國立交通大學

電子工程學系電子研究所

博士論文

低功耗前瞻靜態隨機記憶體之測試方法與錯誤模型

Test methodology and fault modeling for low-power advanced SRAM

研究生：林政偉

指導教授：趙家佐

中華民國一〇二年八月

低功耗前瞻靜態隨機記憶體之測試方法與錯誤模型

Test methodology and fault modeling for low-power advanced SRAM

研究生：林政偉　　Student：Chen-Wei Lin

指導教授：趙家佐　　Advisor：Mango Chia-Tso Chao

國立交通大學

電機學院　電子工程學系　電子研究所

博士論文

A Dissertation
Submitted to Department of Electronics Engineering
and Institute of Electronics
College of Electrical and Computer Engineering
National Chiao-Tung University
in partial Fulfillment of the Requirements
for the Degree of
Doctor of Philosophy
in
Electronics Engineering
August 2013
Hsinchu, Taiwan, Republic of China

中華民國一〇二年八月

低功耗前瞻靜態隨機記憶體之測試方法與錯誤模型

研究生：林政偉　　　　　指導教授：趙家佐

國立交通大學

電機學院　電子工程學系　電子研究所

摘要

對於新開發的低功耗和前瞻靜態隨機存記憶體(SRAM)，製造缺陷在其上所造成的錯誤行為往往比對於在傳統SRAM上造成之錯誤行為來說較為複雜。在相關的研究並未被充分的討論的情況下，許多針對傳統 6T SRAM 之一般性的測試方式都未被驗證，也因此無法滿足對於在製造與設計堅固和可靠的低功耗SRAM與未來的製程技術配合的測試需要。在這篇論文中，我討論了對於各種已在科技文獻中發表的低功耗SRAM設計的測試。針對不同的記憶胞結構我進行了分類並分析相對應的錯誤行為，並開發至少四個可以處理的不同的低功耗 SRAM 測試需要的測試方法。除了討論各種記憶胞結構，我也延伸了從傳統的平面 CMOS 元件的討論，到目前極具前瞻性的 FinFET 元件，以及特殊的薄膜電晶體元件。對於該些前瞻的 SRAM，我進行了元件等級的 TCAD 模擬、SPICE 模型提取和針對多晶矽通道模型建立之應用等，以進行驗證所提出的測試方法與最佳化前瞻 SRAM 設計的相關參數。

Test methodology and fault modeling for low-power advanced SRAM

Student：Chen-Wei Lin　　Advisor：Mango Chia-Tso Chao

Department of Electronics Engineering
and Institute of Electronics
National Chiao-Tung University

Abstract

For the new-developed low-power and advanced SRAMs, the fault behaviors due to manufacture defects are often relatively complicated when being compared to the traditional 6T SRAM. And the complete analysis has not been fully discussed. As a result, the test effectiveness of conventional test methods for the 6T SRAM may not satisfy the need for producing robust and reliable low-power SRAMs with future technologies.

In this thesis, I have discussed the testing of various low-power SRAM designs which have been published in literatures. By categorizing the different cell structures and analyzing the corresponding faulty behaviors, I have developed at least four new test methods which can deal with the diverse needs of the low-power SRAM testing. In addition to including the various cell structures, I also extend the discussion to the SRAM which comes with the specific peripheral write-assist circuitry. For the data-aware write-assist SRAM,

a high-fault-coverage and time-efficient test method is proposed. Finally, the discussions of the special Gate-Oxide Short defects at the traditional planar bulk CMOS and the promising FinFET technology are also covered. For those advanced SRAMs, device-level TCAD simulation, SPICE model extraction, and circuit-level defect model establishing were proceeded to either verify the proposed test methods or to achieve high yield optimized advanced SRAM designing.

Acknowledgements

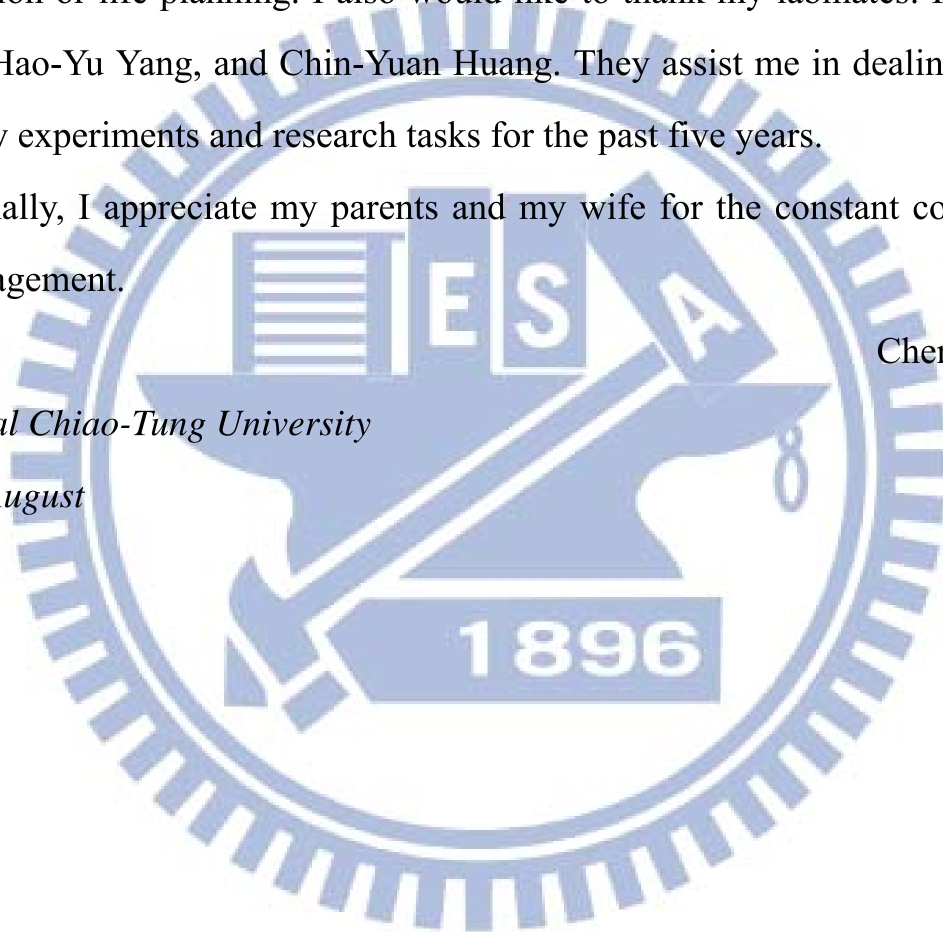
First and foremost, I would like to express my greatest appreciation to my advisor, Professor Chia-Tso Chao (Mango, 趙家佐) for his suggestions and guidance. He is extremely patient to me in everything no matter for the research discussion or life planning. I also would like to thank my labmates: Hung-Hsin Chen, Hao-Yu Yang, and Chin-Yuan Huang. They assist me in dealing with the weighty experiments and research tasks for the past five years.

Finally, I appreciate my parents and my wife for the constant comfort and encouragement.

Chen-Wei Lin

National Chiao-Tung University

2013, August



Contents

中文摘要 I

English Abstract II

Acknowledge IV

List of Tables IX

List of Figures XII

1 Introduction........1

2 Fault Models and Test Methods for Subthreshold SRAMs........3

2.1 Categorization of Subthreshold SRAM Designs........5

2.2 Test Methods for Stability Faults........6

2.2.1. Background of Stability Faults........6

2.2.2. Read Equivalent Stress........7

2.2.3. Severe Write........7

2.2.4. High-V-Write/Low-V-Read........9

2.3 Analysis of Open Defects in Type-A Subthreshold SRAMs........9

2.3.1 Design Overview of Type-A Subthreshold SRAMs........9

2.3.2 Impact of Open Defects on Type-A Subthreshold SRAMs........11

2.3.3 Effectiveness of Test Methods for Type-A Designs........13

2.4 Analysis of Open Defects in Type-B Subthreshold SRAMs........14

2.4.1 Introduction of Type-B Subthreshold SRAMs........14

2.4.2 Impact of Open Defects on Type-B Subthreshold SRAMs........16

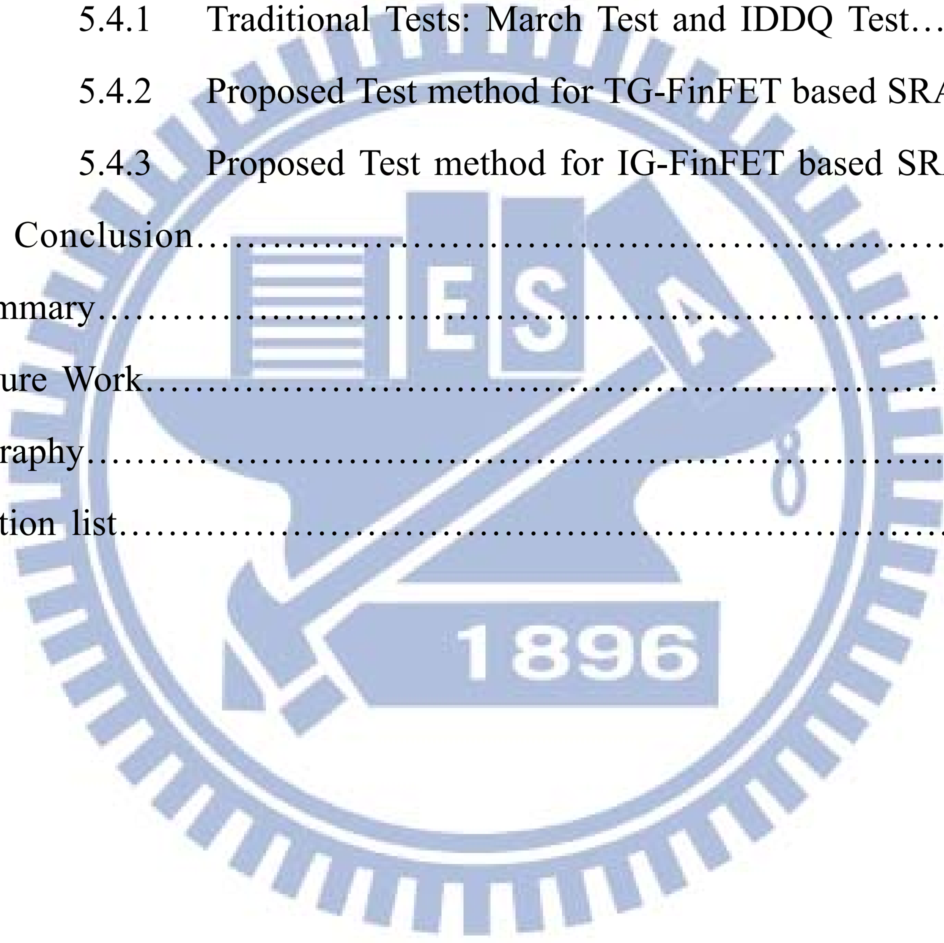
2.4.3 Effectiveness of Test Methods for Type-B Designs........17

2.5 Analysis of Open Defects in Type-C Subthreshold SRAMs........19

2.5.1 Introduction of Type-C Subthreshold SRAMs.............19
2.5.2 Impact of Open Defects on Type-C Subthreshold SRAMs..21
2.5.3 Effectiveness of Test Methods for Type-C Designs......22
2.6 Address Decoder Faults in Subthreshold SRAMS........................24
2.6.1 Type-A Subthreshold SRAM................................24
2.6.2 Type-B Subthreshold SRAM................................26
2.6.3 Type-C Subthreshold SRAM................................26
2.6.4 Address-Decoder Faults with Sequential Behavior......30
2.7 Fault Models for Sense Amplifier under Subthreshold Operations...30
2.7.1 Open Defects...30
2.7.2 V_{th} Mismatch...31
2.8 Impact of Temperature at Test..34
2.9 Conclusion...35
3 Testing of a Low-V_{MIN} Data-Aware Dynamic-Supply 8T SRAM...........37
3.1 Preliminary of the Low-V_{MIN} Date-Aware Dynamic-Supply 8T SRAM...39
3.2 Using March C- to Detect Open Defects in the Low-V_{MIN} DADS 8T SRAM...41
3.3 Test Methods for the Open Defects at the Cross-Couple Inverters.......44
3.3.1 Floating Bit-Line Attacking Method........................44
3.3.2 Proposed Method: Self-Loop Attacking Method.........46
3.3.3 Test Methods Comparison..................................49
3.4 Testing for The Open Defects in the DADS Circuitry..................50
3.5 Conclusion...52
4 A Novel Circuit-level Model for Gate Oxide Short and its Testing Method in

SRAMs……………………………………………………………………54
4.1 Experimental Setup for TCAD and HSPICE……...............................57
4.2 Previous Circuit-Level GOS Models…………………………………59
4.2.1 Bi-dimensional model…………………………………59
4.2.2 Nonlinear non-split model 1 - JET_03............................60
4.2.3 Nonlinear non-split model 2 - IDT_09…………………62
4.3 Proposed GOS Model and The Comparison With Previous Works.......64
4.3.1 Proposed GOS model……………………………………64
4.3.2 Simulation comparisons on DC characteristics…………66
4.3.3 Simulation comparisons on transient characteristics……67
4.3.4 Modeling GOS defects with different locations…………70
4.4 Testing GOS in SRAMs………………………………………………71
4.4.1 Previous test methods……………………………………72
4.4.2 Proposed DFT write operation…………………………75
4.4.3 Detailed simulation result for detecting GOS…………76
4.4.4 Finding valid setting for C_T and ΔV……………………78
4.4.5 Comparison with other GOS models in use……………79
4.5 Implementation of The Proposed Test Method & Its Optimization…80
4.5.1 Implementation and area overhead of DfT……………80
4.5.2 Optimum (C_T, ΔV) for minimizing area overhead………82
4.5.3 Maximization of tolerable ΔV range against the process variation……………………………………………………84
4.6 Conclusion…………………………………………………………85
5 Investigation of Gate Oxide Short in FinFETs and the Test Methods for FinFET SRAMs…………………………………………………………87
5.1 GOS Fault Behaviors in Planar Bulk MOSFETs……………………89

5.2 Experiment Setup……………………………………………………90
5.3 GOS Fault Behaviors in FINFETs……………………………………91
5.3.1 Tied-Gate FinFET……………………………………….....91
5.3.2 Independent-Gate FinFET……………………………….....92
5.3.3 Short Summary……………………………………………..95
5.4 Testing of GOS in FinFET SRAMs…………………………………95
5.4.1 Traditional Tests: March Test and IDDQ Test…………96
5.4.2 Proposed Test method for TG-FinFET based SRAM…...98
5.4.3 Proposed Test method for IG-FinFET based SRAM…100
5.5 Conclusion………………………………………………………………103
6 Summary………………………………………………………………………104
7 Future Work……………………………………………………………………105
Bibliography…………………………………………………………………………..106
Publication list………………………………………………………………………...118



List of Tables

Table 2-I Categorization of subthreshold SRAM designs…………………....5

Table 2-II Faulty behavior of open defects on Type-A designs (Fig. 2-2, 2-3, 2-4)…………………………………………………………………..12

Table 2-III Effectiveness of test methods for detecting STFs in Type-A designs……………………………………………………………………13

Table 2-IV Effectiveness of test methods for detecting fail-to-write gate open defects in Type-A designs……………………………………………………14

Table 2-V Faulty behavior of open defects on Type-B designs (Fig. 2-5, 2-6)…………………………………………………………………..17

Table 2-VI Effectiveness of test methods for detecting STFs in Type-B designs……………………………………………………………………18

Table 2-VII Effectiveness of test methods for detecting fail-to-write gate open defects in Type-B designs……………………………………………………18

Table 2-VIII Impact of using different backgrounds when testing fail-to-read open defects in Type-B designs…………………………………………19

Table 2-IX Faulty behavior of open defects on Type-C designs (Fig. 2-7, 2-8)……………………………………………………………………22

Table 2-X Effectiveness of test methods for detecting STFs in Type-C designs…………………………………………………………………....22

Table 2-XI Impact of using different write voltages during LVW-HVR for Type-C designs……………………………………………………….....23

Table 2-XII Impact of using different backgrounds when testing fail-to-read open defects in Type-C designs…………………………………….......23

Table 2-XIII Setting of WL1 and WL2 for Type-C design……………….27
Table 2-XIV Faulty behavior of address decoder faults on Type-C designs (Fig. 2-8)………………………………………………………………..28
Table 2-XV Minimum detectable resistance for open defects on a differential sense amplifier……………………………………………………………32
Table 2-XVI Minimum detectable resistance for open defects on a single-ended sense amplifier……………………………………………32
Table 3-I Control Signals for the 8T SRAM cell…………………………40
Table 3-II Test Results of March C- for Open Defects……………………….41
Table 3-III Control Signals of the FBA Test Method………………………....45
Table 3-IV Test Results of the FBA Test Method…………………………46
Table 3-V Final States of Defect-Free Cells and Test Results of Open Defects of SLA under Process Corner TT……………………………………………48
Table 3-VI Final States of Defect-Free Cells under Various Process Corners..49
Table 3-VII Test Methods Comparison of Test Efficacy and Test Time………50
Table 3-VIII Test Results of SLA for Open Defects in The DADS Circuitry……………………………………………………………...51
Table 3-IX Simulation Results of Testing the Defects at M11 and M12 but under FF Process Corner……………………………………………......52
Table 4-I DC Characteristics and Parasitic Capacitances of a Defect-Free nMOS and pMOS……………………………………………………....57
Table 4-II DC Fitting Errors Resulting From Different GOS Models………67
Table 4-III Fitting Errors of An Inverter's Delay Caused by Different GOS Models……………………………………………………………………69
Table 4-IV IDVD Fitting Errors of GOS Models With Different Defect Locations………………………………………………………………71

Table 4-V Fitting Errors of Inverter's Delay of GOS Models with Different Defect Locations……………………………………………………………71
Table 4-VI Result of Applying March C- Test to Detect A GOS in 6T SRAM...72
Table 4-VII IDDQ Sensitivity Caused by A GOS Based on A 1.2V 32KX32 SRAM……………………………………………………………………73
Table 4-VIII Comparison Between The Two Test Configurations P1 & P2 Labeled in Figure 4-27……………………………………………………85
Table 5-I Comparison of GOS Fault Behaviors of Different Transistors…95
Table 5-II Experiment Results of Applying Operation Pairs to Detect GOS in FinFET SRAMs……………………………………………………………...97
Table 5-III IDDQ Sensitivity of FinFET SRAMs with GOS (Array Size is 162Mb [5-5])……………………………………………………………98
Table 5-IV Test Efficacy Comparison of The Methods To Detect The GOSs in IG-FinFET SRAM…………………………………………………………103

List of Figures

Fig. 2-1 Illustration of severe write……………………………………………8
Fig. 2-2 First Type-A subthreshold SRAM design [2-9]………………………10
Fig. 2-3 Second Type-A subthreshold SRAM design [2-10]…………………10
Fig. 2-4 Third Type-A subthreshold SRAM design [2-11]……………………11
Fig. 2-5 First Type-B subthreshold SRAM design [2-12]……………………15
Fig. 2-6 Second Type-B subthreshold SRAM design [2-13]…………………16
Fig. 2-7 First Type-C subthreshold SRAM design [2-14]……………………20
Fig. 2-8 Second Type-C subthreshold SRAM design [2-15]…………………21
Fig. 2-9 (a) Conventional march sequence for detecting ADFs; (b) Types of address decoder faults……………………………………………………24
Fig. 2-10 Schematics of single-ended and differential sense amplifiers……31
Fig. 2-11 99th percentile of the largest input voltage offset versus V_{th} mismatch for a differential SA operating at 0.4V and 1.2V, respectively…………………………………………………………33
Fig. 2-12 99th percentile of the largest (smallest) input voltage for read 1(0) for a single-ended SA operating at 1.2V……………………………34
Fig. 2-13 99th percentile of the largest (smallest) input voltage for read 1(0) for a single-ended SA operating at 0.4V……………………………34
Fig. 2-14 (a) Cycle time versus temperature and (b) Power consumption versus temperature for a 128x32 subthreshold SRAM array…………35
Fig. 3-1 Schematic of the low-VMIN data-aware dynamic-supply 8T SRAM [3-7]……………………………………………………………...39
Fig. 3-2 Illustration of how background cells affect the testing of the defect at

M6……………………………………………………………..43
Fig. 3-3 An example of floating bit-line attacking method: using floating-0 on bit-line for detecting the open defects R1……………………………45
Fig. 3-4 Configuration of the proposed self-loop attacking test method………47
Fig. 3-5 Valid SL range under different process corners………………………51
Fig. 4-1 Cross-section view of a GOS-impacted MOSFET……………………55
Fig. 4-2 IDVD curves of a nMOS (a) without and (b) with a gate-to-channel GOS……………………………………………………………...56
Fig. 4-3 Representation of a 3D GOS-impacted MOSFET in TCAD…………58
Fig. 4-4 Comparison of an inverter's transient response between the TCAD simulation and the HSPICE simulation with extracted model cards…59
Fig. 4-5 An exemplary Bi-dimensional model with 5x5 internal points [4-11]..60
Fig. 4-6 The nonlinear non-split GOS model, JET_03, proposed in [4-21]……61
Fig. 4-7 The GOS fitting results by using JET_03 model……………………61
Fig. 4-8 The nonlinear non-split GOS model, IDT_09, proposed in [4-22]…62
Fig. 4-9 The GOS fitting results by using IDT_09 model……………………63
Fig. 4-10 Schematic of the proposed GOS model…………………………..64
Fig. 4-11 TCAD C-V simulation results for (a) nMOS and (b) pMOS, with and without a GOS……………………………………………………65
Fig. 4-12 DC fitting results by using the proposed GOS model……………66
Fig. 4-13 An inverter's transfer function resulting from TCAD and different GOS models when a 5nm-radius GOS locates at the nMOS…………68
Fig. 4-14 Transient response of an inverter with a GOS on nMOS resulting from TCAD and different GOS models……………………………68
Fig. 4-15 Transient response of an inverter with a GOS on pMOS resulting from TCAD and different GOS models……………………………69

Fig. 4-16 GOS defect with nine different locations……………………………70
Fig. 4-17 Design-for-test (DfT) of weak write test mode [4-45]……………74
Fig. 4-18 Concept of the proposed DFT write operation for detecting GOS at pull-up pMOS and pass-gate nMOS……………………………………75
Fig. 4-19 V_S/V_{SN}, $I_1/I_2/I_3$, and V_{BL}/V_{BLB} (a) without and (b) with a GOS at pull-up pMOS when applying the proposed write operation…………76
Fig. 4-20 V_S/V_{SN}, $I_1/I_2/I_3$, and V_{BL}/V_{BLB} (a) without and (b) with a GOS at pass-gate nMOS when applying the proposed write operation………77
Fig. 4-21 Finding valid (C_T, ΔV) combinations for detecting GOS at pull-up pMOS……………………………………………………………78
Fig. 4-22 Finding valid (C_T, ΔV) combinations for detecting GOS at pull-up pMOS or pass-gate nMOS…………………………………………79
Fig. 4-23 Finding valid (C_T, ΔV) combinations by using different GOS models……………………………………………………………80
Fig. 4-24 Mutually-inverse boosting circuitry modified from [4-32]………81
Fig. 4-25 Capacitor-occupied area of each valid (C_T, ΔV)…………………83
Fig. 4-26 Capacitor-occupied area of the preferred (C_T, ΔV)s in Figure 4-25……………………………………………………………84
Fig. 4-27 Tolerable ΔV range versus differently selected C_T with two recommended configurations for high-process-variation-immunity and low-area-cost purposes, respectively……………………………………85
Fig. 5-1 FinFET structure and images of the manufacture defects……………88
Fig. 5-2 IDVD curves of a planar bulk n-type MOSFET (a) without and (b) with a gate-to-channel GOS…………………………………………90
Fig. 5-3 Perspective view of the built FinFETs…………………………………91
Fig. 5-4 IDVD behaviors of TG FinFETs without/with GOS…………………92

Fig. 5-5 IDVD behaviors of IG FinFETs without/with GOS at the front-gate dielectric……………………………………………………………93
Fig. 5-6 Electron density distribution of an n-type IG FinFET with/without GOS at front-gate dielectric…………………………………………………94
Fig. 5-7 IDVD behaviors of IG FinFETs without/with GOS at the back-gate dielectric……………………………………………………………94
Fig. 5-8 FinFET SRAM designs applied for experiments……………………96
Fig. 5-9 Configuration of the Proposed_TG test method for detecting GOS in TG-FinFET SRAM…………………………………………………....99
Fig. 5-10 Minimum operatable ΔV-C_T for SRAMs passing the Proposed_TG…………………………………………………………100
Fig. 5-11 Test results of applying Proposed_TG for detecting GOS in the IG-FinFET SRAM…………………………………………………………101
Fig. 5-12 Configuration of the Proposed_IG test method for detecting GOS in IG-FinFET SRAM…………………………………………………………101
Fig. 5-13 Simulation details of the Proposed_IG detecting front-gate/back-gate GOS at the same time (C_T=0.2fF, ΔV=0.62V)...102
Fig. 6-1 Flowchart showing the proposed cross-layer simulation framework...105
Fig. 6-2 V_{th} distribution of unit-size (1μm/1μm) TFTs: (a) Mean and max V_{th} of TFTs with different Si-grain size, and (b) V_{th} distribution of Si-grain size at 300nm, 600nm, and 1000nm……………………………….....107
Fig. 6-3 V_{th} distribution for different-size (W=1μm, 1.2μm, 1.5μm, and 2μm) of TFTs: (a) Si-grain size at 300nm, and (b) Si-grain size at 1000nm…107
Fig. 6-4 Mean and standard deviation of the characteristics of a unit-size TFT SRAM (all the TFTs are with width=length=1μm) when different

Si-grain sizes are applied: (a) SNM, (b) Read-SNM, (c) Write time, and (b) Read time……………………………………………………108
Fig. 6-5 Individual yields and comprehensive yield for an SRAM design…109
Fig. 6-6 Case study 1: Yield of different SRAM designs versus chosen Si-grain size for process……………………………………………………..111
Fig. 6-7 Case study 1 (cont'd): Optimal SRAM designs under different Si-grain size (each point indicates a certain design scenario): (a) 300 nm, (b) 1000nm……………………………………………………………..111
Fig. 6-8 Case study 1 (cont'd): Globally optimized SRAM designs including transistor sizing and Si-grain size of process……………………112
Fig. 6-9 Case study 2: For low-power operation, the step to filter off the designs with insufficient yield under different supply voltage (a) 3 volt, (b) 2.7 volt……………………………………………………………………113
Fig. 6-10 Case study 2 (cont'd): Find the most energy saving designs (with corresponding process parameter and VDD) among the robust designs in Fig. 6-9………………………………………………………..114

Chapter 1
Introduction

Due to the increasing demand of low-power system, a great amount of research effort has been spent in the past to develop the effective and economic SRAM designs which may operate with low supply voltage or even in the subthreshold region. The test methods regarding those newly developed SRAM designs have not yet been fully discussed. In Chapter 2, I have discussed the subthreshold SRAMs testing for which I have 1) categorized the various subthreshold SRAM designs and 2) studied the open defects in SRAM cells, address decoders, and sense amplifiers. Three test methods have been proposed to cope with the different faulty behaviors of each type of SRAMs. The discussion of device variation and temperature affection to designing and testing of the subthreshold SRAMs is also included.

In addition to the subthresold SRAMs, the 3rd chapter covers the testing of a special 8T SRAM design which operates at super-threshold region but with low supply voltage. The utilized data-aware write-assist technique of the cell makes the previous general SRAM testing method unable to detect the defects in the special structure, and hence I study the specific faulty syndrome and propose an effective and extremely time-saving test method.

Chapter 4 turns the viewpoints of testing from the SRAM cell structures to the hard-to-detect manufacture defect: gate-oxide short (GOS). GOS has become a common defect for advanced technologies as the gate-oxide thickness of a MOSFET is greatly reduced. However, the behavior of a GOS-impacted MOSFET is complicated and difficult to be accurately modeled at the circuit level. In this chapter, I 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, I propose a novel circuit-level GOS model which provides a higher accuracy of its DC characteristics than any of the previous

models while being able to represent a minimum-size GOS-impacted MOSFET. Also, 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 the previous works. Last, I utilize the 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.

Based on the work in Chapter 4, I further extend the discussion of GOS testing to FinFET SRAM in Chapter 5 which has become the promising technology for future VLSI. I investigate the fault behaviors of the gate oxide short in FinFETs. The investigation includes both tied-gate and independent-gate FinFETs. Based on the TCAD mixed-mode simulations, I discover that the gate oxide short in the two types of FinFETs causes different fault behaviors from each other. Compared to planar bulk MOSFETs, the fault behaviors are even more complex. In addition to the discussion at device level, I also discuss the corresponding SRAM testing. For detecting gate oxide short in FinFET SRAMs, I propose two new test methods. By using TCAD transient simulations, I prove the two methods' test efficacy of detecting the gate oxide shorts uncovered by traditional test methods.

Finally, in Chapter 6, I discuss the non-conventional technology for SRAM: low temperature poly silicon thin-film transistor (LTPS-TFT). Operation characteristics of LTPS-TFT based systems vary significantly with design choices and parameters (i.e., process, device, circuit and system). Due to the lack of cross-layer simulation tool, conventional designs only optimize the design layers in isolation, leading to sub-optimal solutions. I present a cross-layer simulation framework for the design of LTPS-TFT SRAM. The proposed simulation framework optimizes design parameters considering the entire design space and hence, greatly reduces design complexity and efforts. The benefits of the proposed framework are illustrated by case studies.

Chapter 2

Fault Models and Test Methods for Subthreshold SRAMs

Lowering supply voltage is the most straightforward but effective method to reduce circuit's overall power consumption, which is especially suitable for those portable, power-limiting, and not-timing-critical applications such as wireless sensor systems and implanted biomedical chips. Previous works [2-1] [2-2] have shown that the most power-saving supply voltage falls around the subthreshold region for CMOS digital circuits and some subthreshold digital circuits have already been demonstrated in silicon successfully. Also, the performance degradation imposed by the subthreshold operations can be compensated by using proper parallel architecture [2-3] [2-4], which further extends the application of a subthreshold system.

In the process of developing a robust subthreshold system, operating SRAMs at a subthreshold voltage is more challenging than operating digital circuits. Under subthreshold operations, the typical 6T SRAM design needs to face the following two major problems: (1) decrease of the static noise margin and (2) decrease of the write margin [2-5] [2-6]. It means that a 6T SRAM bit-cell operating at subthreshold region is more vulnerable to the noise and at the same time harder to write. The detailed reasons of the above phenomenon were explicitly discussed in [2-6]. Also, in order to increase the write margin, the size of the pass transistors in a 6T SRAM bit-cell needs to be increased, which may further jeopardize the static noise margin. Thus, for a 6T SRAM bit-cell, a proper combination of the 6 transistors' sizes are extremely hard to obtained under subthreshold operations, especially when the local process variation of advanced process technologies may significantly change the device characteristics and in turn break the fragile balance between the currents of the 6

transistors for read, write, and hold operations. Previous results [2-7] have shown that the minimum supply voltage for operating a 6T SRAM design is 0.7V based on a bulk CMOS 65nm technology [2-8] and a dynamic-double-gate SOI technology.

To overcome the above two problems and successfully operate a SRAM at subthreshold region, several new SRAM bit-cell designs [2-9]-[2-16] were proposed. Tackling the weak static noise margin, [2-9]-[2-11] [2-14] [2-15] utilized an extra read path (in addition to the original pass transistors) in their SRAM designs to isolate the cross-coupled inverters from the bit-lines during a read operation, which can effectively avoid potential half select or deceptive read destruction. Tackling the inability to write, techniques were utilized to either strengthen the driving capability of the pass transistors or loose the hold ability of the cross-couple inverters during the write operation. To achieve the former one during a write operation, [2-9] specified a boosted word-line voltage to access the pass transistors and [2-16] designed the pass transistor in a way that its reverse short channel effect can be utilized under subthreshold operations. To achieve the latter one during a write operation, [2-13] broke the loop of the cross-coupled inverters with additional transistors and [2-12][2-14][2-15] destroyed the functionality of one or both inverters by adjusting the voltage at its virtual ground and/or virtual VDD.

