Novel Circuit-Level Model for Gate Oxide Short and its Testing Method in SRAMs

Chen-Wei Lin, Mango C.-T. Chao, and Chih-Chieh Hsu

***Abstract*— Gate oxide short (GOS) has become a common defect for advanced technologies as the gate oxide thickness of a MOSFET is greatly reduced. The behavior of a GOS-impacted MOSFET is, however, complicated and difficult to be accurately modeled at the circuit level. In this paper, we first build a golden model of a GOS-impacted MOSFET by using technology CAD, and identify the limitation and inaccuracy of the previous GOS models. Next, we propose a novel circuit-level GOS model which provides a higher accuracy of its dc characteristics than any of the previous models and being is able to represent a minimum-size GOS-impacted MOSFET. In addition, the proposed model can fit the transient characteristics of a GOS by considering the capacitance change of the GOS-impacted MOSFET, which has not been discussed in previous work. Last, we utilize our proposed GOS model to develop a novel GOS test method for SRAMs, which can effectively detect the GOS defects usually escaped from the conventional IDDQ test and March test.**

***Index Terms*— Defect modeling, gate-oxide short, SRAM, testing.**

I. Introduction

GATE oxide short (GOS) is a device defect occurring at the gate insulator of a MOSFET, which may significantly reduce the impedance between the gate and the channel (or source/drain) and in turn affect the electrical behavior of the MOSFET. The root causes of a GOS defect include the oxide rupturing induced by voltage stress, lithographic particles, deviation of oxide growing, or unexpectedly large gate tunneling leakage [1]–[3]. As the gate oxide thickness continually shrinks and the process variation incessantly increases for the CMOS technologies, the probability of having a GOS defect on a manufactured MOSFET becomes much higher than before. Therefore, how to effectively model a GOS defect for test evaluation and further conduct a GOS test is an inevitable and challenging task for advanced technologies [1], [3], [4]–[7].

The GOS defects can be classified into the following two types: the gate-to-source (drain) GOS and the gate-to-channel GOS. A gate-to-source GOS forms a low-impedance path from the gate to the source, which may result from a pinhole of the gate oxide locating in between the gate and the source as shown in Fig. 1(a). Such a gate-to-source GOS can be properly modeled as a resistive short between the gate and the source of the MOSFET [2], [8]–[10], and can be effectively and efficiently detected by using the conventional stuck-at-fault test in logic and general March algorithm in SRAM. Therefore, the gate-to-source GOS is currently considered as an easy-to-detect defect and has hardly attracted any research interest from the testing community during the past decade.

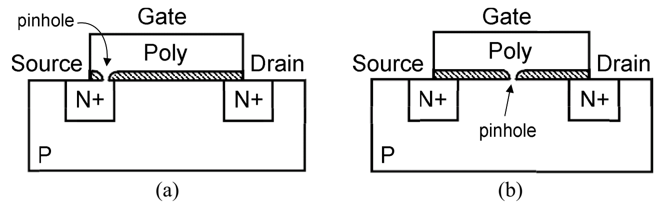


Fig. 1. Cross-sectional view of GOS-impacted MOSFETs. (a) Gate-to-source GOS. (b) Gate-to-channel GOS.

On the other hand, the gate-to-channel GOS forms a low-impedance path from the gate to the channel, as the gate oxide pinhole shown in Fig. 1(b) which was commonly used in previous works [8], [11], [12]. When a gate-to-channel GOS exists in a MOSFET: 1) its gate current is exponentially proportional to its gate voltage; 2) its driving ability is significantly reduced; 3) negative I_D (current flows from channel to drain) exists; and 4) the resistance between the gate and source/drain is gate-voltage controlled. The IDVD curves without and with a gate-to-channel GOS are shown in Fig. 2. Therefore, such faulty behavior of a gate-to-channel GOS is much more complicated to model at the circuit level. In addition, a weak gate-to-channel GOS cannot be detected by the conventional stuck-at-fault test for logics or by the March algorithm for SRAMs, and hence requires IDDQ test to cover [2], [12]–[19].

To properly estimate the effectiveness of the GOS test, several circuit-level models of a gate-to-channel GOS were proposed in the previous works and can be divided into three types: 1) the bi-dimensional model [11]; 2) the split model [2], [12], [14]; and 3) the nonlinear nonsplit model [20]–[22]. The bi-dimensional model contains a mutually connected MOSFET array and hence, requires higher computation time during simulation. In addition, it cannot represent a minimum-size MOSFET and hence, cannot be applied to the most advanced technologies. A split model simplifies the

Manuscript received June 29, 2012; revised January 1, 2013 and April 14, 2013; accepted June 9, 2013. Date of publication July 3, 2013; date of current version May 20, 2014.

C.-W. Lin and M. C.-T. Chao are with the Department of Electronics Engineering, Institute of Electronics, National Chiao Tung University, Hsinchu 30010, Taiwan (e-mail: eeer.ee97g@nctu.edu.tw; mango@faculty.nctu.edu.tw).

C.-C. Hsu is with the Graduate School of Engineering Science and Technology, National Yunlin University of Science and Technology, Yunlin 41349, Taiwan, and also with the Department of Electronic Engineering, National Yunlin University of Science and Technology, Yunlin 64002, Taiwan (e-mail: cchsu@yuntech.edu.tw).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TVLSI.2013.2268984

1063-8210 © 2013 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

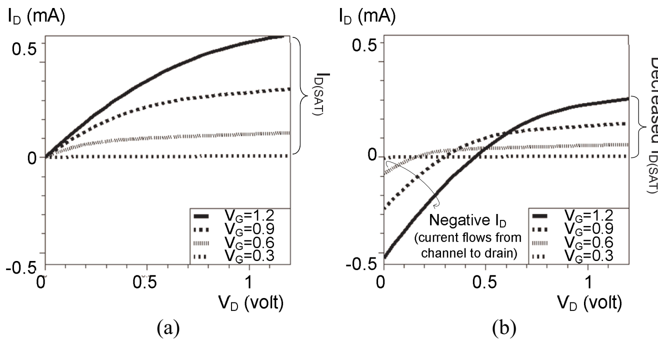


Fig. 2. IDVD curves of an nMOS (a) without and (b) with a gate-to-channel GOS. (a) Defect-free. (b) GOS defective.

TABLE I
DC CHARACTERISTICS AND PARASITIC CAPACITANCES OF A DEFECT-FREE nMOS AND pMOS

DC characteristics:			
MOSFET	V_{th} (volt)	$I_{D(off)}$ (A)	$I_{D(SAT)}$ (A)
nMOS	0.38	3.5E-10	6.4E-4
pMOS	-0.35	-2.1E-9	-2.6E-4
Parasitic capacitance:			
MOSFET	C_G (F)	C_D/C_S (F)	C_B (F)
nMOS	7.71E-17	8.57E-17	1.38E-16
pMOS	7.71E-17	8.75E-17	1.34E-16

complexity of the bi-dimensional model by splitting the MOSFET into two serially-connected MOSFETs and adding a resister in the middle. However, a split model still cannot represent a minimum-size MOSFET. A nonlinear nonsplit model can represent a minimum-size MOSFET by using only one MOSFET in between the source and drain while adding other MOSFETs, resistors, or current sources on the side. However, same as the bi-dimensional model and split model, a nonlinear nonsplit model fails to represent the transient characteristics of a GOS, and hence cannot be applied to SPICE transient simulation.

Note that the bi-dimensional model is considered as the golden model for GOS dc characteristics. The later split model and the nonlinear nonsplit model both tried to match their dc characteristics to the bi-dimensional model. The concept of all the above models makes sense from physics' point of view. However, no previous work validated the correctness of the proposed GOS model through silicon or technology CAD (TCAD) simulation for process technology under 1.5 μm [2], which leaves the precision of the above models unclear. In addition, none of the previous works discussed the transient characteristics of a GOS, which is often required when verifying the effectiveness of a GOS test. Especially several previous works suggested to use delay test for detecting a GOS in logics [23]–[25], which relies on an accurate transient GOS model to evaluate the effectiveness of the test in transient simulation.

In this paper, we propose a novel nonlinear nonsplit model for gate-to-channel GOS defects, which can represent a minimum-size MOSFET and provide a higher accuracy for the GOS dc characteristics. In addition, the proposed model can accurately represent the GOS transient characteristics by considering the capacitance change imposed by the GOS defect, which was not discussed in any of the previous works. A series of experiments will prove the superiority of the proposed model on both the dc and transient characteristics fitting, and all the results will be directly compared with a 3-D TCAD simulation of a GOS-impacted MOSFET instead of to the traditional bi-dimensional model. Next, we will apply our proposed GOS model to evaluate the effectiveness of several previous test methods on detecting GOS defects in SRAMs, identify the limitation of the previous test methods, and then propose a novel GOS test method for SRAMs. We also demonstrate the difference if other GOS models are used in the test evaluation process.

II. EXPERIMENTAL SETUP FOR TCAD AND HSPICE

In this paper, we first run a TCAD simulation to obtain the dc and transient characteristics of a defect-free MOSFET and a GOS-impacted MOSFET, and then use this TCAD simulation result as a golden model to validate the accuracy of each proposed circuit-level GOS model. The TCAD-simulation tool in use is Synopsys' Sentaurs TCAD [26] with a 3-D–structure representation. Note that a gate oxide pinhole in a 2-D structure [as shown in Fig. 1(b)] will form an oxide trench above the channel across the entire channel width, and hence cannot successfully describe a GOS defect. Therefore, the TCAD tool in use must be with a 3-D-structure representation to simulate a GOS-impacted MOSFET. The channel length and width of either a nMOS or a pMOS are both 65 nm for the experiments of this paper. Table I lists the dc characteristics and the parasitic capacitance of a defect-free nMOS and pMOS, respectively.

With the above defect-free devices, a GOS defect can be injected by removing a pinhole of the gate oxide layer in the middle of the channel and filling in with the poly-silicon, where the newly filled poly-silicon pinhole forms a low impedance path between the gate and the channel. The 3-D representation of such a GOS-impacted MOSFET is shown in Fig. 3. The level of a GOS is determined by the radius of the injected pinhole. The larger the injected pinhole, the severer the shorting between the gate and the channel. The pinhole radius shown in Fig. 3 is 2.5 nm.

