A Novel Low-overhead Delay Testing Technique for Arbitrary Two-Pattern Test Application

Swarup Bhunia, Hamid Mahmoodi, Arijit Raychowdhury, and Kaushik Roy
School of Electrical and Computer Engineering, Purdue University, IN 47907, USA
{bhunias, mahmoodi, araycho, kaushik}@ecn.purdue.edu

Abstract—With increasing process fluctuations in nano-scale technology, testing for delay faults is becoming essential in manufacturing test to complement stuck-at-fault testing. Design-for-testability techniques, such as enhanced scan are typically associated with considerable overhead in die-area, circuit performance, and power during normal mode of operation. This paper presents a novel test technique, which can be used as an alternative to the enhanced scan based delay fault testing method, with significantly less design overhead. Instead of using an extra latch as in the enhanced scan method, we propose using supply gating at the first level of logic gates to hold the state of a combinational circuit. Experimental results on a set of ISCAS89 benchmarks show an average reduction of 33% in area overhead with an average improvement of 71% in delay overhead and 90% in power overhead during normal mode of operation, compared to the enhanced scan implementation.

I. INTRODUCTION

Delay faults in a circuit occur when a net functions properly but fails to meet timing requirement. Delay faults are sometimes caused by defects that are not large enough to cause a stuck-at failure by changing logic level, but affect the signal propagation time. However, an emerging cause of delay failure is the uncertainty in circuit design due to process fluctuations, limitation of timing models and static timing analysis tools etc. With growing impact of process variation in sub-100nm technology regime, designers face more uncertainty in circuit design [1] and delay faults become more likely. Therefore, it is becoming mandatory for manufacturing test to include delay testing along with stuck-at tests [7] [8].

Scan architectures provide an efficient way to test for delay faults with good fault coverage. Scan-based structural delay testing not only helps detection but also diagnosis of delay faults [7] and, hence, is a popular choice for delay fault testing. However, testing for delay faults usually require launching a transition at the first pattern, and capturing the response of the circuit at rated clock. Although it is easier to apply a transition at the primary inputs of the CUT, and capturing the response of the circuit at rated clock. Although it is easier to apply a transition at the primary inputs of the CUT, and capturing the response of the circuit at rated clock. Although it is easier to apply a transition at the primary inputs of the CUT, and capturing the response of the circuit at rated clock. Although it is easier to apply a transition at the primary inputs of the CUT, and capturing the response of the circuit at rated clock.

In this paper, we propose a delay fault testing technique, which allows enhanced scan-like test application, but comes at a much lower hardware overhead. The technique, referred as First Level Hold (FLH) employs the principle of collecting the state of the combinatorial logic in response to the first pattern by gating the VDD and GND of the first level logic gates. Test application remains as in enhanced scan approach, except that the control for holding state is now moved from the hold latches to the gating control of the first level of logic.

FLH does not require any extra control signals and does not change the test generation/application process. Moreover,
Unlike enhanced scan test, it does not introduce extra level of logic in the timing path of a circuit and hence, the delay overhead reduces greatly compared to the enhanced scan. We have compared FLH technique with enhanced scan method and a possible MUX-based alternative [13]. Experiments performed on a set of ISCAS89 benchmarks show superior results with FLH in terms of area, delay and power overhead compared to the alternative methods. It is worth noting that FLH also maintains the power-saving advantage of enhanced scan in the test mode, since it prevents redundant switching in the combinational block by isolating it from the activity in scan register.

The rest of the paper is organized as follows: Section II illustrates the proposed gating technique for delay testing. Section III presents experimental results in terms of area, delay and power for a set of benchmark circuits. Section IV describes important test issues associated with the proposed technique. Section V describes ways to further reduce DFT overhead and Section VI concludes the paper.

II. FIRST LEVEL HOLD FOR DELAY FAULT TEST

The requirement of enhanced scan based delay fault testing is to apply a transition at the state inputs of a combinational block by holding its output state in response to the initial pattern before applying the second pattern. This can be achieved by adding a hold latch as in the enhanced scan or a MUX at the input of the combinational circuit (Fig. 1). We have observed that, interestingly, we can achieve holding the state of the combinational logic by "gating" the supply lines of the first level logic gates. Fig. 2 shows first level supply gating for an inverter chain. If the output of the first level logic gates (OUT1 in Fig. 2) can hold their state in the sleep (i.e. gated) mode, logic gates in their fanout cones can also retain their states.

