A Circuit Level Fault Model for Resistive Shorts of MOS Gate Oxide

Xiang Lu†, Zhuo Li†, Wangqi Qiu‡, D. M. H. Walker‡ and Weiping Shi†

†Dept. of Electrical Engineering
Texas A&M University
College Station, TX 77843-3124, USA
wshi@ee.tamu.edu

‡Dept. of Computer Science
Texas A&M University
College Station, TX 77843-3112, USA
walker@cs.tamu.edu

ABSTRACT

Previous researchers in logic testing focused on shorts in MOS gate oxides that have zero-resistance. However, most shorts are resistive and may cause delay faults. In this paper, we propose a simple and realistic delay fault model for gate oxide shorts. A reasonably accurate method is proposed to compute delay change due to resistive shorts. We also enumerate all possible fault behaviors and present the relationship between input patterns and output behaviors, which is useful in ATPG.

1. INTRODUCTION

Delay test detects small manufacturing defects that do not cause functional failure but affect the speed of the circuits. In this paper, we consider the fault model of resistive gate oxide shorts, which can be used for delay test.

Banerjee and Abraham first considered gate oxide shorts. They assumed the shorts are zero-resistance and modeled the shorts as stuck-at faults [1]. Hawkins and Soden [2] presented data to show the short resistance is in the range of 100's to 1000's of ohms. Syrzycki [3] considered different structures of gate oxide shorts and proposed lumped-element models. Gaitonde and Walker [4] studied some problems faced in mapping spot defects including gate oxide shorts to changes in the nominal circuit. Hao and McCluskey [5] studied the logic and delay behavior of resistive shorts. Their circuit level fault model is shown in Figure 1, which covers the external behaviors of gate-to-source shorts, gate-to-drains short and gate-to-channel shorts. Note that because the polysilicon gate is generally doped with n type dopant, there is a diode in series with the short resistor for PMOS transistors. Hao and McCluskey also showed that faults caused by gate oxide shorts can be dependent on input signals of both the gate containing the fault and other gates [6]. And they found that Very Low Voltage testing would detect such pattern dependent faults caused by resistive shorts [7]. In [8], we proposed a circuit level fault model for resistive shorts in the interconnect. However, resistive shorts in MOS gates are more complex in certain cases.

Modeling gate oxide shorts as delay faults helps delay test to detect more shorts. Previous models lack in explicit relationship between delay changes and resistive shorts. In this paper we propose a physically realistic and economical fault model for delay test. Resistive gate oxide shorts are modeled as delay faults as well as logic faults according to the short resistance. We also enumerate all possible fault behaviors and present the relationship between input patterns and output behaviors. The fault model is reasonably accurate and easy to implement for delay fault simulation in large industrial circuits.

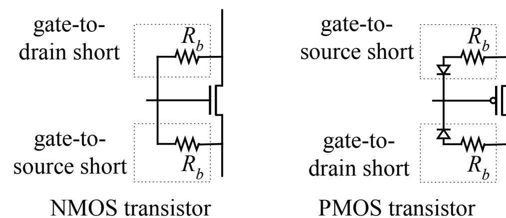

Figure 1. Gate-to-source and gate-to-drain shorts for NMOS and PMOS transistors.

2. FAULT MODEL

The objective of the gate oxide short model is to transform the effect of the resistive short into a delay fault or a logic fault, and compute the delay change as a function of short resistance R_b.

To analyze electrical behaviors of R_b in the circuit level, consider the circuit in Figure 2. There are two cells: the driving cell with output node A and the faulty cell with output node B. In the figure, R_1 and R_2 are interconnect parasitic resistance, C_1 and C_2 are lumped interconnect parasitic and gate input capacitance. In Figure 2 the defect is a gate-to-source short inside an NMOS transistor. Circuits for other defects are similar, except that the faulty transistor may be PMOS or the short may be gate-to-drain.

The faulty cell contains one faulty transistor, and the rest of transistors are good. In this paper, we only consider the short at input transistors, i.e. transistors driven by external signals. Shorts at other transistors can be analyzed similarly by partitioning the faulty cell to make the faulty transistor driven by an external signal. For

example, an AND cell can be partitioned into a NAND cell driving an inverter, in order to analyze the short at one of the inverter transistors.