As our objective is to evaluate the accuracy of a circuit-level GOS model, we first need to make sure that the used SPICE model of a MOSFET can match the behavior of the defect-free MOSFET's TCAD simulation. In our experiments, we use Silvaco's UTMOST [27] to extract the BSIM model card [28] for a MOSFET based on its IDVD, IDVG, gds, gm, and parasitic capacitances obtained from the TCAD simulation. In Fig. 4, we compare the transient response of an inverter obtained by the HSPICE simulation with the extracted BSIM model card to that obtained by the TCAD simulation. The zoom-in result on the output's falling edge and rising edge, respectively, are further shown in Fig. 4(a) and 4(b). As this figure shows, the HSPICE result represented by the red curves

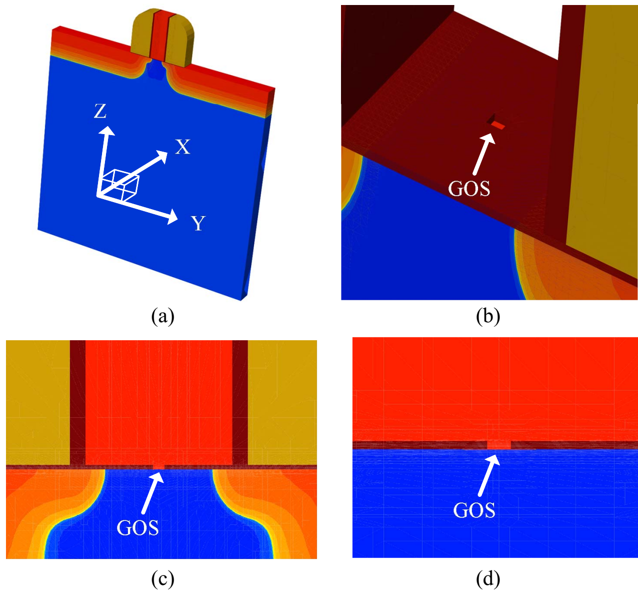


Fig. 3. Representation of a 3-D GOS-impacted MOSFET in TCAD. (a) Whole MOSFET. (b) Gate area with a poly-Si pinhole. (c) X-direction cross section at center. (d) Y-direction cross section at center.

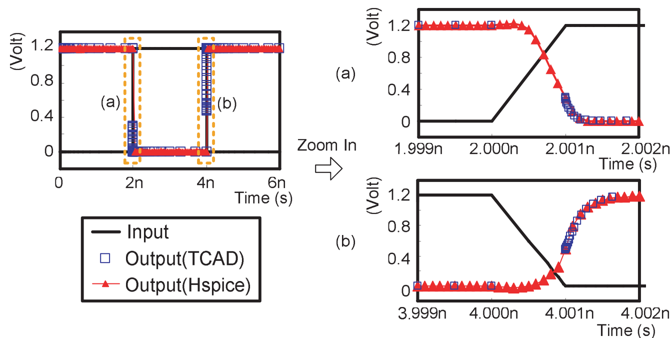


Fig. 4. Comparison of an inverter's transient response between the TCAD simulation and the HSPICE simulation with extracted model cards. (a) Pulling down output response. (b) Pulling-up output response.

fits the TCAD result represented by the blue curves quite precisely.

III. Previous Circuit-Level GOS Models

In this section, we will introduce three circuit-level GOS models: the bi-dimensional model [11] and the two nonlinear nonsplit models [21], [22]. The bi-dimensional model is the first GOS model and the two nonlinear nonsplit models [21], [22] are proposed for representing the minimum-size devices. For both the nonlinear nonsplit models [21], [22], we validate their accuracy by comparing with the TCAD-simulation results and analyze the corresponding limitations.

A. Bi-Dimensional Model

The schematic representation of a 5 × 5 bi-dimensional GOS model, which contains an MOSFET array with all the gates connected together is shown in Fig. 5. The sources of the left-most five MOSFETs are connected together as the source terminal of the GOS-impacted MOSFET, while the drains of the right-most five MOSFETs are connected together as the

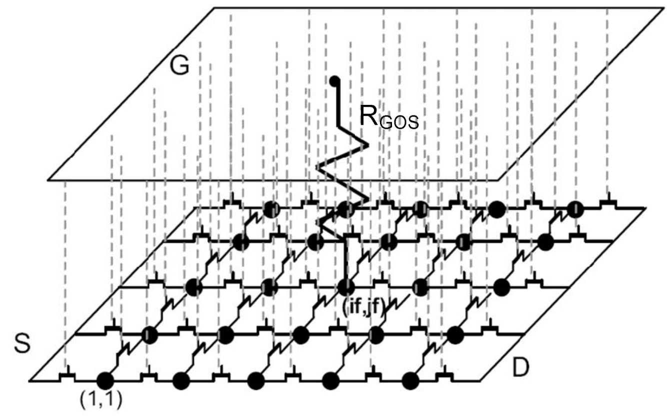


Fig. 5. Exemplary bi-dimensional model with 5 × 5 internal points [11].

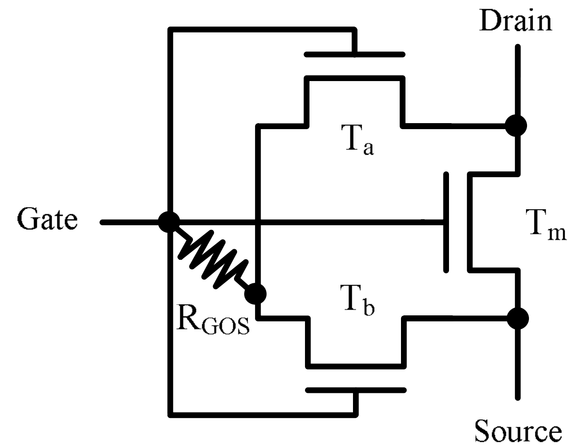


Fig. 6. Nonlinear nonsplit GOS model, JET_03, proposed in [21].

drain terminal. The shorting effect caused by the GOS is represented by the resistor, denoted as R_{GOS}, which connects the gate terminal to the center of the connected MOSFET array. As this model is composed of multiple minimum-size MOSFETs, their combined effect cannot represent a single minimum-size MOSFET with a GOS, which limits the application of this GOS model on a relatively old technologies.

B. Nonlinear Nonsplit Model 1–JET_03

The nonlinear nonsplit model JET_03 proposed in [21], which utilizes three MOSFETs (named T_m, T_a, and T_b in Fig. 6) and a resistor (named R_{GOS} in Fig. 6) to describe a GOS-impacted MOSFET is shown in Fig. 6. According to the fitting guide provided in [21], the size of T_m was tuned to fit the reduced saturation drain current ($I_{D(\mathrm{SAT})}$). The T_a and T_b are tuned to fit the negative drain current and the gate current, respectively. The resistor R_{GOS} is used to refine the fitting.

The IDVD fitting results by using JET_03 model to represent a nMOS with a 2.5-nm-radius GOS are shown in Fig. 7(a). As the result shows, the maximum $I_{D(\mathrm{SAT})}$ and the negative $I_{D(\mathrm{OFF})}$ can be described quite well by JET_03 model in Fig. 7(a), but the other IDVD curves with different V_G cannot. Next, the HSPICE transient simulation of an inverter whose

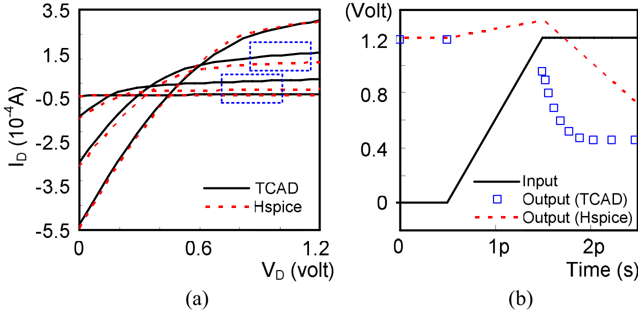


Fig. 7. GOS fitting results by using JET_03 model. (a) DC: IDVD curves. (b) Transient: Inverter response.

nMOS contains the same size GOS modeled by JET_03 model as that in Fig. 7(a) and (b). As the result shows, the inverter's response modeled by JET_03 model cannot match the TCAD simulation result especially when the inverter's response starts to fall. This is caused by the large parasitic capacitance at drain terminal contributed by T_m and T_a. In addition, the parasitic capacitance seen from the gate terminal of the GOS-impacted MOSFET in JET_03 model combines the gate capacitance of three MOSFETs and hence is also larger than a real GOS-impacted MOSFET.

JET_03 model was claimed to be able to describe a minimum-size MOSFET in [21] as its source and drain is connected by only one MOSFET, T_m. However, when fitting the reduced $I_{D(SAT)}$, we have to increase the length of T_m, which turns T_m no longer a minimum-size MOSFET. Similar situation occurs when tuning the size of T_m, T_a, and T_b for fitting other dc curves. Therefore, the size of all three MOSFETs is larger than the minimum size, which results in a further larger gate capacitance of the GOS-impacted MOSFET.

C. Nonlinear Nonsplit Model 2–IDT_09

The nonlinear nonsplit GOS model IDT_09 proposed in [22], which uses only one MOSFET along with an extra three current sources and is more suitable for representing a minimum-size GOS-impacted MOSFET is shown in Fig. 8. Different from JET_03 model, the reduced saturation drain current caused by the GOS is modeled by the "$(1-a)I_D$" current source. The negative drain current and the gate current caused by the GOS are modeled by the other two current sources iGD and iGS, respectively, where iGD and iGS are represented by a third-order polynomial of V_{GD} and V_{GS} as shown in (1) and (2), respectively.

$$iGD = a_1 \cdot V_{GD}^3 + b_1 \cdot V_{GD}^2 + c_1 \cdot V_{GD} + d_1 \quad (1)$$
$$iGS = a_2 \cdot V_{GS}^3 + b_2 \cdot V_{GS}^2 + c_2 \cdot V_{GS} + d_2. \quad (2)$$

The results of applying IDT_09 GOS model to the same experiment is shown in Fig. 9 as in Fig. 7. Most of the IDVD curves match the TCAD result quite well are shown in Fig. 9(a). However, when V_G is low, the corresponding IDVD curve may deviate from the TCAD result at large V_D, which will lead to a high drain current in SPICE simulation when the GOS-impact MOSFET is supposed to be turned off. With further analysis, this fitting error results from the limitation of using a polynomial to represent the current source iGD, even

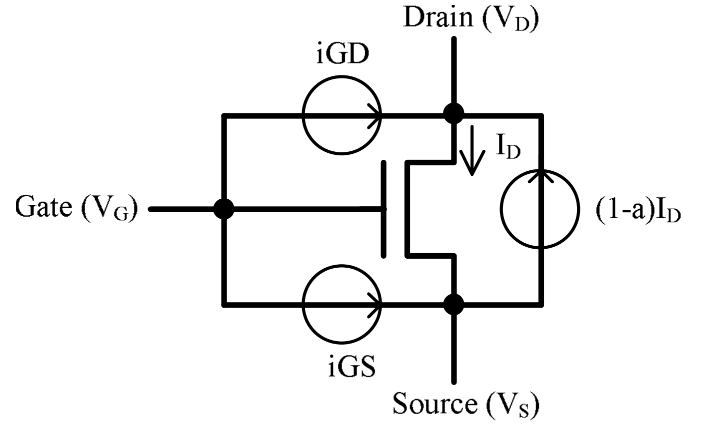


Fig. 8. Nonlinear nonsplit GOS model, IDT_09, proposed in [22].

