Non-intrusive Online Distributed Pulse Shrinking Based Interconnect Testing in 2.5D IC

Tianming Ni, *Member, IEEE*, Hao Chang, Tai Song, Qi Xu, Zhengfeng Huang, Huaguo Liang, Aibin Yan, and Xiaoqing Wen, *Fellow, IEEE*

***Abstract*—In this brief, a non-invasive online solution for 2.5D IC based on distributed pulse shrinking is proposed to test the faults of interconnects. Furthermore, a regression model based on artificial neural network (ANN) is proposed in order to judge whether the interconnects are faulty and quantify the degree of the faults in real time by online monitoring the delay of interconnects. Experiments on defect detection are presented through HSPICE simulation with realistic models for 45nm CMOS technology. The results show that the proposed method has features including: high resolution, low area overhead, high robustness, and be able to predict the class and the fault size.**

***Index Terms*—2.5D IC, pulse shrinking, interconnect, TSV, ANN.**

I. Introduction

Three dimensional integrated circuits (3D-ICs) are invented to address the scaling challenge by stacking 2D dies and connecting them vertically with through silicon vias (TSVs). However, volume production and commercial exploitation of 3D ICs are not feasible before pressing concerns about manufacturing yield are adequately addressed [1-4]. At present, interposer-based 2.5D ICs are being advocated as a precursor to 3D ICs [5]. In 2.5D ICs, multiple active dies are not vertically stacked; rather, they are placed side-by-side on the silicon interposer.

However, the reliability of these interconnects faces great challenges. It has been reported in [6] that interconnects could experience resistance increase (resistive open fault) due to several adverse effects such as thermal-mechanical stress, and electro-migration, which could contribute to an unexpected wear-out situation after certain time of use in the field, and thereby causing failure. In order to ensure the reliability of 2.5D ICs, it is necessary to test the delay of RDL interconnects. By online monitoring of interconnects, the failure of interconnects can be detected at an early stage, and the faulty interconnects can be replaced by the corresponding self-repair mechanisms for reliability improvement.

Online delay testing methods for interconnects can be divided into intrusive and non-intrusive. Intrusive testing methods must halt the normal operation to enter the BIST test mode [5, 6]. The disadvantage is that it affects the work of the chip and fails to detect faults in time. On the contrary, several non-intrusive delay testing methods were proposed [7, 8]. However, these methods have some drawbacks, they require a complete test circuit for each interconnect, thus, it is not suitable for 2.5D IC with long and large number of interconnects. In addition, the existing methodologies experience fault detection resolution challenges. Furthermore, the reliability of the existing designs needs deeper consideration since they didn't consider the robustness against supply voltage variations. Last but not the least, few schemes have used the machine learning approach to judge whether the interconnects are faulty and quantify the degree of the faults in real time by online monitoring the delay of interconnects.

This brief presents a novel distributed test method for interconnects, using a technique named pulse shrinking test (or PS test for short). It can generate a corresponding digital code to compare with an expected value of fault-free. Furthermore, we also propose a method to create a regression model based on artificial neural networks. The regression model can classify whether the interconnect under test is faulty, and it can predict the fault size represented by digital codes, whose input is obtained from PS simulation.

The remainder of the paper is organized as follows. In Section II, we review the electrical model of interconnect, principle of non-intrusive delay testing, and a brief principle of the pulse shrinking method. In Section III, we propose a new distributed delay testing method. Section IV describes the regression model used for fault classification and prediction. In Section V, we present the experimental results. And this brief is concluded in Section VI.

Manuscript received September 19, 2019; revised November 16, 2019, and. Date of publication; date of current version. This work was supported in part by the National Natural Science Foundation of China under grant Nos. 61904001, 61904047, 61974001，61704001, and 61874156; Anhui Provincial Natural Science Foundation under Grant Nos. 1808085QF196 and 1908085QF272; the Key projects of Natural Science Research of Universities in Anhui Province under Grant Nos. KJ2016A001 and KJ2019A0163; and Anhui Polytechnic University Research Startup Foundation under Grant No.2018YQQ007. *(Corresponding authors: Qi Xu.)*

Tianming Ni is with the Key Laboratory of Advanced Perception and Intelligent Control of High-end Equipment, Ministry of Education, College of Electrical Engineering, Anhui Polytechnic University, Wuhu 241000, China (e-mail: timmyni126@126.com).