Although a significant amount of research effort has been put into the area of developing an effective and economic subthreshold SRAM design, however, the testing methodologies for those new subthreshold SRAM designs have not been fully discussed in the literature yet. In this paper, we will first categorize the new subthreshold SRAM designs into three types based on their design characteristics. For each type of subthreshold SRAM designs, we will then discuss the fault models associated with open defects and identify the faults which may or may not be easily detected by a traditional SRAM test algorithm. We will further discuss the corresponding test methodologies for each of the above hard-to-detect faults. Also, we will discuss the faulty behavior of address decoder faults on those new subthreshold SRAMs and show their difference to the address decoder faults on the traditional 6T SRAM. Next, we will discuss the impact of open defects and V_{th} mismatch on sense amplifiers and compare their differences between subthreshold operations and superthreshold operations. A short discuss about the test temperature is provided as well. All the experimental results are collected from the simulation using an UMC 65nm low-leakage process technology.

2.1 Categorization of Subthreshold SRAM Designs

The fault models of a subthreshold SRAM design is associated with its bit-cell structure, and so are their test methodologies. In this section, we categorize the subthreshold designs [2-9]-[2-15] based on the following two criteria regarding the bit-cell structure (Q1 and Q2). The later discussion about the fault behaviors will be based on the result of this categorization.

- Q1: Is its read path different from its write path?
- Q2: Does the design use a single-ended sense amplifier?

Based on Q1 and Q2, the subthreshold SRAM designs can be divided into Type A, B, C, and D as shown in Table 2-I. In fact, Type D represents the bit-cell sharing the read/write paths and utilizing a differential sense amplifier, i.e., the traditional 6T SRAM design. Thus, our later discussion will focus on the fault models and test methods only for the designs in Type A, B, and C. Note that the reason why Q1 and Q2 are used for categorization is because these two criteria can divide the subthreshold SRAM designs into categories that result in similar faulty behaviors.

TABLE 2-I
Categorization of subthreshold SRAM designs

Type	Q1	Q2	Sub-V_{th} SRAM designs
A	Yes	Yes	[2-9]-[2-11]
B	No	Yes	[2-12][2-13]
C	Yes	No	[2-14][2-15]
D	No	No	Typical 6T SRAM

In order to analyze their fault models, we used a UMC 65nm low-leakage process to implement each of the above bit-cell designs in a 128x32 array (128 bit-cells at a bit-line and 32 bit-cells at a word-line), including write drivers and sense amplifiers. Each row contains only one word and the word size is 32-bit. Under the defect-free condition, we first identified the minimum required cycle time for correct read or write operations at the TT corner and 25◦C, and then defined the cycle time as 20% longer than the minimum required cycle time for each bit-cell design. On top of a defect-free design, we will later inject open defects and simulate whether the faulty design can function correctly within the

defined cycle time. A defect is detected if the result of the sense amplifier reports the wrong value.

2.2 Test Methods for Stability Faults

2.2.1 Background of Stability Faults

A stability fault defined in [2-17]-[2-20] refers to a small open defect on the source/drain of the four cross-coupled transistors, which may not fail a read or write operation under a typical operating condition but may fail under some corner conditions (such as significant IR drop, noise, or soft error). As a result, a stability fault may decrease the reliability of the SRAM but may not be easily detected by a conventional march sequence. Therefore, testing stability faults has become one of the most challenging tasks in current SRAM testing. Several test methods were proposed to detect the stability faults with as small resistance as possible [2-17]-[2-20].

For traditional 6T SRAMs, the past research effort mainly focused on the stability faults located on the source/drain of the pull-up pMOS transistors (such as M_{T2} and M_{T4} in Figure 2-1) and ignored the stability faults locating on the pull-down nMOS transistors (such as M_{T3} and M_{T5} in Figure 1), which can be detected relatively easily by a read operation because the bit-lines in general SRAMs are pre-charged to VDD during a read operation. If the nMOS transistors cannot successfully pull down a bit-line due to the open defects, then the pre-charged value (floating 1) will be read out, which is opposite to the expected value. On the other hand, if the pMOS transistors cannot successfully pull up the bit-line due to an open defect, then the pre-charged value (floating 1) just happens to be the expected value and hence the open defect cannot be detected.

However, for subthreshold SRAM designs, the read path can be separated from the write path, meaning that the weak pull-down ability of nMOS transistors will not directly affect the voltage at RBL during a read operation. Therefore, the importance of detecting the stability faults on the pull-down nMOS transistors (M_{T2} and M_{T4}) become more significant for subthreshold SRAM design than that for traditional 6T SRAMs. In this paper, we will validate the effectiveness of the following test methods for defecting the stability faults locating on both the pMOS and nMOS transistors of subthreshold SRAMs.

These testing methods include: (1) read equivalent stress, (2) severe write, and (3) low-V-write/high-V-read.

2.2.2 Read Equivalent Stress

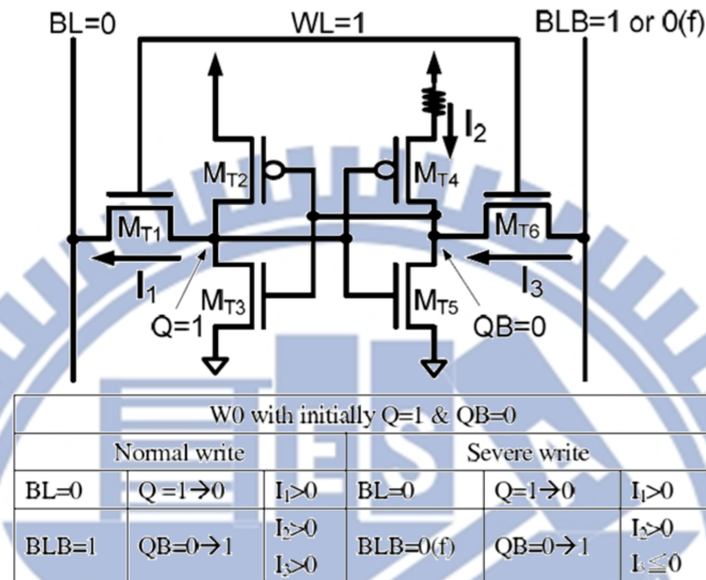
The idea of the read equivalent stress in the 6T SRAM design is to perform consecutive read operations to a designated bit-cell such that its word-line kept opened and its data stored by the cross-coupled inverters can be constantly attacked by the precharged VDD (floating 1) at bit-lines [2-17] [2-21]. However, for the subthreshold SRAMs which utilizes a different read path from its write path (such Type-A and Type-C), a read operation will turn on only its read word-line but not its write word-line. Such a read operation cannot attack the stored data and detect stability faults. Thus, to be able to apply read equivalent stress for Type-A and Type-C subthreshold SRAMs, specialized DFT circuit is required to turn on the write word-line and apply floating 1 at write bit-lines during a read operation at the test mode.

2.2.3 Severe Write

The idea of severe write in the 6T SRAM design is to perform a write operation by setting BL and BLB to floating 0 and strong 0 at the test mode, instead of strong 1 (or floating 1) and strong 0 at the normal mode (as shown in Figure 2-1) [20]. With such a write operation, successfully writing in data becomes more difficult since the floating 0 is opposite to the target value at Q or QB. As a result, if an open defect falls on the source/drain of pMOS transistors (such as M_{T2} and M_{T4}) and weakens the pull-up ability of an inverter, then the severe-write operation will fail to write the correct data and hence detect the open defect. Figure 2-1(a) illustrates how a severe write helps to detect an open defect on the pMOS transistor MT4.

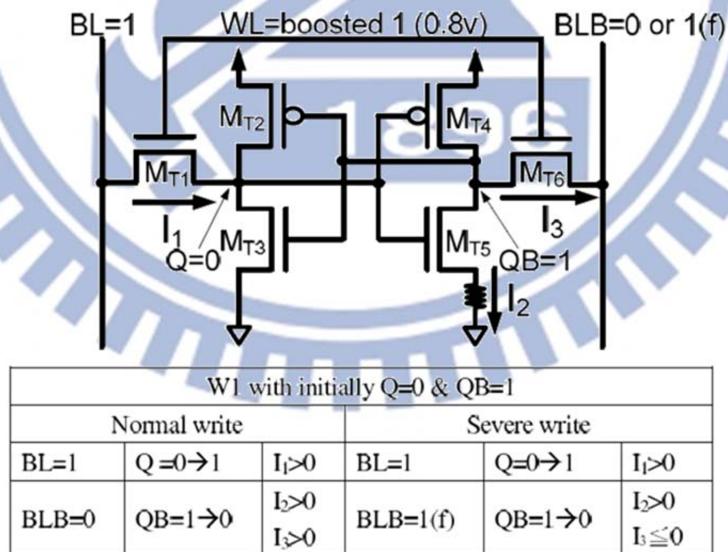
In fact, the above severe write (floating 0 and strong 0) can only detect open defects on pMOS transistors. To detect the stability faults on nMOS transistors, a severe write should set BL and BLB to floating 1 and strong 1. However, the nMOS pass transistors (M_{T1} and M_{T6}) are not suitable for passing a value 1, especially when operating at the subthreshold region (0.4V in our cases). Such a severe write cannot correctly write a data even when no defect exists in the subthreshold SRAM. Therefore, in order to use a severe write to detect stability

faults on nMOS transistors, we need to boost the voltage at WL by another V_t (0.8V in our case) to enhance the ability of passing a value 1 through the nMOS pass transistors during the test mode, which also requires extra DFT circuitry to realize. Figure 2-1(b) illustrates how this refined version of severe write can help the detection of an open defect on the nMOS transistor M_{T5}.



W0 with initially Q=1 & QB=0					
Normal write			Severe write		
BL=0	Q =1→0	I₁>0	BL=0	Q=1→0	I₁>0
BLB=1	QB=0→1	I₂>0 I₃>0	BLB=0(f)	QB=0→1	I₂>0 I₃≦0

(a) Severe write for testing pMOS stability fault



W1 with initially Q=0 & QB=1					
Normal write			Severe write		
BL=1	Q =0→1	I₁>0	BL=1	Q=0→1	I₁>0
BLB=0	QB=1→0	I₂>0 I₃>0	BLB=1(f)	QB=1→0	I₂>0 I₃≦0

(b)Severe write for testing nMOS stability fault

Fig. 2-1. Illustration of severe write.

2.2.4 High-V-Write/Low-V-Read

The idea of low-V-write/high-V-read is similar to the severe write, which increases the difficulty of a write operation such that the degradation of pull-up or pull-down capability caused by an open defect may fail to write the correct data. At the same time, we also need to make sure that this difficult condition for write will not fail the design without any defect. It means that the low operating voltage for write cannot be too far away from the normal voltage. Also, changing the operating voltage on test equipment takes a significant amount of time (around 10 micro seconds in our experience). Thus, we need to apply the low-V write to each word, change the operating voltage to normal, and then read each word. A high-V read immediately after a low-V write is not allowed due to its large overhead on test-application time.

2.3 Analysis of Open Defects in Type-A Subthreshold SRAMs

2.3.1 Design Overview of Type-A Subthreshold SRAMs

According to the categorization, Type-A subthreshold SRAM designs utilize a single-ended sense amplifier for read and build an extra read path in addition to the traditional 6T SRAM, which can protect the value stored in the cross-coupled inverters during read operations and improve its read SNM to the same level as its hold SNM. Figure 2-2 shows the first Type-A subthreshold SRAM design [2-9], where M_{A1} to M_{A6} represent the transistors in the traditional 6T SRAM and MA7 to MA10 represent the transistors in the read path. In this design, the original word-line (WL), bit-line (BL), and bit-line-bar (BLB) are only used for write operations. The new read word-line (RWL) and single-ended read bit-line (RBL) are only used for read operations. During a read operation, the value stored at QB (Q bar) will determine the value at QBB (Q bar bar) through an inverter (formed by M_{A7}, M_{A9}, and M_{A10}), and then determine the value at RBL. Also, the value of QBB is kept at 1 (VDD) or floating during the hold mode to reduce the leakage current of M_{A8} to RBL.

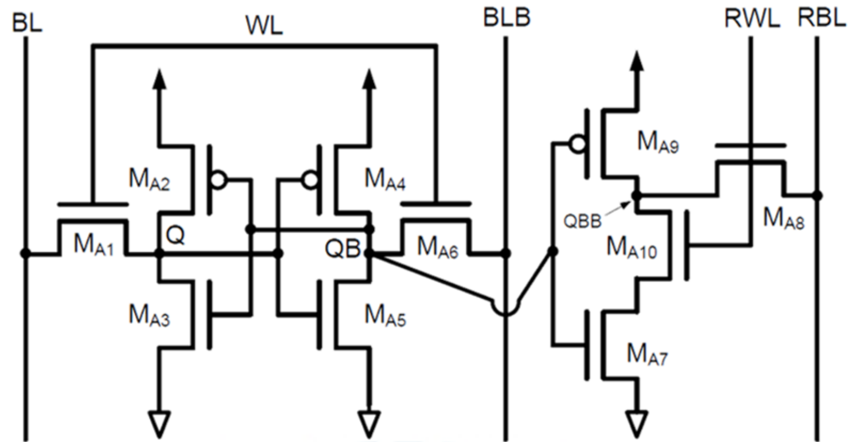


Fig. 2-2. First Type-A subthreshold SRAM design [2-9].

Figure 2-3 shows the second Type-A subthreshold SRAM design [2-10]. Similar to [2-9], [2-10] also use four transistors (M_{A7} to M_{A10}) to build an extra read path. However, its QBB is always kept at 1 during the hold mode since the M_{A9} in [2-10] is controlled by RWL instead of QB. When reading a value 0 out, QBB is pulled down through the path formed by M_{A7} and M_{A10}. However, when reading a value 1 out, QBB is floating since M_{A9} is turned off by RWL. As a result, the pre-charged floating 1 at RBL will be read out.

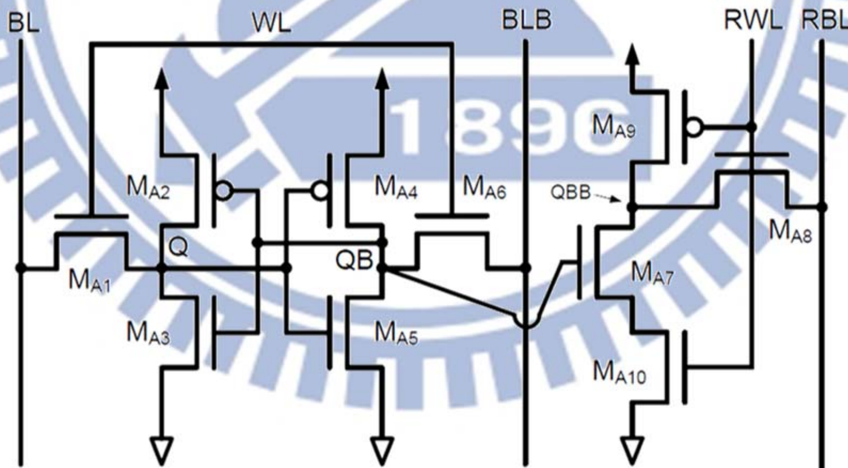


Fig. 2-3. Second Type-A subthreshold SRAM design [2-10].

Figure 2-4 shows the third Type-A subthreshold SRAM design [2-11], which uses two transistors and one extra signal (named buffer-foot) to build the extra read path. During read, the signal buffer-foot is set to GND and hence its read mechanism is the same as [2-10]. It means that QBB is 0 and floating when reading 0 and 1, respectively. During hold, the signal buffer-foot is set to VDD,

meaning that QBB is either 1 or floating based on the value of QB.

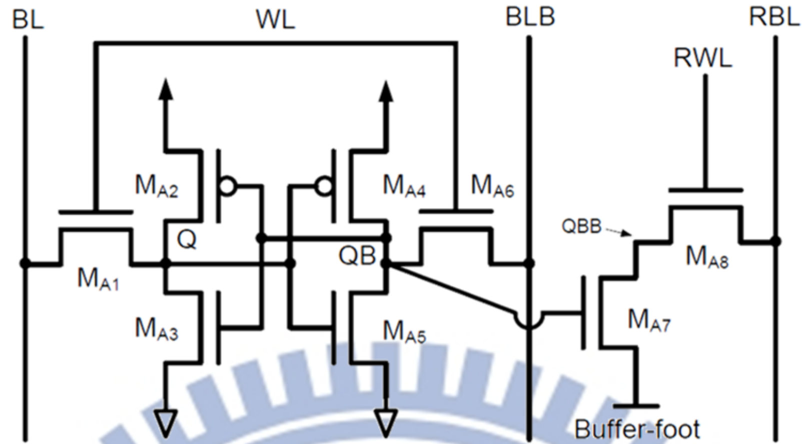


Fig. 2-4. Third Type-A subthreshold SRAM design [2-11].

2.3.2 Impact of Open Defects on Type-A Subthreshold SRAMs

In the following experiments, we inject an open defect with different resistances on each terminal (gate or source/drain) of each transistor and report the minimum resistance which can cause a failure on a read operation or a write operation for Type-A subthreshold SRAM designs. Table 2-II lists the minimum detectable resistance of each open defect (in Column 5) and the operation which the defect cause a failure at (in Column 4). Note that the result reported in Table 2-II is obtained based on the first Type-A design [2-9] at the TT corner and 25◦C. A similar result can be obtained for the other two Type-A designs [2-10] [2-11]. In addition, once the a defect can generate a read failure or write failure, this defect can be easily detected by a conventional SRAM march sequence. Therefore, we only need to consider the open defects with a faulty resistance less than the minimum detectable resistance.

As Table 2-II shows, the open defects locating on the original 6T bit-cell (M_{A1} to M_{A6}) all fail on a write operation. The open defects locating on the source/drain of the four cross-coupled transistors (M_{A2} to M_{A5}) are first highlighted by a gray background color in Table 2-II. Those defects are classified as a stability fault in Section 2.2. Opposite to traditional 6T superthreshold SRAMs, no stability faults on the nMOS transistors (M_{A3} and M_{A5}) can be detected, but the stability faults on the pMOS transistors can be detected with a 60MΩ minimum detectable resistance for Type-A designs. This result demonstrates that detecting the stability faults on nMOS transistors is

more critical than that on pMOS transistors for Type-A designs. Also, all open defects on the gate of the six transistors (M_{A1} to M_{A6}) have a minimum detectable resistance larger than 370MΩ, and hence are also relatively hard to detect.

TABLE 2-II
Faulty behavior of open defects on Type-A designs (Fig. 2-2, 2-3, 2-4).

Transistor property	Transistor name	Transistor terminal	Faulty behavior	Min detectable resistance
Write pass transistor	M_{A1}	G	W0 fail	482MΩ
		S/D	W0 fail	3.8MΩ
	M_{A6}	G	W1 fail	500MΩ
		S/D	W1 fail	3.2MΩ
Pull-up pMOS	M_{A2}	G	W0 fail	900MΩ
		S/D	W1 fail	60MΩ
	M_{A4}	G	W1 fail	800MΩ
		S/D	W0 fail	60MΩ
Pull-down nMOS	M_{A3}	G	W1 fail	370MΩ
		S/D	-	∞
	M_{A5}	G	W0 fail	370MΩ
		S/D	-	∞
Read pass transistor	M_{A8}	G	R0 fail	200MΩ
		S/D	R0 fail	16.9MΩ
Read-path pull-down1	M_{A7}	G	R0 fail	440MΩ
		S/D	R0 fail	5.1MΩ
Read-path pull-down2	M_{A10}	G	R0 fail	240MΩ
		S/D	R0 fail	5.1MΩ
Read-path QBB set	M_{A9}	G	R0 fail	2GΩ
		S/D	-	∞

On the other hand, the open defects locating on the extra read path (M_{A7} to M_{A10}) all fail on a read-0 operation. Also, the open defects on both gate and source/drain of M_{A9} are almost undetectable even though those open defects may reduce the ability of pulling up QBB. However, the read-1 operation does not rely on M_{A9} to pull up RBL and hence the malfunction of M_{A9} can hardly fail a read operation. For M_{A7}, M_{A8}, and M_{A10}, the open defects on their gate is harder to detect than those on their source/drain.

2.3.3 Effectiveness of Test Methods for Type-A Designs

In the following experiment, we attempt to reduce the minimum detectable resistance of each stability fault by applying (1) read equivalent stress (denoted as RES), (2) severe write, and (3) low-V-write/high-V-read (denoted as LVW-HVR) to Type-A subthreshold SRAM designs. Note that the read equivalent stress performed in this experiment will not stop repeating read operations until the minimum detectable resistance can hardly be decreased, which usually takes less than 10 repeated read operations. Also, the operating voltage for write and read in low-V-write/high-V-read is 0.36V and 0.4V, respectively. Table 2-III reports the minimum detectable resistance achieved by each test method. In Table 2-III, the test method W+R means a simple read operation after a write operation, which will actually achieve the same minimum detectable resistance as listed in Table 2-II.

TABLE 2-III
Effectiveness of test methods for detecting STFs in Type-A designs.

Transistor property	Transistor name	W+R	RES	Severe W	LVW-HVR (0.36V)
Pull-up pMOS	M_{A2} (S/D)	60MΩ	∞	6.6MΩ	39.4MΩ
	M_{A4} (S/D)				
Pull-down nMOS	M_{A3} (S/D)	∞	790MΩ	4.3MΩ	∞
	M_{A5} (S/D)				

As Table 2-III shows, severe write outperforms the other two test methods by achieving a 6.6MΩ minimum detectable resistance for pMOS stability faults and a 4.3MΩ minimum detectable resistance for nMOS stability faults. Meanwhile, read equivalence stress cannot detect any pMOS stability faults and its minimum detectable resistance for nMOS stability faults is still high (790MΩ). Note that the read equivalence stress performs even worse than the simple read after write (W+R) for pMOS stability faults. This is because the W+R fails at its write operation but the read equivalent stress assumes that its initial value can be successfully written. Also, the low-V-write/high-V-read cannot detect any nMOS stability faults. In fact, if the boosted WL used in severe write is set to 0.7V, the minimum detectable resistances will be further decreased to the order of hundred-kΩ. However, if the boosted WL is set to 0.6V, no data can be written into the bit-cell even when no defect exists. Thus,

defining a proper boosted voltage at WL is a critical factor when using severe write.

In addition, the severe write and LVW-HVR can also help to reduce the minimum detectable resistance at the gate of M_{A1} to M_{A6}, while read equivalent stress cannot. Table 2-IV shows the corresponding results, in which LVW-HVR achieves a lower minimum detectable resistance at the gate of write pass transistors and pull-up pMOS transistors (M_{A1}, M_{A2}, M_{A4}, and M_{A6}) while the severe write achieves a lower minimum detectable resistance at the gate of pull-down nMOS transistors (M_{A3} and M_{A5}). Overall, severe write is still the most effective test method for Type-A designs since it can cover open defects at the most places.

TABLE 2-IV
Effectiveness of test methods for detecting fail-to-write gate open defects in Type-A designs.

Transistor property	Transistor name	W+R	RES	Severe W	LVW-HVR (0.36V)
Write pass transistor	M_{A1} (G)	482MΩ	∞	350MΩ	32.4MΩ
	M_{A6} (G)	500MΩ	∞	420MΩ	29.9MΩ
Pull-up pMOS	M_{A2} (G)	900MΩ	∞	180MΩ	60MΩ
	M_{A4} (G)	800MΩ	∞	200MΩ	60MΩ
Pull-down nMOS	M_{A3} (G)	370MΩ	∞	110MΩ	260MΩ
	M_{A5} (G)	370MΩ	∞	230MΩ	290MΩ

2.4 Analysis of Open Defects in Type-B Subthreshold SRAMs

2.4.1 Introduction of Type-B Subthreshold SRAMs

According to the categorization shown in Table 2-I, a Type-B subthreshold SRAM design utilizes a single-ended sense amplifier for read and its read operations share the same path with its write operations. Such a bit-cell structure implies that its write operation is performed through a single bit-line as well, which further increases the difficulty of a write operation. Thus, in order to successfully write data through a single bit-line, Type-B subthreshold SRAM designs heavily rely on the design techniques which can effectively reduce the

hold ability of the cross-coupled inverters during the write operation.

Figure 2-5 shows the first Type-B subthreshold SRAM design [2-12], which can adjust the hold ability of the cross-coupled inverters by controlling the voltage at virtual VDD (VirVDD) and virtual GND (VirGND). During a read operation or the hold mode, VirVDD and VirGND are set to VDD and GND as general SRAMs. During a write operation, VirVDD and VirGND will become an offset lower and an offset higher, respectively, which can break the outside inverter (formed by M_{B3} and M_{B4}) and allows the voltage at Q to be directly affected by BL. Also, this design [2-12] utilizes a pMOS pass transistor (M_{B2}) in addition to a normal nMOS pass transistor (M_{B1}) simultaneously, such that both 1 and 0 can effectively passed through either M_{B2} or M_{B1}.

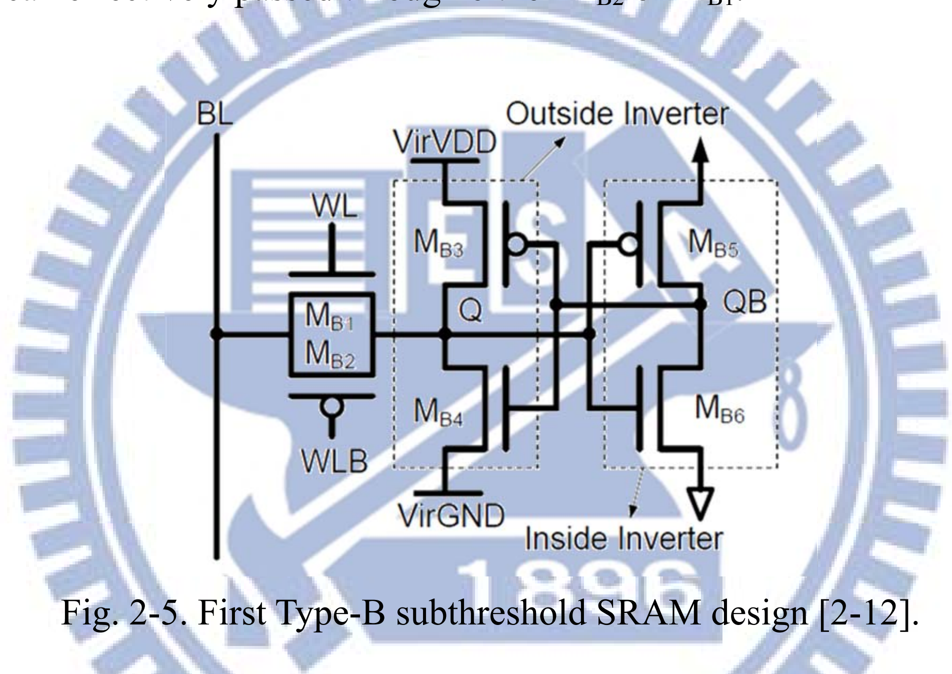


Fig. 2-5. First Type-B subthreshold SRAM design [2-12].

Figure 2-6 shows the second Type-B subthreshold SRAM design [2-13], which decreases the hold ability during a write operation by breaking the loop of the cross-coupled inverters through the control signals Wri and WriB (at M_{B8} and M_{B7}). Once the loop is broken, the value at BL can be easily written into the bit-cell. After the write operation, the loop of the cross-coupled inverters will be recovered as normal.

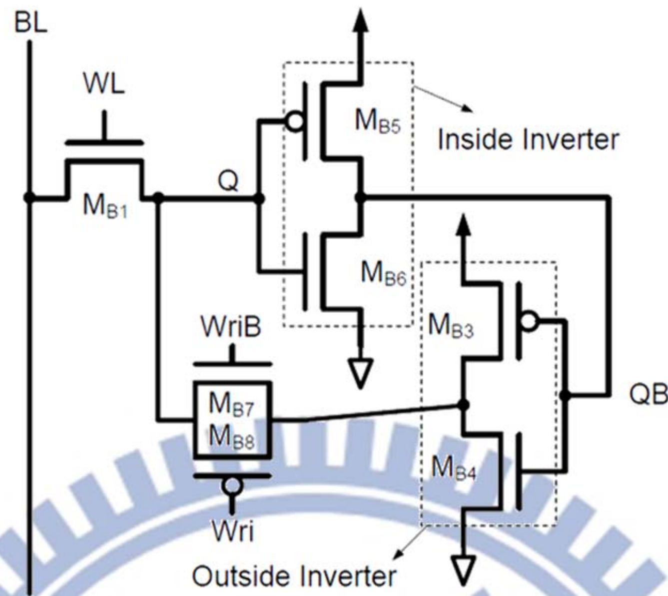


Fig. 2-6. Second Type-B subthreshold SRAM design [2-13].

2.4.2 Impact of Open Defects on Type-B Subthreshold SRAMs

Table 2-V lists the minimum detectable resistance and the corresponding faulty behavior of each open defect in Type-B designs. As Table 2-V shows, the open defect at the source/drain of M_{B4} does not cause a stability fault since the open defect falls on the path of read-0 and can be easily detected by a read-0 operation (with a 900kΩ minimum detectable resistance). Also, the stability fault at the outside pull-up pMOS M_{B3} is harder to detect than that at the inside transistors M_{B5} and M_{B6}. This is because the outside inverter is either destroyed or disconnected during a write operation, so that the value at Q is always correct. Even if a defect occurs on the outside pMOS M_{B3}, its weak pull-up ability will not lead to a wrong value at Q since the value at Q is already set by BL. However, if a defect occurs on the inside inverter, its weak pull-up or pull-down ability may delay the signal at QB and in turn result in a conflict at Q.

Table 2-V also shows that the open defects on the gate and source/drain of M_{B8} can hardly be detected, implying that the design [2-13] may not really need a pMOS transistor to pass a value 1 at the outside inverter's output to Q when the cross-coupled loop is reconnected right after a write operation. In addition, the minimum detectable resistance at each transistor's gate is still high and hence the corresponding open detect is also hard to detect.

TABLE 2-V
Faulty behavior of open defects on Type-B designs (Fig. 2-5, 2-6).

Transistor property	Transistor name	Transistor terminal	Faulty behavior	Min detectable resistance
Write pass transistor	M_{B1}	G	W0/R0 fail	2GΩ
		S/D	R0 fail	6.4MΩ
	M_{B2}	G	W1 fail	590MΩ
		S/D	W1 fail	7.6MΩ
Outside pull-up pMOS	M_{B3}	G	W0 fail	4GΩ
		S/D	-	∞
Inside pull-up pMOS	M_{B5}	G	W0 fail	870MΩ
		S/D	W0 fail	160MΩ
Outside pull-down nMOS	M_{B4}	G	W1 fail	2GΩ
		S/D	R0 fail	900KΩ
Inside pull-down nMOS	M_{B6}	G	W0 fail	970MΩ
		S/D	W1 fail	120MΩ
Cross-coupled loop switch	M_{B7}	G	W1 fail	2GΩ
		S/D	R0 fail	45.8MΩ
	M_{B8}	G	W0 fail	29GΩ
		S/D	-	∞

2.4.3 Effectiveness of Test Methods for Type-B Designs

Table 2-VI reports the minimum detectable resistance achieved by each test method for each stability fault in Type-B designs. Note that the severe write can only be applied to the design utilizing differential write mechanism (with BL and BLB), and hence cannot be applied to Type-B designs, which uses only one bit-line for write. As Table 2-VI shows, only read equivalent stress can detect the most hard-to-detect stability fault (at M_{B3}) in Type-B designs. This is because, by breaking the hold ability of the cross-coupled inverters, write 1 to Q is easy. As a result, detecting stability fault at MB3 cannot be achieved by using a weak write. We can only rely on read operations to detect it. Also, read equivalent stress can reduce the minimum detectable resistance of the other two stability

faults. In addition, LVW-HVR cannot effectively reduce the minimum detectable resistance at transistors' gate for Type-B designs as it does for the Type-A designs. Table 2-VII shows the corresponding result at each transistor's gate. Therefore, read equivalent stress is more preferable than LVW-HVR for Type-B designs overall.