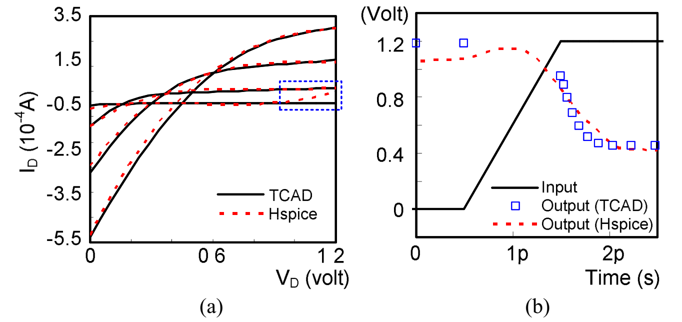


Fig. 9. GOS fitting results by using IDT_09 model. (a) DC: IDVD curves. (b) Transient: Inverter response.

though the order of the polynomial is increased to be more than three.

As the transient-simulation result shown in Fig. 9(b), using IDT_09 model can fit the TCAD result better than using JET_03 model. However, the response of the GOS-impacted inverter based on IDT_09 model is still significantly lower than the TCAD result when the input is low. In addition, the falling slope of the inverter's response based on IDT_09 model is slower than the TCAD result. This error results from the capacitance change induced by the GOS that is not considered in the modeling. Note that this error in the transient simulation will be more significant if the GOS defect is on a pMOS or the size of the GOS is larger. The corresponding experimental results will also be shown in Section IV-C.

IV. Proposed GOS Model and the Comparison With Previous Works

A. Proposed GOS Model

The schematic representation of our proposed GOS model, which uses only one MOSFET along with the three current sources (iGD, iGS, and iSD) and two voltage-controlled capacitors (C_{GS} and C_{GD}) is shown in Fig. 10. Therefore, the proposed GOS model can represent a minimum-size GOS-impacted MOSFET. The three current sources iGD, iGS, and iSD are represented by (3), (4), and (5), respectively, which are different from the equations used in IDT_09 model. First, (3) uses a V_{GD} shift parameter γ and a minimum limitation

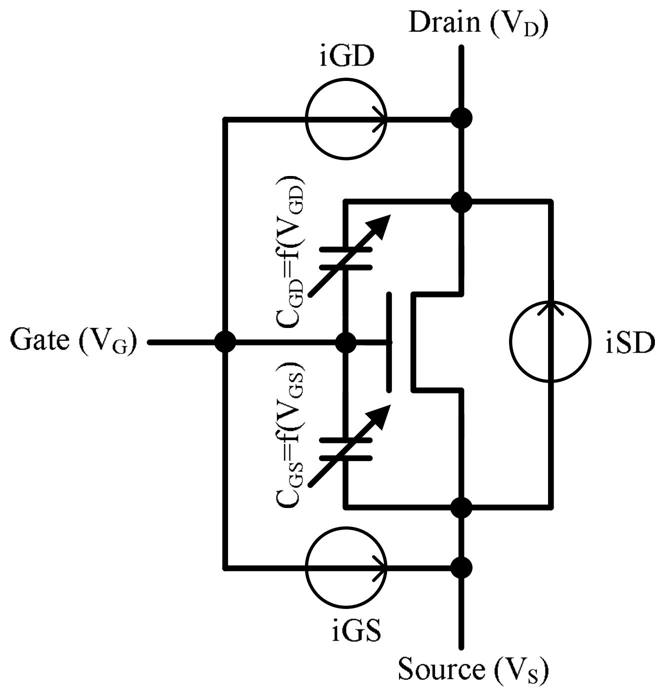


Fig. 10. Schematic representation of proposed GOS model.

for iGD to prevent the overly large drain current when V_G is low and V_D is high, as shown in Fig. 9(a). Second, iSD is actually V_{GS} controlled, not a function of I_D as used in the IDT_09 model, and hence (5) uses a third-order polynomial of V_{GS} to represent iSD. Third, in (4), iGS is simplified to a second-order polynomial of V_{GS} since it can generate the same accuracy as a higher-order one based on our experiments

$$iGD = \left[a_1 \cdot (V_{\mathrm{GD}} + \gamma)^3 + b_1 \cdot (V_{\mathrm{GD}} + \gamma)^2 + c_1 \cdot (V_{\mathrm{GD}} + \gamma) + d_1\right]_{\mathrm{min}=0} \quad (3)$$

$$iGS = a_2 \cdot V_{\mathrm{GS}}^2 + b_2 \cdot V_{\mathrm{GS}} + c_2 \quad (4)$$

$$iSD = a_3 \cdot V_{\mathrm{GS}}^3 + b_3 \cdot V_{\mathrm{GS}}^2 + c_3 \cdot V_{\mathrm{GS}} + d_3. \quad (5)$$

In addition to the three current sources relating to the dc characteristics of a GOS-impacted MOSFET, two voltage-controlled capacitors, C_{GS} and C_{GD}, are utilized in the proposed model to enhance its transient characteristics. C_{GS} and C_{GD} are a piecewise linear function of V_{GS} and V_{GD}, respectively, which is a default expression supported by HSPICE. The CV curves for a defect-free MOSFET, a MOSFET with a 2.5-nm-radius GOS, and a MOSFET with a 5-nm-radius GOS, obtained from our TCAD simulation are shown in Fig. 11. As shown in the figure, $C_{\mathrm{GS/GD}}$ of a GOS-impacted MOSFET is significantly different from that of a defect-free MOSFET and may vary with different $V_{\mathrm{GS/GD}}$. The difference is up to 1.5E-17 F for nMOS and 2.3E-17 F for pMOS, respectively, at $|V_{\mathrm{GS/GD}}| = 1.2$ V. Note that defect-free MOSFET has its gate/source/drain terminal capacitance around 8E-17 F as shown in Table I. In other words, a GOS may lead to 19% and 29% capacitance reduction for nMOS and pMOS, respectively. Therefore, the two voltage-controlled capacitors added in our proposed model are necessary for fully representing the transient characteristics of a GOS-impacted MOSFET.

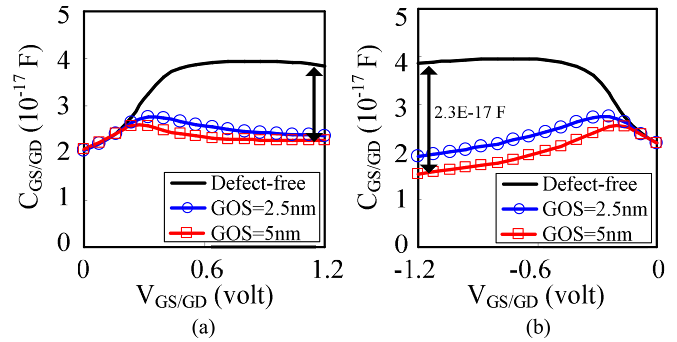


Fig. 11. TCAD CV simulation results for (a) nMOS and (b) pMOS with and without a GOS.

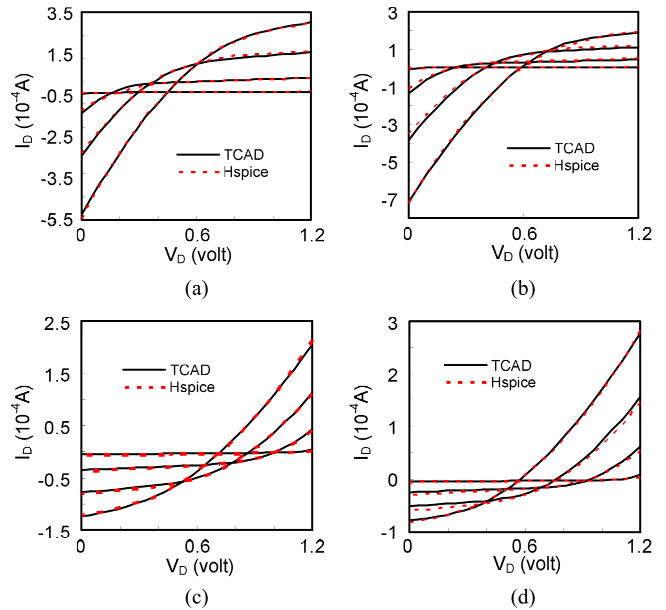


Fig. 12. DC fitting results by using the proposed GOS model. (a) nMOS w/2.5 nm-radius GOS. (b) nMOS w/5 nm-radius GOS. (c) pMOS w/2.5 nm-radius GOS. (d) pMOS w/5 nm-radius GOS.

B. Simulation Comparisons on dc Characteristics

Fig. 12 first shows the IDVD fitting results of the proposed model for different sizes of a GOS on nMOS and pMOS, respectively. As the result shows, the proposed model can closely fit the TCAD-simulation result, which demonstrates the high accuracy of the proposed model on the dc characteristics.

In Table II, we compare the dc fitting errors of the proposed model with those of the bi-dimensional model, JET_03 model, and IDT_09 model. The comparison includes IDVD and IGVG on a GOS-impacted nMOS and pMOS, respectively. The reported error is the root mean square of the difference to each data point of TCAD simulation. As in Table II, all GOS models can fit the IGVG curves fairly well except the bi-dimensional model, which is another evidence why bi-dimensional model cannot represent a minimum-size MOSFET. As to IDVDs, IDT_09 model has less error for most cases when compared with JET_03 model. However, IDT_09 model and JET_03 model may result in 31%∼124% more error than that of our

TABLE II
DC FITTING ERRORS RESULTING FROM DIFFERENT GOS MODELS

GOS MODELs	nMOS			
	GOS radius=2.5nm		GOS radius=5nm	
	IDVD	IGVG	IDVD	IGVG
Bi-dimensional	57.43%	43.55%	56.60%	63.27%
JET_03	59.77%	1.42%	68.98%	0.95%
IDT_09	44.89%	4.30%	81.87%	3.93%
Proposed	5.35%	1.92%	11.78%	2.97%

GOS MODELs	pMOS			
	GOS radius=2.5nm		GOS radius=5nm	
	IDVD	IGVG	IDVD	IGVG
Bi-dimensional	42.62%	15.30%	31.17%	29.26%
JET_03	107.09%	4.97%	137.97%	9.26%
IDT_09	36.40%	4.22%	45.02%	5.56%
Proposed	3.95%	0.46%	13.19%	1.93%

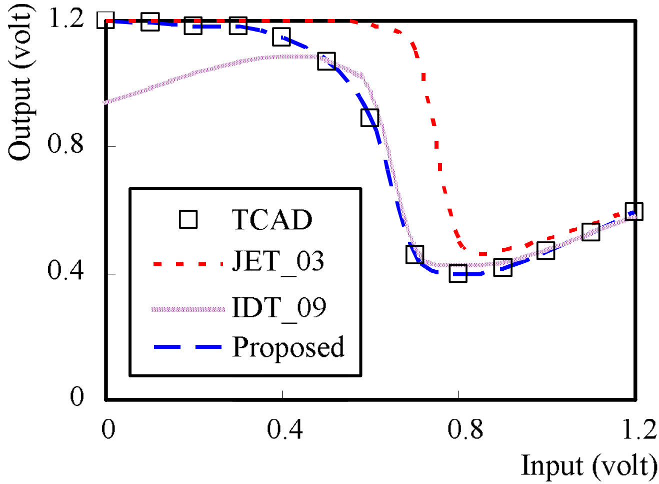


Fig. 13. Inverter's transfer function resulting from TCAD and different GOS models when a 5-nm-radius GOS locates at the nMOS.

proposed model which can limit the IDVD error only from 3.95% to 13.19%. This result demonstrates the better fitting precision of proposed model over the previous ones.