Hao Chang is with the Department of Computer Science and Technology, Anhui University of Finance and Economics, Bengbu, China. (e-mail: 007changhao@163.com).

Tai Song, Qi Xu *(Corresponding author)*, Zhengfeng Huang, and Huaguo Liang are with the School of Electronic Science & Applied Physics, Hefei University of Technology, Hefei, China (e-mail: baikediguo@163.com, xuqi@hfut.edu.cn, huangzhengfeng@139.com, huagulg@hfut.edu.cn).

Aibin Yan is with the School of Computer Science and Technology, Anhui University, Hefei, China (email: abyan@mail.ustc.edu.cn)

Xiaoqing Wen is with the Department of Creative Informatics, and the Graduate School of Computer Science and Systems Engineering, Kyushu Institute of Technology, Fukuoka 8208502, Japan (e-mail: wen@cse.kyutech.ac.jp)

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCSII.2019.2962824, IEEE Transactions on Circuits and Systems II: Express Briefs

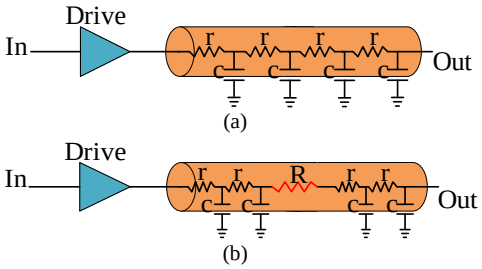

Fig. 1. Electrical model of an interconnect. (a) Fault-free; (b) Faulty.

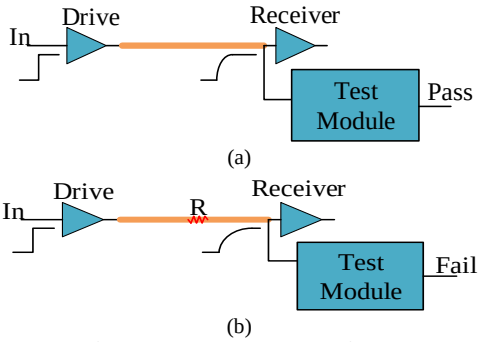

Fig. 2. Delay testing for an interconnect. (a) Fault-free; (b) Faulty.

II. Preliminary

A. Electrical Model of Interconnects

As shown in Fig. 1(a), an interconnect can be viewed as a multi-segment RC network in terms of electrical characteristics. Several adverse effects such as thermal-mechanical stress, and electro-migration may cause additional parasitic resistance R on the interconnect, as shown in Fig. 1(b). For a given interconnect, its delay refers to overall delay from *In* to *Out*.

We use the following terminologies in this brief:

(1) *r*: The resistance of an interconnect with unit segment.

(2) *c*: The capacitance of an interconnect with unit segment.

(3) *R*: The additional parasitic resistance.

(4) R_{wire}: The lumped series resistance of the interconnect. It is split into *L* segments, then $R_{wire} = r * L$.

(5) C_{wire}: The lumped series capacitance of the interconnect. It is split into *L* segments, then $C_{wire} = c * L$.

B. The principle of Non-intrusive Delay Testing

The principle of non-intrusive delay testing is shown in Fig. 2. When the testing signal *In* (rise/fall) transfers through an interconnect, the rise/fall time becomes longer at the other end of the interconnect, as shown in Fig. 2 (a). The test module monitors the delay in real time without affecting the normal operation of the interconnect. As shown in Fig. 2(b), if the interconnect is defective, creating additional resistance, then the interconnect delay will increase. And the test module will find the problem in time and output the judgement Fail.