TABLE 2-VI
Effectiveness of test methods for detecting STFs in Type-B designs.

<table>
<tr><th rowspan="2">Transistor property</th><th rowspan="2">Transistor name</th><th rowspan="2">W+R</th><th rowspan="2">RES</th><th colspan="2">LVW-HVR</th></tr>
<tr><th>0.38V-W</th><th>0.36V-W</th></tr>
<tr><td rowspan="2">Pull-up pMOS</td><td>M_{B3} (S/D)</td><td>∞</td><td>300KΩ</td><td>∞</td><td><0</td></tr>
<tr><td>M_{B5} (S/D)</td><td>160MΩ</td><td>160MΩ</td><td>150MΩ</td><td><0</td></tr>
<tr><td>Pull-down nMOS</td><td>M_{B6} (S/D)</td><td>120MΩ</td><td>62MΩ</td><td>43.7MΩ</td><td><0</td></tr>
</table>

TABLE 2-VII
Effectiveness of test methods for detecting fail-to-write gate open defects in Type-B designs.

<table>
<tr><th>Transistor property</th><th>Transistor name</th><th>W+R</th><th>LVW-HVR (0.38V-W)</th></tr>
<tr><td rowspan="2">Write pass transistor</td><td>M_{B1} (G)</td><td>2GΩ</td><td>2GΩ</td></tr>
<tr><td>M_{B2} (G)</td><td>590MΩ</td><td>410MΩ</td></tr>
<tr><td rowspan="2">Pull-up pMOS</td><td>M_{B3} (G)</td><td>4GΩ</td><td>3GΩ</td></tr>
<tr><td>M_{B5} (G)</td><td>870MΩ</td><td>790MΩ</td></tr>
<tr><td>Outside pull-down nMOS</td><td>M_{B4} (G)</td><td>2GΩ</td><td>430MΩ</td></tr>
<tr><td>Inside pull-down nMOS</td><td>M_{B6} (G)</td><td>970MΩ</td><td>410MΩ</td></tr>
<tr><td rowspan="2">Cross-coupled loop switch</td><td>M_{B7} (G)</td><td>2GΩ</td><td>3GΩ</td></tr>
<tr><td>M_{B8} (G)</td><td>29GΩ</td><td>190GΩ</td></tr>
</table>

In Table 2-V, open defects on the source/drain of M_{B1}, M_{B4}, and M_{B7} may result in a read-0 fail. Since Type-B designs use a single read path and BL is pre-charged to floating 1 for a read operation, a read-1 operation will never fail by an open defect on the bit-cell. In fact, the worst case of performing a read-0 operation occurs when the value of all other bit-cells at the same BL is set to 1, such that the leakage current from M_{B1} and M_{B2} can prevent the BL from being pulled down by the accessed bit-cell. Also, the devices need to be in the FF

corner and operated at a high temperature. Such a condition can result in a more significant leakage current, even though the pull-down capability of the targeted read path is also increased at a higher temperature (will discuss more in Section 2.8).

In the following experiment, we attempt to observe the impact of setting the data of all other bit-cells at the same BL to the same value (0) or the opposite value (1) to the accessed bitcell when performing a read-0 operation in Type-B designs. Table 2-VIII lists the minimum detectable resistance of the three read-0-fail open defects with both background settings. The simulation is conducted based on the FF corner at 75˚C. As the result shows, with the same data background, a large open defect may not be even detectable since the leakage at the same BL can help to pull down the data. With the opposite background, the minimum detectable resistance can be significantly reduced. Note that we have tried a similar experiment to Type-A designs but its difference of using different backgrounds is limited.

TABLE 2-VIII

Impact of using different backgrounds when testing fail-to-read open defects in Type-B designs.

Transistor name	Same background	Opposite background
M_{B1} (S/D)	∞	8.1MΩ
M_{B4} (S/D)	∞	90KΩ
M_{B7} (S/D)	150MΩ	20.4MΩ

To apply this all-1 background for a read-0 operation at each bit-cell, the march sequence in use needs to include the march element (w0, r0, w1). This march element can generate a read 0 out of an all-1 BL background and then recover the target bit-cell to 1, such that the background can remain all 1 when moving to the next address. Note that the march element (w0, r0, w1) is not included in a conventional SRAM march sequence, such as March C-.

2.5 Analysis of Open Defects in Type-C Subthreshold SRAMs

2.5.1 Introduction of Type-C Subthreshold SRAMs

According to the categorization shown in Table 2-I, a Type-C subthreshold

SRAM design utilizes a differential sense amplifier for read and its read path is different from its write path. It means that each of Q and QB needs to be read out through a different extra read path to BL or BLB instead of through the pull-up or pull-down paths of the cross-coupled inverters. Once the read paths are independent from the cross-coupled inverter, the read static noise margin can be protected. Also, Type-C subthreshold SRAM designs utilize a virtual GND to destroy the original stored data and improve its write ability.

Figure 2-7 shows the first Type-C subthreshold SRAM design [2-14], which embeds a 6T-SRAM design (with M_{C2}, M_{C4}, M_{C5}, M_{C6}, M_{C7}, and M_{C8}) in the center and one extra read path on a side to read out the value of Q (with M_{C1} and M_{C3}) or QB (with M_{C9} and M_{C10}). Also, two word-lines (WL1 and WL2) are used in this design. During a read operation, WL1 is set to 0 and WL2 is set to 1. Then the pre-charged BL will be pulled down by M_{C3} if Q = 1 and will remain floating 1 if Q = 0, meaning that the value read out from BL (or BLB) is different from the value at Q (or QB). During a write operation, both WL1 and WL2 are set to 1 and virtual GND is pulled up to VDD, which changes the original stored value at Q and QB to a voltage around 0.5 VDD and provides a weaker initial value at the cross-coupled inverters for write. After the write operation, the virtual GND will be pulled down to GND, which separates the voltages at Q and QB further apart. During the hold mode, both WL1 and WL2 are set to 0.

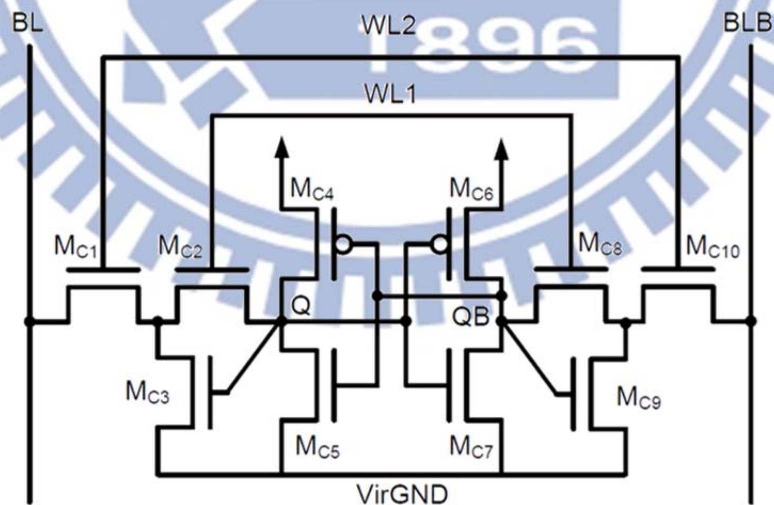


Fig. 2-7. First Type-C subthreshold SRAM design [2-14].

Figure 2-8 shows the second Type-C subthreshold SRAM design [2-15], which further improves the first Type-C design [2-14] with the following modification. In [2-15], its BL is connected to the output of the inverter formed

by M_{C6} and M_{C7} (through M_{C1} and M_{C2}) instead of that by M_{C4} and M_{C5}. Similarly, its BLB is connected to the output of the inverter formed by M_{C4} and M_{C5} (through M_{C8} and M_{C10}). As a result, the value read out at BL will be the same as the value at Q. Also, during its hold mode, WL2 is set to 0 but WL1 is set to 1. Under this setting of word-lines, M_{C3} or M_{C9} can help to pull down QB or Q during the hold mode, which can further increase its hold ability. In addition, because the value at Q equals to the value at BL during a read operation, the leakage of M_{C2} in [2-15] can be significantly reduced when compared to [2-14]. Similar situation applies to the leakage of MC8 during a read operation. Since [2-15] is a more refined version of [2-14], we will only consider the case of [2-15] in our later discussion regarding Type-C subthreshold SRAM designs.

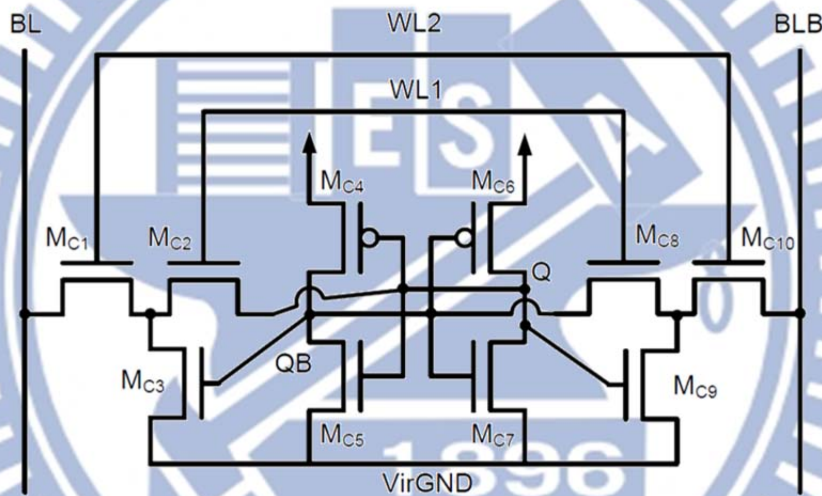


Fig. 2-8. Second Type-C subthreshold SRAM design [2-15].

2.5.2 Impact of Open Defects on Type-C Subthreshold SRAMs

Table 2-IX lists the minimum detectable resistance and the corresponding faulty behavior of each open defect in Type-C designs. As Table 2-IX shows, the stability faults on the nMOS transistors M_{C5} and M_{C7} cannot be detected at all. However, the stability faults on the pMOS transistors M_{C4} and M_{C6} are relatively easy to detect (with 11MΩ minimum detectable resistance), even compared to other stability faults in Type-A and Type-B designs. This is because the write mechanism in Type-C design relies on M_{C4} (or M_{C6}) to strongly hold the value 1 at QB (or Q) at the end of a write-0 operation, while VirGND just turns from VDD to GND. Thus, a small open defect on the source/drain of M_{C4} or M_{C6} may

fail the write operation. In addition, the open defect at a transistor's gate is also relatively easier to detect when compared to that in Type-A and Type-B designs.

TABLE 2-IX
Faulty behavior of open defects on Type-C designs (Fig. 2-7, 2-8).

Transistor property	Transistor name	Transistor terminal	Faulty behavior	Min detectable resistance
Write only pass transistor	M_{C2} & M_{C8}	G	W1 fail	58MΩ
		S/D	W1 fail	16MΩ
Write/read pass transistor	M_{C1} & M_{C10}	G	R1 fail	39MΩ
		S/D	R1 fail	2MΩ
Pull-up pMOS	M_{C4} & M_{C6}	G	W1 fail	64MΩ
		S/D	W0 fail	11MΩ
Pull-down nMOS	M_{C5} & M_{C7}	G	W1 fail	410MΩ
		S/D	-	∞
Read-path pull-down	M_{C3} & M_{C9}	G	W1 fail	170MΩ
		S/D	R1 fail	3MΩ

2.5.3 Effectiveness of Test Methods for Type-C Designs

Table 2-X reports the minimum detectable resistance achieved by each test method for each stability fault in Type-C designs. As the result shows, only LVW-HVR can detect the stability faults on nMOS transistors M_{C5} and M_{C7} while both RES and severe write cannot. However, the write voltage for LVWHVR need to be carefully assigned such that the nMOS stability faults can be detected and the fault-free design can still correctly function.

TABLE 2-X
Effectiveness of test methods for detecting STFs in Type-C designs.

Transistor property	Transistor name	W+R	RES	Severe write	LVW-HVR (0.26v)
Pull-up pMOS	M_{C4} (S/D)	11MΩ	17MΩ	6MΩ	930KΩ
	M_{C6} (S/D)				
Pull-down nMOS	M_{C5} (S/D)	∞	∞	∞	16MΩ
	M_{C7} (S/D)				

Table 2-XI shows the corresponding result of applying different write voltages to LVW-HVR. As the result shows, LVW-HVR cannot detect nMOS stability faults until the write voltage is reduced to 0.26V. However, if we further lower the write voltage to 0.24V, the minimum detectable resistance of pMOS and nMOS stability faults will be reduced to 2KΩ and 45KΩ. Such a low minimum detectable resistance kills almost all design margin for tolerating small detects during the test mode and in turn may result in an over-testing. Therefore, setting a proper write voltage is critical when applying LVW-HVR.

TABLE 2-XI

Impact of using different write voltages during LVW-HVR for Type-C designs.

Transistor property	Transistor name	LVW-HVR with different write voltage			
		(0.30v)	(0.28v)	(0.26v)	(0.24v)
Pull-up pMOS	M_{C4} (S/D)	4MΩ	3MΩ	930KΩ	2KΩ
	M_{C6} (S/D)				
Pull-down nMOS	M_{C5} (S/D)	∞	∞	16MΩ	45KΩ
	M_{C7} (S/D)				

Similar to Table 2-VIII, Table 2-XII reports the minimum detectable resistance obtained by applying the same background and the opposite background for all read-fail open defects in Type-C designs. The simulation is also conducted based on the FF corner at 75˚C. As the result shows, the opposite data background can effectively help to detect those read-fail open defects (with an acceptable minimum detectable resistance) while the same data background may fail to detect a large open defect, which again shows the effectiveness of setting an opposite background for detecting a read-fail open defect.

TABLE 2-XII

Impact of using different backgrounds when testing fail-to-read open defects in Type-C designs.

Transistor name	Same background	Opposite background
M_{C1} (G)	∞	310MΩ
M_{C1} (S/D)	78MΩ	7MΩ
M_{C3} (S/D)	∞	4MΩ

2.6 Address Decoder Faults in Subthreshold SRAMS

Address decoder faults (ADFs) in memories have been studied in the past [2-22] [2-23] [2-24], and it is proven in [2-24] that all the gross ADFs (not including the faults with sequential behavior and the small timing defect in the address decoder) can be detected by a march algorithm as long as the two march elements in Figure 2-9(a) are included. Figure 2-9(b) shows the four gross ADFs defined in [2-24]. In Figure 2-9(b), A_m represents the word-line signal of the address m, and C_m represents the physical memory cell indexed by the address m. Also, both m and n represent addresses.

Note that the above march algorithm is derived based on the assumption that only one word-line is used for both read and write operations, which is the case of the traditional 6T SRAM design. However, some subthreshold SRAM designs utilize multiple word-lines for read and write operations. Thus, the above march algorithm may not be able to detect all ADFs for all subthreshold SRAM designs. In the following subsections, we will briefly discuss the impact of the ADFs for each type of the subthreshold SRAM designs. Also, only the single ADF model is considered.

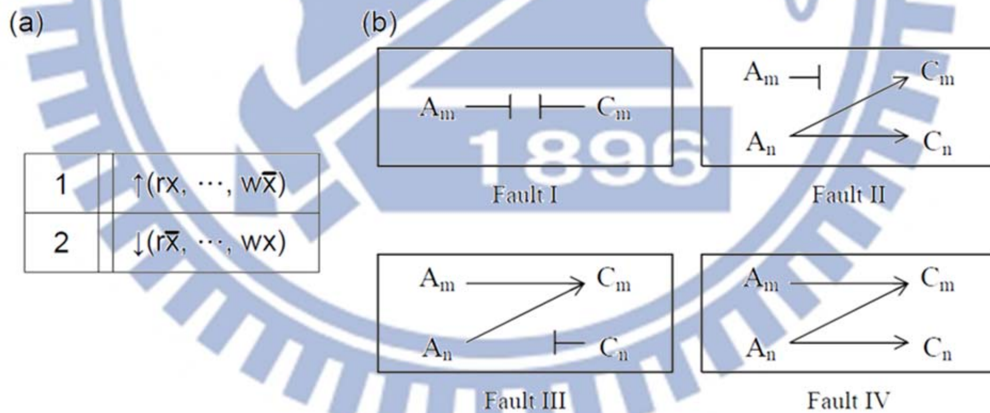


Fig. 2-9. (a) Conventional march sequence for detecting ADFs; (b) Types of address decoder faults.

2.6.1 Type-A Subthreshold SRAM

Type-A subthreshold SRAM designs use separate read wordline and write word-line (denoted as RWL and WWL in Figure 2-2 to 2-4) for read operations and write operations, respectively. Each ADF shown in Figure 2-9 may occur on

each of these two wordlines, and hence we need to consider total 8 cases of ADFs (4 types of ADFs on two word-lines). In the following paragraphs, the 8 cases of ADFs would be discussed.

1) Fault-I: When Fault-I exists and occurs on the WWL, the cell C_m (refer Figure 2-9(b)) would be unaccessible when it should be written, but accessible for reading. The sense amplifier (SA), when reading Cm, would thus always output the same value as the pre-stored data in C_m. The SAF-like behavior can be easily tested by the march in Figure 2-9(a). In the other case of Fault-I occurring on the RWL, since the RWL of C_m will never be triggered, the voltage on RBL (refer Figure 2-2 to 2-4) when reading C_m will always keep high regardless of the value in C_m. The faulty behavior is just like SA1 and can also be tested by the Figure 2-9(a) march.

2) Fault-II: When Fault-II occurs on the WWL, C_m could not be written by system operation ”Write C_m” but by the “Write C_n”. The faulty behavior can be tested by Figure 2-9(a) march. It’s because, in either $\uparrow(\mathrm{rx}, \dots, \mathrm{w}\bar{x})$ or $\downarrow(\mathrm{r}\bar{x}, \dots, \mathrm{wx})$ where C_n is earlier accessed than C_m, the ”Read C_m” will output the inverse value since the previous ”Write C_n” operation changes the value stored in C_m. In the other case of occurring on the RWL, C_m is unaccessible for read operation and thus the SA output of operation ”Read C_m” will always keep high as Fault-I on RWL. The SA1-like behavior is testable by the Figure 2-9(a) march.

3) Fault-III: In the faulty behavior of Fault-III occurring on WWL, C_m would be written by operation ”Write C_n” just like Fault-II on WWL. Thus in either $\uparrow(\mathrm{rx}, \dots, \mathrm{w}\bar{x})$ or $\downarrow(\mathrm{r}\bar{x}, \dots, \mathrm{wx})$ where C_n is earlier accessed than C_m, the SA output of ”Read C_m” will be the inverse value written by operation ”Write C_n”. Figure 2-9(a) march is still useful for Fault-III on WWL. For Fault-III on RWL, Figure 2-9(a) is still useful but uses the different test element from on-WWL case. In the march element in which C_m is earlier accessed than C_n, the ”Read C_n” will read the value in C_m, which is changed by previous operation ”Write C_m”, rather than unchanged value in C_n.

4) Fault-IV: Fault-IV on WWL is just like Fault-II/III on WWL which can be tested by the march element in which C_n is earlier accessed than C_m. The detail can be referred in previous paragraph. For Fault-IV occurring on RWL, its write operation works correctly but when reading cell n, both C_m and C_n will be read out at the same time. Assuming that m > n in the ADF Fault-IV (i.e., $\uparrow$ will visit n earlier than m), the march element $\uparrow(\mathrm{rx}, \dots, \mathrm{w}\bar{x})$ march element cannot detect the ADF Fault-IV because both C_n and C_m store the same value x when reading C_n. Also, an address-decreasing march element $\downarrow(\mathrm{rx}, \dots, \mathrm{w}\bar{x})$ may not

necessarily detect the ADF Fault-IV. For example, even though the march element $\downarrow(\mathrm{r0},\dots,\mathrm{w1})$ can create the situation that C_n stores 0 and C_m stores 1 when reading C_n, the RBL remains the good value 0 because the read bit-line will not be pulled up by the value 1 of cell m for designs [2-10] and [2-11]. The read-1 mechanism in [2-10] and [2-11] is to turn off the pull-down path at the read bit-line and leave the read bit-line floating 1. Thus, only the march element $\downarrow(\mathrm{r1},\dots,\mathrm{w0})$ can detect the ADF Fault-IV in this case. Note that the above discussion is based on the assumption that m > n in the ADF Fault-IV. To cover the case that that m < n, another march element $\uparrow(\mathrm{r1},\dots,\mathrm{w0})$ is also required.

5) Short Summary: After the analysis, most cases of ADFs can be detected by the march algorithm shown in Figure 2-9(a). However, the case of Fault-IV occurring on the RWL needs both $\downarrow(\mathrm{r1},\dots,\mathrm{w0})$ and $\uparrow(\mathrm{r1},\dots,\mathrm{w0})$ Therefore, a march algorithm which can cover four ADFs for Type-A SRAM designs needs to include three march elements. The two possible combinations of the three march elements are (1) $\downarrow(\mathrm{r1},\dots,\mathrm{w0})$, $\uparrow(\mathrm{r1},\dots,\mathrm{w0})$, and $\downarrow(\mathrm{r0},\dots,\mathrm{w1})$, and (2) $\downarrow(\mathrm{r1},\dots,\mathrm{w0})$, $\uparrow(\mathrm{r1},\dots,\mathrm{w0})$, and $\uparrow(\mathrm{r0},\dots,\mathrm{w1})$.

2.6.2 Type-B Subthreshold SRAM

Type-B subthreshold SRAM designs utilize WL and WL to access a bit-cell for both read and write operations. In general, these two signals (WL and $\overline{WL}$) come from the same address decoder but with the difference of an inverter. Thus, once an ADF falls in the address decoder, the signal at both WL and $\overline{WL}$ will be affected. As a result, the impact of an ADF fault in Type-B Subthreshold SRAM designs is exactly the same as that in a 6T SRAM design, and hence the march algorithm shown in Figure 2-9(a) is sufficient to detect all the ADFs for Type-B Subthreshold SRAM designs.

2.6.3 Type-C Subthreshold SRAM

The analysis of ADFs in Type-C subthreshold SRAM design [2-15] is more complicated than that in Type-A or Type-B designs since the Type-C design uses the combination of the values at WL1 and WL2 to determine the operation mode of a cell. Table 2-XIII shows the value of WL1 and WL2 at its hold, read, and write mode, respectively.

TABLE 2-XIII
Setting of WL1 and WL2 for Type-C design

Operation	WL1	WL2
Hold	1	0
Read	0	1
Write	1	1

A full analysis of ADFs in the Type-C design should include the impact of each ADF on each word-line (total 4 ADFs for 2 word-lines). For each ADF on each word-line, we need to enumerate the value at each word-line caused by the ADF based on different operation modes of the two faulty cells, which includes four effective combinations: C_m/C_n= (1) Read/Hold, (2) Write/Hold, (3) Hold/Read, and (4) Hold/Write. Note that we eliminate the cases of simultaneous Read and/or Write (i.e. Read/Read, Read/Write, Write/Read, and Write/Write) since the subthreshold SRAM is a single-port SRAM.

Table 2-XIV lists the complete analysis results of WL1 and WL2 values of the Type-C design [2-15] under the four C_m/C_n operations when each of the ADFs in Figure 2-9(b) occurs on WL1 and WL2 separately. According to the setting in Table 2-XIII, the values of WL1 and WL2 will lead to the corresponding behavior listed in the "Behavior" columns of Table 2-XIV. If the corresponding behavior is different from the supposed one, we highlight the faulty behavior with a gray background in Table 2-XIV. Note that we view the combination WL1=WL2=0 as a Hold operation since this configuration also enables the Type-C design [2-15] to hold the data but just without the extra assistance of M_{C3} and M_{C9}.

The faulty behaviors in Table 2-XIV are categorized into four groups (FB1-USR, FB2-UA, FB3-AR, and FB4-AW). In the paragraphs below, we will detail how each faulty behavior performs and give a short summary for testing ADFs in the Type-C design at the end.

1) FB1-USR (UnSafe Read) : The faulty behavior FB1-USR means that a cell is supposed to be read out, but its value may be attacked during the read operation. As shown in Table 2-XIII, only WL2 should be turned on during a read operation such that the turned off WL1 can protect the cross-coupled inverters from BL/BLB's direct accessing (as illustrated in Figure 2-8). The cell with the faulty behavior FB1-USR would have both its word-lines turned on during a read operation, and thus the stored data (Q and QB) would be affected by the pre-charged BL/BLB just as the typical 6T SRAM would. In other words, the designed extra-read path in the Typc-C subthreshold SRAM is disabled and

TABLE 2-XIV
Faulty behavior of address decoder faults on Type-C designs (Fig. 2-8).

		Faulty-WL1						Faulty-WL2					
		C_m			C_n			C_m			C_n		
		WL1	WL2	Behavior	WL1	WL2	Behavior	WL1	WL2	Behavior	WL1	WL2	Behavior
Fault-I	Read C_m	1	1	FB1-USR	-	-	-	0	0	FB2-UA	-	-	-
	Write C_m	1	1	Write	-	-	-	1	0	FB2-UA	-	-	-
Fault-II	Read C_m	1	1	FB1-USR	1	0	Hold	0	0	FB2-UA	1	0	Hold
	Write C_m	1	1	Write	1	0	Hold	1	0	FB2-UA	1	0	Hold
	Read C_n	0	0	Hold	0	1	Read	1	1	FB3-AR	0	1	Read
	Write C_n	1	0	Hold	1	1	Write	1	1	FB4-AW	1	1	Write
Fault-III	Read C_m	0	1	Read	1	0	Hold	0	1	Read	1	0	Hold
	Write C_m	1	1	Write	1	0	Hold	1	1	Write	1	0	Hold
	Read C_n	0	0	Hold	1	1	FB1-USR	1	1	FB3-AR	0	0	FB2-UA
	Write C_n	1	0	Hold	1	1	Write	1	1	FB4-AW	1	0	FB2-UA
Fault-IV	Read C_m	0	1	Read	1	0	Hold	0	1	Read	1	0	Hold
	Write C_m	1	1	Write	1	0	Hold	1	1	Write	1	0	Hold
	Read C_n	0	0	Hold	0	1	Read	1	1	FB3-AR	0	1	Read
	Write C_n	1	0	Hold	1	1	Write	1	1	FB4-AW	1	1	Write

can no longer help the cell to avoid the potential read disturb. To detect the faulty behavior FB1-USR, we need to apply consecutive read operations to the same cells in the test sequence.

2) FB2-UA (UnAccessible) : The faulty behavior FB2-UA means that a cell is unaccessible by either a read or a write operation. This fault can already be detected by the conventional march sequence shown in Figure 2-9(a), and thus needs no further discussion.

3) FB3-AR (Attacked Read) : As shown as Table 2-XIV, the faulty behavior FB3-AR occurs when ADF II, III or IV occurs on WL2, where C_m and C_n should originally be hold and read, respectively. However, both word-lines of C_m in this case are turned on instead. If C_m and C_n locate at different columns, C_m will be attacked by the un-selected, pre-charged bit-lines just like the cell suffering FB1-USR, which can be detected by the consecutive read operations as discussed in "Section 1) FB1-USR". On the other hand, if C_m and C_n locate at the same column, the read operation on C_n will be affected by the value stored in C_m as well since both word-lines of Cm are turned on. To trigger this fault, we need C_m and C_n to store the inverse data when C_n is read. The march sequence shown in Figure 2-9(a) satisfies this criterion. However, based on our simulation result, we found that the sensed output of this read fail with both BL and BLB pulled-down (one by C_m, and the other by C_n) is actually determined by the favored value of the sense amplifier in use. Thus, in order to cover different favored values of the sense amplifier, we should apply the march sequence shown in Figure 2-9(a) twice, one with x = 1 and the other with x = 0. In other words, the march elements $\downarrow(r1, \ldots, w0)$, $\uparrow(r0, \ldots, w1)$, $\downarrow(r0, \ldots, w1)$, and $\uparrow(r1, \ldots, w0)$ should be included in the march algorithm.

4) FB4-AW (Attacked Write) : The faulty behavior FB4-AW is similar to FB3-AR, where C_m should originally be hold with WL1/WL2 = 1/0 but both its word-lines are unexpectedly turned on instead. The only difference is that a write operation (instead of a read operation) is applied to C_n for FB4-AW when both word-lines of C_m are unexpectedly turned on. If C_m and C_n are at the same column, our simulation result shows that the value stored in C_m will not be over-written by the value writing into C_n since the VirGND of C_m still remain low (unlike a normal write operation keeping VirGND high). Also, the value of C_m will not prevent the original write operation to C_n from successfully performed even when their values are different. Thus, FB4-AW is more difficult to detect than FB3-AR. Fortunately, as shown in Table 2-XIV, an ADF causing FB4-AW must cause FB3-AR as well, meaning that FB4-AW can also be detected as long as FB3-AR can be detected through the methods described in

"Section 3) FB3-AR". Therefore, we only need to focus on detecting FB3-AR when designing the test algorithm.

5) Short Summary : To detect the ADFs occurring on WL1, we need to use consecutive read operations to cover FB1-USR as shown in "Section 1) FB1-USR". As to the ADFs on WL2, Fault-I, Fault-II, and Fault-III all cause FB2-UA, such that conventional march sequence shown in Figure 2-9(a) can already detect them. In the case that Fault-IV occurs on WL2, we can apply consecutive read operations and the march sequence {⇓(r1, … , w0), ⇑(r0, … , w1), ⇓(r0, … , w1), ⇑(r1, … , w0)} to detect FB3-AR.

2.6.4 Address-Decoder Faults with Sequential Behavior

Note that the march algorithm shown in Figure 2-9(a) can detect all the ADFs for conventional 6T SRAM only under the assumption that no ADF has sequential behavior. If an ADF has sequential behavior, we need to apply the non-linear test algorithm proposed by [2-31] [2-32] to detect it in conventional 6T SRAM. However, the effect of ADFs with sequential behavior in subthreshold SRAMs still remains unclear, which could also be a potential research topic.

2.7 Fault Models for Sense Amplifier under Subthreshold Operations

2.7.1 Open Defects

In this subsection, we first attempt to observe the impact of a single open defect which falls on a single-ended or differential sense amplifier operating under the subthreshold operations, and then compare the results to that under the normal superthreshold operations. In the following experiment, we will inject a single open defect with different resistances to different terminals inside the sense amplifier and check whether the injected defect can cause a failure of a read-0 operation (denoted as R0) or a read-1 operation (denoted as R1). However, the result of this experiment may depend on the setting of the cycle time. In order to make fair comparison of sense amplifiers between the

subthreshold operations and the super-threshold operations, we operate the same bit-cell design at both 0.4V and 1.2V, and then set the cycle time by adding extra 20% to the minimum required cycle for both designs under subthreshold and superthreshold operations.

Figure 2-10 (a) and (b) illustrate the schematic of the single-ended and differential sense amplifiers used in our experiment. Also, we label the terminals where an open defect may be injected in Figure 2-10. Table 2-XV first reports the minimum detectable resistance of each possible open defect based on the operating voltage of both 0.4V and 1.2V, respectively, for the differential sense amplifier. As the result shows, the minimum detectable resistance of almost all open defects under 0.4V operations is at least one order higher than that under 1.2V operations. A similar result can also be observed from the single-ended sense amplifier, whose result is reported in Table 2-XVI. Therefore, we can conclude that the sense amplifiers under subthreshold operations is more immune to the open defects than that under super-threshold operations.

