V. S. Iyengar, B. K. Rosen, and J. A. Waicukauski, "On Computing the Sizes of Detected Delay Faults," *IEEE Trans. CAD*, vol. 9, pp. 299-312, Mar. 1990.

[3] A. K. Pramanick and S. M. Reddy, "On the Fault Coverage of Gate Delay Fault Detecting Tests," *IEEE Trans. Computer-Aided Design*, vol. 16, no. 1, pp. 78-94, Jan. 1997.

[4] S. Bose and V. D. Agrawal, "Delay Test Quality Evaluation using Bounded Gate Delays," in *Proc. 25th IEEE VLSI Test Symp.*, May 2007, pp. 23-28.

[5] S. Chakraborty, D. L. Dill, and K. Y. Yun, "Min-Max Timing Analysis and an Application to Asynchronous Circuits," *Proc. IEEE*, vol. 87, no. 2, pp. 332-346, Feb. 1999.

[6] N. Ishiura, Y. Deguchi, and S. Yajima, "Coded Time Symbolic Simulation Using Shared Binary Decision Diagrams," in *Proc. Design Automation Conference*, June 1990, pp. 130-135.

[7] M. Linderman and M. Leeser, "Simulation of Digital Circuits in the Presence of Uncertainty," in *Proc. International Conf on Computer Aided Design*, November 1994, pp. 248-251.

[8] E. G. Ulrich, K. P. Lentz, S. Demba, and R. Razdan, "Concurrent Min-Max Simulation," in *Proc. Design Automation Conference*, June 1991, pp. 554-557.

[9] K. N. Lalgudi, D. Bhattacharya, and P. Agrawal, "Architecture of a Min-Max Simulator on MARS," *Proc.* International Conference on VLSI Design, January 1993, pp. 246-249.

[10] J. W. Bierbauer, J. A. Eiseman, F. A. Fazal, and J. J. Kulikowski, "System Simulation with MIDAS," *AT&T Tech. J.*, vol. 70, no. 1, pp. 36-51, January 1991.