The transfer function of a GOS-impacted inverter simulated with TCAD, JET_03 model, IDT_09, and the proposed model is shown in Fig. 13. The defective inverter contains a minimum-size pMOS and nMOS with a 5-nm-radius GOS on its nMOS. As the result shows, the proposed model closely matches the defective inverter's transfer function obtained by TCAD, while the other previous models do not. The error resulting from JET_03 model is caused by the mismatch of the IDVD curves for different V_G as shown in Fig. 7(a). The error resulting from the IDT_09 model is caused by the overly large turnoff drain current when V_D is high, and hence mostly occurs when the inverter's input voltage is low.

C. Simulation Comparisons on Transient Characteristics

The transient-simulation result of an inverter with a 2.5-nm-radius and 5-nm-radius GOS on its nMOS, respectively, based on TCAD and different GOS models is shown in Fig. 14(a) and 14(b). Note that we only show the inverter's

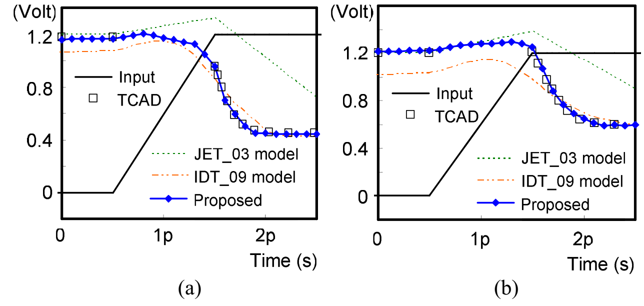


Fig. 14. Transient response of an inverter with a GOS on nMOS resulting from TCAD and different GOS models. (a) 2.5 nm-radius GOS. (b) 5 nm-radius GOS.

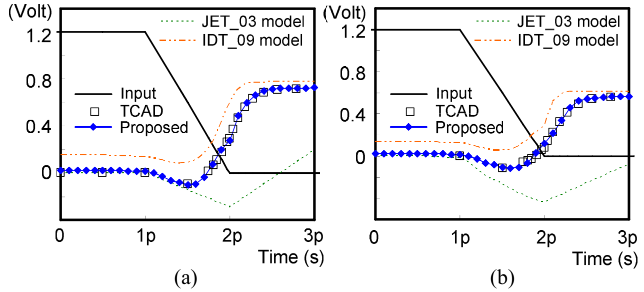


Fig. 15. Transient response of an inverter with a GOS on pMOS resulting from TCAD and different GOS models. (a) 2.5 nm-radius GOS. (b) 5 nm-radius GOS.

response when its input changes from 0 to 1, not from 1 to 0. This is because the GOS-impacted nMOS affects the inverter's response more significantly during the period. As the result shows, the transient-simulation result obtained with the proposed model can closely match the TCAD result while JET_03 model and IDT_09 model cannot. The error caused by JET_03 model is larger when the size of the GOS is smaller. The error caused by IDT_09 model is larger when the size of the GOS is larger.

Similar experiment is applied to the transient simulation of an inverter with the GOS on pMOS, and the result is shown in Fig 15. Again, the same conclusion can be drawn as the proposed model can closely match the TCAD result.

Table III further summarizes the error of an GOS-impacted inverter's output delay time for all the above cases in Figs. 14 and 15, where the output delay is defined as the time period between the 50% of the input signal switch and the 50% of the output signal switch. As the result shows, the error on the inverter's delay caused by the proposed model is less than 1% for all the cases while the other models can result in an error from 17% to 203%. All the above results demonstrate the accuracy of the proposed model on representing the GOS transient characteristics and the importance of considering the GOS-imposed capacitance change in a circuit-level GOS model.

D. Modeling GOS Defects With Different Locations

In this subsection, we discuss the precision of different GOS models when the location of a GOS is not in the middle of the MOSFET's channel. The potential nine locations of a GOS used in the following experiment, denoted from a to i are

TABLE III
FITTING ERRORS OF AN INVERTER'S DELAY CAUSED BY DIFFERENT GOS MODELS

Defect location	Defect radius	TCAD delay (ps)	Models' fitting error		
			JET_03	IDT_09	Proposed
nMOS	defect-free	0.331	-		
	2.5nm	0.689	180%	17%	<1%
	5nm	1.130	203%	27%	<1%
pMOS	defect-free	0.529	-		
	2.5nm	0.689	437%	-29%	<1%
	5nm	0.836	755%	-28%	<1%

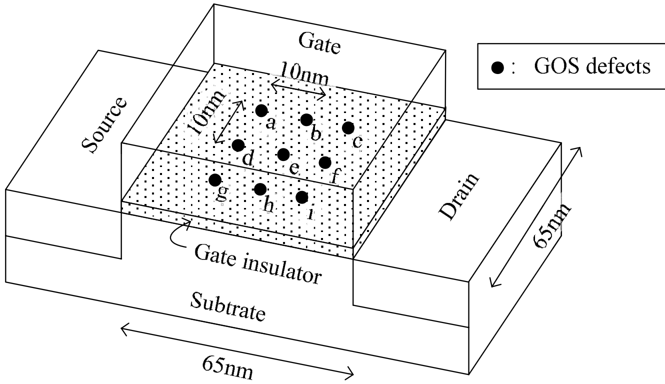


Fig. 16. GOS defect with nine different locations.

TABLE IV
IDVD FITTING ERRORS OF GOS MODELS WITH DIFFERENT DEFECT LOCATIONS

GOS model	Fitting errors (%) with defect locations in Fig. 16						Avg.
	a & g	b & h	c & i	d	e	f	
JET_03	47.5	59.2	64.3	47.5	59.8	64.0	57.1
IDT_09	59.3	55.5	60.1	58.5	44.9	61.0	56.6
Proposed	8.2	11.6	13.9	8.2	5.4	14.5	10.3

shown in Fig. 16. Among the nine locations, e locates at the center of the MOSFET channel, and the rest defect locations surround e with a minimum distance 10-nm of each other. At each location, a 2.5-nm-radius GOS will be injected, and the corresponding dc and transient behaviors will be extracted from TCAD simulation.

Table IV first lists the dc-IDVD fitting errors of each defective nMOS resulting from each GOS model. With JET_03 in use, the fitting error is 57.1% in average, ranging from 47.5% to 64.3%, respectively. With IDT_09 in use, the fitting error is 56.6% in average, ranging from 44.9% to 61.0%, respectively. With our proposed GOS model in use, the fitting error is 10.3% in average, ranging from 5.4% to 14.5%, which is significantly smaller than both previous works. Note that g, h, and i are electrically symmetric to a, b, and c, respectively, and hence their TCAD simulation results are the same in Fig. 16.

From the transient aspect, Table V lists the fitting errors of an inverter's delay time obtained by different GOS models based on different GOS locations. As shown in the table, when

TABLE V
FITTING ERRORS OF INVERTER'S DELAY OF GOS MODELS WITH DIFFERENT DEFECT LOCATIONS

GOS model	Fitting errors (%) with defect locations in Fig. 16						Avg.
	a & g	b & h	c & i	d	e	f	
JET_03	169	229	213	152	180	205	191
IDT_09	16	12	21	16	17	17	17
Proposed	<1	<1	<1	<1	<1	<1	<1

TABLE VI
RESULTS OF APPLYING MARCH C- TEST TO DETECT A GOS IN 6T SRAM

GOS location	GOS radius	
	2.5nm	5nm
Pull-down nMOS	detected	detected
Pull-up pMOS	not detected	detected
Pass-gate nMOS	not detected	not detected

the defective nMOS is modeled by JET_03, all the errors are larger than 150%. With IDT_09 in use, the error is around 17% in average. With the proposed model in use, all the fitting errors can be controlled to less than 1%.

The above experimental results illustrated that the proposed model can model the GOS effect much more accurately no matter where the defect location is within the MOSFET channel.

V. TESTING GOS IN SRAMs

Testing GOS in SRAMs was studied in several previous works [3], [16], [29]. As reported, serious GOS will cause SRAM read/write fail [3] and can be tested by general March tests. As for the weak GOS, IDDQ test is usually recommended for the detection [16], [29]. In this section, we will first evaluate the test effectiveness and limitation of the conventional test methods, March test, IDDQ test, and weak write test mode, for detecting GOS in SRAMs. Next, we will present a novel test method, which utilizes the techniques of floating bit-line writing and bit-line voltage adjustment, to effectively detect GOS in SRAMs. All the experiments and comparisons are made through circuit-level simulation of a 32 K $\times$ 32 SRAM with 1.2 V supply VDD and the cycle time which is 35% longer than the minimum required one of a defect-free cell. GOS injection is based on the proposed GOS model with defect location at channel center for simplicity.

A. Previous Test Methods

1) Conventional March Test: Table VI reports the result of applying the March C-algorithm to detect the GOS with different defect sizes (radius of 2.5 and 5 nm) on different locations. As the result shows, the GOS on pull-down nMOS can be directly detected by the March test for both the defect sizes. They are detected as stuck-at faults, and hence no more discussion is needed for the easy-to-detect defects.

If the GOS is on a pull-up pMOS, only the GOS with larger defect size (radius of 5 nm) can be detected. For defect size of 2.5 nm, although the SRAM has reduced S-SN voltage

TABLE VII
IDDQ SENSITIVITY CAUSED BY A GOS BASED ON A 1.2 V 32K × 32 SRAM

Device	GOS radius	IDDQ at hold	IDDQ at write
Pull-up pMOS	2.5nm	1.1%	1.2%
Pass-gate nMOS	2.5nm	≅ 0	3.8%
	5nm	≅ 0	4.3%

different at 40% of VDD, the voltage on S and SN can keep correct and never flips even during a read. Hence, a small GOS on a pull-up pMOS would escape from the general March test.

If a GOS locates at a pass-gate nMOS, the defective SRAM cell can operate write and read successfully without being detected. During write, the steady voltage supply on BL/BLB provides sufficient current through the pass-gate nMOS to flip the previous data. During read, the pull-down nMOS can generate the current that is 5.1× ∼ 6.9× larger than the GOS induced leakage current. Therefore, the voltage difference on BL/BLB remains recognizable to the sense amplifier.

Note that the GOS defects escaping from a March test in Table VI may become a source of defect level due to the reliability issues and hence are the target to cover in our proposed test method.