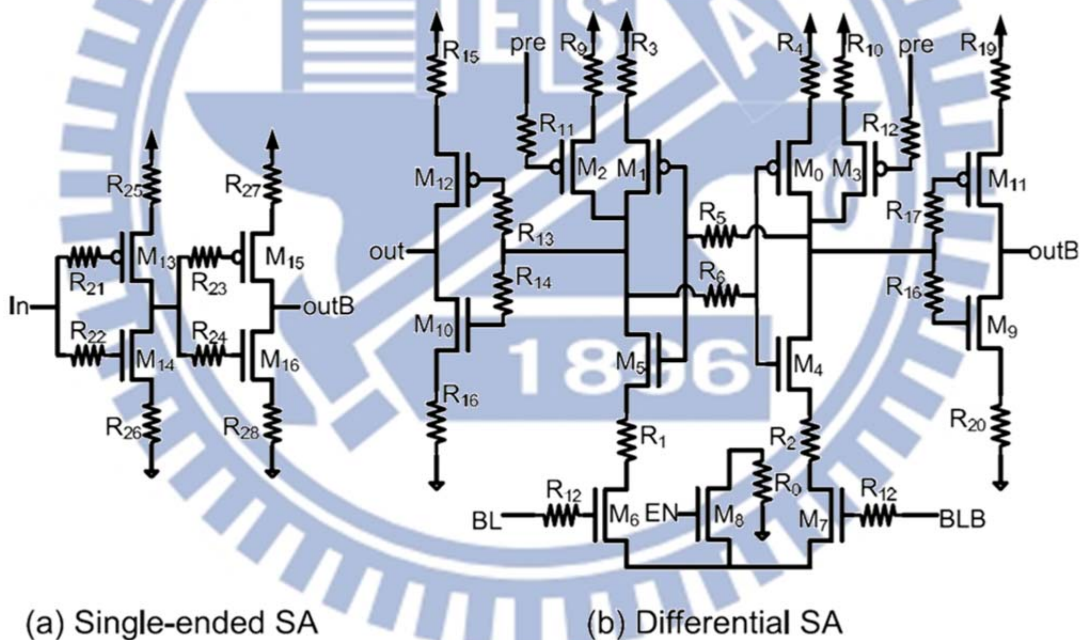


Fig. 2-10. Schematics of single-ended and differential sense amplifiers.

2.7.2 V_{th} Mismatch

The sensing ability of a sense amplifier can be significantly affected by the mismatch of device's V_{th} [2-25] [2-26], especially when the local process variation has continually increased in advanced process technologies [2-5] [2-6]. V_{th} mismatch may result in a larger input offset voltage for a differential sense

amplifier. Also, V_{th} mismatch may result in a higher input voltage to read 1 (or a lower input voltage to read 0) for a single-ended sense amplifier. In the following experiment, we attempt to observe the impact of different levels of V_{th} mismatch on differential and single-ended sense amplifiers operating at 1.2V and 0.4V, respectively.

TABLE 2-XV
Minimum detectable resistance for open defects on a differential sense amplifier.

Differential SA					
Open defect	0.4 V	1.2 V	Open defect	0.4 V	1.2 V
R0	280KΩ	10KΩ	R11	360MΩ	660KΩ
R1	890KΩ	10KΩ	R12	∞	130KΩ
R2	∞	10KΩ	R13	8.83MΩ	60KΩ
R3	∞	∞	R14	100MΩ	190KΩ
R4	∞	∞	R15	1.9MΩ	10KΩ
R5	30MΩ	0.1MΩ	R16	8.53GΩ	6.41GΩ
R6	∞	50KΩ	R17	∞	∞
R7	30MΩ	20KΩ	R18	∞	∞
R8	∞	30KΩ	R19	∞	∞
R9	290MΩ	390MΩ	R20	∞	∞
R10	300MΩ	90KΩ			

TABLE 2-XVI
Minimum detectable resistance for open defects on a single-ended sense amplifier.

Single-ended SA					
Open defect	0.4 V	1.2 V	Open defect	0.4 V	1.2 V
R21	50MΩ	1.1MΩ	R25	60MΩ	105KΩ
R22	520MΩ	1.8MΩ	R26	60MΩ	1.6MΩ
R23	360MΩ	∞	R27	110MΩ	10.8MΩ
R24	100MΩ	2MΩ	R28	22MΩ	500KΩ

We first model the V_{th} of each device in an independent normal distribution and specify a 3-sigma value to the normal distributions to represent the level of V_{th} mismatch. Based on the specified V_{th} distributions for all devices, we then randomly sample 10000 device configurations for a sense amplifier and collect the 99th percentile of its largest input voltage offset among the 10000

configurations. Figure 2-11 plots the 99% percentile of its largest input offset voltage versus the 3-sigma value of V_{th} distributions for the differential sense amplifier operating at both 0.4V and 1.2V, respectively. Note that the 3-sigma value of V_{th} distributions in Figure 2-11 is represented by its percentage to the mean value of V_{th}, i.e., 0.4V in this technology. For example, a 10% 3-sigma value shown in Figure 2-11 means 10% of 0.4V, i.e, 40mV.

For the result of 0.4V operations in Figure 2-11, the input offset voltage of the differential sense amplifier dramatically increases and exceeds its operating voltage 0.4V when the 3-sigma value of V_{th} distributions is larger than 9%. On the other hand, the input offset voltage under 1.2V operations increases more slowly based on the same level of V_{th} mismatch. This result shows that a differential sense amplifier under subthreshold operations is more vulnerable against V_{th} mismatch than that under super-threshold operations. In other words, the probability to have a faulty differential sense amplifier is higher in subthreshold SRAMs, compared to the traditional SRAMs under super-threshold operations.

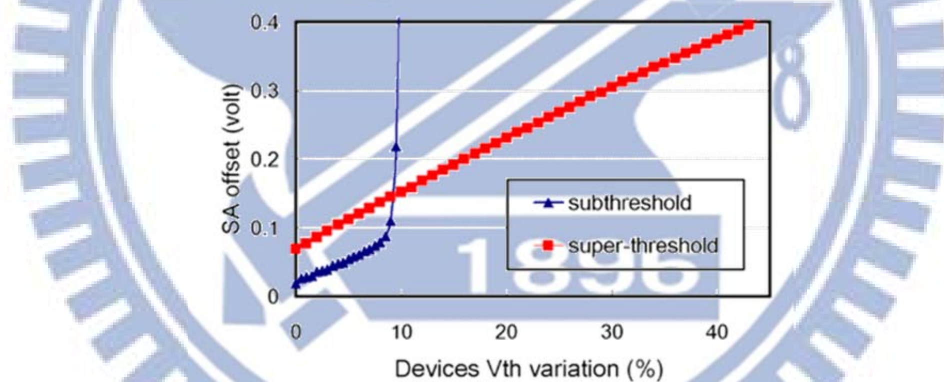


Fig. 2-11. 99th percentile of the largest input voltage offset versus V_{th} mismatch for a differential SA operating at 0.4V and 1.2V, respectively.

Figure 2-12(a) plots the 99th percentile of the largest input voltage to successfully read 1 from a single-ended sense amplifier under 1.2V operations. Figure 2-12(b) plots the 99th percentile of the smallest input voltage to successfully read 0 from a single-ended sense amplifier under 1.2V operations. Figure 2-13 plots the same result under 0.4V operations. By comparing Figure 2-12 and Figure 2-13, we can find that the change in the largest input voltage for read 1 (or the smallest input voltage for read 0) under 1.2V operations is similar to that under 0.4V operations based on the same level of V_{th} mismatch, if we scale the result of 0.4V to 1.2V. Also, under 0.4V operations, the input-offset

change of a single-ended sense amplifier increases more slowly than that of a differential sense amplifier when V_{th} mismatch increases. This result shows that a single-ended sense amplifier may tolerate a more significant process variation than a differential sense amplifier.

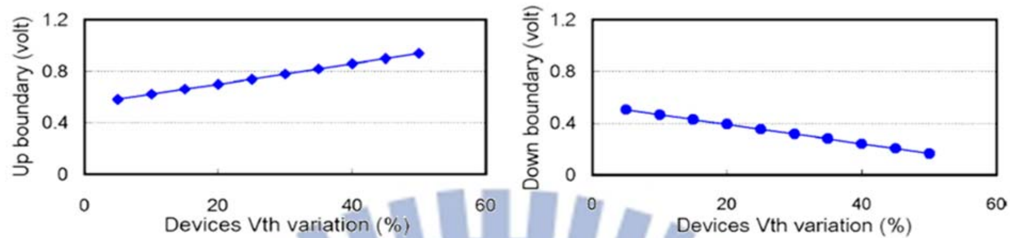


Fig. 2-12. 99th percentile of the largest (smallest) input voltage for read 1(0) for a single-ended SA operating at 1.2V.

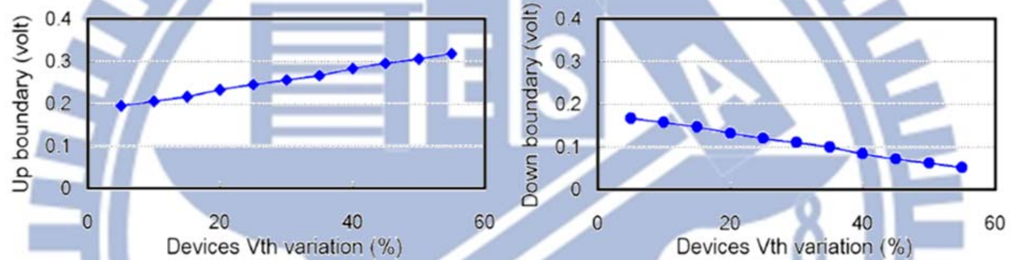


Fig. 2-13. 99th percentile of the largest (smallest) input voltage for read 1(0) for a single-ended SA operating at 0.4V.

2.8 Impact of Temperature at Test

When operating at a super-threshold voltage (e.g. 1.2V), the on-current of a transistor decreases along with the increase of temperature [2-27], meaning that the performance of a circuit also decreases. At the same time, the power consumption of a circuit increases along with the increase of temperature as well due to the lower V_{th} at a higher temperature [2-27] [2-28]. Therefore, we in general test a circuit (including logic and memory) more often at a high temperature than that at a low temperature since operating a circuit at a high temperature can exercise a worse corner of both the circuit's performance and power consumption, which in turn can examine the circuit's marginality and reliability. In addition, such a stressed condition at a high temperature can further speed up the aging of circuits and identify the infant mortality of circuits (such as burn-in).

However, the above property may not hold for subthreshold SRAMs (or general subthreshold logic circuits). Figure 2-14(a) first plots the minimum required cycle time for a subthreshold SRAM array [2-9] (used in our previous experiment) at different temperatures. As the figure shows, the cycle time decreases when the temperature increases under 0.4V operations, which is opposite to the case under 1.2V operations. On the other hand, Figure 2-14(b) plots the power consumption of the same subthreshold SRAM array at different temperatures and shows that the power consumption of a subthreshold SRAM array still increases when the temperature increases. The same trend about circuit performance and power consumption also holds for the subthreshold logic circuits [2-29] [2-30].

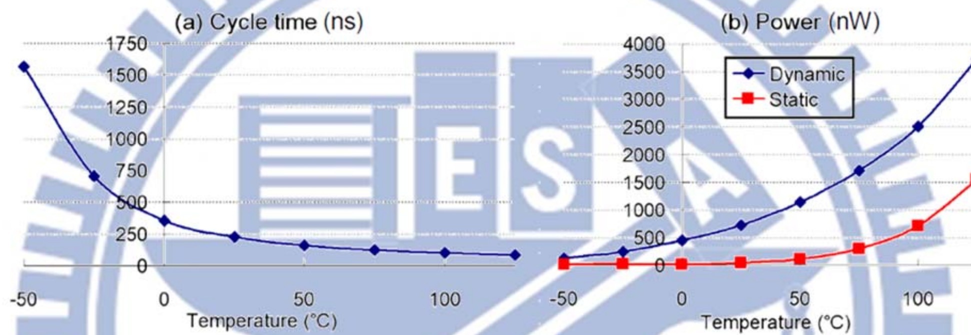


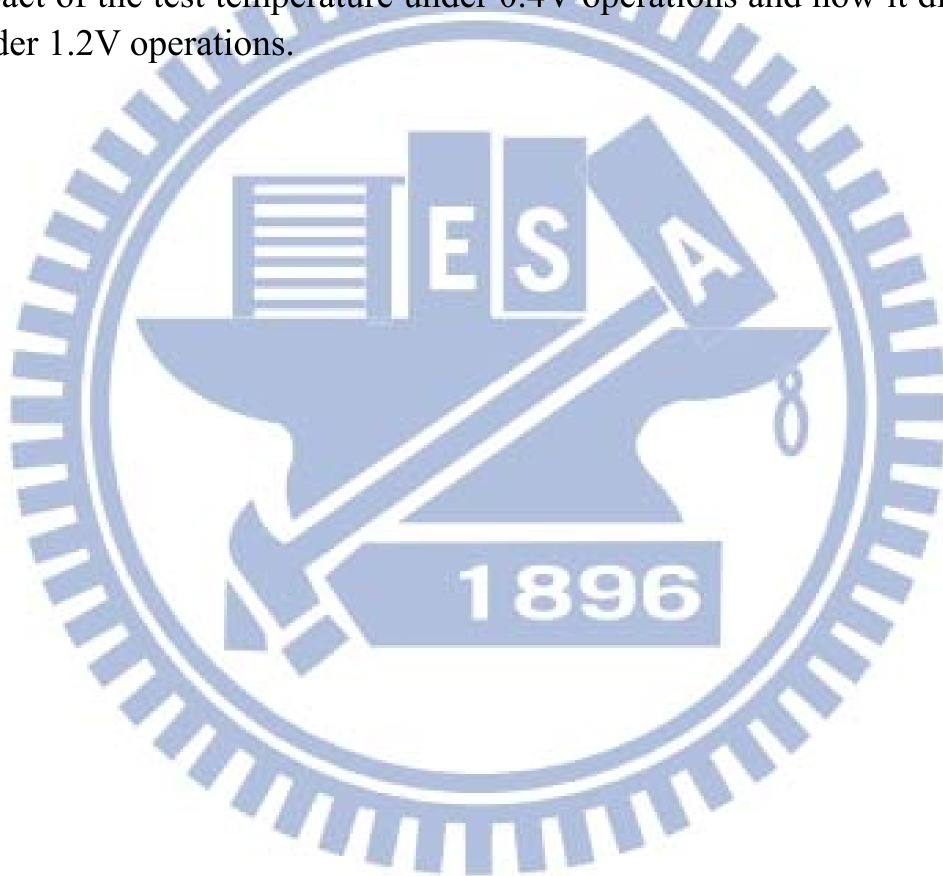
Fig. 2-14. (a) Cycle time versus temperature and (b) Power consumption versus temperature for a 128x32 subthreshold SRAM array.

As a result, testing a subthreshold SRAM at a high temperature can exercise a worse corner only for its power consumption. To exercise a worse corner for its performance, testing the subthreshold SRAM at a low temperature is required. This result also implies that the effectiveness of the traditional burn-in test may need to be reevaluated for subthreshold circuits.

2.9 Conclusion

In this chapter, we first validated the effectiveness of three different test methods on detecting stability faults through simulation and found that (1) only severe write can cover all stability faults for Type-A designs, (2) only read equivalent stress can cover all stability faults for Type-B designs, and (3) only low-V-write/high-V-read can cover all stability faults for Type-C designs. We also discussed the effectiveness of using opposite background for detecting a

fail-to-read open defect for each type of designs and found that this background works for Type-B and Type-C designs. Next, we discussed the faulty behavior of address decoder faults for each type of designs and found that (1) the detection of ADFs in Type-A designs requires a different march sequence from the traditional one and (2) the detection of ADFs in Type-C designs requires consecutive read operations and a specialized march sequence. Next, we studied the impact of open defects and Vth mismatch on sense amplifiers and found that (1) sense amplifiers under 0.4V operations are more immune to open defect and (2) differential sense amplifiers under 0.4V operations are more vulnerable to Vth mismatch, when compared that under 1.2V operations. At last, we discuss the impact of the test temperature under 0.4V operations and how it differs from that under 1.2V operations.



Chapter 3

Testing of a Low-V_{MIN} Data-Aware Dynamic-Supply 8T SRAM

Low power has been one of the most critical design issues for current electronics products, especially for those portable and battery-limited applications. Among all the low-power design techniques, lowing supply voltage is the most straight-forward and effective method. In [3-1] [3-2], the most power-saving supply voltage has been demonstrated falling around the device's threshold voltage. However, when operating CMOS circuits at such low supply voltage, the traditional 6T SRAM which occupies most of the area will encounter two problems: 1) the decreased read static noise margin (SNM) and 2) the decreased write margin [3-3][3-4]. The two problems indicate that the 6T SRAM is vulnerable during read and also difficult to write at the same time. As a result, finding the balance among each transistor's size in the 6T SRAM cell is extremely difficult, especially under the large process variation [3-5][3-6]. To vanquish the difficulties and develop effective low-power system, proposing new SRAM cell structures along with new SRAM design techniques has become necessary and essential.

In [3-7], the author proposed a new low-V_{MIN} data-aware dynamic-supply 8T SRAM. The SRAM is composed of a cross-point 8T SRAM (published in [3-8]) and a data-aware dynamic-supply circuitry (applied U.S. patent [3-9]). As validated through silicon chips, the SRAM design can operate at VDD=0.6 volt with operating frequency 209MHz. To increase the read-SNM, the 8T SRAM includes the commonly used and verified technique of independent read path [3-10]–[3-17]. With the path, the data is transmitted to BL indirectly for the read, which prevents BL's directly accessing internal nodes. As a result, the SRAM cell being read can keep the data very well as in the hold period. As to

improving the write margin, the SRAM applies the data-aware dynamic-supply (DADS) technique. Different from traditional techniques of strengthening pass gate's driving ability [3-10] [3-18] or breaking written-SRAM's cross-coupled loop [3-17] [3-19], the DADS technique configures the SRAM cell in an asymmetric manner. It separates the SRAM cell into two halves and assigns different and data-dependent supply to each half. The written-cell hence becomes inclined and suitable for accepting the new coming data. The write operation then completes with ease, and less power is consumed consequently. In addition to above features, the SRAM also have eliminated Write-Half-Select (WHS) disturb and single-BL driving scheme which will be introduced in detail in the following section.

For testing non-conventional SRAMs, [3-20] has categorized the various designs into types and proposed the test method for each type respectively. However, the categorization in [3-20] is made based on two simple design criteria, which as well as the test methods cannot cover all the new SRAM cell designs proposed later on. For the new DADS 8T SRAM discussed in this paper, the recommended test method for the corresponding type of SRAM needs both BL and BLB for the testing configuration. However, the new DADS 8T SRAM has only one BL, which makes the test method never applicable. Besides, in previous works [3-20] [3-21] [3-22], the testing of SRAM mostly focuses on the SRAM cell. For the DADS circuitry which is beyond the cell design, the corresponding testing has not been discussed yet. Since the data-aware write-assist technique based SRAMs are being the promising designs [3-23]–[3-26], the corresponding testing method becomes more and more important for the future production.

In this paper, we focus on testing the open defects for the new low-V_{MIN} DADS 8T SRAM. Open defects are the common defects in the manufacturing process and would reduce the circuit's reliability [3-21] [3-22]. To detect the defects, we firstly apply March test. The corresponding test efficacy and the undetectable cases will be shown. Then, for the undetectable defects in cell, we modify the floating bit-line attacking (FBA) method [3-20] for the 8T SRAM. The limitation of FBA is also discussed. Finally, we propose a test method which utilizes the design feature of the 8T SRAM cell. The proposed method not only detects all the undetected defects of March/FBA in both the cell and the DADS circuitry but also achieves lower detectable resistance than the two previous methods. Besides, the test time required by the proposed method is only 33%~44% of the FBA's, which does greatly speed up the defect detection.

3.1 Preliminary of the Low-VMIN Date-Aware Dynamic-Supply 8T SRAM

In this section, we introduce the low-V_{MIN} data-aware dynamic-supply 8T SRAM [3-7] with its operations and control signals. Figure 3-1 shows the schematic of the SRAM. The cell is as the lower part in the figure composed of M1–M8, and the data-aware dynamic-supply circuitry is as the upper part with M9–M12. The SRAM cell holds data by the cross-coupled inverter M1–M4. Read operation relies on the independent path M7 and M8. By using the path, the BL does not directly access the Q/QB during the read. To write-0, BL is set to zero and connects the Q through ”M5 and M7”. If write-1, BL is still set to zero but QB on the contrary will be connected to through “M6 and M7”. Based on this write scheme, only one BL is needed and the designing of BL driver is simple.

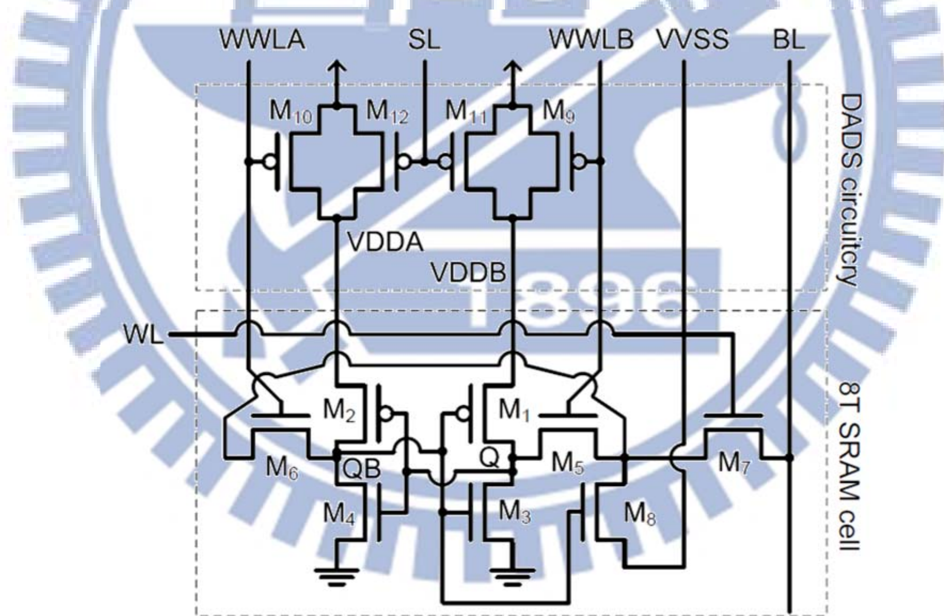


Fig. 3-1. Schematic of the low-VMIN data-aware dynamic-supply 8T SRAM [3-7].

Table 3-I summarizes the controls for the 8T SRAM including the row-based WL and column-based WWLA/WWLB, BL, and VVSS. The read-write word-line (WL) turns on for both read and write, but WWLA/WWLB turn on only for write. Besides, depending on the written data, only one of WWLA/WWLB is on during the write. BL is floating-1 when read and 0 for

write. The VVSS, mainly used for read operation, is suggested to follow WWLA to prevent the write disturb of background cells [3-8].

TABLE 3-I
Control Signals for the 8T SRAM cell

Operation	Control signals				
	WL	WWLA	WWLB	BL	VVSS
Read	1	0	0	1 (floating)	0
Write-0	1	0	1	0	0
Write-1	1	1	0	0	1
Hold	0	0	0	x	0

The data-aware dynamic-supply (DADS) circuitry, shown as M9–M12 in Figure 3-1, controls the supply voltage (VDDA and VDDB) for the cells in column. With WWLA and WWLB as the inputs, either M9 or M10 would be turned off when one of the cells in column is being written. For example, when a cell in the column is at write-1, WWLA is high for BL accessing and pulling down the QB. The M10 is then turned off, and pMOS M2 will get lower supply current from VDDA since only M12 provides that. The DADS then assists the write-1 operation because QB can be pulled down easily. For write-0, WWLB is high and M9 is turned-off on the contrary. The DADS circuitry assists the write-0 by leaving only M11 supporting VDDB. The supply-level (SL) which controls M11 and M12 is set to provide a minimum supply current for the background cells at write periods. Thus, there is a voltage upper bound for SL. SL still has a voltage lower bound. It's because the SL with voltage too low will turn on M11/M12 too much, which makes VDDA/VDDB always with high supply capability. The data-aware write-assist function would then be canceled.

In our experiments, we apply a 256Kb DADS 8T SRAM with eight 32Kb blocks. Each block is composed by 256 rows and 128 columns. Thus a DADS circuitry drives 256 cells in column. The SL is set to 0.5*VDD (VDD=0.6V) which has been verified that the 8T SRAM can operate correctly from process corner SS to FF. For most of our experiments, we run the simulation under TT corner. Only for our proposed test method, we consider all the process corners to prove the method's validity. This will be shown in the sections later.

3.2 Using March C- to Detect Open Defects in the Low-VMIN DADS 8T SRAM

To detect open defects, March C- algorithm is firstly applied in our experiment as representative. The defects are simulated by injecting a resister into each MOSFET of the SRAM with resistance swept from high (100GΩ) to low. For each single defect, we record the resistance when the SRAM's sense amplifier reports error. If the defect does not fail the SRAM even with 100GΩ, it is considered undetectable. Table II lists the simulation results of the test. The first two columns are the device type and the device name at which the defect is injected. The third and fourth columns are the detecting operation and the minimum detectable resistance.

TABLE 3-II
Test Results of March C- for Open Defects

Device type	Device name	Detecting operation	Min-detectable resistance
Pull-up pMOS	M_1	Write-1	800MΩ
	M_2	Write-0	500MΩ
Pull-down nMOS	M_3	undetectable	-
	M_4	undetectable	-
Write pass-gate	M_5	Write-0	8MΩ (Background = 1) 3MΩ (Background = 0)
	M_6	Write-1	3MΩ (Background = 1) 800KΩ (Background = 0)
Read-write pass-gate	M_7	Read-0	300KΩ
Read path	M_8	Read-0	50KΩ
Data-aware dynamic-supply circuitry	M_9	undetectable	-
	M_{10}	undetectable	-
	M_{11}	undetectable	-
	M_{12}	undetectable	-

As shown in the table, the defects at pull-down nMOSs and in the DADS circuitry are undetectable. For the pull-down nMOSs (M3 and M4), the defects

fail neither the write nor the read. In write, Q or QB can always be successfully pulled down by the BL even if the pull-down nMOSs are defective. In read, the independent read path would take charge to transmit the data to the floating-1 on BL in place of the original pull-down nMOSs in 6T SRAM. Thus, normal operations cannot detect the defects. As to the DADS circuitry, when the defect occurs at M9 and M10, the fault is masked. It's because M9 and M10 are originally set to off when write. And even when read/hold, the defect-free M11 and M12 will help support VDDA and VDDB that the stored data never flip due to the defect at M9 and M10. While defects occur at M11 and M12, and certain cell in the column is written, the background cells do get weak supply from VDDA (or VDDB). But the simulation results show the background cells can hold the data correctly during the write period and until M9 or M10 is re-turned on at the end of the write. Thus, the defects at M11 and M12 are also undetectable.

In addition to the pull-down nMOSs and the DADS circuitry, the defects at pull-up pMOSs (M1 and M2) are also belonged to hard-to-detect ones. Although the defects are detected by write operations according to Table 3-II, the min-detectable resistance is at 500MΩ~800MΩ. This implies the defect could be detected only if the resistance is large. Hence, we still need other test methods for lowering the detectable resistance.

Open defects at the pass-gates (M5, M6, and M7) and read-path transistor (M8) can be easily detected by write and read operations respectively. The minimum detectable resistance is at 3~8MΩ or below. Note that the minimum detectable resistance of the defects at M5 and M6 will vary depending on the data stored in the background cells. As shown in Table 3-II, when all the background cells store 0, the write operation in the March can detect lower open resistance for both the cases. However, the reasons for the two are different. For the defect at M5, when the defective cell is being tested by the write-0 operation, the Q is being pulled down by the BL, and QB is being pulled up by pMOS M2 with VDDA supply (see Figure 3-1). While all the background cells store 0, most cells share the VDDA since the corresponding QBs are 1. Hence, the background setting with all 0 causes the most severe write operation for the defective cell. Lower resistance is then detected.

As to the defect at M6, the background cells affect the testing via VVSS unlike the VDDA in the previous case. Figure 3-2 shows the detail of the affection. In the figure, the above cell is the defective one being tested by the write-1 operation. The initial data in the cell is Q/QB=0/1. The cell below represents the 255 background cells also with Q'/QB'=0/1. In the figure, all the

turned-off pass-gates are ignored. Firstly, the BL is accessing and attempting to pull down the QB through M7 and M6. Before the write operation completes, the M8 remains turned on since the QB is originally 1. The turned-on M8 connects VVSS and VP. The voltage on VP is then not a perfect zero since it's not only driven by the BL through M7 but also driven by the VVSS driver through M8. While the background cells all store 0 as shown in the figure, VVSS further connects to VDDA through the M8', M6', and M2' in the background cells. The connection of VVSS-VDDA then raises the voltage on VP much more. As a result, to succeed the write-1 becomes harder and lower resistance is detected.

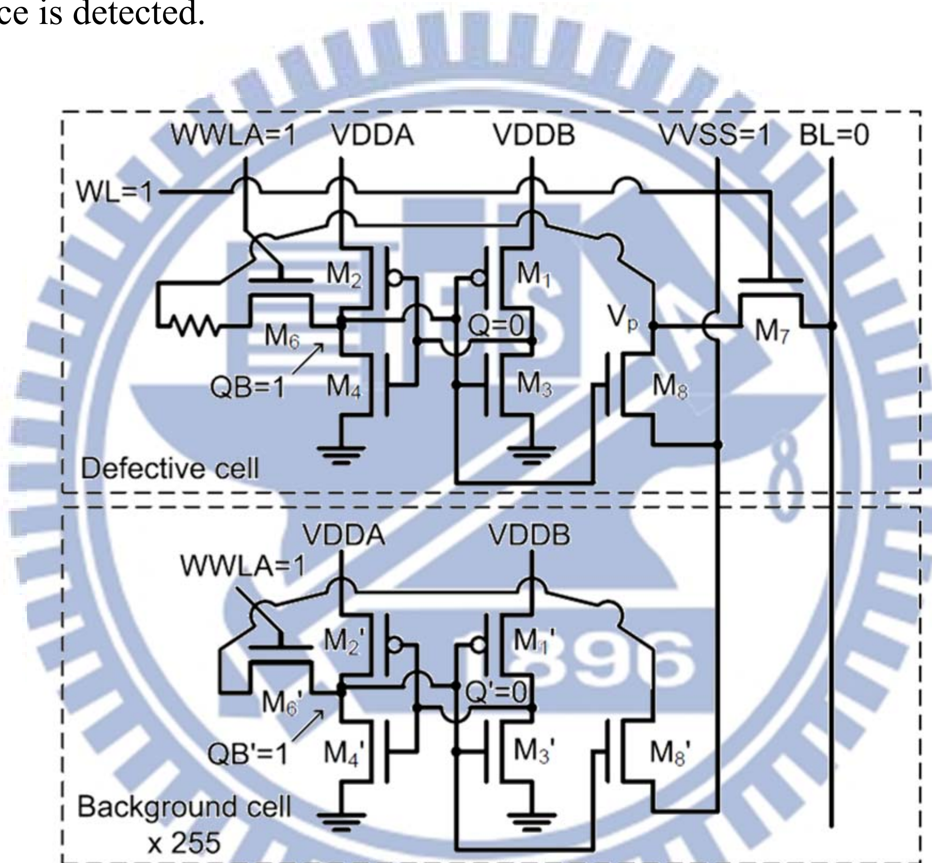


Fig. 3-2. Illustration of how background cells affect the testing of the defect at M6.