C. The principle of Pulse Shrinking

Fig. 3 shows the concept of principle diagram of the proposed pulse shrinking cell. The dimensions of inverter 1 and inverter 3 are the same, and only that of the inverter 2 is different. The inhomogeneous dimension of the inverter makes the input pulse undergo different rising and falling time at the

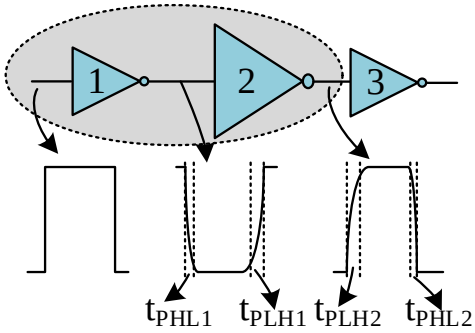

Fig. 3. One stage pulse shrinking delay cell

interface boundaries among the inverter. This mechanism can be used to accurately control the pulse shrinking time.

According to the literature [10], due to the asymmetrical transfer characteristics, when the pulse signal pass through the first inverter, the pulse width will be shrunk for $t_{PHL1} - t_{PLH1}$. The t_{PHL1} defines the response time of the first inverter for a high to low output transition, while t_{PLH1} refers to a low to high transition. For the same reason, the shrunken pulse from former inverter will be shrunk for t_{PHL2}–t_{PLH2} again when passing through the latter inverter. Thus, for one shrinkage cell the pulse width will be shrunk for:

$$\Delta W = [(t_{PHL1} - t_{PLH1}) + (t_{PLH2} - t_{PHL2})]/2 \quad (1)$$

From formula (1), the pulse signal will be shrunk by ΔW when passing through one stage pulse shrinkage cell. With multiple shrinkage cells connected forming a delay line, the pulse with width T will vanish after T/ΔW cells. The resolution is dependent on ΔW that is much smaller than the single cycle of the delay line, thus it can reach high resolution.

III. Proposed Distributed Testing Method

A. Basic Concept of Distributed Delay Testing Architecture

In order to solve large area overhead problems existing in the existing methods, this brief presents a test method based on distributed pulse shrinking time-to-digit converter (TDC). For better understanding, only four interconnects are used to illustrate it, as shown in Fig. 4. The interconnect terminal T*i* is connected to the TDC cell *i*, and another input of TDC cell *i* is the output of TDC cell *i*-1. TDC cell will select the input signal according to the need.

Taking interconnect 1 as an example, signal In1 is propagated to interconnect 1 through a driver and then to T1 through a receiver. Node T1 is connected to TDC cell 1. At this

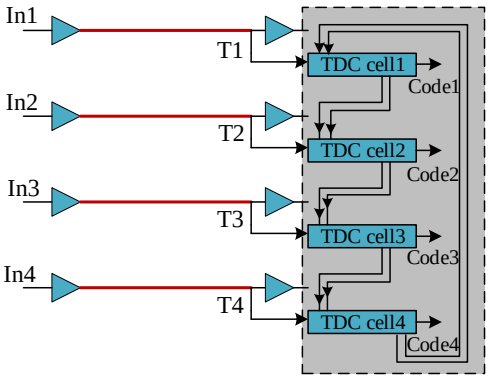

Fig. 4. One stage pulse shrinking delay cell

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCSII.2019.2962824, IEEE Transactions on Circuits and Systems II: Express Briefs

time, TDC cell 1 chooses T1 as input signal, that is, TDC cell 1 as the first stage of pulse shrinking circuit, and the output of TDC cell *i*-1 is chosen as the input of TDC cell *i*, and the final output is in the form of digital code Code1 Code2 Code3 Code4.

When interconnect 2 is tested, TDC cell 2 is the first stage of the pulse shrinking circuit, and the final test result is Code2Code3Code4Code1. When the ring pulse shrinking TDC is designed, the internal circuit of TDC cell will be different, and the test result digital code will be changed to Code1Code2…Code1...

B. *Circuit of TDC Cells*

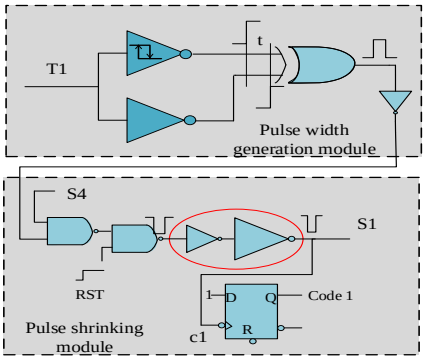

Fig. 5. Circuit of TDC cell 1 based on pulse shrinking

The purpose of our approach is to test interconnect defect by the variations of the delay parameter. These variations will be converted to pulse width, then be measured using pulse shrinking time to digital converter technology. The overall test architecture is composed of pulse width generation and pulse shrinking modules as shown in Fig.5. And the meaning if signals in TDC cells is illustrated in Table I.