2) IDDQ Test: In [16] and [29], IDDQ test was recommended for detecting the GOS escaped from general March tests. Therefore, we only discuss the three escaped cases of GOS shown in Table VI for IDDQ test. The corresponding IDDQ sensitivity [30] which is defined as the ratio of the extra current imposed by the defect over the overall current of the targeted defect-free circuit is shown in Table VII. In addition, we measure the IDDQ sensitivity not only at the hold mode but also at the write mode as recommended by [29]. As the result shows, the IDDQ sensitivity for a GOS at pull-up pMOS is about 1.1% for both hold and write while 3.8% ∼ 4.3% for a GOS at pass-gate nMOS. Such small IDDQ sensitivity is not sufficient for an IDDQ test to effectively detect the defects in practice, especially for an advanced process technologies. Following are the reasons.

First, the IDDQ sensitivity calculated in Table VII is a ratio over the current of the SRAM cell array only. All the peripheral circuits (such as decoders or sense amplifiers) are not included. Second, the power mesh of an SRAM macro is usually shared with other logic blocks or other SRAM macros. Therefore, the IDDQ measurable IDDQ sensitivity in practice will be much smaller. To increase the IDDQ sensitivity, each SRAM macro needs to have an its own independent power mesh for IDDQ measurement, which may introduce tremendous area overhead and design effort as an system-on-a-chip chip can easily contain more than hundreds of SRAM macros. More importantly, the current calculated in Table VII is for the TT corner. The process variation in an advanced process technology can have large impact on device's leakage current. For example, a device's leakage current at the FF corner can be around 45× of that at TT corner for a UMC 65-nm process. Then the corresponding IDDQ sensitivity for the cases in Table VII can be reduced to less than 0.1%. All the above facts make the IDDQ test a less preferable solution to detect GOS in SRAMs.

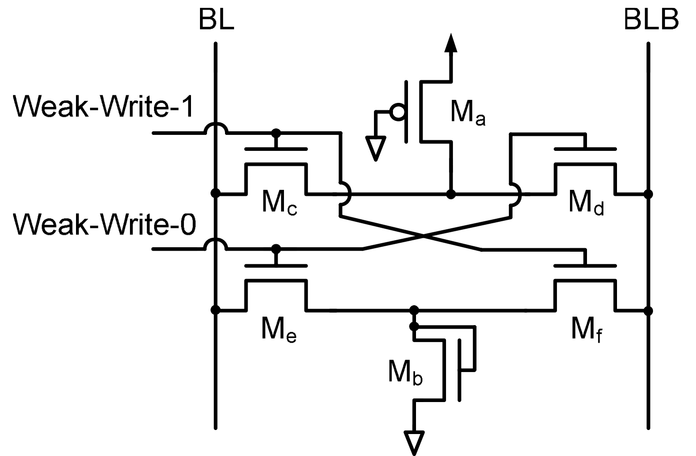


Fig. 17. DFT of weak write test mode [45].

3) Weak Write Test Mode: For detecting data retention faults in the SRAM, weak write test mode (WWTM) [45] is a very common test method. The method sets the weak test write operation on cells by using the BL/BLB voltage setting scheme as shown in Fig. 17. During the weak write operation, BL/BLB connects to VDD/GND through at least two pass-gate MOSFETs. Therefore, the voltage on BL/BLB with logic 1 is lower than the supply voltage. The voltage with logic 0 is higher than the GND. According to the weak write setup, the healthy cell remains the original data, but defective cell would be written into the new data.

In our experiment, we also applied WWTM to test GOS in the SRAM. With all the pass gates $M_a \sim M_f$ in Fig. 17 set to the minimum size (65/65 nm), the WWTM detects the GOSs at pull-down nMOS as well as the large-size GOS at pull-up pMOS. For small-size GOS at pull-up pMOS and GOSs at pass-gate nMOS, the WWTM has to increase the device length of the pass gate to induce the write fail. Nonetheless, the increasing of the device length leads to enormous device characteristic variation. Only few nanometer (1 ∼ 2 nm) of device length range for the pass gate is applicable to distinguish a defective cell from a healthy one. Therefore, the narrow tunable window exceedingly limits the test effectiveness of WWTM for detecting GOS under various process corners.

B. Proposed Design-for-Test Write Operation

To detect GOS in SRAM, we propose a new design-for-test (DFT) write operation in the test mode, which contains two special configurations: 1) floating bit-line (BL and BLB) and 2) adjustable voltage difference between the bit-line pair. The first configuration is to increase the difficulty for a successful write operation while the second one adjusts the test effectiveness and the sensitivity to GOS. The concept of this proposed test method, where the GOS could be on a pull-up pMOS M_1 or a pass-gate nMOS M_2 is shown in Fig. 18. To detect the GOS at either M_1 or M_2, the BL and BLB are both set floating with a capacitance C_T. The C_T may come

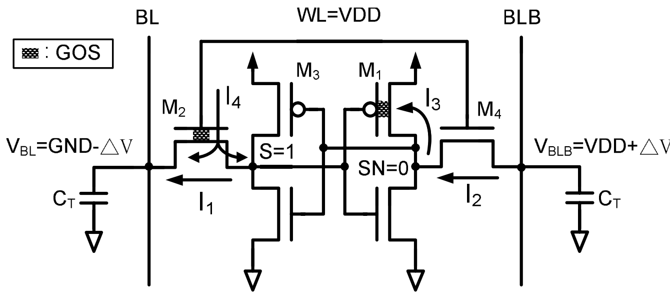


Fig. 18. Concept of the proposed DFT write operation for detecting GOS at pull-up pMOS and pass-gate nMOS.

from the parasitic capacitance of the original bit-line or our intentional added one through the DFT implementation. In this test write operation, the cell is supposed to be written with (S,SN) = (0,1). The voltage at BL and BLB is then set to GND − ΔV and VDD + ΔV, respectively, where the ΔV is used to increase the voltage difference between BL and BLB.

When a defect-free SRAM cell is being written by the proposed DFT write operation, the data stored at S and SN is flipped by only I_1 and I_2, which are caused by their charge-sharing to the capacitance at BL and BLB (C_T). Therefore, this DFT write operation can still successfully write the data if the combination of C_T and the BL-BLB voltage difference can provide sufficient charges to S and SN through I_1 and I_2. However, when a GOS locates at the pull-up pMOS M_1, an extra current I_3 will exist to share significant charges from C_T, which are originally consumed only by I_2, and further fail the DFT write operation. In addition, when a GOS locates at the pass-gate nMOS M_2, a gate current I_4 supplied by the word-line voltage will exist and write data 1 to S, which will fight with I_1 writing data 0 to S. Therefore, the DFT write operation may fail as I_4 is constantly supplied by the word-line voltage while I_1 is supplied by limited charges at C_T. Similar concepts can be applied to detect the GOS locating at the other pull-up pMOS M_3 and pass-gate nMOS M_4 by writing the opposite data.

C. Detailed Simulation Result for Detecting GOS

How the GOS at pull-up pMOS (M_1 in Fig. 18) is detected by the proposed DFT write operation by comparing the simulation results with and without the GOS defect is shown in Fig. 19. The waveforms of $V_S/V_{\rm SN}$, $I_1/I_2/I_3$, and $V_{\rm BL}/V_{\rm BLB}$ when applying the DFT write operation to a defect-free SRAM cell are shown in Fig. 19(a). The corresponding waveforms when applying the DFT write operation to a SRAM cell with a 2.5-nm-radius GOS at M_1 are shown in Fig. 19(b).

As V_S and $V_{\rm SN}$ flip when $V_{\rm BL}$ ($V_{\rm BLB}$) is increased (decreased) to a certain level for the defect-free case, where I_3 is 0 μA is shown in Fig. 19(a),. However, for the defective case shown in Fig. 19(b), $V_{\rm BL}$ ($V_{\rm BLB}$) continues to increase (decrease) since V_S and $V_{\rm SN}$ cannot be flipped due to the existence of I_3 (around 40 μA), which significantly consumes the charges at C_T that is originally used for pulling up $V_{\rm SN}$ through I_2. Once $V_{\rm BL}$ exceeds $V_{\rm BLB}$, the write operation has no chance to be successfully performed.

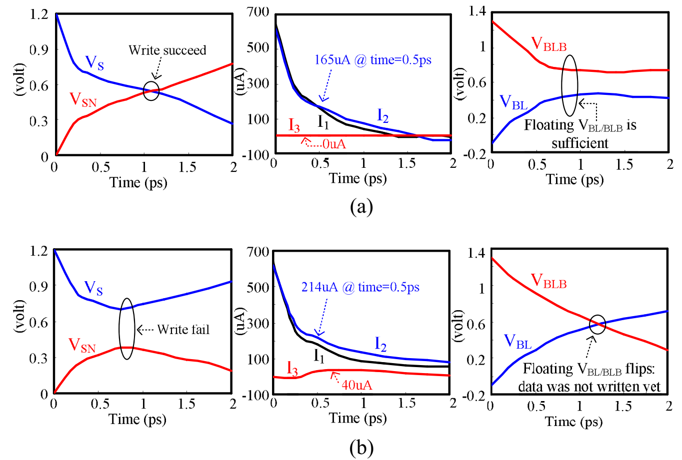


Fig. 19. $V_S/V_{\rm SN}$, $I_1/I_2/I_3$, and $V_{\rm BL}/V_{\rm BLB}$ (a) without and (b) with a GOS at pull-up pMOS when applying the proposed write operation. (a) Defect-free SRAM cell. (b) Defective SRAM cell with pMOS suffering GOS.

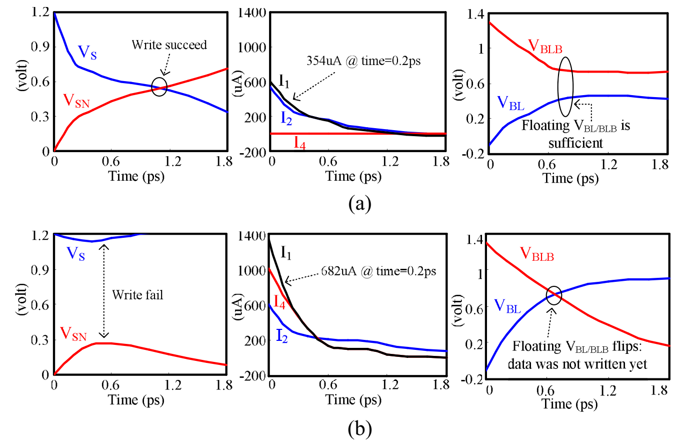


Fig. 20. $V_S/V_{\rm SN}$, $I_1/I_2/I_3$, and $V_{\rm BL}/V_{\rm BLB}$ (a) without and (b) with a GOS at pass-gate nMOS when applying the proposed write operation. (a) Defect-free SRAM cell. (b) SRAM cell with pass-gate nMOS suffering GOS.