To summarize, the March C- detects the defects at M5-M8. For M5–M6, the minimum detectable resistance can be further lowered if all the background cells are set to 0 for write-0 and write-1 respectively. The algorithm {↕(w0); ↕(w1,r1,w0,r0)} as example can achieve the goal. As to the defects at the cross-couple inverters M1–M4 and the DADS circuitry M9–M12, they are either undetectable or hard-to-detect. In the following section, we will introduce the test methods for the detects.

3.3 Test Methods for the Open Defects at the Cross-Couple Inverters

To detect the defects at the cross couple inverters of the 8T SRAM, we apply two test methods. The first one is the floating bitline attacking method. The method is modified from the previous work [3-20]. The second one is a new proposed test method. The method is designed to utilize the feature of the 8T SRAM, which achieves high test efficacy but requires much less test time. At the end of this section, we will compare the two methods together with the March C-.

3.3.1 Floating Bit-Line Attacking Method

According to the non-conventional SRAM categorization in [3-20], the 8T SRAM should be categorized to ”Type-A”. However, the corresponding test method recommended for the defects at cross-coupled inverters requires dual bit-lines during the test. For the 8T SRAM which has only one bit-line, the method is not applicable. We select the floating bit-line attacking (FBA) method which is for another type of SRAM in [3-20] but can be modified for single-bit-line based SRAMs. FBA uses the floating voltage pre-set on bit-line to access and attack the Q (or QB) with inverse logic in the cell. If the data stored in the cell flips, the following read operation can then detect the defect. Figure 3-3 shows an example of using floating-0 as the attacking source on BL. The open defect R1 at M1 is the target to test. As shown in the figure, the Q/QB need to be 1/0 initially. The BL with floating-0 is accessing/attacking the Q through the turned-on M5 and M7. If the open defect makesM1 unable to maintain the 1 on Q, the sense amplifier will output 0 in the following read operation. The defect is then detected. Note that BL with floating-1 can also be applied for detecting R1 if QB is accessed/attacked through M7 and M6 instead. Table 3-III lists the complete control signals of FBA for the four open defect locations. According to the table, WL turns on for every cases. The initial Q/QB value is set depending on the defect locations: 1/0 for M1/M4 and 0/1 for M2/M3. As to WWLA/WWLB, they control the attacking source accessing the node with inverse logic.

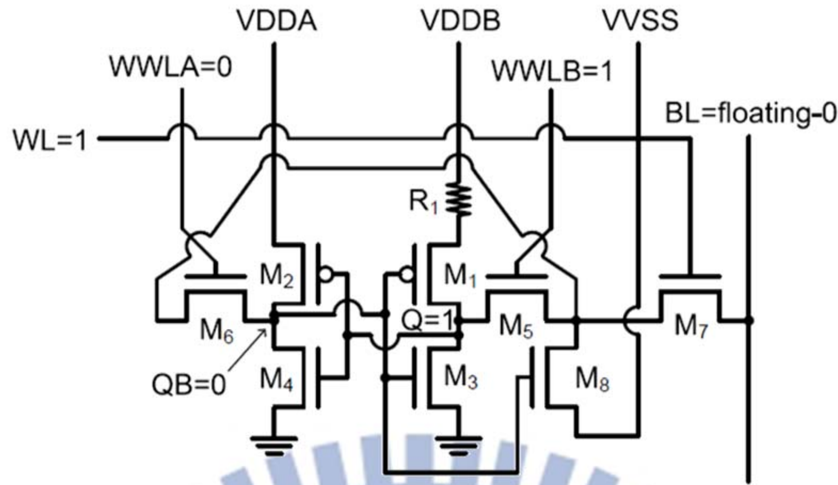


Fig. 3-3. An example of floating bit-line attacking method: using floating-0 on bit-line for detecting the open defects R1.

Table 3-IV shows the test results of FBA. The first column is the attacking source on bit-line. The second and the third columns are the background cells' data and the VVSS logic during the test. The rest of the table lists the test results of the four open defect locations respective. As shown in the table, there are three different results: detected, overtest, and undetectable. For the detected results, the table lists the minimum detectable resistance. For overtests marked "x", the stored data will flip even the SRAM cell is defect-free. Hence, it is inapplicable as well as the undetectable cases marked "-".

TABLE 3-III
Control Signals of the FBA Test Method

Defect Location	Attacking source on BL	Control signals				
		WL	Q	QB	WWLA	WWLB
M_1 & M_4	Floating-0	1	1	0	0	1
	Floating-1	1	1	0	1	0
M_2 & M_3	Floating-0	1	0	1	1	0
	Floating-1	1	0	1	0	1

According to the results, the defect at M1 in never detected since floating-0 in FBA will cause overtest, and floating-1 does not detect any data flipping. However, to M2, floating-0 in FBA with VVSS=1 is applicable for detecting the defect. The minimum detectable resistance is 1MΩ~3MΩ. For nMOSs M3 and M4, floating-1 is more appropriate than floating-0. The minimum detectable

resistance of the defect is 5MΩ~30MΩ and 10MΩ~30MΩ for M3 and M4 respectively.

TABLE 3-IV
Test Results of the FBA Test Method

Attacking source on BL	Environments during test		Test results & Minimum detectable resistance			
	BC	VVSS	M_1	M_2	M_3	M_4
Floating-0	0	0	x	x	-	-
		1	x	1MΩ	-	-
	1	0	x	x	x	x
		1	x	3MΩ	-	x
Floating-1	0	0	-	-	-	x
		1	-	-	30MΩ	10MΩ
	1	0	-	-	-	30MΩ
		1	-	-	5MΩ	30MΩ

BC: data of the 255 background cells x: overtest -: undetectable

For test time, FBA requires three operations for each defect on each cell. The first one is to write the initial data into the tested cell. The second one is the floating bit-line attacking. The last one is a simple read. Here we ignore the global background setting since the {↕ (w0)} or {↕ (w1)} has been achieved by applying a powerful BL driver with turning on multiple word-lines in the current SRAM design. The total number of operations for the complete FBA is hence 3Nx3 for defects at M2–M4. If the defect at M1 is still tried to use FBA for detection, the test time will be 3Nx4. Note that, although floating-1 in FBA which detects the defects at M3 and M4 are with the same BC and VVSS as shown in Table 3-IV, the actual setups for the two are different. As shown in Table 3-III, the initial Q/QB set for M3 and M4 should be different. Therefore, detections for M3 and M4 require test time 3N individually.

3.3.2 Proposed Method: Self-Loop Attacking Method

We introduce a new test method named self-loop attacking (SLA) for the 8T SRAM. The test method detects the open defects not only in the cross-coupled inverters of the cell but also in the DADS circuitry. In this section, we firstly

discuss the detection for the cross-coupled inverters. The part of the DADS circuitry will be in next section.

The SLA method utilizes the specific dual-write-pass-gate structure of the 8T SRAM cell. By controlling the word-lines, the method creates an internal attacking loop of Q/QB inside the cell without BL's accessing. The Q and QB with self-attacking each other will go to a final state which depends on the initial Q/QB, background cells, VVSS, process corners, and the most important one: the existence of defects. If the defects result in a different final state from that of defect-free cells, the following read operation can then detect the faults. Figure 3-4 shows the configuration of the method. Before the test, Q and QB store an initial state. During the test, M7 is off and M5/M6 are on so that Q and QB inside the cell will attack each other. After the test, WWLA/WWLB is off, and the final state of the cell is read. Note that, since the BL is not used during the "self-loop attacking", the test operation hence has chances to be done for the whole array at one time.

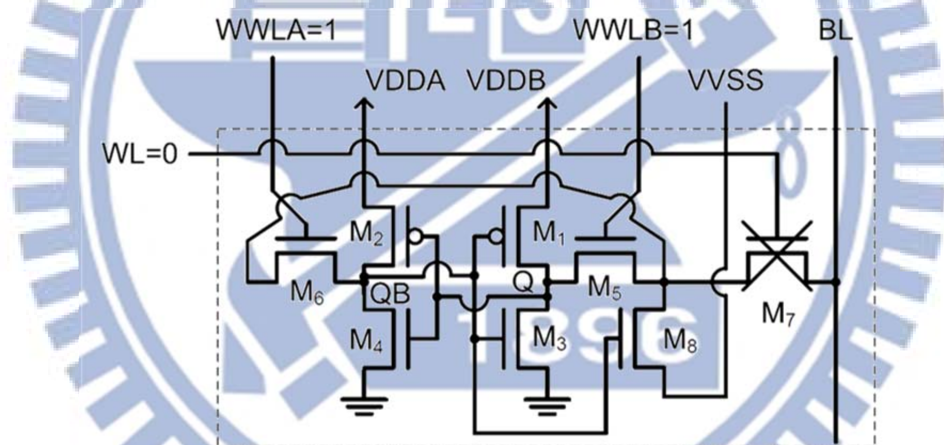


Fig. 3-4. Configuration of the proposed self-loop attacking test method.

Table 3-VI shows the simulations results of the test under process corner TT. The first two columns are the eight possible configurations. The TC_0 means the initial state of the target cell under the SLA test. BC_0 is the initial data stored in the background cells. The third column shows the final states of defect-free SRAMs with the corresponding configuration at left. For example, in Config. 1, the target cell and background cells before the test are both set initially to 0. VVSS is also 0. After test, the final states of target cell and background cells both become 1. The background cells have their data changed because their WWLA/WWLB are shared with the target cell. The turned-on WWLA/WWLB also cause the Qs and QBs in background cells attacking each other. Based on

the final states in the third column, the open defects causing different ones will be detected by the read operation afterwards. The rest of the table lists the minimum detectable resistance of the defected open defects.

TABLE 3-V
Final States of Defect-Free Cells and Test Results of Open Defects of SLA under Process Corner TT

Config.	Logic setup (TC_0, BC_0, VVSS)	Final states (TC, BC)	Min-detectable resistance (Ω)			
			M_1	M_2	M_3	M_4
1	0/0/0	1/1	10M	-	-	10K
2	0/0/1	1/1	300M	-	-	-
3	0/1/0	0/1	-	-	-	-
4	0/1/1	1/1	300M	-	-	300M
5	1/0/0	0/1	-	5M	1M	-
6	1/0/1	0/1	-	10K	8K	-
7	1/1/0	1/1	800M	-	-	-
8	1/1/1	1/1	800M	-	-	-

$TC_{(0)}$: (initial) data of target cell -: undetectable
$BC_{(0)}$: (initial) data of the 255 background cells

According to the results, Config. 1, 2, 4, 7, and 8 can be applied for testing the defect at M1. For M2 and M3, Config. 5 and 6 can detect both the defects. As for M4, Config. 1 and 4 are the appropriate ones. In addition to above simulations run under TT corner, we further examine the final states of defect-free cells under various process corners. It's because only if the final states keep equal for various process corners, the feasibility of above configurations can be verified. Table 3-VI shows the results. In the table, we show process corners FF and SS as examples since they have most different final states. According to the table, Config. 3, 4, 6, and 7 have different final states when test is under SS or FF. The four configurations are hence not feasible for test. As a result, for M1, just Config. 1, 2, and 8 remain applicable. For M2/M3 and M4, only Config. 5 and 1 are applicable respectively. To summarize, we will use the Config. 1 for detecting the defects at both M1 and M4. For M1, it's the one detecting minimum resistance. For M4, it's the only applicable configuration. For M2 and M3, even though Config. 6 detects lower resistance under TT corner, we can only choose the Config. 5 which covers various process corners.

TABLE 3-VI
Final States of Defect-Free Cells under Various Process Corners

Config.	Logic setup (TC_0, BC_0, VVSS)	Final states (TC, BC) under process corners:		
		TT	SS	FF
1	0/0/0	1/1	-	-
2	0/0/1	1/1	-	-
3	0/1/0	0/1	1/0	-
4	0/1/1	1/1	-	0/0
5	1/0/0	0/1	-	-
6	1/0/1	0/1	1/1	1/1
7	1/1/0	1/1	1/0	0/0
8	1/1/1	1/1	-	-

-: the same as the TT corner

The test time of SLA is much less than the previous methods. For M1 and M4 using Config. 1, since the initial states of TC and BC cells are the same, only 1 self-loop attacking operation is needed for the whole array. We just need to turn on WWLA/WWLB of all cells in the array at a time. Then, by an 1N read operation for reading final states of each cell, the defection can complete. The test time for M1 and M4 is hence 1N+1. Here, we also ignore the global background setting as for FBA. As to M2 and M3 using Config. 5, since the initial states of TC and BC are different, we need 1N for writing initial data into the tested cell, 1N for SLA operation, and 1N for the following read. The test time of SLA for M2 and M3 is then 3N. The entire test time for M1–M4 is 4N+1.

3.3.3 Test methods comparison

Table 3-VII summarizes the test methods of test efficacy and corresponding test time. For test efficacy, we list the minimum detectable resistance. Firstly, March C- as the typical test method can only detect the open defects at M1 and M2. The detectable resistance is above 500MΩ. The floating bit-line attacking (FBA) can detect the defects at M3 & M4 and lower the detectable resistance for M2. But the defect at M1 becomes undetectable. The self-loop attacking (SLA) method as the proposed one detects all the open defects at M1–M4. The detected

resistance for M1, M3, and M4 is further lower than that of FBA. For M2, although no lower resistance is detected, the SLA can still achieve 5MΩ near the 1MΩ of FBA.

In addition to the capability of detecting all the defects, the proposed method uses much less test time than the previous methods. As shown in the table, SLA method needs 4N+1 while the FBA needs 9N. Note that the FBA needs 12N if M1 is included. Taking the 256Kb SRAM for example, the N is 8192 (4 words x 256 rows x 8 blocks). The FBA needs 73728 to 98304 operations. SLA needs only 32769 operations. The test time of SLA is 33%~44% of FBA's.

TABLE 3-VII
Test Methods Comparison of Test Efficacy and Test Time

Test method	Min-detectable resistance (Ω)				Test time
	M_1	M_2	M_3	M_4	
March C-	800M	500M	-	-	10N
FBA	-	1M	5M	10M	9N
SLA	10M	5M	1M	10K	4N+1

3.4 Testing for The Open Defects in the DADS Circuitry

In section 3.2, the open defects in the DADS circuitry have been shown undetectable by the March C- test. When trying to apply FBA discussed in previous section for test, the FBA method actually does not detect the defects as well. It's because FBA only triggers one cell in the column by which the requested supply current is limited. The DADS circuitry even with open defects can always provide such supply current that no data is flipped in the cell under FBA. On the contrary, the proposed SLA method triggers all the cells in the column by turning on both the WWLA/WWLB. Much more supply current is requested by the cells in the column. Since the high current loading may not be satisfied, the SLA method therefore has chances to detect the defects in the DADS circuitry. Table 3-VIII shows the test results of SLA. In the table, we only apply Config. 1, 2, 5, and 8 of SLA which can cover the various process corner as listed in Table 3-VI.

TABLE 3-VIII
Test Results of SLA for Open Defects in The DADS Circuitry

Config.	Logic setup (TC_0, BC_0, VVSS)	Min-detectable resistance			
		M_9	M_{10}	M_{11}	M_{12}
1	0/0/0	1MΩ	-	-	-
2	0/0/1	1MΩ	-	-	-
5	1/0/0	-	1MΩ	-	-
8	1/1/1	10MΩ	-	-	-

For M9, SLA with Config. 1, 2, and 8 can detect the defect. The minimum detectable resistance is 1MΩ. For M10, the defect is detected but with Config. 5 instead. The min-detectable resistance is also 1MΩ. As to M11 and M12, the defects do not cause faults and are still undetected. In fact, it's because the function of M11 and M12 for the SRAM operated under process corner TT, as for above experiments, is not crucial. When the SRAM operates under the other process corner that M11 and M12 are critical, the defects would then cause faults and should be detected. Figure 5 as example can help determine the process corner under which the function of M11 and M12 is critical.

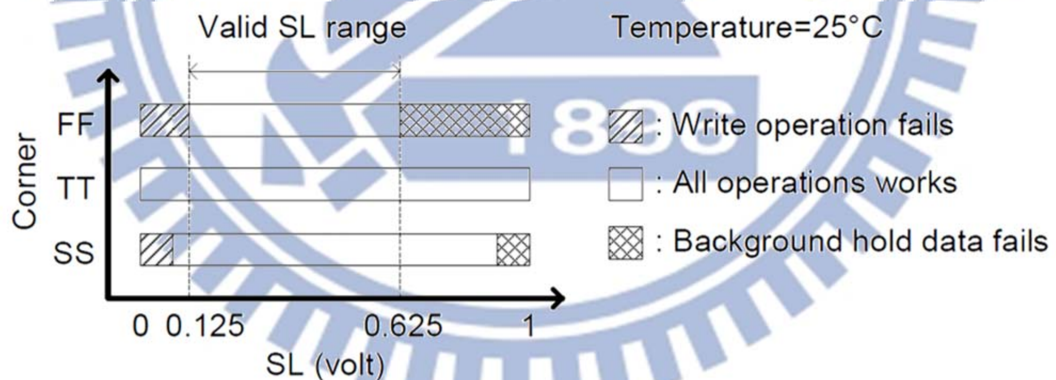


Fig. 3-5. Valid SL range under different process corners.

In Figure 3-5, the x-axis is the voltage of SL which is the control of M11 and M12. The three bars indicate the SRAM's response under each process corner. When SL is very low (as shown in left side of the bars), the write operation may fail depending on the process corner. When SL is too high (right side of the bars), the background cells may fail holding the data. As a result, here is a valid SL range with lower and upper bounds. As shown in the figure, the valid SL ranges for various process corners are different. For FF corner, it is

most limited, especially with lowest upper bound. This indicates the SRAM under FF corner is most sensitive to the turning off of M11 and M12 (High SL voltage turns off M11 and M12). Since open defects cause similar effect as turning off a MOSFET, they most likely cause faults for the SRAM under FF corner. Based on the fact, we further simulate the SLA test method for the defects at M11 and M12 under FF corner. The voltage of SL keeps 0.5*VDD as for previous experiments. March C- and FBA are also applied for comparison. Table 3-IX shows the results. According to the results, March C- and FBA are still unable to detect the defects, but the proposed SLA method with Config. 2, 5, and 8 can. The minimum detectable resistance for M11 and M12 is 1MΩ and 10MΩ respectively.

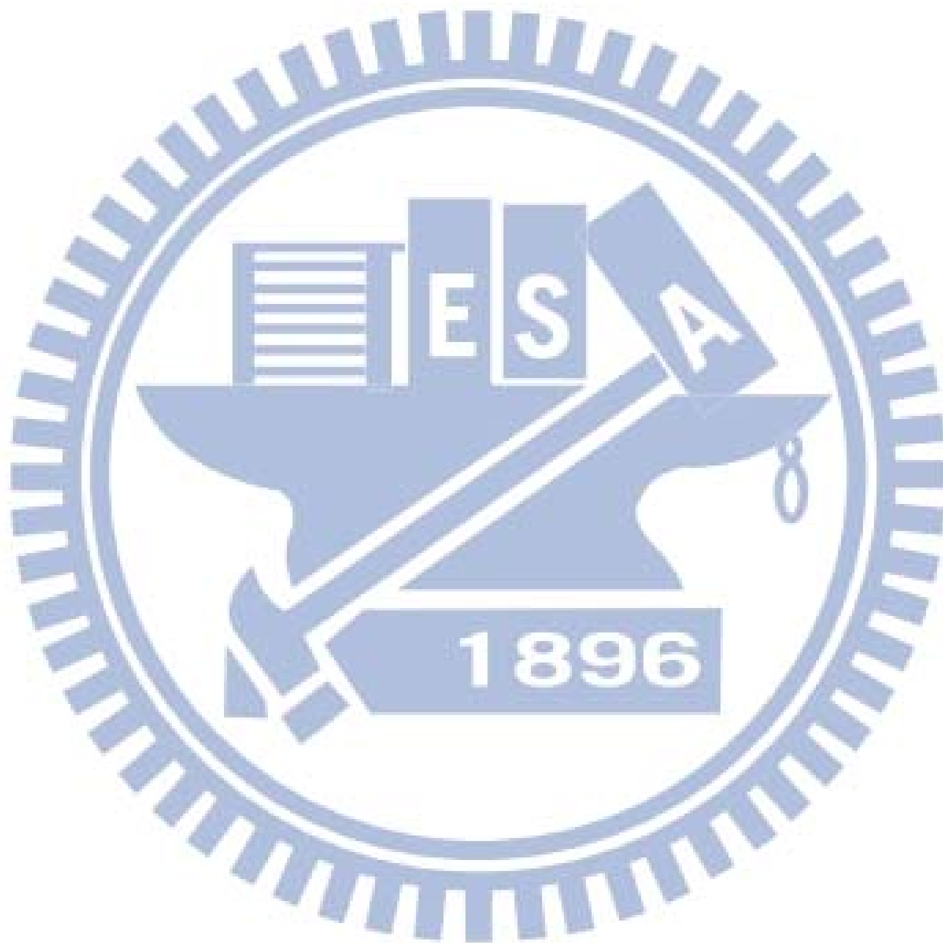
TABLE 3-IX
Simulation Results of Testing the Defects at M11 and M12 but under FF Process Corner

Test method (under FF corner)		Min-detectable resistance	
		M_{11}	M_{12}
March C-		undetectable	undetectable
FBA		undetectable	undetectable
SLA	Config. 1	undetectable	undetectable
	Config. 2	1MΩ	10MΩ
	Config. 5	undetectable	10MΩ
	Config. 8	1MΩ	undetectable

3.5 Conclusion

In this chapter, we discuss the fault behavior and the testing of open defects in an existing low-VMIN DADS 8T SRAM. Firstly, by applying the conventional March C-, we discover the defects at the read/write pass-gates and the read-path can be detected. But for the cross-couple inverters in cell and the DADS circuitry, the defects occurring there are undetected. For the detectable defects, we further propose a March algorithm that the detectable resistance can be even lowered. As to the undetected ones, we firstly analyze and explain why the defects are undetected by normal operations. Then, we propose a novel test

method by utilizing the specific dual-write-word-line structure of the 8T SRAM. The proposed method creates an internal attacking loop inside the cell. By self-attacking, all the undetected defects in cell are detected. Besides, since BL is not required during the test, the self-attacking can operate for the whole array at one time. The method hence requires much less test time. When compared to the conventional FBA method which still leaves one defect undetected, the test time of the proposed method is only 33%~44% of FBA's. In addition to detecting the defects in cell, the proposed method is also proved able to detect all the defects in the DADS circuitry.



Chapter 4
A Novel Circuit-level Model for Gate Oxide Short and its Testing Method in SRAMs

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 [4-1] [4-2] [4-3]. 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 has been an inevitable and challenging task for advanced technologies [4-1] [4-3] [4-4] [4-5] [4-6] [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 illustrated in Figure 4-1(a). Such a gate-to-source GOS can be properly modeled as a resistive short between the gate and the source of the MOSFET [4-2] [4-8] [4-9] [4-10], and can be effectively and efficiently detected by using conventional stuck-at-fault test in logic and general March algorithm in SRAM. As a result, 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.

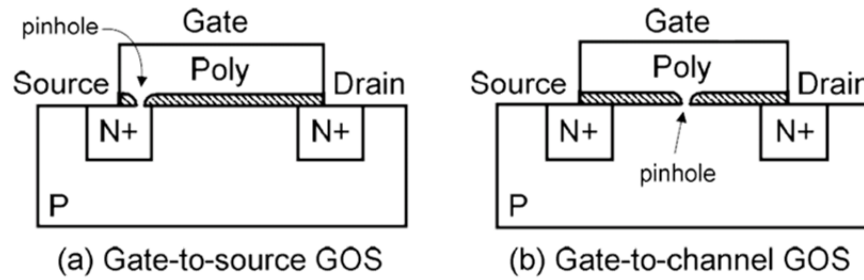


Fig. 4-1. Cross-section view of a GOS-impacted MOSFET.

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 Figure 4-1(b) which was commonly used in previous works [4-8] [4-11] [4-12]. 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 ID (current flows from channel to drain) exists, and (4) the resistance between the gate and source/drain is gate-voltage controlled. Figure 4-2 shows the IDVD curves without and with a gate-to-channel GOS. As a result, such faulty behavior of a gate-to-channel GOS is much more complicated to model at the circuit level. Furthermore, 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 [4-2] [4-12] [4-13] [4-14] [4-15] [4-16] [4-17] [4-18] [4-19].

In order to properly estimate the effectiveness of the GOS test, several circuit-level models of a gate-to-channel GOS have been proposed in the previous works and can be divided into three types: (1) the bi-dimensional model [4-11], (2) the split model [4-2] [4-14] [4-12], and (3) the nonlinear non-split model [4-20] [4-21] [4-22]. The bi-dimensional model contains a mutually connected MOSFET array and hence requires higher computation time during simulation. Also, it cannot represent a minimum-size MOSFET and hence cannot be applied to the most advanced technologies. A split model simplifies the 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 non-split 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 non-split model fails to represent the transient characteristics of a GOS, and hence cannot be applied to

SPICE transient simulation.

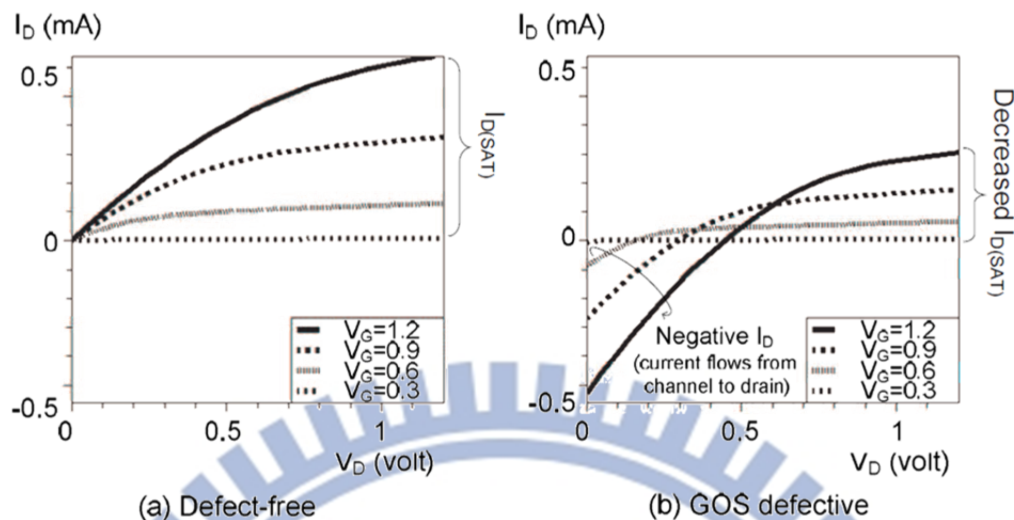


Fig. 4-2. IDVD curves of a nMOS (a) without and (b) with a gate-to-channel GOS.

Fig. 4-2. IDVD curves of a nMOS (a) without and (b) with a gate-to-channel GOS. Note that the bi-dimensional model was considered as the golden model for GOS DC characteristics. The later split model and nonlinear non-split 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 has validated the correctness of the proposed GOS model through silicon or Technology-CAD simulation for process technology under 1.5μm [4-2], which leaves the precision of the above models unclear. Furthermore, 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 have suggested to use delay test for detecting a GOS in logics [4-23] [4-24] [4-25], which relies on an accurate transient GOS model to evaluate the effectiveness of the test in transient simulation.

In this paper, we will propose a novel nonlinear non-split model for gate-to-channel GOS defects, which can represent a minimum-size MOSFET and provide higher accuracy for the GOS DC characteristics. Also, the proposed model can accurately represent the GOS transient characteristics by considering the capacitance change imposed by the GOS defect, which has not been discussed in any of the previous works. A series of experiments will prove the superiority of the proposed model on both DC and transient characteristics fitting, and all the results will be directly compared to a 3D Technology-CAD

(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 will also demonstrate the difference if other GOS models are used in the test evaluation process.

4.1 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 [4-26] with 3D-structure representation. Note that a gate-oxide pinhole in 2D structure (as shown in Figure 4-1(b)) will form an oxide trench above the channel across the entire channel width, and hence cannot successfully describe a GOS defect. As a result, the TCAD tool in use must be with 3D-structure representation in order to simulate a GOS-impacted MOSFET. The channel length and width of either a nMOS or a pMOS are both 65nm for the experiments of this paper. Table I reports the DC characteristics and the parasitic capacitance of a defect-free nMOS and pMOS, respectively.

TABLE 4-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

Based on 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. Figure 4-3 shows the 3D representation of such a GOS-impacted MOSFET. The level of a GOS here is determined by the radius of the injected pinhole. The larger the injected pinhole, the severer the shorting between the gate and the channel will be. The pinhole radius shown in Figure 4-3 is 2.5nm.

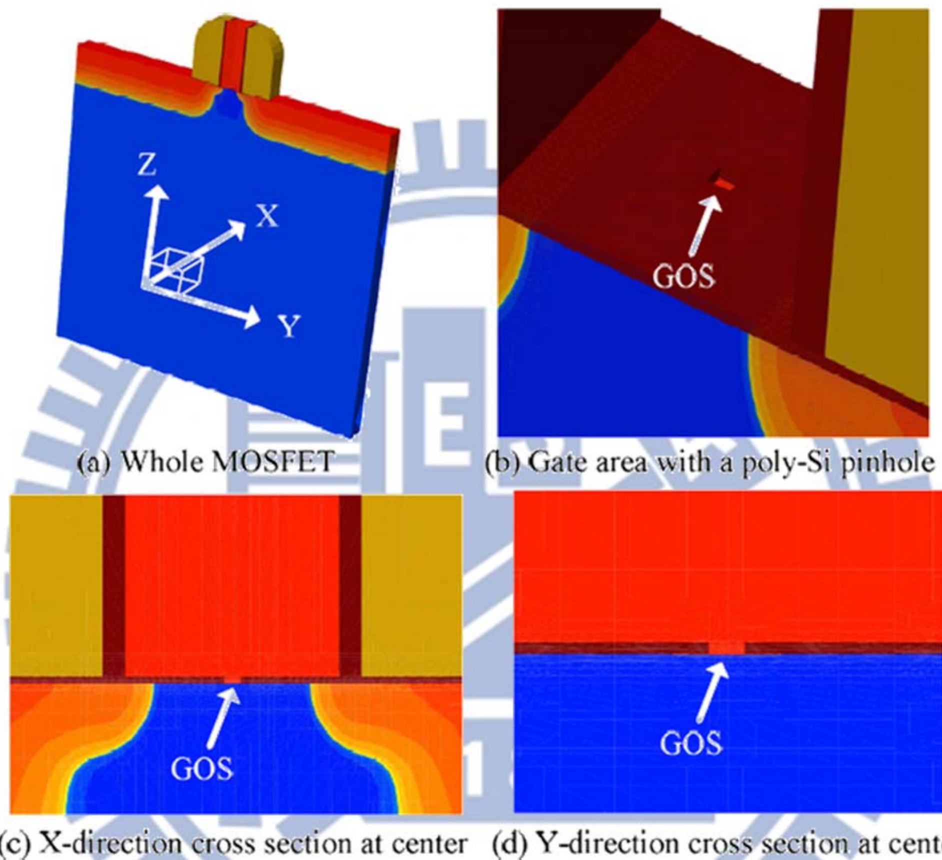


Fig. 4-3. Representation of a 3D GOS-impacted MOSFET in TCAD.

Since 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 [4-27] to extract the BSIM model card [4-28] for a MOSFET based on its IDVD, IDVG, gds, gm, and parasitic capacitances obtained from the TCAD simulation. In Figure 4-4, 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. Figure 4-4(a) and Figure 4-4(b) further show the zoom-in result on the output's falling edge and rising edge, respectively. As the figure shows, the HSPICE result represented by the red curves fits the TCAD result represented by the blue

curves quite precisely.