Let us consider the circuit in Fig. 2 with IN at ‘0’ and OUT1 at ‘1’ when the gating control or SLEEP signal is applied. When the SLEEP signal is ‘1’, the node OUT1 is floated since there is no path to VDD or GND from this node. In this case, the voltage of OUT1 can remain at ‘1’ due to the charge that is held in that node. However, since OUT1 is floated, the charge held in OUT1 node can leak due to leakage of transistors connected to that node, which can result in a change in the state of OUT1 node. This is particularly aggravated if IN switches to ‘1’ in the sleep mode and stays at ‘1’ for a long enough time. This scenario is simulated in Hspice for the circuit shown in Fig. 2 using the 70nm Berkeley Predictive Technology Models [14]. We have observed that the voltage of OUT1 falls below 600mV in less than 100ns. Assuming a scan chain with a length 1000 flip-flops and a scan frequency of 1GHz, the scan time is 1µs which is much longer than 100ns. As OUT1 slowly decays below Vdd−Vth, in the second inverter (Fig. 2), both the PMOS and NMOS transistors get turned ON causing static short circuit current to flow through the second inverter. Consequently, the output of the second inverter (OUT2) rises resulting in static current on the third inverter (Idd3). If OUT1 decays below the trip point of the second gate, a switching occurs in the second gate, which results in a change in the state of the circuit. In addition to leakage, crosstalk noise or transient effects due to soft error can also easily change the voltage of a
floated output. Crosstalk noise can particularly occur in this circuit because the switching of input (IN) can couple to OUT1 through the gate-to-drain capacitances of both PMOS and NMOS transistors of the first level gate. Moreover, the switching of the inputs can result in charge sharing between the floated output node and intermediate nodes of the NMOS or PMOS network in complex gates resulting in change of the output voltage.

In order to avoid floated nodes in the sleep mode and ensure hold capability, the outputs of the first level gates need to be forced to VDD or GND, depending on their initial logic state. This can be achieved by adding a latch element (cross-coupled inverters) at the output node. The latch element needs to be enabled only in the sleep mode to hold the output state of the first level gate. The general scheme of the proposed supply gating scheme is shown in Fig. 3. The two inverters, INV1 and INV2, form a cross-coupled inverter loop if the transmission gate is closed. In the sleep mode (TC='0'), the transmission gate is closed and the inverter loop holds the state of the output node. In the normal mode (TC='1'), however, the transmission gate is open and the gate can control its output. Therefore, in this scheme, the output of the gate never gets floated and there cannot be any static short circuit current on the next stage gates in the sleep mode. The proposed scheme is called “First Level Hold (FLH)” since only the first stage is set in the hold mode. The inverters (INV1 and INV2) and the transmission gate can use minimum-sized transistors to minimize their impact on area, circuit delay, and power during normal mode of operation. Minimum sized inverters are large enough to be able to hold the state of the output node in the hold mode despite the presence of leakage and noise. Use of minimum sized transistors for the latch element reduces loading on the outputs of first level gates, resulting in minimal delay and power penalty. The size of the supply gating transistors can be optimized for delay under the given area constraint. Fig. 4 shows the simulated waveforms of the FLH scheme applied to the inverter chain in Fig. 2. As observed from the waveforms, the circuit can strongly hold its state (OUT1, OUT2, and OUT3) despite the switching at the input (IN).

A. Scan Design Using FLH

Fig. 5 shows the proposed FLH technique applied to a general sequential circuit. FLH does not require any extra timing control signals. It only uses the test control (TC) signal, that is used in conventional scan-based testing, and its complement (TC̅). Enhanced scan method requires two control signals, TC and HOLD. The timing diagram during test application is shown in Fig. 5 (b). During scan-in, TC is set to '0' to prevent activity in the scan chain affecting the combinational circuit. Once scan-shifting is completed for the

Fig. 3. Proposed supply gating scheme with output hold capability

Fig. 4. Simulated waveforms of proposed supply gating scheme applied to circuit in Fig. 2

Fig. 5. (a) Modified scan architecture with holding logic at first level gates; (b) Timing diagram for delay testing with FLH
first pattern (V1), it is applied to the combinational circuit by turning the gating transistors on, while the primary input (PI) bits are applied to PI. After the combinational circuit stabilizes, the second pattern (V2) is scanned-in while V1 is held since the gating transistors in the first level gates are turned off. Next, the transition is launched by activating TC and applying the PI bits and the results are latched after one rated clock period.