A similar experiment is applied to an SRAM cell with a 2.5-nm-radius GOS at the pass-gate nMOS M_2. The corresponding waveforms without and with the GOS at M_2, respectively are shown in Fig. 20(a) and 20(b). As $V_{\rm BL}$ exceeds $V_{\rm BLB}$ faster than that for the case of Fig. 19(b) while V_S ($V_{\rm SN}$) decreases (increases) very limited is shown in Fig. 20(b). This is because I_4 induced by the GOS at M_2 is much higher than I_3 induced by the GOS at M_1 and can quickly pull up V_S that is supposed to be pulled down. Therefore, the GOS at a pass-gate nMOS is relatively easier to detect than the GOS at a pull-up pMOS by the proposed DFT write operation.

D. Finding Valid Setting for C_T and ΔV

The ability of writing a cell with the proposed DFT write operation is determined by two factors: 1) the BL/BLB capacitance of C_T and 2) the ΔV that adjusts the voltage difference between BL and BLB. The larger the C_T or ΔV, the easier the DFT write operation can be successfully performed. If the (C_T, ΔV) combination is too large, both defect-free and defective cases would pass the DFT write operation, which

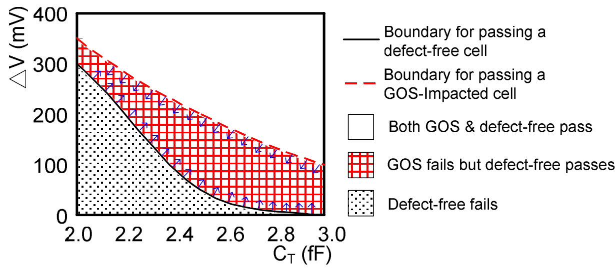


Fig. 21. Finding valid (C_T, ΔV) combinations for detecting GOS at pull-up pMOS.

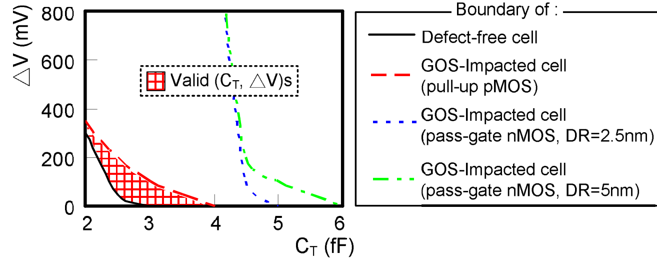


Fig. 22. Finding valid (C_T, ΔV) combinations for detecting GOS at pull-up pMOS or pass-gate nMOS.

results in test-escape. If the (C_T, ΔV) combination is too small, both defect-free and defective cases would fail the DFT write operation, which is over-kill. Therefore, our objective is to find a valid value for both C_T and ΔV such that the designed DFT write operation can pass for a defect-free cell and fail for a GOS-impacted cell. In the following experiment, we attempt to find a proper combination of C_T and ΔV for the proposed DFT write operation based on the HSPICE simulation using the proposed GOS model.

Simulation result for a 2.5-nm-radius GOS at a pull-up pMOS is shown in Fig. 21. The black solid line represents the boundary of the (C_T, ΔV) combinations that can pass a defect-free cell, where any (C_T, ΔV) combination above the line can successfully write a defect-free cell. The red dashed line represents the boundary of the (C_T, ΔV) combinations that can pass the GOS-impacted cell, where any (C_T, ΔV) combination below the line cannot successfully write the GOS-impacted cell. A (C_T, ΔV) combination falling in the intersection of the two regions (high-lighted by red) is a valid setting for the proposed DFT write operation.

Fig. 22 further includes the (C_T, ΔV) boundaries for the case where the GOS locates at a pass-gate nMOS with two radius sizes. As the result shows, the boundaries for the GOS at a pass-gate nMOS are all far above the boundary for the GOS at a pull-up pMOS, meaning that it is easier to find a (C_T, ΔV) combination that can detect the GOS at a pass-gate nMOS (more combinations below the boundary). This result also shows that once a (C_T, ΔV) combination is valid for detecting the GOS at a pull-up pMOS, the combination is also valid for detecting the GOS at a pass-gate nMOS. Therefore, when designing the proposed DFT write operation, we only need to consider the case for the GOS at a pull-up pMOS.

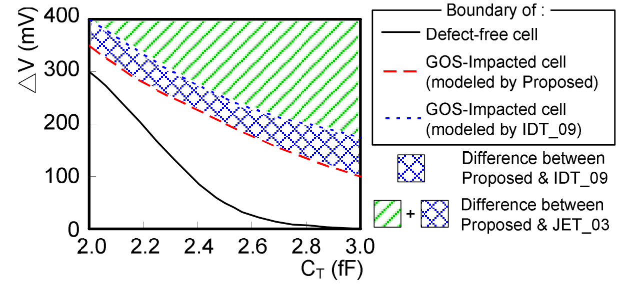


Fig. 23. Finding valid (C_T, ΔV) combinations by using different GOS models.

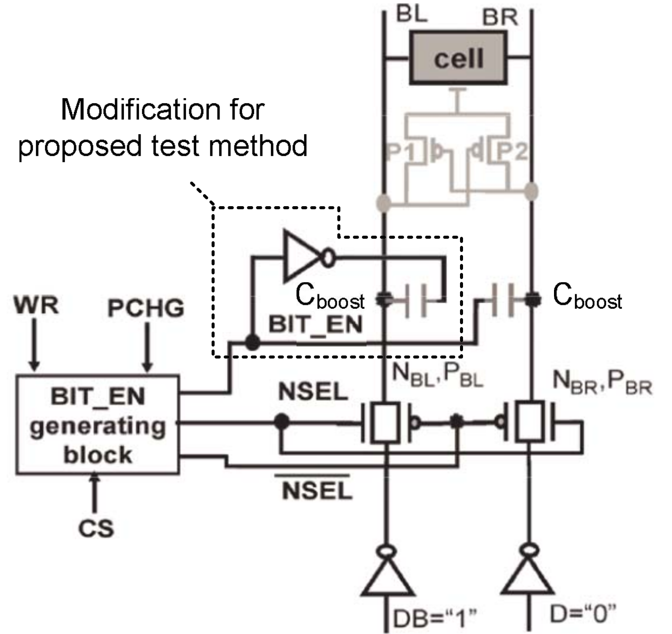


Fig. 24. Mutually inverse boosting circuitry modified from [32].

E. Comparison With Other GOS Models in Use

To obtain an accurate valid (C_T, ΔV) combination through simulation, an accurate circuit-level GOS model is the key. Fig. 23 further shows the boundaries of the (C_T, ΔV) combinations that can detect the corresponding GOS as shown in Fig. 21 when other GOS models, such as IDT_09 model and JET_03 model, are used instead of the proposed GOS model. As the result shows, the blue region represents the (C_T, ΔV) combinations that are considered as valid by using IDT_09 model but not valid by using the proposed GOS model. This significant blue area indicate the risk of using IDT_09 model since the experiment in Section IV demonstrated the accuracy of the proposed GOS model. If JET_03 model is used, all the green area plus the blue region are considered as valid (C_T, ΔV) combinations, which may lead to an incorrect conclusion of a valid (C_T, ΔV) combination even more easily. A false-valid (C_T, ΔV) combination in the above cases may lead to a test escape in reality since both defect-free and defective cases can pass the proposed DFT write operation with the larger (C_T, ΔV) setting.

VI. Implementation of the Proposed Test Method and Its Optimization

To realize the proposed write operation, two DFT hardware components need to be included into the SRAM design. The first one is the C_T adjusting scheme for BL and BLB. The other is the positive (or negative) voltage boosting circuit which is used for generating the VDD + ΔV (and GND − ΔV). In this section, we first discuss the implementation of the DFT components and the approximate area overhead. Next, we attempt to find the optimum (C_T, ΔV) for minimizing the DFT area overhead while keeping the test effectiveness. Finally, the discussion of properly test configuration setting against the process variation is given.

A. Implementation and Area Overhead of DFT

To implement the C_T adjusting scheme, we use the common metal-insulator-metal capacitor (MIMCAP) for the extra capacitance on BL and BLB. As a valid C_T ranges from 2 to 4 fF is shown in Fig. 22. A bit-line in our SRAM design connects to 128 cells and its parasitic capacitance is 13 fF. For the adopted 65nm technology, the MIMCAP is 5 fF/μm^2 and the area of a SRAM cell is 0.69 μm^2, which are the same parameters used by some previous publications [42], [43]. In this case, the cell area of a column is 128*0.69 μm^2, i.e, 88.32 μm^2. If C_T ranges from 2 to 4 fF, the resulting capacitor area ranges from 0.8 to 1.6 μm^2 (including BL's and BLB's). Therefore, the area overhead of the extra capacitance C_T is 0.9% to 1.8% of the cell area, which is relatively low.

Another area overhead of the proposed test method is the bit-line boost circuitry, which is already a popular design technique used in SRAM to improve the writability under low supply voltage and regain the design margin eaten by the continually increasing process variation. References [31], [32], and [46]–[51] are some of the recent publications (from 2008 to 2012) using boost circuitry in SRAM designs. Most of the publications are from major semiconductor companies, such as IBM, ARM, Renesas, and Freescale, and are validated with silicon result. As reported by [31], [32], [46], and [47], the area overhead of a boost circuitry can be reduced to as low as 5% (ranging from 5% to 7%). In addition, once a bit-line boost circuitry is used, its writability and design margin can be improved at the same time. Then this 5% area overhead can be shared by the other design purposes.

Note that most of the above boost circuitries [31], [32] [46]–[51] can be easily modified for the use of the proposed test method. For example, How the bit-line booster in [32] can be turned into the use of the proposed test method by simply adding an inverter is shown in Fig. 24.

In addition to the two DFT components discussed above, we also recommend two design skills to improve the test quality. First, pass-gate switches should be added between BL/BLB and C_T as well as BL/BLB and the voltage boosting circuitry. The pass-gate switches should turnoff to isolate the BL/BLB from the DFT components after the test operation. By doing so, the bit-line capacitance loading would restore, and the sense amplifier at the following operation can work appropriately to read the test results without being affected by the DFT components. The second recommended skill is to apply the mirror symmetry in the SRAM cell layout. By applying the symmetric layout style, each BL is adjacent to one BL and one BLB rather than two BLBs. The purpose here is to reduce the possible coupling between BL and BLB when they are being boosted to VDD + ΔV and GND − ΔV, respectively.