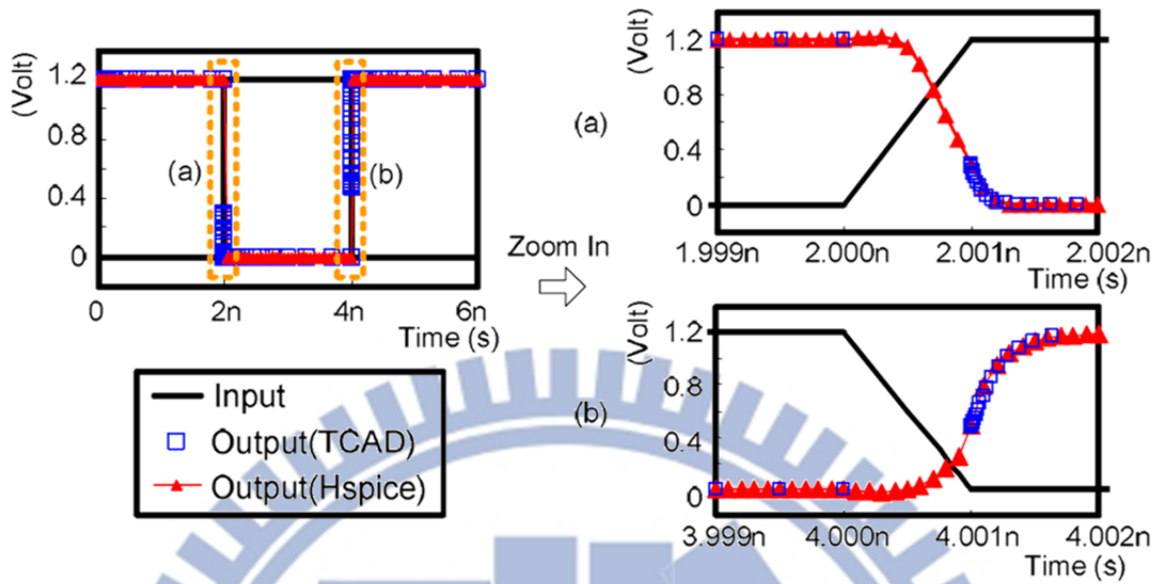


Fig. 4-4. Comparison of an inverter's transient response between the TCAD simulation and the HSPICE simulation with extracted model cards.

4.2 Previous Circuit-Level GOS Models

In this section, we will introduce three circuit-level GOS models: the Bi-dimensional model [4-11] and two nonlinear non-split models ([4-21] and [4-22]). The Bi-dimensional model is the first GOS model and the two nonlinear non-split models [4-21] [4-22] were proposed for representing minimum-size devices. For both the nonlinear non-split models [4-21] [4-22], we will validate their accuracy by comparing to the TCAD-simulation results and analyze the corresponding limitations.

4.2.1 Bi-dimensional model

Figure 4-5 shows the schematic of a 5x5 Bi-dimensional GOS model, which contains an MOSFET array with all the gates connected together. 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 drain terminal. The shorting effect caused by the GOS is represented by the resistor, denoted as RGOS, which connects the gate

terminal to the center of the connected MOSFET array. Since 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 relatively old technologies.

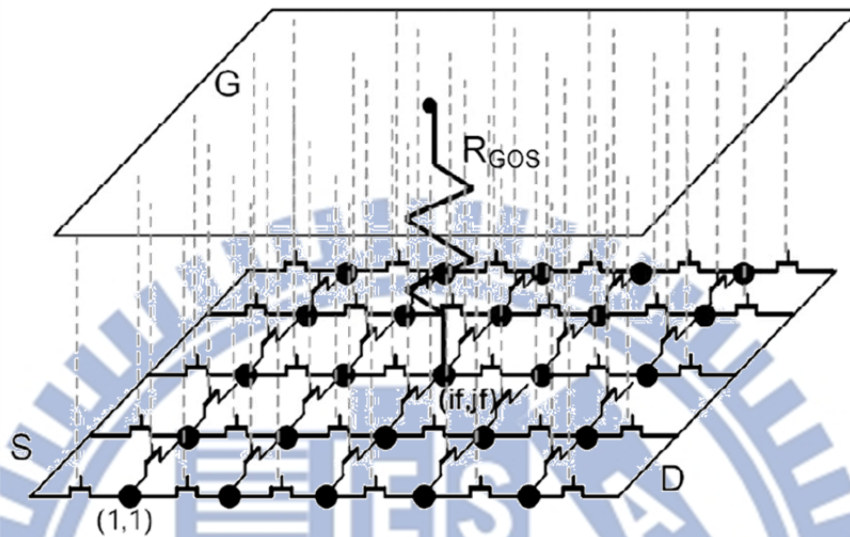


Fig. 4-5. An exemplary Bi-dimensional model with 5x5 internal points [4-11].

4.2.2 Nonlinear non-split model 1 - JET_03

Figure 4-6 shows the nonlinear non-split model JET_03 proposed in [4-21], which utilizes three MOSFETs (named Tm, Ta, and Tb in Figure 4-6) and a resistor (named RGOS in Figure 4-6) to describe a GOS-impacted MOSFET. According to the fitting guide provided in [4-21], the size of Tm is tuned to fit the reduced saturation drain current ($I_{D(SAT)}$). The Ta and Tb are tuned to fit the negative drain current and the gate current, respectively. The resistor R_{GOS} is used to refine the fitting.

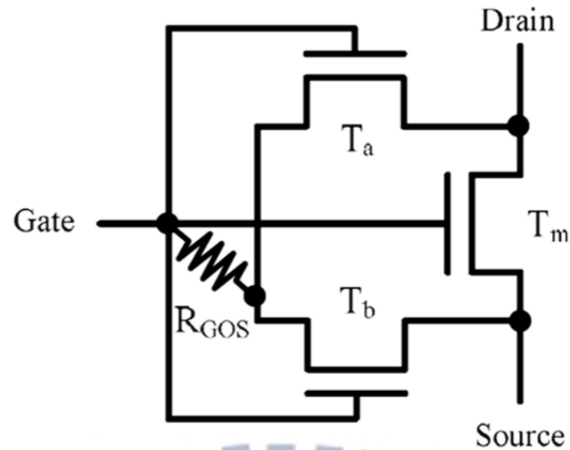


Fig. 4-6. The nonlinear non-split GOS model, JET_03, proposed in [4-21].

Figure 4-7(a) shows the IDVD fitting results by using JET_03 model to represent a nMOS with a 2.5nm-radius GOS. As the result shows, the maximum $I_{D(SAT)}$ and the negative $I_{D(off)}$ can be described quite well by JET_03 model in Figure 4-7(a), but the other IDVD curves with different V_G cannot. Next, Figure 4-7(b) shows the HSPICE transient simulation of an inverter whose nMOS contains the same size GOS modeled by JET_03 model as that in Figure 4-7(a). 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 Tm and Ta. 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.

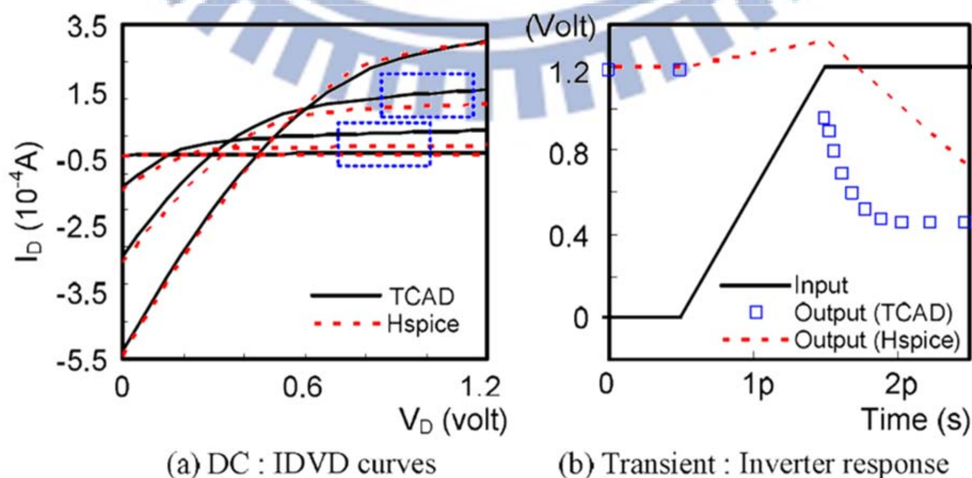


Fig. 4-7. The GOS fitting results by using JET_03 model.

JET_03 model was claimed to be able to describe a minimum-size MOSFET in [4-21] since its source and drain is connected by only one MOSFET, Tm. However, when fitting the reduced $I_{D(SAT)}$, we have to increase the length of Tm, which turns Tm no longer a minimum-size MOSFET. Similar situation occurs when tuning the size of Tm, Ta, and Tb for fitting other DC curves. As a result, 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.

4.2.3 Nonlinear non-split model 2 - IDT_09

Figure 4-8 shows the nonlinear non-split GOS model IDT_09 proposed in [4-22], which uses only one MOSFET along with extra three current sources and is more suitable for representing a minimum-size GOS-impacted MOSFET. 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 3rd-order polynomial of V_{GD} and V_{GS} as shown in Equation 4-1 and Equation 4-2, respectively.

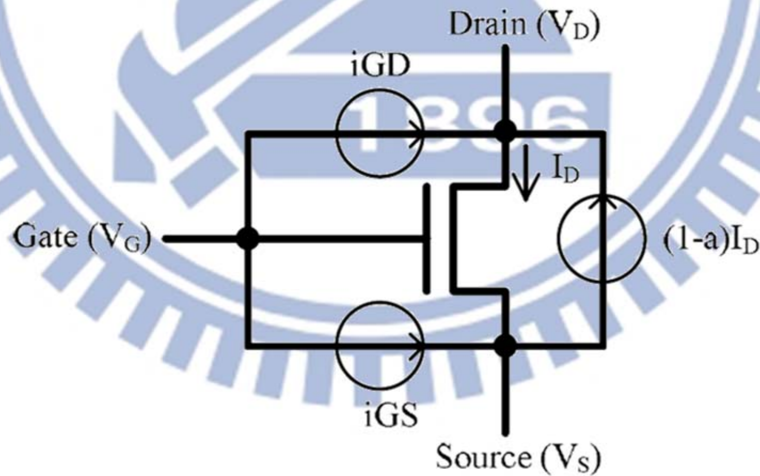


Fig. 4-8. The nonlinear non-split GOS model, IDT_09, proposed in [4-22].

$$iGD = a_1 \cdot V^3_{GD} + b_1 \cdot V^2_{GD} + c_1 \cdot V_{GD} + d_1 \quad (4\text{-}1)$$
$$iGS = a_2 \cdot V^3_{GS} + b_2 \cdot V^2_{GS} + c_2 \cdot V_{GS} + d_2 \quad (4\text{-}2)$$

Figure 4-9 shows the results of applying IDT_09 GOS model to the same experiment as in Figure 4-7. As Figure 4-9(a) shows, most of the IDVD curves

match the TCAD result quite well. 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. Based on further analysis, this fitting error results from the limitation of using a polynomial to represent the current source iGD, even though the order of the polynomial is increased to more than 3.

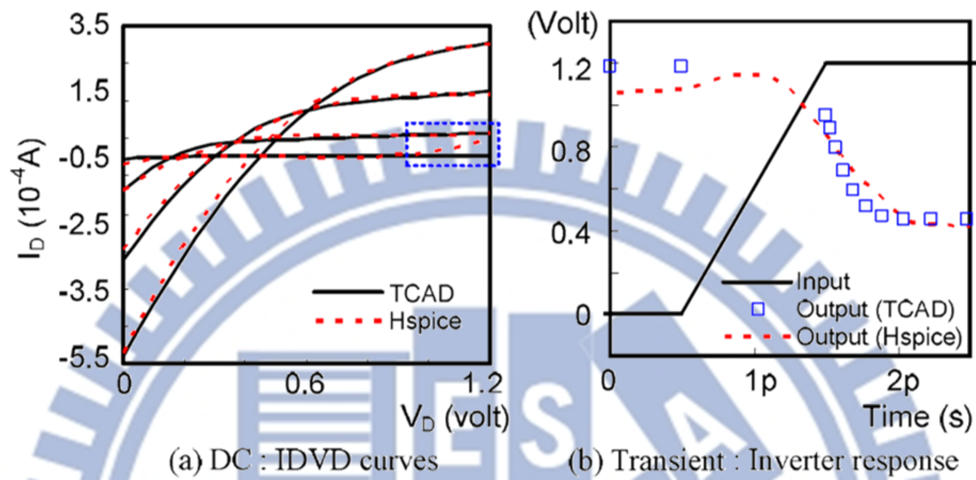


Fig. 4-9. The GOS fitting results by using IDT_09 model.

As the transient-simulation result shown in Figure 4-9(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. Also, 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 has not been 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 4.3.3.

4.3 Proposed GOS Model and The Comparison With Previous Works

4.3.1 Proposed GOS model

Figure 4-10 shows the schematic of our proposed GOS model, which uses only one MOSFET along with three current sources (iGD, iGS, and iSD) and two voltage-controlled capacitors (C_{GS} and C_{GD}). As a result, the proposed GOS model can represent a minimum-size GOS-impacted MOSFET. The three current sources iGD, iGS, and iSD are represented by Equation 4-3, Equation 4-4, and Equation 4-5, respectively, which are different from the equations used in IDT_09 model. First, Equation 4-3 uses a V_{GD} shift parameter and a minimum limitation for iGD to prevent the overly large drain current when V_G is low and V_D is high, as shown in Figure 4-9(a). Second, iSD is actually V_{GS} controlled, not a function of I_D as used in the IDT_09 model, and hence Equation 4-5 uses a 3rd-order polynomial of V_{GS} to represent iSD. Third, in Equation 4-4, iGS is simplified to a 2_{nd}-order polynomial of V_{GS} since it can generate the same accuracy as a higher-order one based on our experiments.

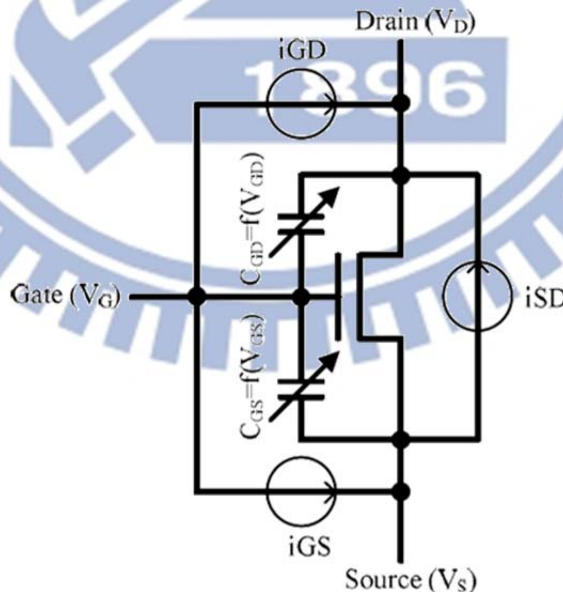


Fig. 4-10. Schematic of the proposed GOS model.

$$iGD = [a_1 \cdot (V_{GD}+ \gamma)^3 + b_1 \cdot (V_{GD}+\gamma)^2+c_1 \cdot (V_{GD}+\gamma) + d_1]_{min=0} \quad (4\text{-}3)$$

$$iGS = a_2 \cdot V^2_{GS} + b_2 \cdot V_{GS} + c_2 \quad (4\text{-}4)$$

$$iSD = a_3 \cdot V^3_{GS} + b_3 \cdot V^2_{GS} + c_3 \cdot V_{GS} + d_3 \quad (4\text{-}5)$$

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 piece-wise linear function of V_{GS} and V_{GD} respectively, which is a default expression supported by HSPICE. Figure 4-11 shows the C-V curves for a defect-free MOSFET, a MOSFET with a 2.5nm-radius GOS, and a MOSFET with a 5nm-radius GOS, obtained from our TCAD simulation. As shown in the figure, $C_{GS/GD}$ of a GOS-impacted MOSFET is significantly different from that of a defect-free MOSFET and may vary with different $V_{GS/GD}$. The difference is up to 1.5E-17 F for nMOS and 2.3E-17 F for pMOS, respectively, at $|V_{GS/GD}|$=1.2 volt. Note that defect-free MOSFET has its gate/source/drain terminal capacitance around 8E-17 F as shown in Table 4-I. In other words, a GOS may lead to 19% and 29% capacitance reduction for nMOS and pMOS, respectively. As a result, the two voltage-controlled capacitors added in our proposed model are necessary for fully representing the transient characteristics of a GOS-impacted MOSFET.

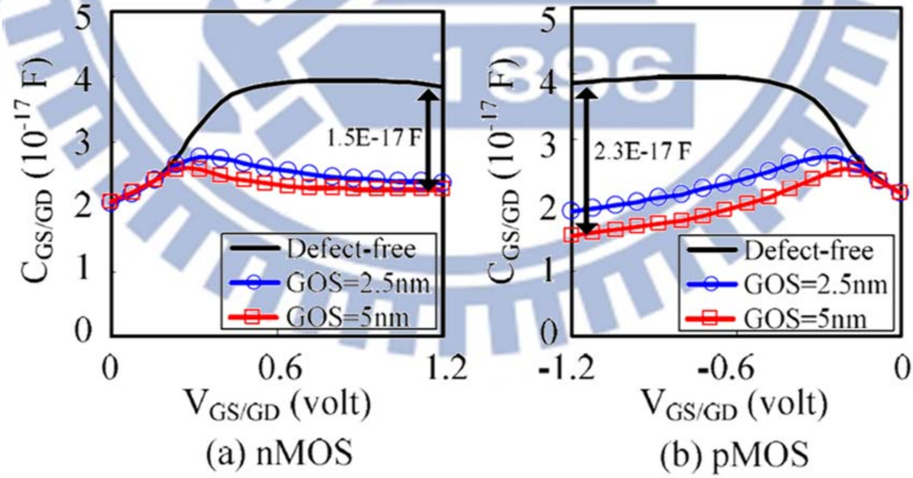


Fig. 4-11. TCAD C-V simulation results for (a) nMOS and (b) pMOS, with and without a GOS.

4.3.2 Simulation comparisons on DC characteristics

Figure 4-12 first shows the IDVD fitting results of 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.

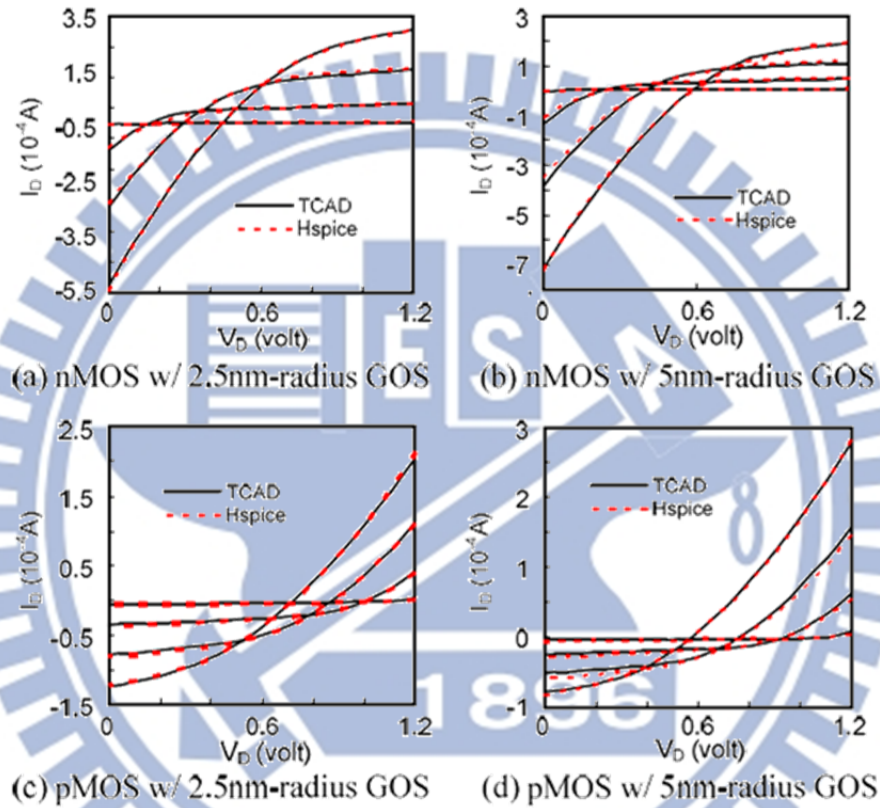


Fig. 4-12. DC fitting results by using the proposed GOS model.

In Table 4-II, we compare the DC fitting errors of the proposed model with those of 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 4-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 to JET_03 model. However, IDT_09 model and JET_03 model may result in 31%~124% more error than that of our 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.

TABLE 4-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%

Figure 4-13 shows the transfer function of a GOS-impacted inverter simulated with TCAD, JET_03 model, IDT_09, and the proposed model. The defective inverter contains a minimum-size pMOS and nMOS with a 5nm-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 VG as shown in Figure 4-7(a). The error resulting from IDT_09 model is caused by the overly large turn-off drain current when V_D is high, and hence mostly occurs when the inverter's input voltage is low.

4.3.3 Simulation comparisons on transient characteristics

Figure 4-14(a) and Figure 4-14(b) show the transient-simulation result of an inverter with a 2.5nm-radius and 5nm-radius GOS on its nMOS, respectively, based on TCAD and different GOS models. Note that we only show the inverter's 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.

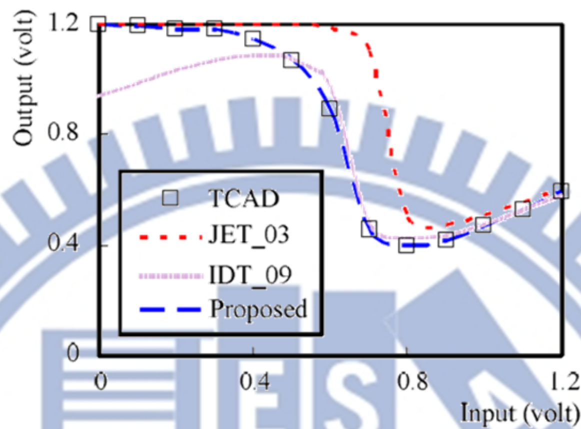


Fig. 4-13. An inverter's transfer function resulting from TCAD and different GOS models when a 5nm-radius GOS locates at the nMOS.

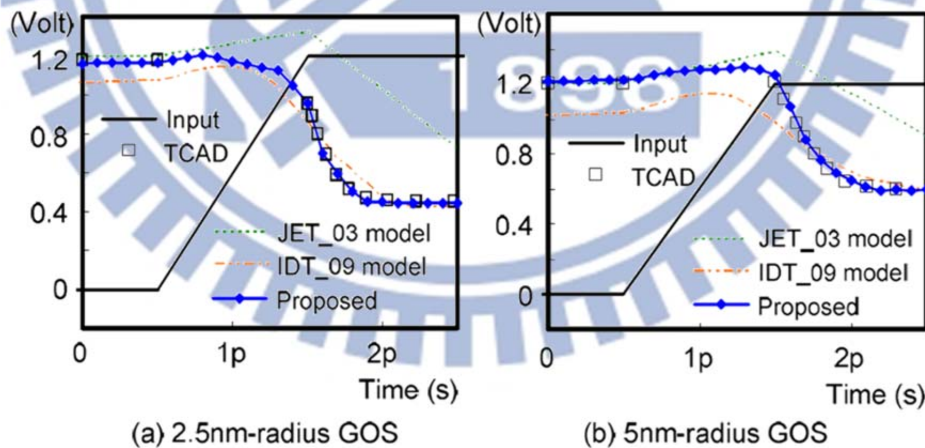


Fig. 4-14. Transient response of an inverter with a GOS on nMOS resulting from TCAD and different GOS models.

A similar experiment is applied to the transient simulation of an inverter with the GOS on pMOS, and the result is shown in Figure 4-15. Again, the same conclusion can be drawn as the proposed model can closely match the TCAD result.

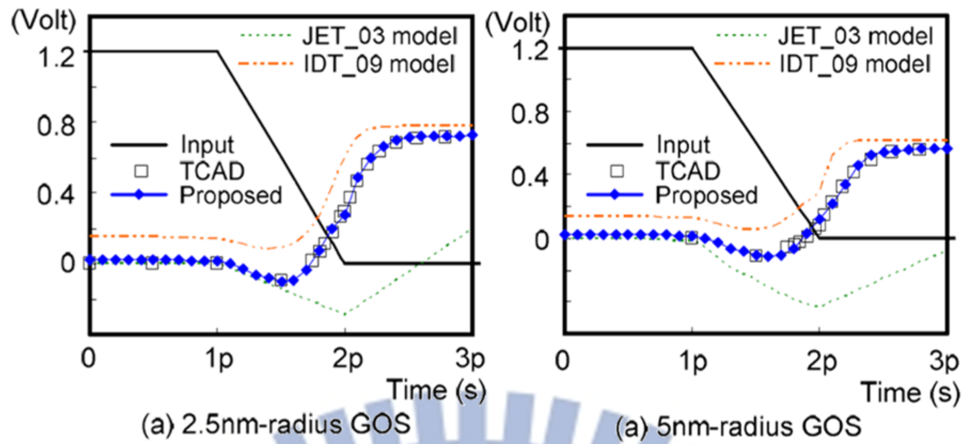


Fig. 4-15. Transient response of an inverter with a GOS on pMOS resulting from TCAD and different GOS models.

Fig. 4-15. Table 4-III further summarizes the error of an GOS-impacted inverter's output delay time for all the above cases in Figure 4-14 and Figure 4-15, 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 GOS transient characteristics and the importance of considering the GOS-imposed capacitance change in a circuit-level GOS model.

TABLE 4-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%

4.3.4 Modeling GOS defects with different locations

In this subsection, we will discuss the precision of different GOS models when the location of a GOS is not in the middle of the MOSFET's channel. Figure 4-16 illustrates the potential nine locations of a GOS used in the following experiment, denoted from a to i. Among the nine locations, ”e” locates at the center of the MOSFET channel, and the rest defect locations surround ”e” with minimum distance 10nm of each other. At each location, a 2.5nm-radius GOS will be injected, and the corresponding DC and transient behaviors will be extracted from TCAD simulation.

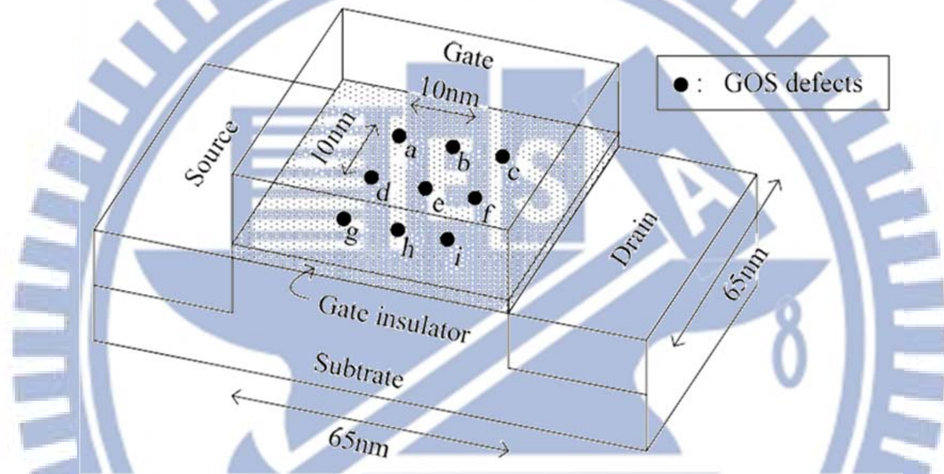


Fig. 4-16. GOS defect with nine different locations.

Table 4-IV firstly 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 Figure 4-16.

TABLE 4-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

From the transient aspect, Table 4-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 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%.

TABLE 4-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

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

4.4 Testing GOS in SRAMs

Testing GOS in SRAMs has been studied in several previous works [4-3] [4-16] [4-29]. As reported, serious GOS will cause SRAM read/write fail [4-3] and can be tested by general March tests. As for the weak GOS, IDDQ test is usually recommended for the detection [4-16] [4-29]. 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 32Kx32 SRAM with 1.2-volt 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.

4.4.1 Previous test methods

1) Conventional March test : Table 4-VI reports the result of applying the March C- algorithm to detect the GOS with different defect sizes (radius of 2.5nm and 5nm) 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.

TABLE 4-VI
Result 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

If the GOS is on a pull-up pMOS, only the GOS with larger defect size (radius of 5nm) can be detected. For defect size of 2.5nm, although the SRAM has reduced S-SN voltage 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 which is 5.1X~6.9X larger than the GOS induced leakage current. As a result, the voltage different on BL/BLB remains recognizable to the sense amplifier.

Note that the GOS defects escaping from a March test in Table 4-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 [4-16] [4-29], 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 4-VI for IDDQ test. Table 4-VII shows the corresponding IDDQ sensitivity [4-30] which is defined as the ratio of the extra current imposed by the defect over the overall current of the targeted defect-free circuit. Also, we measure the IDDQ sensitivity not only at the hold mode but also at the write mode as recommended by [4-29]. 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 advanced process technologies. Following are the reasons.

TABLE 4-VII

IDDQ Sensitivity Caused by A GOS Based on A 1.2V 32KX32 SRAM

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

First, the IDDQ sensitivity calculated in Table 4-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. As a result, 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 since an SoC chip can easily contain more than hundreds of SRAM macros. More importantly, the current calculated in Table 4-VII is for the TT corner. The process variation in 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 45X of that at TT corner for a UMC 65nm process. Then the corresponding IDDQ sensitivity for the cases in Table 4-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.

3) Weak write test mode : For detecting data retention faults in the SRAM, weak write test mode (WWTM) [4-45] 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 Figure 4-17. During the weak write operation, BL/BLB connects to VDD/GND through at least two pass-gate MOSFETs. As a result, the voltage on BL/BLB with logic 1 is lower than 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.

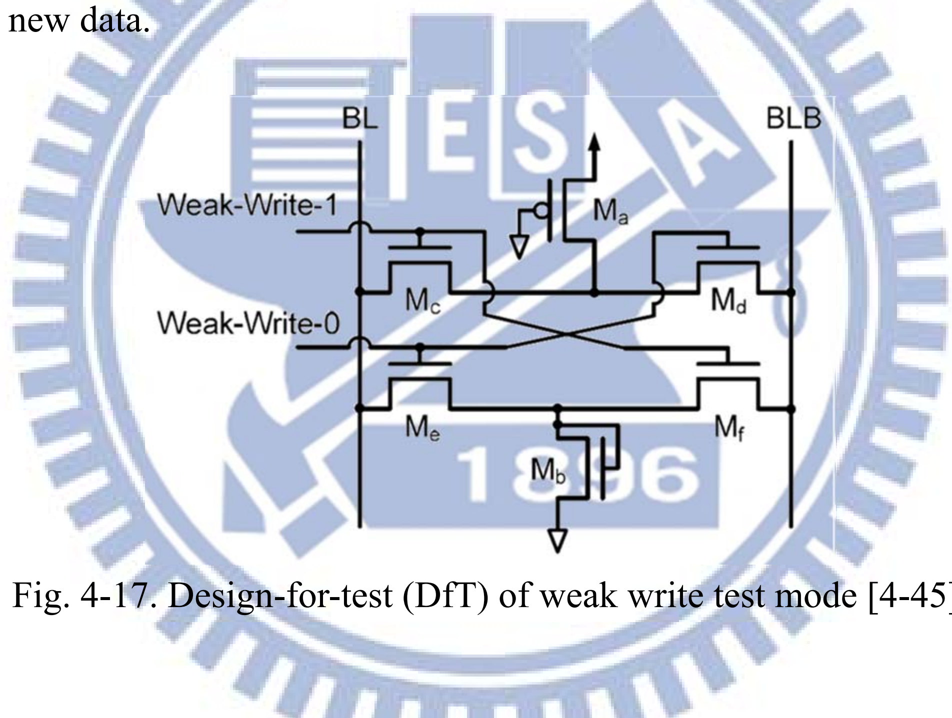


Fig. 4-17. Design-for-test (DfT) of weak write test mode [4-45].