The structure of the rise time and fall time is converted to pulse width as shown in Fig. 5. At T1 node, signal enters an inverter and a Schmidt-trigger, respectively. Due to the hysteresis of Schmidt-trigger, the time interval between the two output signals is equivalent to the rising/falling time of the input signals, and thus a pulse width proportional to TSV delay can be generated after a XOR gate. Then, this pulse can be converted into digital code through several pulse shrinking cells.

Pulse shrinking module is similar to Section II.C. The shrinkage value of this shrinking element is equal to the test resolution. The output of each shrinking element is connected to the clock terminal of the D flip flop. The Data terminal of D flip flop (DFF) is constantly tied to '1', thus the DFF will latch '1' if a pulse exists. The DFF will be reset when the pulse is transferred to the next element. That is to say only one DFF is set to '1' before the pulse disappeared.

TABLE I
SIGNALS IN TDC CELLS

Type	Name	Meaning
Input	In *i*	input signal under testing
	S *i-1*	from the output of (*i*-1)-*th* PS cell
	RST	reset
Output	S *i*	the output of *i-th* PS cell
	Code *i*	digital code of TDC cell *i*

For the overall *N* TDC cells, the state of the DFFs is decoded as a final digital code.

$$\text{Digital code} = Nc * Ns + Nd \quad (2)$$

Where Nc is the number of circulation captured by the output counter, Ns is the number of shrinking elements of cyclic delay line (Ns=4 in Fig. 5), and Nd is the first number of DFF to be set to '1' at the end.

The actual delay time can be calculated according to the following formula (3), where offset is the measurement offset to the setup and hold of the DFF.

$$\text{Delay} = \text{Digital code} * \text{resolution} + \text{offset} \quad (3)$$

IV. REGRESSION MODEL BASED ON ARTIFICIAL NEURAL NETWORKS

In this work, we propose a regression model based on artificial neural networks (ANNs). The number of input variables is 4×K (i.e., K different supply voltages), which makes it infeasible to manually find an empirical regression model. Therefore, we use ANNs to build a regression model that can accurately diagnose an interconnect based on measured data.

Fig. 6 shows the architecture of the regression model. Let K be the number of different voltage levels, at which the rise time and fall time are measured. As inputs for the regression model, we provide K sets of four variables within a TDC cell:

(1) $T_{rise,s}$ is the rise time after a Schmidt-trigger;
(2) T_{rise} is the rise time after an inverter;
(3) $T_{fall,s}$ is the fall time after a Schmidt-trigger;
(4) T_{fall} is the fall time after an inverter.

All 4K inputs feed into three ANNs. The first network, Class-net, is a classification network with two binary outputs that determine whether the interconnect under test is fault-free (class fault-free), or open fault (class open). According to the 4K datasets obtained by simulation, the label "open" or "fault-free" can be received for each case. And this network is trained on a large set of simulation data that covers a range of open faults.

The other two networks, R-net and Code-net, are fitting functions that output the open resistance *R*, and the digital code *Code*, respectively. When Class-net outputs class open = 1, we use the output values of R-net as well as Code-net as a prediction of the open-fault size and the degree of fault.

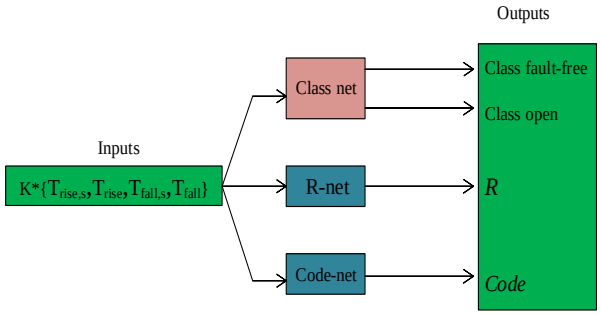

Fig. 6. Regression model based on ANNs.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCSII.2019.2962824, IEEE Transactions on Circuits and Systems II: Express Briefs