B. Optimum (C_T, ΔV) for Minimizing Area Overhead

The (C_T, ΔV)s in Fig. 22 with red background are all the valid ones for detecting the GOS defect. However, the DFT area overhead for each (C_T, ΔV) setting varies from one another because the required sizes of C_T and C_{boost} are different. To find the optimum (C_T, ΔV) for achieving low area overhead while keeping good test efficacy, we use 6 to 9 for calculating the capacitor-occupied area. Equation 6 describes the relationship between the ΔV in Fig. 22 and the C_{boost} in Fig. 24 when the ΔV is generated by the modified voltage boosting scheme. Equation 7 calculates the occupied area by the two extra capacitors C_{boost} and C_T. In the equation, both C_{boost} and C_T are counted twice because one is for BL and the other is for BLB. The γ_c denotes the capacitance of a unit-size (1 μm^2) MIMCAP provided by the process

$$\Delta V = V_{\text{DD}} \cdot \frac{C_{\text{boost}}}{C_{\text{boost}} + C_T + C_{\text{bitline}}} \tag{6}$$

$$\text{Area}_{(C_{\text{boost}} \& C_T)} = \frac{2 \cdot (C_{\text{boost}} + C_T)}{\gamma_c}. \tag{7}$$

We rearrange 6 into 8, and substitute the C_{boost} in 7 with the one in 8. Finally, we acquire the capacitor-occupied area presented as 9, which is a function of C_T, C_{bitline}, ΔV, V_{DD}, and γ_c

$$C_{\text{boost}} = \frac{\Delta V (C_T + C_{\text{bitline}})}{(V_{\text{DD}} - \Delta V)} \tag{8}$$

$$\text{Area}_{(C_{\text{boost}} \& C_T)} = \frac{2 \cdot (V_{\text{DD}} \cdot C_T + \Delta V \cdot C_{\text{bitline}})}{(V_{\text{DD}} - \Delta V) \cdot \gamma_c}. \tag{9}$$

In our experiment, the C_{bitline} is 13 fF, V_{DD} is 1.2 V, and γ_c is 5 fF/μm^2. By defining the range of C_T (2 ∼ 4 fF) and ΔV (0 ∼ 400 mV), the capacitor-occupied area of each valid (C_T, ΔV) can be calculated and drawn as Fig. 25. In Fig. 25, we label four nodes, from A to D, along the middle line between the boundary of defective cells and the boundary of healthy cells. Note that the (C_T, ΔV) represented on the middle line are more preferred since a (C_T, ΔV) too close to the boundary of healthy cells may cause over-test and a (C_T, ΔV) too close to the boundary of defective cells may cause test escape.

The extra capacitor-occupied area resulting from the (C_T, ΔV)s along the targeted middle line is shown in Fig. 26. As the result shows, the minimum overhead of the extra capacitor-occupied area is 1.31 μm^2 and occurs when C_T = 3 fF and ΔV = 20 mV.

C. Maximization of Tolerable ΔV Range Against the Process Variation

The ΔV on the bitline for test is generated by the boosting circuitry. The existence of process variation affects the booster

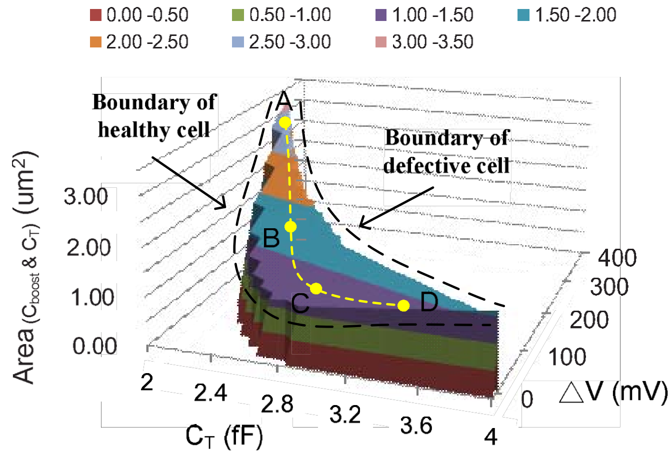


Fig. 25. Capacitor-occupied area of each valid (C_T, ΔV).

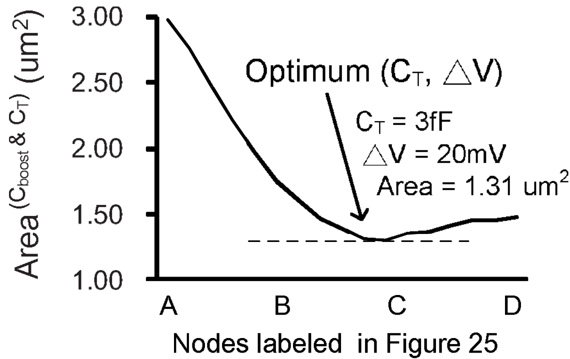


Fig. 26. Capacitor-occupied area of the preferred (C_T, ΔV)s in Fig. 25.

characteristics and leads to the fluctuation of the resulting ΔV value. The variation immunity of the proposed test method depends on the tolerable ΔV range, which is defined as the difference of the ΔV between the boundary of a defective cell and a healthy cell as shown in Fig. 25. As long as variation-shifted ΔV falls in between the two boundaries, the proposed test method can still differentiate a defective cell from a healthy cell. Therefore, a setting with higher tolerable ΔV range can tolerate larger process variation on the booster circuitry.

Instead of minimizing the area overhead of (C_{boost}+C_T) as shown in Section VI-B, we can also select a proper C_T to maximize its tolerable ΔV range. The tolerable ΔV range resulting from different selected C_T is shown in Fig. 27. The largest tolerable ΔV range is 152 mV, which occurs when C_T is 2.6 fF and is labeled by P_1 as shown in Fig. 27. If we try to minimize (C_{boost}+C_T) as shown in Section VI-B, the resulting tolerable ΔV range will be 91 mV, which is labeled by P_2. Note that the ΔV fluctuation reported by the previous boost-circuitry works [31], [32] is around ±25 mV, i.e., a range of 50 mV. Therefore, both P_1 and P_2 are well above this 50 mV ΔV fluctuation, showing that the proposed test method can still be effective under the potential process variation on the boost circuitry.

Table VIII further compares the best settings for maximizing the tolerable ΔV range (P_1) and minimizing the capacitance overhead (P_2). As the result shows, P_1 can result in a 61 mV

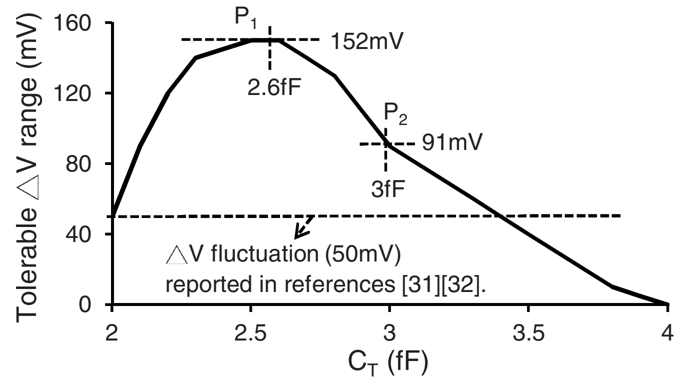


Fig. 27. Tolerable ΔV range versus differently selected C_T with two recommended configurations for high-process-variation-immunity and low-area-cost purposes, respectively.

TABLE VIII
COMPARISON BETWEEN THE TWO TEST CONFIGURATIONS P_1 AND P_2 LABELED IN FIG. 27

Label	Test config.		Area overhead of (C_{boost}+C_T)	Tolerable ΔV range
	C_T	ΔV		
P_1	2.6fF	90mV	1.7% (1.55μm^2)	152mV
P_2	3fF	20mV	1.4% (1.31μm^2)	91mV

higher tolerable ΔV range while P_2 can result in a 0.3% lower area overhead. The designers can balance this tradeoff based on the actual need of the product line.

VII. CONCLUSION

In this paper, we first built a TCAD simulation environment to help us evaluate the accuracy of a circuit-level GOS model. Next, we proposed a novel nonlinear nonsplit GOS model, which can provide higher accuracy on fitting the dc characteristics of a GOS than the previous models while being able to represent a minimum-size GOS-impacted MOSFET. In addition, the proposed GOS model enabled an accurate transient simulation of a GOS-impacted MOSFET by considering the capacitance change imposed by a GOS, which was never discussed in previous works. In addition, based on this GOS model, we developed a novel DFT write operation, which used the techniques of floating-bit-line writing and voltage-difference adjustment at bitlines. The experimental results demonstrated that the proposed DFT write operation can effectively detect the GOS defects that cannot be detected by the conventional March tests or IDDQ tests.

REFERENCES

[1] K. Roy, S. Mukhopadhyay, and H. Mahmood-Meimand, "Leakage current mechanisms and leakage reduction techniques in deep-submicrometer CMOS circuits," *Proc. IEEE*, vol. 91, no. 2, pp. 305–327, Feb. 2003.

[2] J. Segura, C. De Benito, A. Rubio, and C. F. Hawkins, "A detailed analysis of GOS defects in MOS transistors: Testing implications at circuit level," in *Proc. Int. Test Conf.*, Oct. 1995, pp. 544–551.

[3] A. K. Bin, K. S. Sim, and C. K. Siong, "The effects of gate oxide short in 6-transistors SRAM cell," in *Proc. IEEE Int. Semicond. Electron.*, Dec. 2004.

[4] T. W. Chen, K. Kim, Y. M. Kim, and S. Mitra, "Gate-oxide early life failure prediction," in *Proc. VLSI Test Symp.*, Apr./May 2008, pp. 111–118.

[5] J. M. Galliere, F. Azais, M. Renovell, and L. Dilillo, "Influence of gate oxide short defects on the stability of minimal sized SRAM core-cell by applying non-split models," in *Proc. 4th Int. Conf. Design Technol. Integr. Syst. Nanoscale Era*, Apr. 2009, pp. 225–229.

[6] A. Ille, W. Stadler, T. Pompl, H. Gossner, T. Brodbeck, K. Esmark, P. Riess, D. Alvarez, K. Chatty, R. Gauthier, and A. Bravaix, "Reliability aspects of gateoxide under ESD pulse stress," *Microelectron. Rel.*, vol. 49, no. 12, pp. 1407–1416, 2009.

[7] L. Lakshmanan, L. Herlinger, and K. Miller, "A sample preparation technique to localize gate oxide defects in memory arrays using conducting atomic force microscopy," in *Proc. Int. Symp. Test. Failure Anal.*, Nov. 2011, pp. 218–222.

[8] H. Hong and E. J. McCluskey, "Analysis of gate oxide shorts in CMOS circuits," *IEEE Trans. Comput.*, vol. 42, no. 12, pp. 1510–1516, Dec. 1993.

[9] H. Hong and E. J. McCluskey, "On the modeling and testing of gate oxide shorts in CMOS logic gate," in *Proc. Int. Workshop Defect Fault Tolerance VLSI Syst.*, Nov. 1991, pp. 161–174.

[10] H. Hong and E. J. McCluskey, "Very-low-voltage testing for weak CMOS logic ICs," in *Proc. Int. Test Conf.*, Oct. 1993, pp. 275–284.

[11] M. Sytrzycki, "Modeling of gate oxide shorts in MOS transistors," *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst.*, vol. 8, no. 3, pp. 193–202, Mar. 1989.

[12] J. Segura, A. Rubio, and J. Figueras, "Analysis and modeling of MOS devices with gate oxide short failures," in *Proc. IEEE Int. Symp. Circuits Syst.*, Jun. 1991, pp. 2164–2167.