III. EXPERIMENTAL RESULTS AND COMPARISONS

To estimate the effectiveness of the FLH scheme, we simulated a set of ISCAS89 benchmark circuits and obtained area, power, and performance overhead in case of FLH, enhanced scan, and MUX-based approaches. The simulations were performed using the 70nm BPTM models [14] to observe the effect of gating in a sub-100nm scaled technology. For the latch and mux circuitry, we have used optimized implementation obtained from the LEDA library, as shown in the schematic in Fig. 6. The gate-level netlists were first technology-mapped to LEDA 0.25µm standard cell library using Synopsys design compiler by setting the mapping effort to medium. The library contains complex gate types e.g. “aoi” (and-or-invert) and “mux”, and hence, the total number of logic gates is reduced from that in original benchmark. The benchmark circuits are then translated to Hspice netlists and scaled to 70nm. We assumed full-scan implementation of the benchmarks. Power is measured in NanoSim by applying 100 random vectors to the inputs and delay is measured by Hspice simulation of the critical path of a circuit.

Table I shows comparisons of these techniques in terms of area overhead. Since the layout rules for the 70nm node are not available, the measure used for area is the total transistor active area (W * L for a transistor). Enhanced scan circuit has the largest area overhead followed by the MUX-based technique. FLH exhibits the largest area overhead for most benchmark circuits. In both enhanced scan and MUX-based methods, the holding elements (latch and MUX) are inserted at the state inputs of the circuit. This means that there is one gating element per scan flip-flop (Fig. 1 (a)). However, in FLH, gating logic is inserted in all first level gates (Fig. 5), the number of which depends on the number of unique fanout gates of the scan flip-flops. Therefore, for a circuit with large fan-out for state inputs, such as s838, the area overhead in the FLH technique can be more than the others. However, number of fanouts in a circuit are usually not high (2.3 on average per scan flip-flop as can be obtained from column 2 and 3) to satisfy delay constraint of a circuit, since higher fanout means higher load at the output of a gate and hence, higher delay. Number of unique fanouts, i.e. the first level gates (as shown in column 4) is further less (1.8 on average per scan flip-flop) due to overlapping of fanout cones. FLH shows 33% and 26% reduction in area overhead on an average as compared to the enhanced scan and MUX-based techniques, respectively. It is worth noting that FLH does not introduce additional test control signals. Therefore, FLH is expected to have no area penalty over enhanced scan due to routing of test controls.

Table II shows comparison of impact on circuit delay for different benchmark circuits. As observed from Table II, the proposed technique has the least impact (minimal increase) on circuit delay. The MUX-based method shows the largest increase. FLH exhibits reduction of up to 10% in overall circuit delay compared to enhanced scan approach. It is worth noting that the logic depth for the test circuits is fairly high (column 2). Since the original delay of the critical path is very large, the percentage reduction in circuit delay in FLH compared to the others is not very high. However, comparing the percentage reduction in delay overhead in FLH with that in enhanced scan method, an average improvement of 71% is observed. As the logic depth decreases for better performance in sequential circuit, the proposed FLH scheme will show much less delay overhead as compared to enhanced scan.

Table III shows comparison of power in the normal mode of operation. Significant power savings are observed for all the benchmark circuits. In fact, for most benchmark circuits the power dissipations of the FLH circuits are close to the power dissipations of the original circuits. This is because in the proposed technique, the supply gating transistors do not switch in the normal mode. The only source of power overhead is due to switching of the minimum-sized inverters and the diffusion capacitance added to the outputs of the first level gates due to the transmission gate. It is interesting to notice that for a large benchmark circuit such as s13207, the power of the FLH circuit is even less than the power of the original circuit. This can be attributed to two facts: a) the sleep transistor results in active leakage reduction (due to stacking [9]) for the idle gates b) reduced number of switching at the outputs of first level gates compared to the number of switching at scan flip-flop outputs. For a large circuit, at each time instant, there are many idle first level gates during scan shifting. Saving leakage in those gates, hence, reduces overall power. FLH shows an average reduction of 44% overall circuit power compared to the enhanced scan method. However, the percentage reduction in power overhead compared to the enhanced scan is 90% on an average.