In our experiment, we also applied WWTM to test GOS in the SRAM. With all the pass gates MaMf in Figure 4-17 set to the minimum size (65nm/65nm), 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 device length leads to enormous device characteristic variation. Only few nanometer (1~2nm) of device length range for the pass gate is applicable to distinguish a defective cell from a healthy one. As a result, the narrow tunable window exceedingly limits the test effectiveness of WWTM for detecting GOS under various process corners.

4.4.2 Proposed DFT write operation

To detect GOS in SRAM, we propose a new DFT write operation in the test mode, which contains two special configurations: (1) floating bit-line (BL & 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. Figure 4-18 illustrates the concept of this proposed test method, where the GOS could be on a pull-up pMOS M1 or a pass-gate nMOS M2. To detect the GOS at either M1 or M2, the BL and BLB are both set floating with a capacitance C_T . The C_T may come 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.

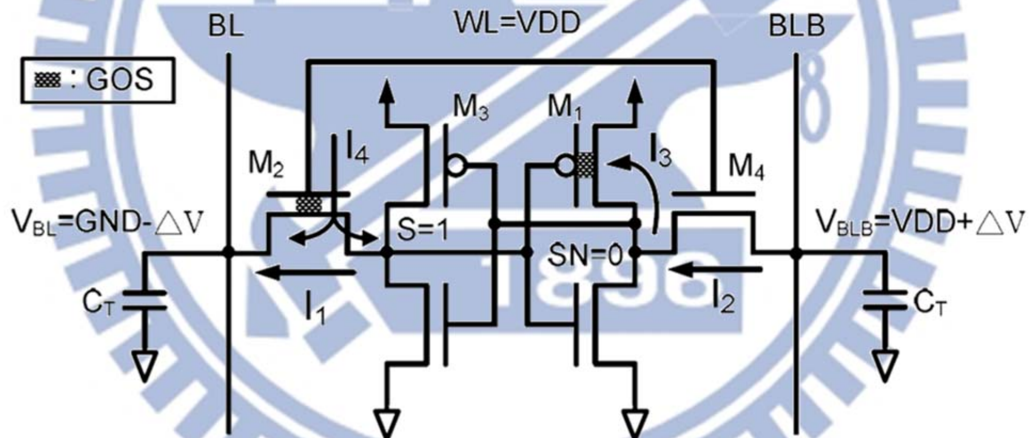


Fig. 4-18. Concept of the proposed DFT write operation for detecting GOS at pull-up pMOS and pass-gate nMOS.

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 M1, 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. Also,

when a GOS locates at the pass-gate nMOS M2, 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. As a result, the DFT write operation may fail since 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 M3 and pass-gate nMOS M4 by writing opposite data.

4.4.3 Detailed simulation result for detecting GOS

Figure 4-19 shows how the GOS at pull-up pMOS (M1 in Figure 4-18) is detected by the proposed DFT write operation by comparing the simulation results with and without the GOS defect. Figure 4-19(a) shows the waveforms of V_S/V_{SN}, $I_1/I_2/I_3$, and V_{BL}/V_{BLB} when applying the DFT write operation to a defect-free SRAM cell. Figure 4-19(b) shows the corresponding waveforms when applying the DFT write operation to a SRAM cell with a 2.5nm-radius GOS at M1.

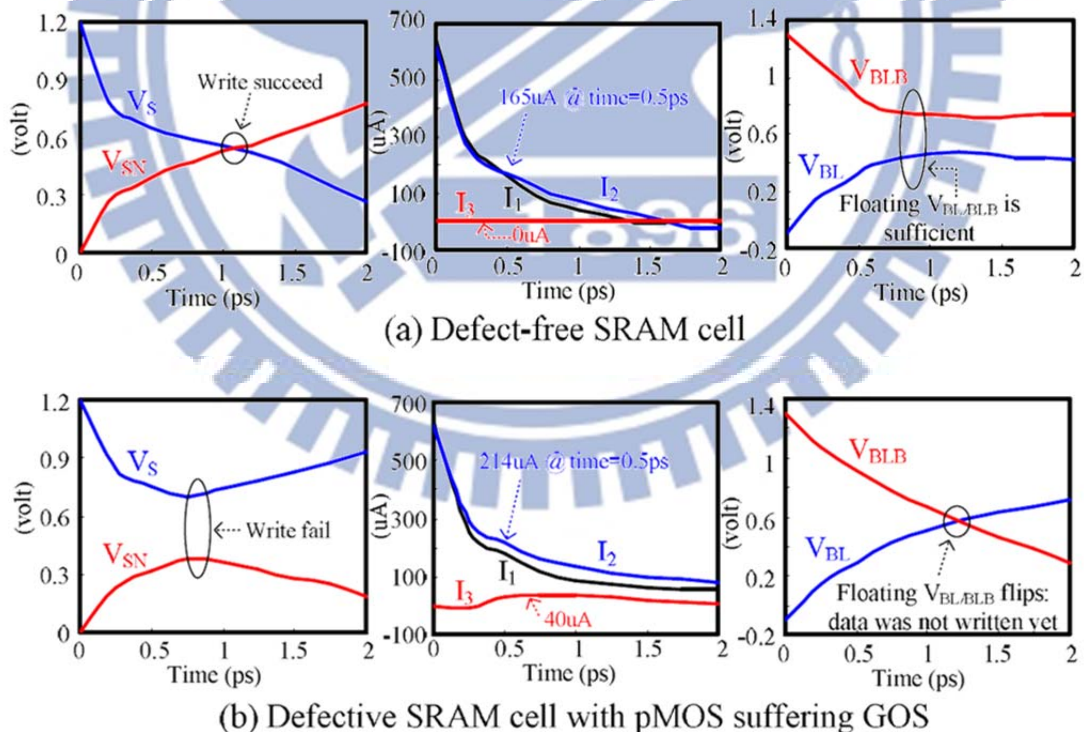


Fig. 4-19. V_S/V_{SN}, $I_1/I_2/I_3$, and V_{BL}/V_{BLB} (a) without and (b) with a GOS at pull-up pMOS when applying the proposed write operation.

As Figure 4-19(a) shows, V_S and V_{SN} flip when V_{BL} (V_{BLB}) is increased (decreased) to a certain level for the defect-free case, where I_3 is 0μA. However, for the defective case shown in Figure 4-19(b), V_{BL} (V_{BLB}) continues to increase (decrease) since V_S and V_{SN} cannot be flipped due to the existence of I_3 (around 40μA), which significantly consumes the charges at C_T that is originally used for pulling up V_{SN} through I_2. Once V_{BL} exceeds V_{BLB}, the write operation has no chance to be successfully performed.

A similar experiment is applied to an SRAM cell with a 2.5nm-radius GOS at the pass-gate nMOS M2. Figure 4-20(a) and Figure 4-20(b) show the corresponding waveforms without and with the GOS at M2, respectively. As Figure 4-20(b) shows, V_{BL} exceeds V_{BLB} faster than that for the case of Figure 4-19(b) while V_S (V_{SN}) decreases (increases) very limited. This is because I_4 induced by the GOS at M2 is much higher than I_3 induced by the GOS at M1 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.

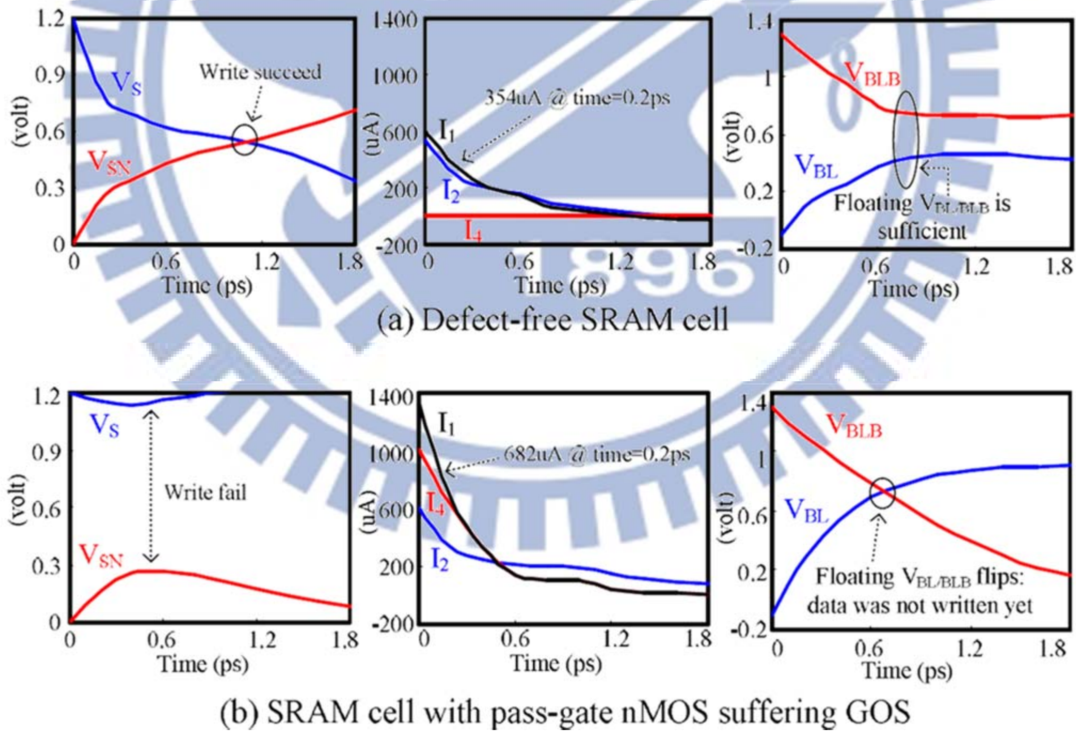


Fig. 4-20. V_S/V_{SN}, I_1/I_2/I_3, and V_{BL}/V_{BLB} (a) without and (b) with a GOS at pass-gate nMOS when applying the proposed write operation.

4.4.4 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 which 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 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.

Figure 4-21 shows simulation result for a 2.5nm-radius GOS at a pull-up pMOS. 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.

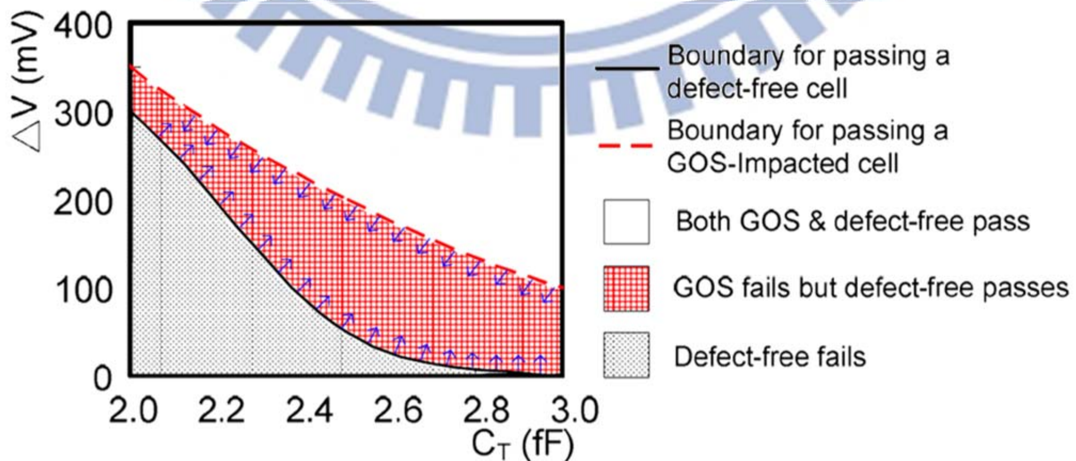


Fig. 4-21. Finding valid (C_T, ΔV) combinations for detecting GOS at pull-up pMOS.

Figure 4-22 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.

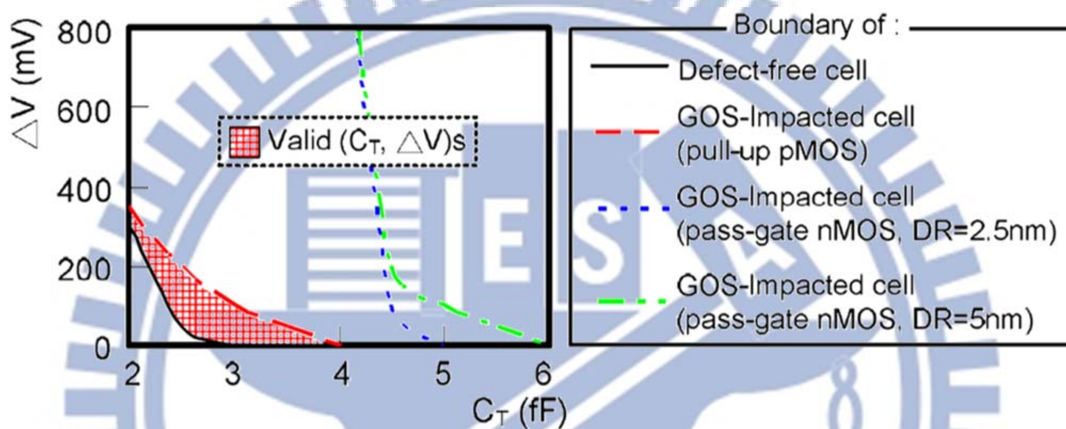


Fig. 4-22. Finding valid (C_T, ΔV) combinations for detecting GOS at pull-up pMOS or pass-gate nMOS.

4.4.5 Comparison with other GOS models in use

In order to obtain an accurate valid (C_T, ΔV) combination through simulation, an accurate circuit-level GOS model is the key. Figure 4-23 further shows the boundaries of the (C_T, ΔV) combinations that can detect the corresponding GOS as shown in Figure 4-21 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 4.3 have 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.

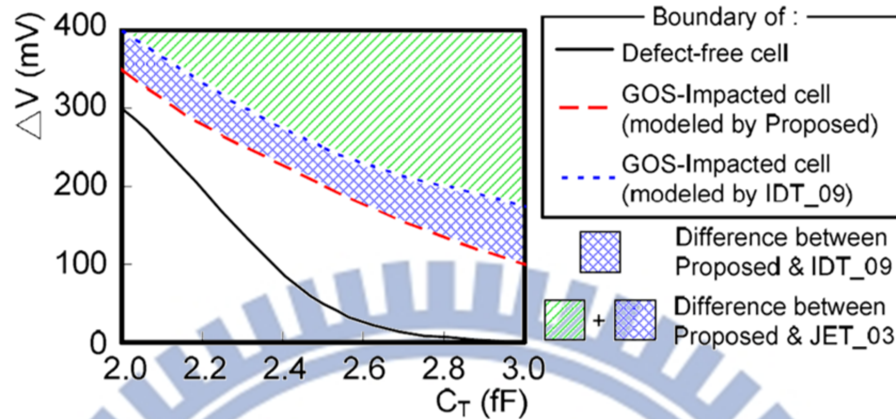


Fig. 4-23. Finding valid (C_T, ΔV) combinations by using different GOS models.

4.5 Implementation of The Proposed Test Method & Its Optimization

To realize the proposed write operation, two Design-for-Test (DfT) hardware components need to be included into the SRAM design. The first one is the C_T adjusting scheme for BL & 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.

4.5.1 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 Figure 4-22 shows, a valid C_T ranges from 2fF to 4fF. A bit-line in our SRAM design connects to 128 cells and its parasitic capacitance is 13fF. For the adopted 65nm technology, the MIMCAP is 5fF/μm^2 and the area of a SRAM

cell is 0.69μm^2, which are the same parameters used by some previous publications [4-42] [4-43]. In this case, the cell area of a column is 128*0.69μm^2, i.e, 88.32μm^2. If C_T ranges from 2fF to 4fF, the resulting capacitor area ranges from 0.8μm^2 to 1.6μm^2 (including BL's & 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 bitline boost circuitry, which has already been 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. [4-31] [4-32] [4-46] [4-47] [4-48] [4-49] [4-50] [4-51] 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 have been validated with silicon result. As reported by [4-31] [4-32] [4-46] [4-47], the area overhead of a boost circuitry can be reduced to as low as 5% (ranging from 5% to 7%). Furthermore, once a bitline 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 [4-31] [4-32] [4-46] [4-47] [4-48] [4-49] [4-50] [4-51] can be easily modified for the use of the proposed test method. For example, Figure 4-24 shows how the bitline booster in [4-32] can be turned into the use of the proposed test method by simply adding an inverter.

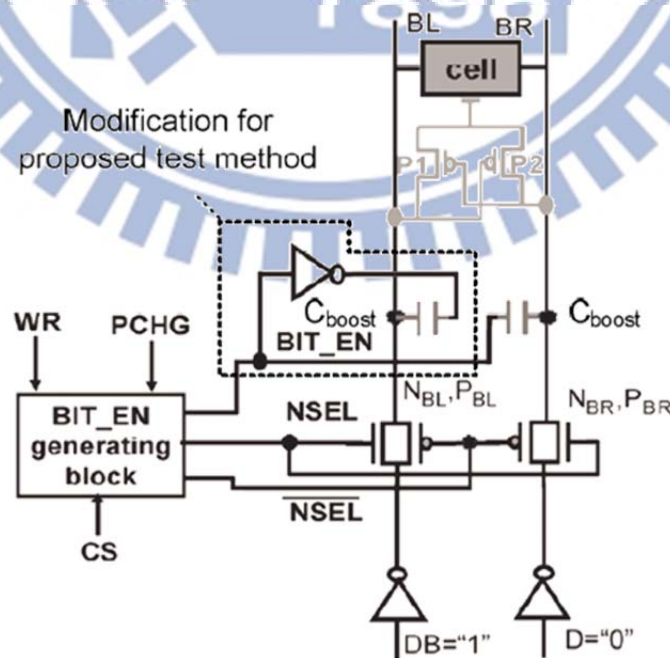


Fig. 4-24. Mutually-inverse boosting circuitry modified from [4-32].

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 turn off 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 & 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.

4.5.2 Optimum (CT, ΔV) for minimizing area overhead

The (C_T, ΔV)s in Figure 4-22 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 Equation 4-6 to 4-9 for calculating the capacitor-occupied area. Equation 4-6 describes the relationship between the ΔV in Figure 4-22 and the C_{boost} in Figure 4-24 when the ΔV is generated by the modified voltage boosting scheme. Equation 4-7 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 ($1\mu m^2$) MIMCAP provided by the process.

$$\Delta V = V_{DD} \cdot \frac{C_{boost}}{C_{boost}+C_T+C_{bitline}} \quad (4\text{-}6)$$

$$Area_{(C_{boost}\&C_T)} = \frac{2\cdot(C_{boost}+C_T)}{\gamma_c} \quad (4\text{-}7)$$

We rearrange Equation 4-6 into Equation 4-8, and substitute the C_{boost} in Equation 4-7 with the one in Equation 4-8. Finally, we acquire the capacitor-occupied area presented as Equation 4-9, which is a function of C_T, $C_{bitline}$, ΔV, VDD, and γ_c.

$$C_{boost} = \frac{\Delta V(C_{bitline}+C_T)}{(VDD-\Delta V)} \tag{4-8}$$

$$Area_{(C_{boost}\&C_T)} = \frac{2\cdot(VDD\cdot C_T+\Delta V\cdot C_{bitline})}{(VDD-\Delta V)\cdot\gamma_c} \tag{4-9}$$

In our experiment, the $C_{bitline}$ is 13fF, VDD is 1.2 volt, and γ_c is 5fF/μm^2. By defining the range of C_T (2~4fF) and ΔV (0~400mV), the capacitor-occupied area of each valid (C_T, ΔV) can be calculated and drawn as Figure 4-25. In Figure 4-25, 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.

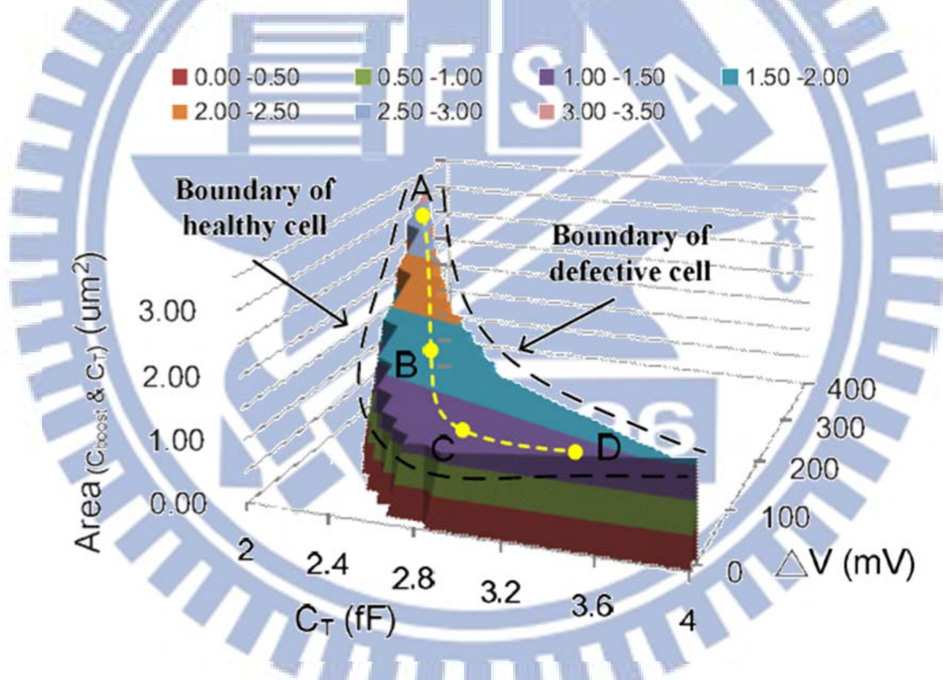


Fig. 4-25. Capacitor-occupied area of each valid (C_T, ΔV).

Figure 4-26 shows the extra capacitor-occupied area resulting from the (C_T, ΔV)s along the targeted middle line. As the result shows, the minimum overhead of the extra capacitor-occupied area is 1.31 μm^2 and occurs when C_T =3fF and ΔV=20mV.

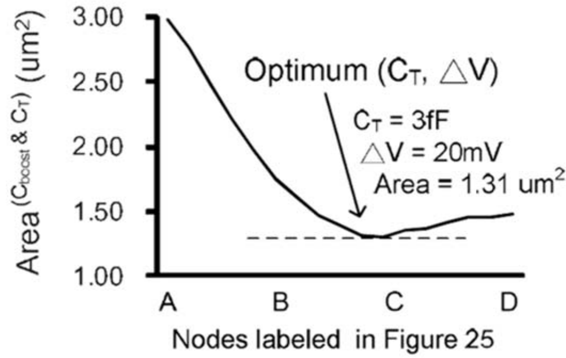


Fig. 4-26. Capacitor-occupied area of the preferred (C_T, ΔV)s in Figure 4-25.

4.5.3 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 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 Figure 4-25. 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 4.5.2, we can also select a proper C_T to maximize its tolerable ΔV range. Figure 4-27 plots the tolerable ΔV range resulting from different selected C_T. As Figure 4-27 shows, the largest tolerable ΔV range is 152mV, which occurs when C_T is 2.6fF and is labeled by P1. If we try to minimize (C_{boost}+C_T) as shown in Section 4.5.2, the resulting tolerable ΔV range will be 91mV, which is labeled by P2. Note that the ΔV fluctuation reported by the previous boost-circuitry works [4-31] [4-32] is around ±25mV, i.e., a range of 50mV. As a result, both P1 and P2 are well above this 50mV ΔV fluctuation, showing that the proposed test method can still be effective under the potential process variation on the boost circuitry.

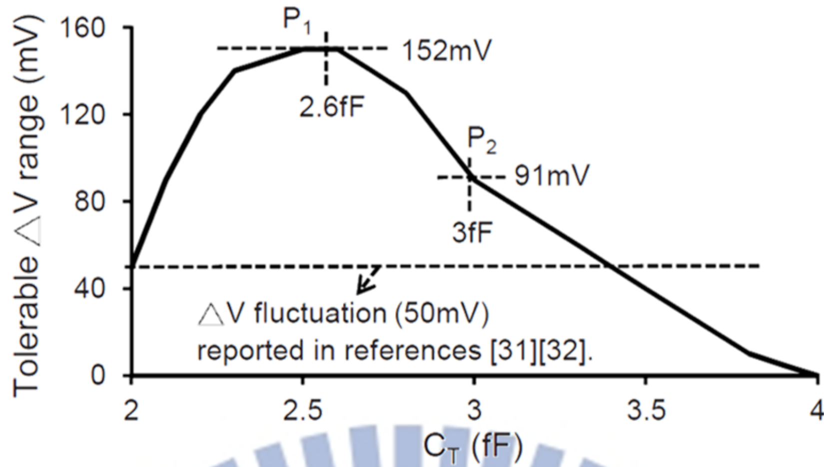


Fig. 4-27. Tolerable ΔV range versus differently selected C_T with two recommended configurations for high-process-variation-immunity and low-area-cost purposes, respectively.

Table 4-VIII further compares the best settings for maximizing the tolerable ΔV range (P1) and minimizing the capacitance overhead (P2). As the result shows, P1 can result in a 61mV higher tolerable ΔV range while P2 can result in a 0.3% lower area overhead. The designers can balance this tradeoff based on the actual need of the product line.

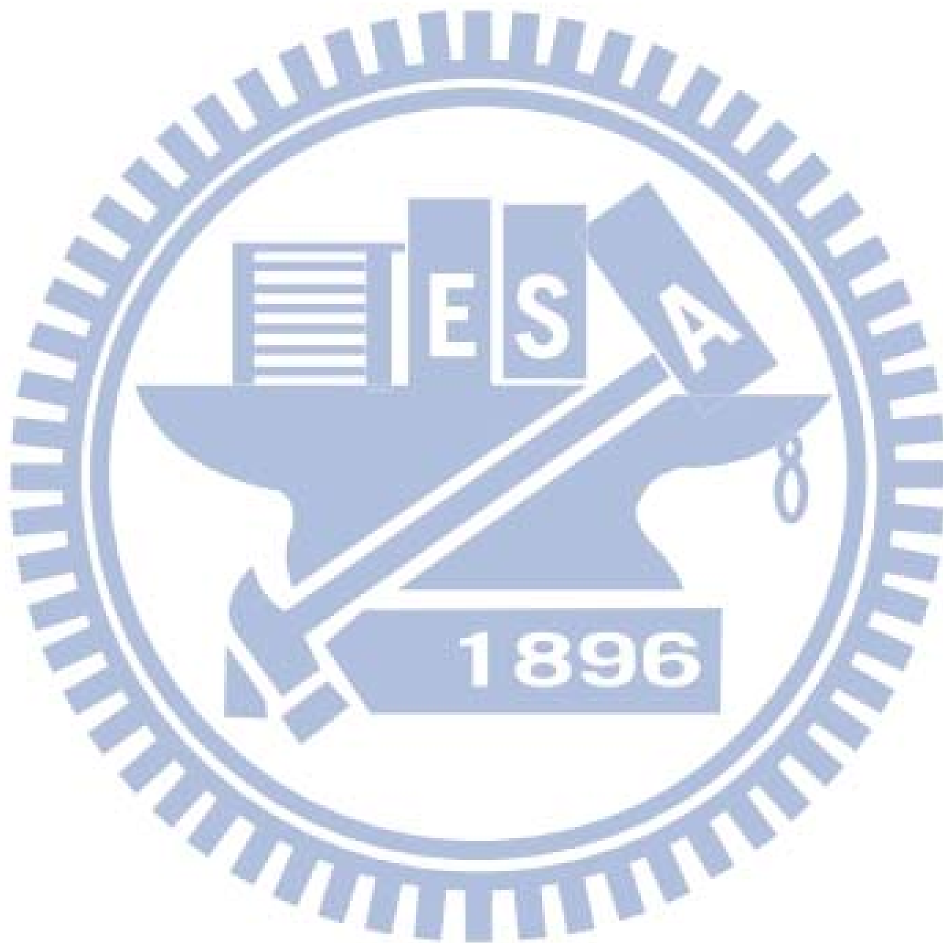
TABLE 4-VIII
Comparison Between The Two Test Configurations P1 & P2 Labeled in Figure 4-27

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

4.6 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 non-split GOS model, which can provide higher accuracy on fitting the DC characteristics of a GOS than previous models while being able to represent a minimum-size GOS-impacted MOSFET. Also, the proposed GOS

model enables accurate transient simulation of a GOS-impacted MOSFET by considering the capacitance change imposed by a GOS, which has never been discussed in previous works. Furthermore, based on this GOS model, we developed a novel DFT write operation, which uses the techniques of floating-bit-line writing and voltage-difference adjustment at bit-lines. 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.



Chapter 5

Investigation of Gate Oxide Short in FinFETs and the Test Methods for FinFET SRAMs

As CMOS technology enters post-22nm era, FinFET becomes the promising device. This is because FinFET possesses the superior electrical characteristics such as reduced short channel effect, good sub-threshold slope, reduced random dopant fluctuation [5-1] [5-2], and high-speed performance [5-3]. The reason FinFET performs differently from the traditional planar bulk MOSFETs is that it utilizes a specific physical structure. As shown in Figure 5-1(a), an upright fin-like silicon with fin width of nanometer-scale forms the channel. While source and drain locate at the fin's alternative edges, the gate with its dielectric covers the fin sidewalls to control the channel's conductivity. In 2012 Intel has demonstrated the 22nm FinFET logics and SRAMs [5-4] [5-5], which almost claims the ready-to-deliver technology.

For FinFETs, one of the major-concerned issues is the fin line-edge roughness (Fin-LER) [5-6] [5-7] [5-8]. This phenomenon depicts the fin sidewalls of the device would be rough after process due to the limitation of lithography and etching. Figure 5-1(b) shows the SEM image of the silicon fin profile with Fin-LER. As a result, device's characteristics and performance would vary from each other. To mitigate the Fin-LER caused device variation, several solutions have been proposed [5-7] [5-9]. The techniques include adjusting devices' number of fins, controlling gate bias voltage, and considering sidewall surface (conducting channel) orientation. However, when the dielectric thickness of FinFETs is scaled to 1~5nm [5-10] [5-11], and the non-conformal deposition of the dielectric layer is reported [5-12], the FinFET with Fin-LER suffers the poor insulator coverage at the sidewalls as shown in Figure 5-1(c). The gate oxide short defects may consequently occur [5-13] [5-14].

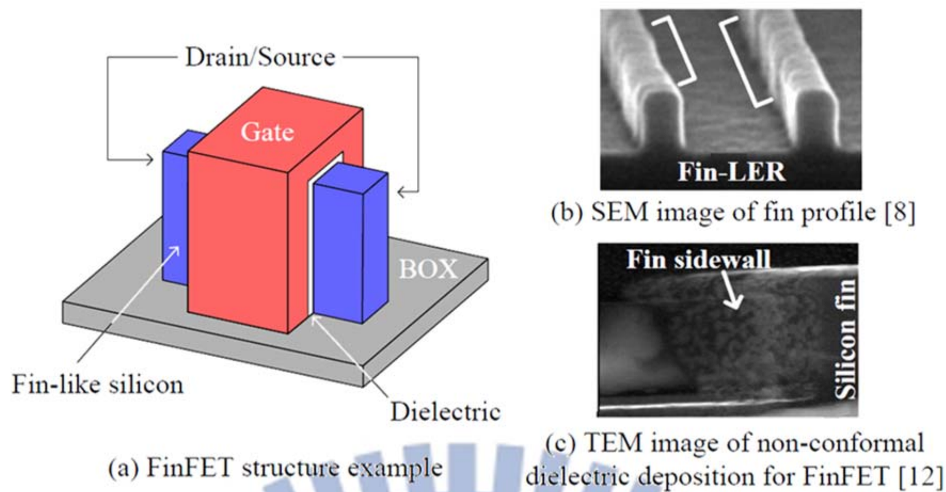


Fig. 5-1. FinFET structure and images of the manufacture defects.