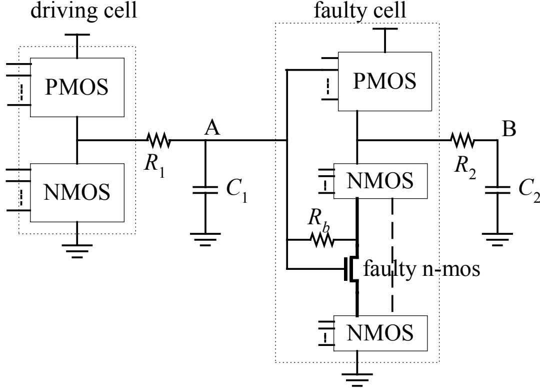

Figure 2. The circuit with an NMOS gate-to-drain short. Circuits with other shorts are similar.

We assume there is only one single input transition on the faulty cell. When a signal transition at the input of the driving cell causes the transition at B, we consider the sum of the driving cell delay and the faulty cell delay. However, when the signal at A is static, and the transition at B is caused by a transition at other input of the faulty cell, we only consider the faulty cell delay. Here delay is computed at the 50% of V_{dd} point on the signal waveform, and denoted as d_{Rb}. When there is no short, the driving cell delay is denoted as D_A, and the faulty cell delay is denoted as D_B. Both D_A and D_B can be computed by any commercial tool, such as SPICE.

2.1 NMOS gate-to-drain short

Static analysis

The circuit in the static analysis is shown in Figure 3. R_{up1} and R_{down1} are pull-up and pull-down resistance of the driving cell respectively. R_{up2} represents the pull-up resistance in the faulty cell. The pull-down path inside the faulty cell is divided into two parts from the faulty transistor, represented by resistors R_{down2} and R_{down3}. The linear resistance of the faulty transistor is included in R_{down2}.

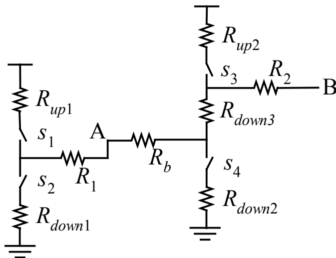

Figure 3. The circuit under static analysis.

When a pull-down path in the driving cell is formed, s_1 = "off", s_2 = "on", s_3 = "on" and s_4 = "off". The voltage at node A is derived as

$$V_{A,down} = \frac{R_1 + R_{down1}}{R_{down1} + R_1 + R_b + R_{down3} + R_{up2}} V_{dd} \,. \quad (1)$$

The voltage at node B, $V_{B,down}$, can be derived similarly.

Define the Bridge Threshold Resistance (BTR) $R_{A,\text{down}}$, such that when $R_b < R_{A,\text{down}}$, $V_{A,\text{down}}$ is larger than the logic low threshold voltage V_{LL}. Then according to (1),

$$R_{A,down} = \frac{(V_{dd} - V_{LL})(R_1 + R_{down1})}{V_{LL}} - R_{down3} - R_{up2} \,. \quad (2)$$

Therefore when $R_b < R_{A,down}$, $V_{A,\text{down}} > V_{LL}$, there is a logic fault at A. When $R_b > R_{A,down}$, there will be no logic fault at A, but there might be a delay fault.

When a pull-up path in the driving cell is formed, similar voltage expressions $V_{A,up}$ at node A, $V_{B,up}$ at node B and corresponding BTRs can be derived.

Transition analysis

We present transition analysis by two cases. In one case, a signal transition propagates through the driving cell to the faulty cell. In the other case, a signal transition happens only at the faulty cell while the signal at node A is static. For each case, we assume R_b is large enough such that there is no logic fault when the circuit is static. We model the resistive short as a delay fault, and propose the method to compute the delay change due to the short.

Case 1

Assume a falling signal transition happens at node A, and accordingly a rising signal transition happens at node B.

During the transition, the faulty NMOS transistor is cut-off and a pull-up path inside the faulty cell in formed. If there is no R_b present, C_1 is discharged only through R_{down1} and C_2 is charged only through R_{up2}. With the present of R_b, a current path through R_b is formed between node A and B. Then C_1 is partially discharged to C_2 because the initial voltage at A is higher than that at B. At the same time, C_1 is also charged by the faulty cell through R_b. The circuit in transition analysis is illustrated in Figure 4.

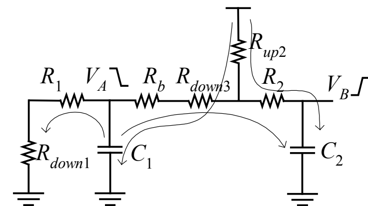

Figure 4. With the present of R_b, a current path is formed between node A and B.