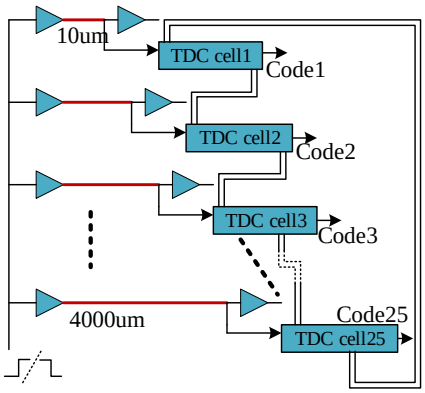

Fig. 7. Pulse shrinking based distributed TDC cells consists of 25 interconnects

V. Experimental Results

To evaluate the performance of the proposed distributed pulse shrinking solution, we perform the simulation with HSPICE and Nangate 45nm Open Cell Library [9], and the proposed method is compared with the ring oscillator (RO) testing method [6], which is the mainstream related work. The shrinking element is composed of two heterogeneous inverters, the device sizes used for the first inverter are 200 nm/50 nm for PMOS and 50 nm/50 nm for NMOS, and the second are 100 nm/50 nm for PMOS and 50 nm/50 nm for NMOS. For other gates, X1 versions are used. And we set the typical supply voltage VDD = 1.1V.

A. Code for Distributed Interconnect PS-based Testing

A group of 25 interconnects were monitored with their lengths ranging from a minimum 10μm to 4000μm, and the resolution is 7ps, as illustrated in Fig. 7. Our architecture is highly scalable in the sense that the same TDC cell can be used for interconnects with even larger delays for longer interconnects.

Fig. 8 shows the obtained digital code of input rising and falling edge in a case of a resistive open fault. As expected, an increase in the resistance R leads to a reduction of digital code from 22 to 11. This indicates that we can detect resistive opens by measuring the digital code directly. When the resistance R increases from 14kΩ to 20kΩ, the digital code does not change and stay at 11, which indicates the interconnect is completely broken. When the resistance R reduces from 0.4kΩ to 0kΩ, the digital code keeps constant and stays at 21, thus open resistances higher than 0.4kΩ can be detected by means of the proposed technique and it is assumed to be no resistive open fault if R is less than 0.4kΩ.

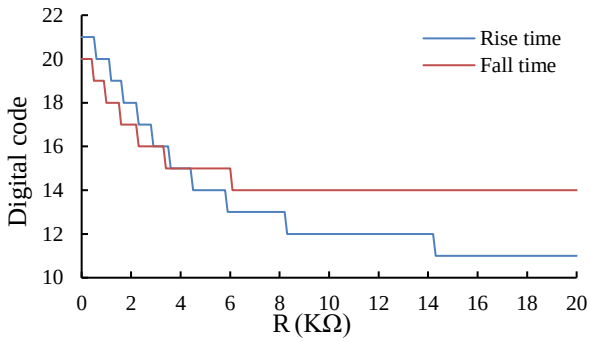

Fig. 8. Digital code measured of rise/fall time in resistive open fault.

B. Impact of TDC-Insertion for Increased Area

The proposed method needs to add a TDC cell at an end of each interconnect. However, the impact of TDC cell "insertion" on the area increase might not be linearly proportional to the number of TDC cells. To prove the scalability of the proposed method, a comparison of before- and after-TDC cell insertion area along the number of inserted TDC cells for RO-based design and the proposed method in terms of increased area ratio to the original area was conducted. It can be found from Fig. 9 that for the proposed method, the area increase ratio is almost linearly proportional to the number of TDC cells, which is much better than the RO design. This shows the scalability of the proposed method.

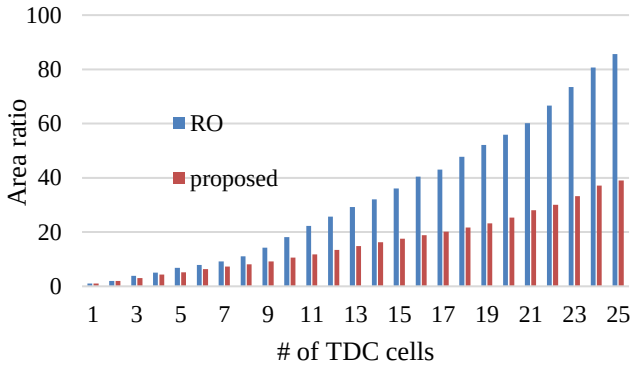

Fig. 9. Comparison between the RO and the proposed methods in terms of increased area ratio linearity along the number of inserted TDC cells.