[13] S. I. Syed and D. M. Wu, "Test generation for gate oxide short in CMOS ICs," in *Proc. IEEE Southeastcon*, Apr. 1990, pp. 1062–1065.

[14] R. Rodriguez-Montanes, J. A. Segura, V. H. Champac, J. Figueras, and J. A. Rubio, "Current vs. logic testing of gate oxide short, floating gate and bridging failures in CMOS," in *Proc. Int. Test Conf.*, Oct. 1991.

[15] D. M. Wu, "Detecting gate-oxide shorts and delay defects in CMOS IC's," in *Proc. IEEE Southeastcon*, Apr. 1991, pp. 16–18.

[16] J. A. Segura and A. Rubio, "GOS defects in SRAM: Fault modeling and testing possibilities," in *Proc. IEEE Int. Workshop Memory Technol.*, Aug. 1994, pp. 66–71.

[17] J. M. Soden and C. F. Hawkins, "Test considerations for gate oxide shorts in CMOS ICs," *IEEE Design Test Comput.*, vol. 3, no. 4, pp. 56–64, Aug. 1986.

[18] S. I. Syed and D. M. Wu, "Defect analysis and test generation for gate oxide shorts in CMOS ICs," in *Proc. IEEE Custom Integr. Circuits Conf.*, May 1990, pp. 28.6/1–28.6/4.

[19] S. I. Syed and D. M. Wu, "Defect analysis, test generation and fault simulation for gate oxide shorts in CMOS ICs," in *Proc. IEEE Int. Symp. Circuits Syst.*, May 1990, pp. 2705–2707.

[20] M. Renovell, J. M. Galliere, F. Azais, and Y. Bertrand, "Modeling gate oxide short defects in CMOS minimum transistors," in *Proc. IEEE Eur. Test Workshop*, May 2002, pp. 15–20.

[21] M. Renovell, J. M. Galliere, F. Azais, and Y. Bertrand, "Modeling the random parameters effects in a non-split model of gate oxide short," *J. Electron. Test.*, vol. 19, no. 4, pp. 377–386, 2003.

[22] S. Chehade, A. Chehab, and A. Kayssi, "Modeling of gate oxide short defects in MOSFETS," in *Proc. 4th Int. Design Test Workshop*, Nov. 2009, pp. 1–4.

[23] M. Renovell, J. M. Galliere, F. Azais, and Y. Bertrand, "Delay testing of MOS transistor with gate oxide short," in *Proc. 12th Asian Test Symp.*, Nov. 2003, pp. 168–173.

[24] X. Lu, Z. Li, W. Qiu, D. M. H. Walker, and W. Shi, "A circuit level fault model for resistive shorts of MOS gate oxide," in *Proc. Fibres Opt. Passive Compon.*, 2005, pp. 97–102.

[25] J. M. Galliere, M. Renovell, F. Azais, and Y. Bertrand, "Delay testing viability of gate oxide short defects," *J. Comput. Sci. Technol.*, vol. 20, no. 2, pp. 195–200, 2005.

[26] (2013). *Technology Computer Aided Design (TCAD): Process and Device Simulation Tools* [Online]. Available: http://www.synopsys.com

[27] (2013). *UTMOST IV: Optimization Module for Compact/Macro-Modeling* [Online]. Available: http://www.silvaco.com

[28] (2013). *BSIM: Berkeley Short-Channel IGFET Model* [Online]. Available: http://www-device.eecs.berkeley.edu/~bsim

[29] J. Segura and A. Rubio, "A detailed analysis of CMOS SRAM's with gate oxide short defects," *IEEE J. Solid-State Circuits*, vol. 32, no. 10, pp. 1543–1550, Oct. 1997.

[30] S. Bhunia, L. Hai, and K. Roy, "A high performance IDDQ testable cache for scaled CMOS technologies," in *Proc. 11th Asian Test Symp.*, Nov. 2002, pp. 157–162.

[31] M. Yabuuchi, K. Nii, Y. Tsukamoto, S. Ohbayashi, Y. Nakase, and H. Shinohara, "A 45 nm 0.6 V cross-point 8T SRAM with negative biased read/write assist," in *Proc. Symp. VLSI Circuits*, Jun. 2009, pp. 158–159.

[32] S. Mukhopadhyay, R.-M. Rao, J.-J. Kim, and C.-T. Chuang, "SRAM write-ability improvement with transient negative bit-line voltage," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 19, no. 1, pp. 24–32, Jan. 2011.

[33] A. Karandikar and K. K. Parhi, "Low power SRAM design using hierarchical divided bit-line approach," in *Proc. Int. Conf. Comput. Design, VLSI Comput. Process.*, Oct. 1998, pp. 82–88.

[34] V. Dabholkar and S. Chakravarty, "Computing stress tests for gate-oxide shorts," in *Proc. Int. Conf. VLSI Design*, Jan. 1998, pp. 378–381.

[35] E. Isern, M. Roca, and J. Segura, "Analyzing the need for ATPG targeting GOS defects," in *Proc. IEEE VLSI Test Symp.*, Apr. 1999, pp. 420–425.

[36] A. Yassine, K. Wieczorek, K. Olasupo, and V. Heinig, "A novel electrical test to differentiate gate-to-source/drain silicide short from gate oxide short," in *Proc. IEEE Int. Integr. Rel. Workshop Final Rep.*, Oct. 2000, pp. 90–94.

[37] M. Renovell, J. M. Galliere, F. Azais, and Y. Bertrand, "Boolean and current detection of MOS transistor with gate oxide short," in *Proc. Int. Test Conf.*, 2001, pp. 1039–1048.

[38] S. A. Kumar, R. Z. Makki, and D. M. Binkley, "I_{DDT} testing of embedded CMOS SRAMs," in *Proc. Design, Autom. Test Eur. Conf. Exhibit.*, 2002, pp. 1530–1591.

[39] C. Y. Moh and A. K. A'ain, "A study on characterization of gate oxide shorts using non-split model," in *Proc. Student Conf. Res. Develop.*, Aug. 2003, pp. 188–192.

[40] A. Chehab, A. Kayssi, and A. Ghandour, "Transient current testing of gate-oxide shorts in CMOS," in *Proc. Int. Design Test Workshop*, Dec. 2007, pp. 77–81.

[41] A. Ghandour, K. Fawaz, A. Chehab, and A. Kayssi, "Wavelet transform-based transient current analysis for detection of gate-oxide shorts in CMOS," in *Proc. 4th Int. Design Test Workshop*, Nov. 2009, pp. 1–6.

[42] STMicroelectronics, "Deep sub-micron processes: 120 nm, 90 nm, 65 nm, 45 nm CMOS," in *Proc. CMP Annu. Users Meeting*, Jan. 2008.

[43] F. Arnaud, F. Boeuf, F. Salvetti, D. Lenoble, F. Wacquant, C. Regnier, P. Morin, N. Emonet, E. Denis, J. C. Oberlin, D. Ceccarelli, P. Vannier, G. Imbert, A. Sicard, C. Perrot, O. Belmont, I. Guilmeau, P. O. Sassoulas, S. Delmedico, R. Palla, F. Leverd, A. Beverina, V. DeJonghe, M. Broekaart, L. Pain, J. Todeschini, M. Charpin, Y. Laplanche, D. Neira, V. Vachellerie, B. Borot, T. Devoivre, N. Bicais, B. Hinschberger, R. Pantel, N. Revil, C. Parthasarathy, N. Planes, H. Brut, J. Farkas, J. Uginet, P. Stolk, and M. Woo, "A functional 0.69 μm^2 embedded 6T-SRAM bit cell for 65 nm CMOS platform," in *Proc. Symp. VLSI Technol., Dig. Tech. Papers*, Jun. 2003, pp. 65–66.

[44] T.-S. Yeoh, N. R. Kamat, R. S. Nair, and S.-J. Hu, "Gate oxide breakdown model in MOS transistors," in *Proc. IEEE Int. Rel. Phys. Symp.*, Apr. 1995, pp. 149–155.

[45] A. Meixner and J. Banik, "Weak write test mode: An SRAM cell stability design for test technique," in *Proc. Int. Test Conf.*, Nov. 1997, pp. 1043–1052.

[46] V. Chandra, C. Pietrzyk, and R. C. Aitken, "On the efficacy of write-assist techniques in low voltage nanoscale SRAMs," in *Proc. DATE*, Mar. 2010, pp. 345–350.

[47] S. Mukhopadhyay, R. M. Rao, J.-J. J. Kim, and C.-T. T. Chuang, "Capacitive coupling based transient negative bit-line voltage (Tran-NBL) scheme for improving write-ability of SRAM design in nanometer technologies," in *Proc. IEEE Int. Symp. Circuits Syst.*, May 2008, pp. 384–387.

[48] D. Kim, G. K. Chen, M. Fojtik, M. Seok, D. T. Blaauw, and D. M. Sylvester, "A 1.85 fW/bit ultra low leakage 10T SRAM with speed compensation scheme," in *Proc. IEEE ISCAS*, May 2011, pp. 69–72.

[49] L. F. Childs, C. D. Gunderson, O. R. Lu, and J. D. Burnett, "Memory having self-timed bit line boost circuit and method therefor," U.S. Patent 0 074 032, Mar. 25, 2010.

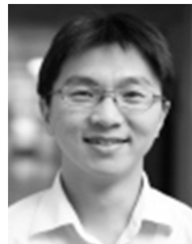
[50] I. Arsovski, H. Pilo, V. Ramadurai, and V. Ramadurai, "Static random access memory (SRAM) write assist circuit with leakage suppression and level control," U.S. Patent 0 140 551, Jan. 29, 2013.

[51] V. Sharma, F. Catthoor, and W. Dehaene, *SRAM Design for Wireless Sensor Networks*. New York, NY, USA: Springer-Verlag, 2013, pp. 31–64.

Chen-Wei Lin received the B.S. degree from the Department of Electronics Engineering, National Chiao-Tung University, Hsinchu, Taiwan, in 2006, and the M.S. degree from the Graduate Institute of Electronics Engineering, National Taiwan University, Taipei, Taiwan, in 2007. He is currently pursuing the Ph.D. degree with the Department of Electronics Engineering, National Chiao-Tung University.

Mango C.-T. Chao received the B.S. and M.S. degrees from the Department of Computer and Information Science, National Chiao-Tung University, Hsinchu, Taiwan, in 1998 and 2000, respectively, and the Ph.D. degree from the Department of Electrical and Computer Engineering, University of California, Santa Barbara, CA, USA, in 2006.

He joined the Department of Electronics Engineering, National Chiao-Tung University, as an Assistant Professor. His current research interests include VLSI testing, TFT circuitry design, and physical design automation.



Chih-Chieh Hsu received the B.S. degree from the Department of Materials Science and Engineering, National Tsing Hua University, Hsinchu, Taiwan, in 2004, and the M.S. and Ph.D. degrees from the Department of Electronics Engineering and Institute of Electronics Engineering, National Tsing Hua University, in 2005 and 2009, respectively.