Our delay model works as follows. Through analyzing the circuit in Figure 4, we compute the driving cell delay $d_{A,Rb}$ and the faulty cell delay $d_{B,Rb}$. Considering both delays are functions of R_b, we compute the ratio between the delay with some certain value of R_b and the delay with $R_b = \infty$, then approximate d_{Rb} as

$$d_{R_b} = \frac{d_{A,R_b}}{d_{A,R_b=\infty}} D_A + \frac{d_{B,R_b}}{d_{B,R_b=\infty}} D_B \ . \quad (3)$$

The computation of $d_{A,Rb}$ is as follows.

First we introduce an effective resistance $R_{A,eff}$ to replace the discharging path R_b-R_{down3}-R_2-C_2. The value of $R_{A,eff}$ is computed by equalizing the average current of the two circuits in Figure 5, during a certain time interval T_d. The value of T_d is the time for the voltage of C_1 in the circuit of Figure 5(a), dropping from its initial voltage V_0 to 50% of the total dropping range. The equivalent current method is also used to calculate effective capacitance in [9]. In this paper we use it to compute $R_{A,eff}$.

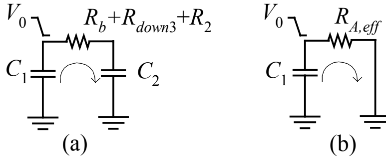

Figure 5. $R_{A,eff}$ is computed by equalizing the average current in the two circuits.

We get

$$R_{A,eff} = \frac{C_2}{C_1 + C_2} \cdot \frac{\ln(2)}{\ln(\frac{2(C_1 + C_2)}{2C_1 + C_2})} \cdot (R_b + R_{down3} + R_2) \ .$$

Second, because we replace path R_b-R_{down3}-R_2-C_2 with $R_{A,eff}$, the circuit becomes a first order linear system and the driving cell delay can be derived as

$$d_{A,R_b} = d_{f1} + \ln(\frac{V_{A,down} - V_{A,up}}{V_{A,down} - 0.5V_{dd}}) \cdot R_A \cdot C_1 \ , \quad (4)$$

where d_{f1} is the intrinsic falling delay of the driving cell, R_A=(R_{down1}+R_1)//(R_b+R_{down3}+R_{up2})//$R_{A,eff}$. The symbol "//" represents the parallel computation of two resistances.

Note when R_b is infinity, $d_{A,Rb=\infty}$ is the sum of intrinsic gate delay and gate load delay under the lumped C delay model.

The computation of $d_{B,Rb}$ is similar. For the rising signal transition at B, path C_1-R_b-R_{down3}-R_2 is a charging path. We insert a resistance $R_{B,eff}$ at A to equalize the charging effect of the path. The value of $R_{B,eff}$ is computed similarly. Here the value of T_d is the time for the voltage of C_2 in the circuit of Figure 6(a), rising to 50% of the total rising range.

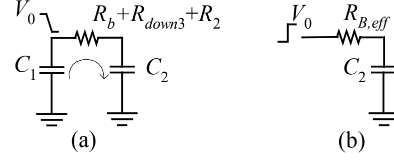

Figure 6. $R_{B,eff}$ is computed by equalizing the average current in the two circuits.

Therefore, we get

$$R_{B,eff} = \frac{C_1}{C_1 + C_2} \cdot \frac{\ln(2)}{\ln(\frac{2(C_1 + C_2)}{C_1 + 2C_2})} \cdot (R_b + R_{down3} + R_2) \ .$$

Then the faulty cell delay is derived as

$$d_{B,R_b} = \alpha \cdot d_{r2} + \ln(\frac{V_{B,down} - V_{B,up}}{V_{B,down} - 0.5V_{dd}}) \cdot R_B \cdot C_2 \ , \quad (5)$$

where d_{r2} is the intrinsic rising delay of the faulty cell, α=(V_{dd} − $V_{B,up}$)/V_{dd} is introduced to scale d_{r2} because the initial voltage for the rising transition at B is not zero, and R_B=(R_{down1}+R_1+$R_{B,eff}$+R_b+R_{down3})//R_{up2}+R_2.

Experiments are performed on a circuit, which consists of an inverter as the driving cell, and an NAND gate as the faulty cell. TSMC 180 nm 1.8V technology is used. Assume V_{LL} = 0.7V, then Bridge Threshold Resistance (BTR) $R_{A,down}$ = 1284Ω according to our model. Results of delay ~ R_b computed by our model and by SPICE simulation are shown in Figure 7. Our model provides reasonably accurate approximation of the delay change with present of R_b. When $R_b < R_{A,down}$, there is a logic fault on node A and when $R_b \geq R_{A,down}$, there is a decreasing delay fault.