C. Robustness

Pulse shrinking element is key to the stability of the proposed test method. Monte Carlo simulations are executed to estimate its robustness against supply voltage variation. When the supply voltage changes by ± 10% with Gaussian distribution, the resolution of a pulse shrinking element, and the delay error under different number of Monte-Carlo simulations were calculated, as shown in Fig. 10 and Fig. 11, respectively. It can be found that for the resolution, the average absolute error is 3.7% and the maximum error is 11.8%. Through up to 1000 numbers of simulation experiments, we can draw a conclusion that if we set the test resolution within 11.8% error bound of the estimated value, the confidence level can reach to 99.99%. For the delay error, the maximum error is 1ps, which is negligible.

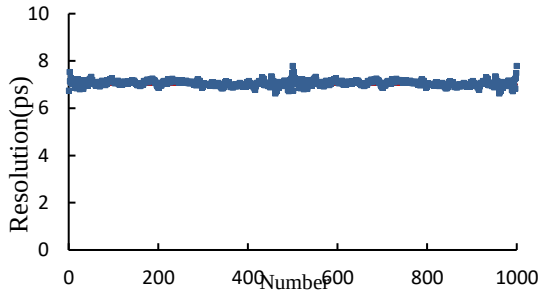

Fig. 10. Variations of the resolution when the power supply variation by ± 10% from its nominal value over one thousand runs.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCSII.2019.2962824, IEEE Transactions on Circuits and Systems II: Express Briefs

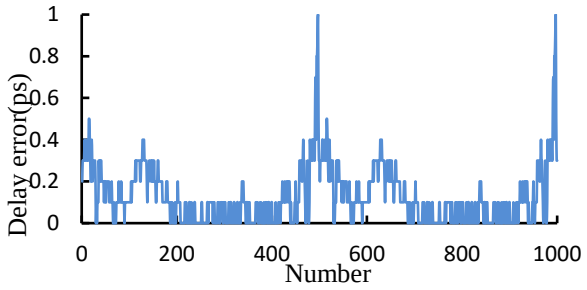

Fig. 11. Variations of the delay error when the power supply variation by ± 10% from its nominal value over one thousand runs.

D. Regression Model for Inferring Fault Size

In order to verify the proposed regression model based on ANNs, we have generated two independent large sets of training and test data with the sample size over 10, 000. Each Monte-Carlo instance was simulated at K = 8 different voltage levels (0.85…. 1.2 V with 50 mV steps), and the corresponding parameters $T_{rise,s}$, T_{rise}, $T_{fall,s}$, and T_{fall} were recorded.

First, we created Class-net in MATLAB and trained it using the training-data set. As the performance evaluation metric, we used the root mean squared error (RMSE). We considered different settings for the ANN architecture and achieved the best performance with the following settings:

(1) One output layer with two neurons (equals the number of output signals);

(2) One hidden layer with ten neurons;

(3) Scaled Conjugate Gradient as training method.

Then, Class-net was evaluated using the test-data set. Fig. 12 shows the confusion matrix from that evaluation. The entry (*i, j*) shows the percentage of samples of type *j* classified as *i* by the network. For instance, the entry (2,1) shows 0.1%, which means that 0.1% of all test points were classified as open, although the actual status is fault-free. According to this matrix, most of the samples were classified correctly (green cells), and only a small percentage of the samples were mis-predicted (red cells), which shows the high accuracy of the classification network.

The network R-net and Code-net were generated using subsets of the generated samples. We used the following settings:

(1) One output layer with one neuron (equals the number of output signals);

(2) One hidden layer with 10-20 neurons (depending on the seed);

(3) Levenberg-Marquardt as training method.