Larger-sized sleep transistors for gates in the critical path can be used to further reduce the delay penalty. It increases the area overhead but does not affect the switching power of the gates. However, upsizing the hold latch and MUX does...
not help much to improve delay since it increases load on the scan flip-flop. Moreover, it comes at the cost of increase in both area and power overhead. Area and power overhead can be further reduced by local fanout optimization under delay constraint, as explained in section V.

### IV. Test Considerations

Fault coverage and fault models remain unaffected with the insertion of FLH logic. During normal mode of operation the gating transistors are turned ON, hence the conventional stuck-at fault model, transition and path delay fault models remain valid. FLH does not require any change in test vectors generated by ATPG tools. Hence, fault coverage for enhanced scan and FLH for a given test set remain unchanged.

In a conventional scan-based circuit, combinational logic suffers from redundant switching in response to changing scan values during the entire period of scan-shifting. Gerstendorfer and Wunderlich [12] have shown that on an average about 78% of energy in the test mode can be saved by preventing redundant switching in combinational logic by using blocking gates at the output of scan flip-flops. It is worth noting that an enhanced scan flip-flop embeds a blocking gate, and thus, isolates combinational logic from activity in the scan register during shift operation. Although FLH does not insert any blocking logic at the output of scan flip-flops, supply gating at the first level logic gates holds the previous output state.
of the gate and prevents propagation of switching. FLH is, thus, equally effective in completely eliminating redundant switching power in the combinational logic.

The proposed technique can be easily applied to scan-based test-per-scan BIST (Built-In Self Test) [11] circuits. A circuit designed with BIST has weighted random pattern generator and output response analyzer built into the circuit. The patterns are applied to both primary inputs and scan cells. If test patterns are applied to the primary inputs serially, as in the scan chain, FLH technique proposed for scan path can be equally used to the fanout logic gates for the primary inputs to provide a transition. Scan insertion with FLH can be easily automated by test synthesis tools by inserting the gating logic of FLH for each scan cell to each of its first level fanout gates. It can be noted that additional logic for FLH (gating transistors and the embedded latch) does not require to modify a logic gate. Hence, it is not necessary to change the standard cell library in case of a cell-based design. However, integrating the gating logic into the layout of a standard-cell element allows more efficient routing and hence, can reduce the area overhead in physical implementation.

V. FURTHER REDUCTION OF AREA/POWER OVERHEAD

Transistor downsizing can be applied to all the methods, including FLH, to reduce the area and power overhead. But narrowing transistor width usually trades off circuit performance by affecting critical path delay. FLH, however, has potential to reduce the area penalty further without compromising delay. We designed a low-complexity local fanout reduction algorithm which targets minimization of first level gates under constraint on critical path time. The algorithm is based on identifying the scan flip flops with higher fanouts and then adding two inverters in cascade between output of the scan flip-flops and their fanout gates. No inverter is added in the critical path of the circuits and maximum circuit delay is kept unaltered. We then try to re-synthesize the second inverter with its fanout gates to reduce area penalty due to the additional inverters. If a scan flip-flop already has an inverter connected to it, we do not need the second inverter. The algorithm utilizes the timing slack available in the non-critical paths. We implemented the algorithm applied it on a set of benchmarks with higher number of scan flip-flops. The result is presented in Table IV. It can be observed that we can get as high as 37% improvement (with an average of 18%) in area overhead with fanout optimization under delay constraint. The power in normal mode remains comparable. It is interesting to note that, for some cases (e.g. s5378), the number of first level gates becomes even lower than the number of scan flip-flops. This is because for these benchmarks, most of the high-fanout scan flip-flops have largely reduced fanouts (1 or 2) after the fanout optimization, while overlap among the fanout cones of the other flip-flops is maintained. It results in total number of first level gates lower than the number of flip-flops.

VI. CONCLUSIONS

This paper presents First Level Hold (FLH), a novel technique based on supply gating, as a low-cost alternative to enhanced scan approach of delay fault testing. The proposed technique does not affect test generation, test application and fault coverage. FLH does not require any extra test control signal. It maintains the power-saving advantage of the enhanced scan method in test mode by suppressing activity in the combinational logic during scan shifting. FLH is more suitable for high-speed applications since it induces significantly less delay in critical path of a circuit. At the same time, it provides the benefit of lower overhead in die-area and power consumption in normal mode of circuit operation.

ACKNOWLEDGMENT

This research was funded in part by MARCO GSRC and by National Science Foundation (NSF).

REFERENCES