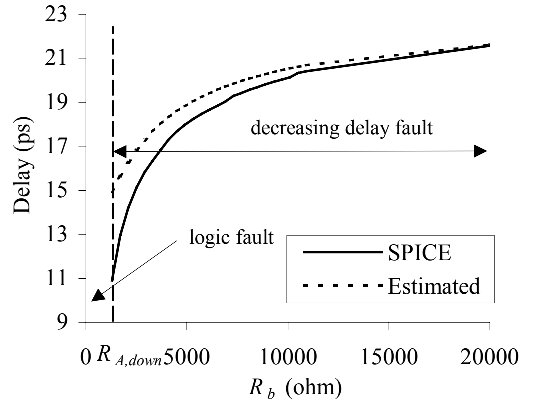

Figure 7. Decreasing delay estimated and computed by SPICE with increasing values of R_b.

Case 2

Consider an NMOS transistor T_g, which is in series with the faulty NMOS transistor T_f. Assume there is a rising

signal driving T_g, resulting in a falling transition at B. The signal at node A, which is driving T_f, is static high.

During the transition, a pull-down path through T_f and T_g is formed. The simplified circuit is shown in Figure 8, where R_{down3} represents the linear resistance of T_f and the NMOS block connecting output of the faulty cell, and R_{down2} represents the linear resistance of the NMOS block connecting T_f and Gnd.

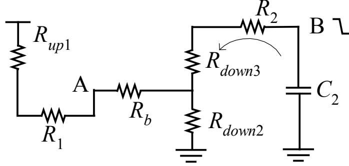

Figure 8. C_2 is discharging through the pull-down path in the faulty cell.

The final voltage at B is dependent on the value of R_b, and can be derived as

$$V_{B,up} = \frac{R_{down2}}{R_{up1} + R_1 + R_b + R_{down2}} V_{dd} \,. \quad (6)$$

Then the faulty cell delay is computed as

$$d_{B,R_b} = d_{f2} + \ln\left(\frac{V_{dd} - V_{B,up}}{50\% V_{dd} - V_{B,up}}\right) \cdot (R_{down2} + R_{down3} + R_2) \cdot C_2 \,,$$

where d_{f2} is the intrinsic rising delay of the faulty cell.

Similarly, we approximate d_{Rb} as

$$d_{R_b} = \frac{d_{B,R_b}}{d_{B,R_b=\infty}} \cdot D_B \,. \quad (7)$$

Experiments results on the same circuit are shown in Figure 9. The BTR for $V_{B,up}$ is $R_{B,up}$ = 1798Ω. The figure shows that when $R_b < R_{B,up}$, there is a logic fault on node A and when $R_b \geq R_{B,up}$, there is a increasing delay fault in the faulty cell.

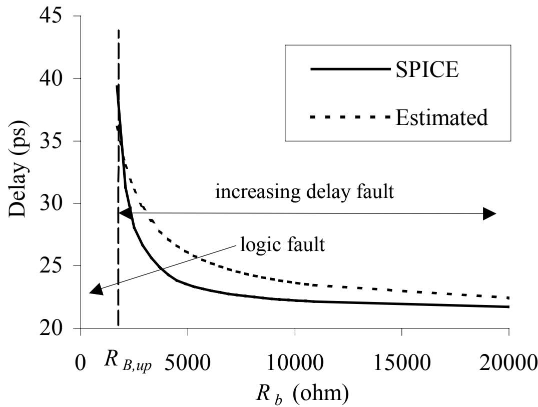

Figure 9. Increasing delay estimated and computed by SPICE with increasing values of R_b.

2.2 NMOS gate-to-source short

In the case of an NMOS gate-to-source short, it may affect the circuit behaviors in two ways. 1) The short acts as a resistor that connects A and Gnd, thus reducing the voltage at A due to voltage division. The value of the resistor is R_b+R_{down}, where R_{down} represents the linear resistance of the NMOS block between the faulty transistor and Gnd. Note that whether the current path through R_b to Gnd is formed depends on the side inputs of the faulty cell. Then the short can be viewed as a resistive bridge between interconnect and GND, and modeled using the method proposed in Li *et al.* [8]. 2) The short connects A and B through the faulty NMOS transistor, and currents flow through A to the faulty cell output, then though some NMOS transistors to Gnd. Therefore the voltage at B is affected and a functional fault possibly happens at B. A delay fault at B possibly happens and the delay change can be derived similarly.

2.3 PMOS gate-to-drain short

The gate-to-drain short in a PMOS transistor can be modeled as a bridge resistance in series with a diode between the interconnect and Vdd. The diode is always forward-biased and is represented by a certain voltage drop across the bridge. The behavior of the diode can be analyzed similarly with the method in Hao and McCluskey [3]. Due to the diode, currents can never flow through A to B.