The RMSE rate of the R-net for the proposed and RO methods were trained to be 0.68% and 2.71% on average, respectively. Meanwhile, the RMSE rate of the Code-net for the proposed and RO methods were trained to be 0.60% and 2.30% on average, respectively, as shown in Table II.

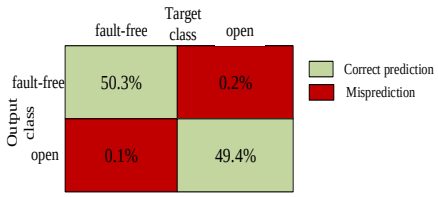

Fig. 12. Confusion matrix for class net.

TABLE II
RMSE RATE UNDER DIFFERENT VOLTAGES

Voltage (V)	RMSE (%)			
	R		Code	
	Proposed	RO	Proposed	RO
0.85	1.43	3.24	1.27	2.98
0.9	1.21	3.15	1.15	2.87
0.95	1.13	3.03	0.96	2.75
1.0	0.71	2.97	0.63	2.54
1.05	0.54	2.75	0.46	2.37
1.1	0.26	2.44	0.18	1.89
1.15	0.13	2.11	0.09	1.64
1.2	0.05	1.97	0.02	1.38
average	0.68	2.71	0.60	2.30

VI. CONCLUSION

In this brief, a non-invasive online solution for 2.5D IC based on distributed pulse shrinking is proposed to test the faults of interconnects. Furthermore, a regression model based on ANN is proposed in order to judge whether the interconnects are faulty and quantify the degree of the faults in real time by online monitoring the delay of interconnects. The experimental results show that the resolution of the proposed method is 7ps, and it can detect open fault at least 0.4kΩ. The area increase ratio is near linearly proportional to the number of TDC cells, and the robustness analysis using Monte-Carlo simulation over one thousand runs shows that power supply variations affect the results by less than 11.8%.

REFERENCES

[1] H. Jooypa, *et al*., "Statistical Strategies to Reproduce the Propagation Delay Time Variability Using Compact Models," *IEEE Trans. Circuits Syst. II, Exp. Briefs*, vol. 66, no. 11, pp. 1880–1884, Nov. 2019.

[2] A. Yan, *et al*., "A Double-Node-Upset Self-Recoverable Latch Design for High Performance and Low Power Application," *IEEE Trans. Circuits Syst. II, Exp. Briefs*, vol. 66, no. 2, pp. 287–291, Feb. 2019.

[3] T. Ni, *et al*., "A region-based through-silicon via repair method for clustered faults," *IEICE Trans. Electron. E100-C (2017) 1108*.

[4] T. Ni, *et al*., "LCHR-TSV: Novel Low Cost and Highly Repairable Honeycomb-Based TSV Redundancy Architecture for Clustered Faults," *IEEE Comput.-Aided Design Integr. Circuits Syst.*, 2019:1-14, early access.

[5] L. R. Huang, *et al*., "Parametric Fault Testing and Performance Characterization of Post-Bond Interposer Wires in 2.5-D ICs," *IEEE Comput.-Aided Design Integr. Circuits Syst.*, vol. 33, no. 3, pp. 476-488, Mar. 2014.

[6] S. Y. Huang, *et al*., "Feedback-Bus Oscillation Ring: A General Architecture for Delay Characterization and Test of Interconnects," *in Proc. of Design. Automation and Test in Europe. Conf*., pp. 924-927, 2015

[7] S. Y. Huang *et al*., "Nonintrusive On-Line Transition-Time Binning and Timing Failure Threat Detection for Die-to-Die Interconnects," *IEEE Trans. Circuits Syst. II, Exp. Briefs*, vol. 34, no. 12, pp. 2039-2048, Feb. Dec. 2015.

[8] S. Y. Huang *et al*., "Versatile Transition-Time Monitoring for Interconnects via Distributed TDC," *IEEE Design & Test*, vol. 33, no. 6, pp. 23-30, Jun. 2016.

[9] Nangate 45 nm Open Cell Library, http://www.nangate.com/openlibr

[10] P. Chen, *et al*., "A CMOS pulse-shrinking delay element for time interval measurement," *IEEE Trans. Circuits Syst. II, Analog Digit. Signal Process*, vol. 47, no. 9, pp. 954–958, Sep. 2000.