2.4 PMOS gate-to-source short

Similar analysis to NMOS gate-to-drain shorts can be performed, except for the voltage drop effect due to the diode. And when a pull-up path is formed in the driving cell, currents cannot flow through A to B due to the diode.

3. CONCLUSIONS

In this work, we propose a circuit level fault model for resistive gate oxide shorts in MOS transistors. The resistive short is modeled as a logic fault when the short resistance is less than the value of bridge threshold resistance (BTR). Otherwise it is modeled as a delay fault.

We derive expressions of BTR for all possible short behaviors. To detect a logic fault, the test pattern should be selected to maximize BTR. With explicit expressions of BTR, it is possible to compare performance of different test patterns analytically. Previous methods, for example the stuck-at fault model [1] and the short resistive behavior analysis in [5], are insufficient for that.

For the case that the short provokes delay changes in the circuit, we propose a fast yet reasonably accurate delay estimation method and provide closed form delay functions in terms of the short resistance. In Table I, for each type of short, we list delay behaviors for different input signals. T_f is the faulty NMOS/PMOS transistor and

T_g is another NMOS/PMOS transistor. The rising/falling signal on T_g or T_f provokes a rising or a falling transition at the output of the faulty cell. In the table, “$T_f | T_g$” means the two transistors are in parallel, “$T_f \sim T_g$” means the two transistors are in series, “decrease” means decreasing delay, “increase” means increasing delay, and “0” means there is no change in delay when the short is present. The table enumerates all input patterns that will cause increasing delay (or decreasing delay) at the output B due to the short defect. The amount of delay change is given by such formulae as (3) and (7). The table and corresponding delay formula can guide ATPG to find the best test patterns that can maximize the detectability.

Table I. Behaviors for each type of short under different input signals.

Short	Input of faulty transistor T_f	Input of other transistor T_g	Delay change
nmos g-s short	rising	static	increase
	falling	static	decrease
	static	rising	0($T_f \| T_g$), increase($T_f \sim T_g$)
	static	falling	0($T_f \| T_g$), decrease($T_f \sim T_g$)
nmos g-d short	rising	static	decrease
	falling	static	decrease
	static	rising	decrease($T_f \| T_g$), increase($T_f \sim T_g$)
	static	falling	increase($T_f \| T_g$), decrease($T_f \sim T_g$)
pmos g-d short	rising	static	decrease
	falling	static	increase
	static	rising	decrease($T_f \| T_g$),0($T_f \sim T_g$)
	static	falling	increase ($T_f \| T_g$),0($T_f \sim T_g$)
pmos g-s short	rising	static	decrease
	falling	static	increase
	static	rising	decrease($T_f \| T_g$),0($T_f \sim T_g$)
	static	falling	increase($T_f \| T_g$),0($T_f \sim T_g$)

REFERENCES

[1] P. Banerjee and J. Abraham, “Generating tests for physical failures in MOS logic circuits,” *ITC* 1983, pp. 554-559.

[2] C. F. Hawkins and J. M. Soden, “Electrical characteristics and testing considerations for gate oxide shorts in CMOS ICs,” *ITC* 1985, pp. 544-555.

[3] M. Syrzycki, “Modeling of gate oxide shorts in MOS transistors,” *IEEE Trans. CAD*, 8(3), 1989, pp. 193-202.

[4] D. Gaitonde and D. M. H. Walker, “Circuit-level modeling of spot defects,” *DFT* 1991, pp. 63-66.

[5] H. Hao and E. J. McCluskey, “On the modeling and testing of gate oxide shorts in CMOS logic gates,” *DFT* 1991, pp. 161-174.

[6] H. Hao and E. J. McCluskey, “‘Resistive shorts’ within CMOS gates,” *ITC* 1991, pp. 292-301.

[7] H. Hao and E. J. McCluskey, “Very-low-voltage testing for weak CMOS logic ICs,” *ITC* 1993, pp. 275-284.

[8] Z. Li, X. Lu, W. Qiu, W. Shi and H. Walker, "A circuit level fault model for resistive opens and bridges", *ACM Trans. on Design Automation of Electronic Systems*, 8(4), 2003, pp. 546-559.

[9] J. Qian, S. Pullela and L. Pillage, “Modeling the “Effective capacitance” for the RC interconnect of CMOS gates,” *IEEE Trans. CAD*, 13(12), 2001, pp. 1526–1535.