Gate oxide short (GOS) is a defect that has been widely discussed in planar bulk MOSFETs. In addition to the defect mechanism mentioned in the previous paragraph, the GOS defect can also be induced by voltage stress, lithographic particles, or unexpectedly large gate tunneling leakage [5-13] [5-15]. The GOS used to be classified into two types: gate-to-source/drain short and gate-to-channel short. The first type depicts the shorting between gate and source/drain and was usually modeled by a resistive short between them. We do not include this type into our discussion for two reasons. Firstly, for FinFETs, only the devices with overlapped gate-source/drain have chance to suffer the defect. In the devices with underlapped gate-source/ drain, the defect rarely appears. Secondly, even for the overlaptype FinFETs, the gate-to-source/drain GOS modeled by a resistance short actually behaves the same as a simple node-to-node short. Since previous works [5-16] [5-17] have already studied the node-to-node shorts for FinFET logics, the corresponding fault behaviors can be quickly referred and hence need no more discussion. We focus on the second type of GOS: gate-to-channel short. The gate-to-channel GOS causes a low impedance from gate to the silicon channel. The corresponding fault behavior is more complex. In planar bulk MOSFETs, many works of defect modeling [5-18] [5-19] and testing [5-14] [5-20] [5-21] have been published. But for FinFETs, the related research paper has not been seen yet to the best of our knowledge.

In this paper, we investigate the gate-to-channel GOS (will be abbreviated to only GOS in the following contents) in FinFETs by using the TCAD mixed-mode simulations. Firstly, we build defect-free FinFETs in the TCAD 3D environment. Based on the built devices, we inject the GOS defect and extract the corresponding faulty characteristics. By comparing the electrical I-V curves

of the defective FinFETs and the defect-free ones, we show how the GOS affects the FinFETs. In addition, we also illustrate and explain the fault behavior difference between FinFETs and planar bulk MOSFETs. Finally, following the above studies, we discuss the corresponding SRAM testing. For detecting GOS in FinFET SRAMs, we introduce two new test methods. One is for tied-gate FinFET based SRAM, and the other is for independent-gate FinFET based one. The test efficacy of both the methods is proved by applying TCAD transient simulations.

5.1 GOS Fault Behaviors in Planar Bulk MOSFETs

In this section, we briefly review the fault behaviors when GOS occurs in the planar bulk MOSFETs. As reported [5-13] [5-22], GOS affects the devices in three aspects. The first one is decreasing the saturation drain current. Due to the defect, the channel inversion becomes weak and the drive ability of the device is accordingly reduced. Secondly, when the drain voltage is low, there will be a negative drain current. The current comes from the gate leakage current flowing through the GOS to the drain. As to the last, the total gate leakage current increases exponentially with the increasing of gate voltage. Below are their summarizations.

- Decreased saturation drain current ($I_{D(SAT)}$)
- Negative drain current at low drain voltage
- Gate voltage dependent gate leakage current

The three fault behaviors are usually demonstrated by the device's IDVD figure. Figure 5-2 shows the example. Figure 5-2(a) is the IDVD of a defect-free planar bulk nMOS with VDD=1.2 volt. Figure 5-2(b) belongs to the same device but with GOS on the contrary. When comparing the right hand sides of the two figures, firstly the "Decreased $I_{D(SAT)}$" occurs obviously. As to the "Negative $I_{D(lowVD)}$", it appears at the left hand side of Figure 5-2(b) when being compared to Figure 5-2(a). The third fault behavior "Leakage I_G" is not directly shown in the figures. However, it can be implied by the negative $I_{D(VD=0)}$ which increases exponentially with V_G.

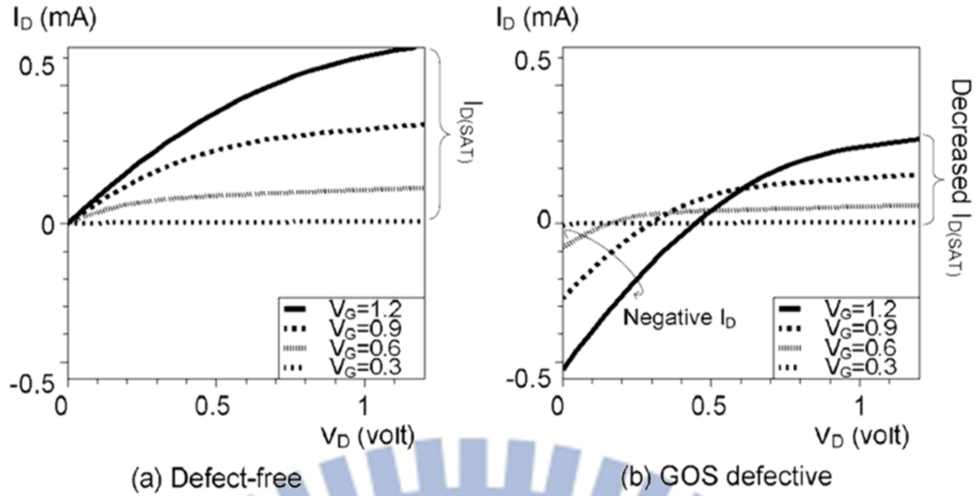


Fig. 5-2. IDVD curves of a planar bulk n-type MOSFET (a) without and (b) with a gate-to-channel GOS.

5.2 Experiment Setup

For experiments, we build four 25nm FinFET devices in the Synopsys Sentaurus TCAD environment. The devices are n/p-type FinFETs with tied-gate (TG) and independent-gate (IG) structures as shown in Figure 5-3. The TG FinFET in Figure 5-3(a) uses only one gate to control the channel. On the other hand, the IG FinFET in Figure 5-3(b) has two gates: front gate and back gate. The front gate usually works to turn-on/off the device while the back gate adjusts the V_{th} of the device. The parameters of all the FinFETs follow the previous work [5-23]: N_a=1x10^{17}cm^{-3}, L_{eff}=25nm, W_{fin}=7nm, H_{fin}=20nm, effective oxide thickness (EOT)=0.65nm, and work function of gate metal=4.55eV.

Based on the above setup, we further inject the GOS into the built devices to simulate defective FinFETs. The GOS injection in this paper utilizes the representation of pinhole [5-24] [5-25]. The following steps are consequently applied. Firstly, choose one sidewall of the FinFETs as the defect location. For the chosen sidewall, remove a tiny cubic of the dielectric layer and leave a pinhole in the center. Then, fill the pinhole with gate material. As a result, the gate contacts the channel via a small pinhole, and the GOS injection thereby completes. In our experiments, we apply the pinhole with diameter 5nm as example.

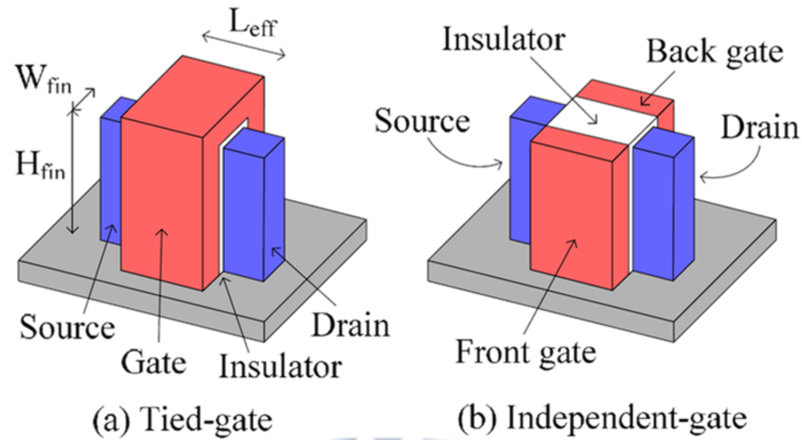


Fig. 5-3. Perspective view of the built FinFETs.

5.3 GOS Fault Behaviors in FINFETs

5.3.1 Tied-Gate FinFET

For TG FinFET shown as Figure 5-3(a), GOS occurring at either sidewall of the device causes equivalent fault behavior. It's because the unique gate terminal controls both the sidewall channels in the same manner. Accordingly, we only need to inject the GOS once for the TG-FinFET unlike the twice injections for IG-FinFET, which will be discussed in the next sub-section. Figure 5-4 shows the experiment results of using an n-type TG-FinFET as example. The supply VDD is 1 volt. In the figure, two sets of IDVD curves are shown. The black-square IDVD curves represent the defect-free FinFET, and red ones represent the defective one. As shown in the figure, GOS causes $I_{D(SAT)}$ decreased as in the planar bulk MOSFETs but just less obviously. As to "Negative $I_{D(lowVD)}$" and "Leakage I_G", the defective FinFET also suffers the two fault behaviors similarly.

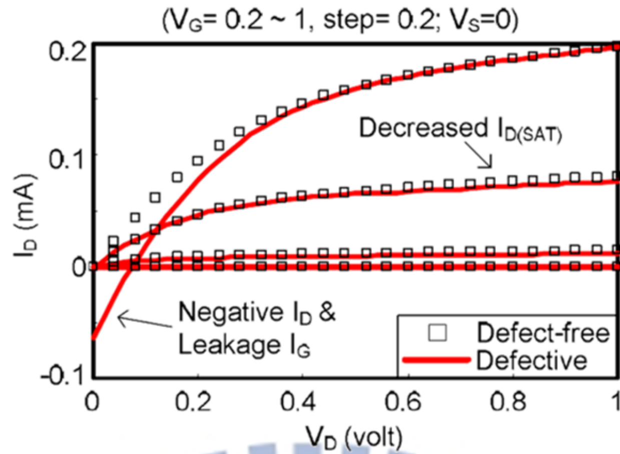


Fig. 5-4. IDVD behaviors of TG FinFETs without/with GOS.

5.3.2 Independent-Gate FinFET

For IG FinFET shown in Figure 5-3(b), GOS occurring at the dielectric layers of front gate and back gate causes different I-V characteristics. It's because the control voltage on the front/back gates was usually set separately for the sake of either performance or low power. For performance, [5-26] [5-27] set the voltage of FinFETs' back gates varied depending on the circuit's status to improve the operating speed. For low power, [5-2] [5-16] on the contrary fix the back gate voltage to GND/VDD to minimize the leakage current. In our discussion, we apply the low-power mode IG-FinFET design since the device status can be defined more clearly. Accordingly, the back gates of n-type IG FinFETs will be connected to GND. For p-type IG FinFETs, the back gates will be connected to VDD.

Figure 5-5 shows the simulation results of an n-type IG FinFET without/with GOS at front gate's dielectric. In the figure, the black-square curves belong to the defect-free FinFET. Red curves are of the defective one. As shown in the figure, when V_D is at low voltage, the defective IG FinFET suffers the "Negative $I_{D(lowVD)}$" and "Leakage I_G" fault behaviors as well as the previous TG FinFET does. However, when V_D is at high voltage, GOS causes the saturation drain current increased, which is opposite to the decreased $I_{D(SAT)}$ of planar bulk MOSFETs and TG FinFETs.

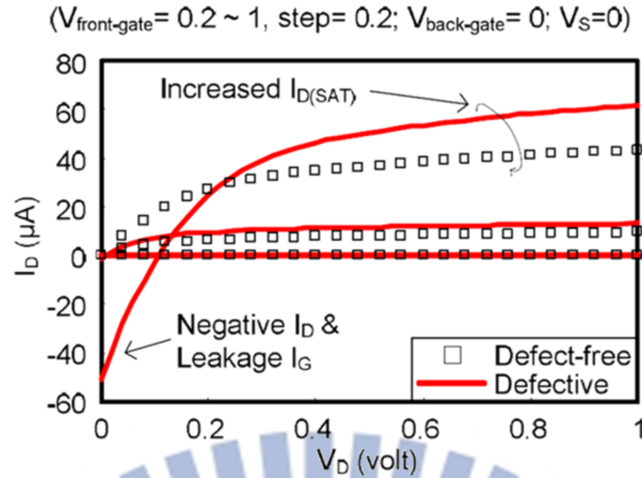


Fig. 5-5. IDVD behaviors of IG FinFETs without/with GOS at the front-gate dielectric.

For analyzing the specific fault behavior, we extract the carrier density of the device under saturation region from the TCAD simulations. We discover the GOS causes much higher carrier density in the channel. The channel thereby becomes more conductive. Figure 5-6 shows the part of electron density distribution of the IG FinFET. For the defect-free FinFET, the electron density is shown as above. For the defective one, the electron density is shown as below. Comparing the two density distributions, GOS causes thicker channel and higher electron density (from 8.4E+20 cm^{-3} to 1.9E+21 cm^{-3}). As to hole, the carrier density is even more tremendously increased at the back-gate side from 1.1E+8 cm^{-3} to 1.4E+21 cm^{-3}. The reasons for the high carrier density in the silicon fin are two: 1) Holes are injected into the silicon fin through GOS from the front gate with positive bias and continuously accumulate in the back-gate side. 2) The holes change the electrons density in the channel and enhance the electrons injection from source, which leads to the V_{th} shifting. The two phenomenons increase the drain current and are similar to the floating body effect (kink effect) for SOI devices [5-28] [5-29].

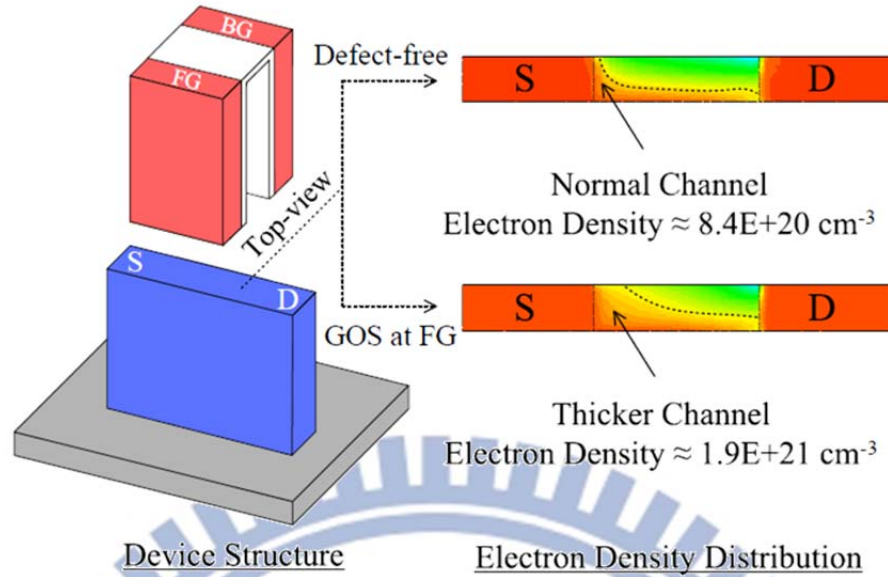


Fig. 5-6. Electron density distribution of an n-type IG FinFET with/without GOS at front-gate dielectric.

As to GOS occurring at the back-gate dielectric, Figure 5-7 shows the simulation results. The results indicate that the defective IG-FinFET only suffers the "Decreased $I_{D(SAT)}$" but neither "Negative $I_{D(lowVD)}$" nor "Leakage I_G". The saturation drain current decreases because the GOS has the back gate turn off more region of the silicon fin, which consequently increases the device's V_{th}. As to the "Negative $I_{D(lowVD)}$" and "Leakage I_G", there is no gate leakage occurring because the back gate connected to GND always turns off the backchannel. Since the back channel cannot conduct the carriers, the back-gate leakage current remains very low as of defect-free FinFETs.

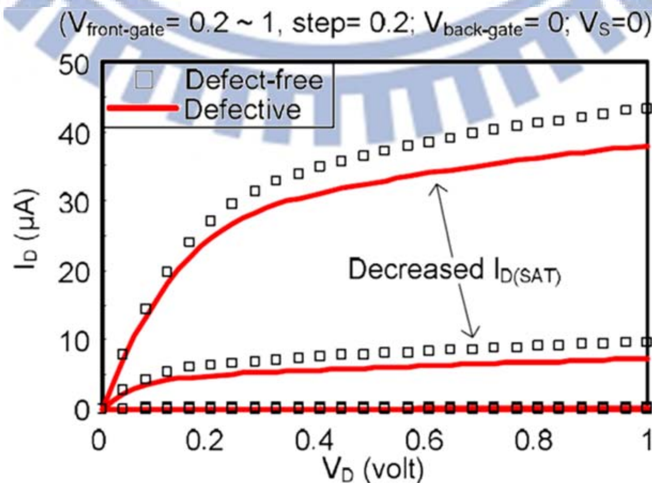


Fig. 5-7. IDVD behaviors of IG FinFETs without/with GOS at the back-gate dielectric.

5.3.3 Short Summary

Table 5-I summarizes the GOS fault behaviors of different MOSFETs. For the planar bulk MOSFET, the fault behaviors include the “Decreased $I_{D(SAT)}$”, ”Negative $I_{D(lowVD)}$”, and ”Leakage I_G” as reviewed in section 5.1. For TG FinFET, the fault behaviors are almost the same as the planar bulk MOSFET, but the fault behavior of “Decreased $I_{D(SAT)}$” is not obvious. As a result, to detect the GOS in TG FinFET, only the test method detecting leakage current can be applied. The test method which intends to detect the reduction of drive ability may be unable to trigger and capture the fault. For IG FinFET, when the GOS occurs at the front-gate dielectric, one of the fault behaviors is different from the two previous cases. The $I_{D(SAT)}$ no longer decreases but increases on the contrary. As to GOS occurring at the back gate, the defective IG FinFET only suffer ”Decreased $I_{D(SAT)}$” but no other fault behaviors related to the gate leakage current. To detect the GOS in IG-FinFET is much more complex. For the GOS at front-gate, only the test method detecting leakage current can be applied. For the GOS at back-gate, however, only the test method detecting the reducing of drive ability is applicable.

TABLE 5-I
Comparison of GOS Fault Behaviors of Different Transistors

Transistor type & defect location (for IG FinFET)		Fault behaviors		
		$I_{D(SAT)}$ variation	Negative I_D (at low V_D)	Leakage I_G
Planar bulk MOSFET		Decreased	Yes	Yes
TG FinFET		Decreased (not obvious)	Yes	Yes
IG FinFET	Front-gate	Increased	Yes	Yes
	Back-gate	Decreased	No	No

5.4 Testing of GOS in FinFET SRAMs

Based on the above investigations of GOS in single FinFETs, in this section, we further discuss the corresponding SRAM testing. Figure 5-8 shows the

SRAMs applied in experiments. Figure 5-8(a) is the typical TG-FinFET based SRAM [5-30] [5-31]. Figure 5-8(b) is the low-power mode IG-FinFET based SRAM [5-2] [5-32]. For detecting GOS in the SRAMs, we firstly apply traditional methods including March test and IDDQ test. The experiment results show that the traditional methods are limited when testing the defect. Hence, we introduce two new test methods. The two proposed methods are for the TG/IG-FinFET based SRAMs respectively and can detect the undetectable GOSs of previous methods.

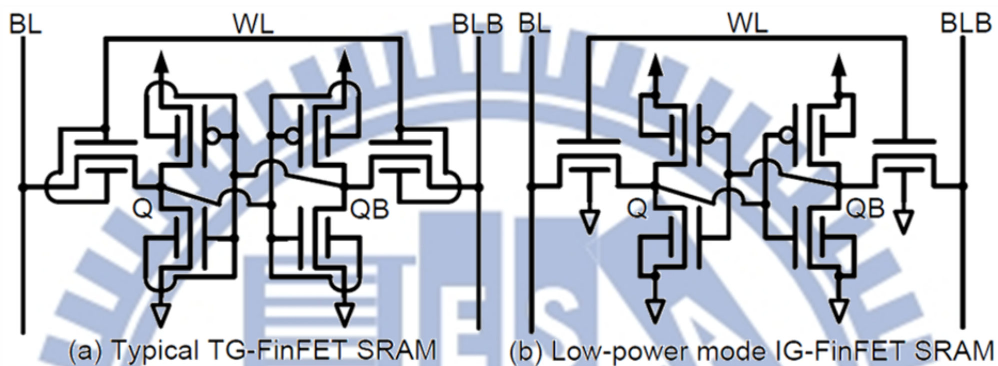


Fig. 5-8. FinFET SRAM designs applied for experiments.

All the related experiments in this section are run under the TCAD transient simulation. To the SRAMs, the setup is as following. The FinFETs are with minimum size (L_{eff}=25nm, W_{fin}=7nm, H_{fin}=20nm) for simplicity [5-26] [5-30]. The operation frequency is determined by using the period length which is 20% more than the minimum operatable one.

5.4.1 Traditional Tests: March Test and IDDQ Test

To test SRAM, March algorithm is the most commonly used method. By applying organized normal operations to the SRAM, the fault is detected if the read output is different from the expected one. In our experiments, we have difficulty to run a complete March for the SRAM. It's because the TCAD transient simulation requires extremely-long computation period. Thus, we apply parallel operation pairs instead to test the SRAM. The operation pairs include Write-Hold, Write-Read, Hold-Read, Hold-Write, Read-Hold, and Read-Write. And the data covers both 0 and 1. If any one of the operations fails due to the GOS, the defect will be deemed as detected. To distinguish the pass/fail of hold and write, we inspect if Q and QB flip. For read, we examine

the BL-BLB voltage difference at the end of the operation. We do not use sense amplifier to judge the read pass/fail because the TCAD limits the number of included FinFETs for each simulation. In our case, only 7~8 FinFETs at most can be applied each time. Referring to [5-33], the sense amplifier with VDD = 1 volt can achieve 99.8% yield if the input voltage difference reaches 55mV. Therefore, we judge the read success if BL-BLB's voltage difference is larger than 55mV. For BL-BLB voltage difference less than 55mV, the read is considered failed.

Table 5-II shows the results of applying operation pairs to detect the GOS. The first two columns are the SRAM type and the possible GOS location. The third column shows if any operation fails due to the GOS at the designated location. If yes, the fourth column shows the detecting operation. As shown in the table, for TG-FinFET based SRAM, only the GOS at the pass-gate nFinFET is detected. The detecting operation is read. For GOS locating at other FinFETs, the TG-FinFET SRAM functions correctly. For IG-FinFET based SRAM, three GOSs are detected. The detected GOSs locate at the front-gate/back-gate of the pass-gate nFinFET and the front-gate of the pull-down nFinFET. The detecting operations are also read.

TABLE 5-II
Experiment Results of Applying Operation Pairs to Detect GOS in FinFET SRAMs

SRAM type	GOS location		Failure occurs	Failed operation
TG-FinFET based	pass-gate nFinFET		yes	read
	pull-down nFinFET		no	-
	pull-up pFinFET		no	-
IG-FinFET based	pass-gate nFinFET	front-gate	yes	read
		back-gate	yes	read
	pull-down nFinFET	front-gate	yes	read
		back-gate	no	-
	pull-up pFinFET	front-gate	no	-
		back-gate	no	-

Apparently, the GOS only affects the FinFET SRAMs' read operations. For hold/write operations, no fault is found. However, according to our experiments,

the undetected GOS at the cross-couple inverters actually affects the storing nodes Q/QB of the SRAM. As a result, the cell's reliability is reduced. For those undetected GOS, we further apply the IDDQ test secondly since it has been commonly recommended for detecting GOS in planar bulk MOSFETs [5-20] [5-21]. Table 5-III shows the IDDQ sensitivity of each defective case. For calculating the sensitivities, we apply the SRAM with array size 163Mb [5-5]. Besides, we also apply write operations for the IDDQ test [5-21] in addition to the hold operation. According to the results, the largest IDDQ sensitivity for the two SRAMs both occur when GOS locates at pull-up pFinFETs. However, the largest IDDQ sensitivity is only $8.6x10^{-4}$% and $6.2x10^{-4}$% respectively which still limits the test efficacy of IDDQ.

TABLE 5-III
IDDQ Sensitivity of FinFET SRAMs with GOS (Array Size is 162Mb [5-5])

SRAM type	GOS location		IDDQ sensitivity (%)	
			Hold	Write
TG-FinFET based	pull-down nFinFET		$7.8x10^{-4}$	$6.5x10^{-4}$
	pull-up pFinFET		$8.6x10^{-4}$	$4.6x10^{-4}$
IG-FinFET based	pull-down nFinFET	BG	$9.7x10^{-7}$	$1.5x10^{-4}$
	pull-up pFinFET	FG	$6.2x10^{-4}$	$6.8x10^{-5}$
		BG	$3.9x10^{-8}$	$9.7x10^{-6}$

FG: front gate BG: back gate

5.4.2 Proposed Test method for TG-FinFET based SRAM

In previous sub-section, the traditional methods are shown limited in testing the GOS at the cross-couple inverters. To detect the GOS in TG-FinFET based SRAM, we propose a new test method notated Proposed_TG. The Proposed_TG applies a write operation to the targeted cell with both BL and BLB floating during the period. The voltage on BL/BLB is set specifically to adjust the test efficacy. Figure 5-9 illustrates the configuration of the Proposed_TG. In the figure, we assume M_a and M_b are the two FinFETs suffering GOS. To detect the GOSs, the floating BL and BLB are with capacitance C_T. And the voltage on them is set to (GND-ΔV) and (VDD+ΔV) respectively to execute a write-0 operation. The original data inside the cell is assumed Q/QB=1/0.

When the SRAM is defect-free, the BLB with voltage (VDD+ΔV) pulls up QB by I_2. BL with voltage (GND-ΔV) pulls down Q by I_1. In this situation, only

I_1 and I_2 exist in the figure. However, when the GOS in M_a occurs, I_3 will appear to share I_2 and consume the stored charge on BLB more rapidly. On the other hand, if the GOS in M_b occurs, I_4 will appear to share the I_1 which intends to pull down Q. As a result, the voltage difference of BL-BLB will decrease much more quickly due to the GOS induced I_3 or I_4. The write-0 operation will consequently fail. While the write-0 fails, the following read will output the original 1/0 instead of the expected 0/1. The GOS is then detected. If M_c or M_d is the FinFET suffering GOS, the Proposed_TG can also detect the defect as long as the voltage on BL/BLB and Q/QB exchanges respectively to execute a write-1 instead.

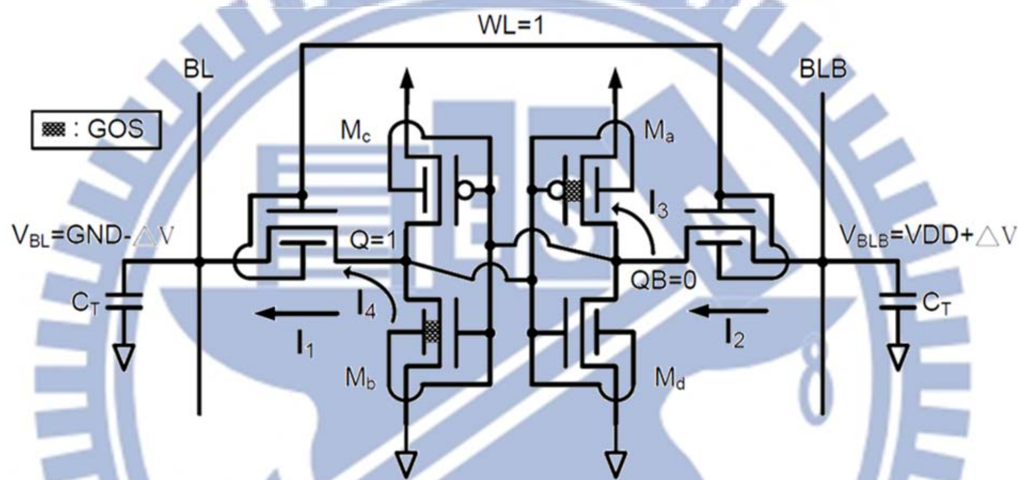


Fig. 5-9. Configuration of the Proposed_TG test method for detecting GOS in TG-FinFET SRAM.

The test efficacy of the Proposed_TG depends on the setup of ΔV and C_T. In Figure 5-10, we show the corresponding experimental results to help find the valid ΔV-C_T for test. In the figure, the three curves represent the minimum operatable ΔV-C_T for each SRAM. The black-square curve is of the defect-free SRAM, and the other two curves belong to the defective SRAMs with different GOS location. For each curve in the figure, the ΔV-C_T below will fail the corresponding SRAM under the Proposed_TG. On the contrary, the ΔV-C_T above allows the corresponding SRAM pass the test write operation. According to the results, there are three regions in the figure. The lowest one causes not only the defective SRAMs but also the defect-free SRAM failed. The highest region on the contrary has all the SRAMs pass the Proposed_TG. Finally, the ΔV-C_T in the middle region fails the defective SRAMs but has the defect-free SRAM pass the Proposed_TG. The ΔV-C_T in the region is thus the valid one for

test.

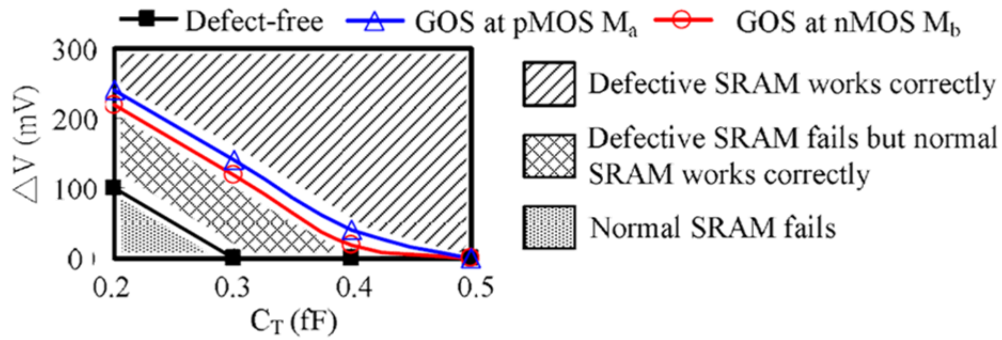


Fig. 5-10. Minimum operatable ΔV-C_T for SRAMs passing the Proposed_TG.

5.4.3 Proposed Test method for IG-FinFET based SRAM

To detect the GOS in the IG-FinFET based SRAM, we propose a test method notated Proposed_IG. The Proposed_IG is actually modified from the Proposed_TG in the previous sub-section. Before introducing the Proposed_IG, we firstly illustrate how and why the Proposed_TG is insufficient for the test here. We repeat the experiment in Figure 5-10, but the tested SRAM is changed to the IG-FinFET based one. Figure 5-11 shows the results. Figure 5-11(a) is with GOS at the front gate of the IG FinFETs. Figure 5-11(b) is with GOS at the back gate on the other hand. According to the results, the Proposed_TG has valid ΔV-C_T for detecting the front-gate GOS as shown in Figure 5-11(a). But for back-gate GOS, the results in Figure 5-11(b) show no valid ΔV-C_T exists. The Proposed_TG detects the front-gate GOS because it triggers the gate leakage current in IG-FinFETs as in the TG-FinFETs. However, while the back-gate GOS causes no fault behavior related to gate leakage current as mentioned in section 5.3.2, the test method has no way to detect the defect.

To detect the back-gate GOS, we include the testing of decreased $I_{D(SAT)}$ into the Proposed_IG. Besides, by being modified from Proposed_TG, the Proposed_IG also inherits the advantage of detecting gate leakage current. The Proposed_IG can detect both the front-gate/back-gate GOS at the same time. Figure 5-12 shows the configuration of the Proposed_IG. In the figure, M_a is assumed the IG FinFET of which the front gate and the back gate may suffer GOS. To test the GOSs, Proposed_IG sets the BL and BLB floating similar to Proposed_TG. The voltage on BL is also (GND-ΔV). But the voltage on BLB is modified from the original (VDD+ΔV) to 0. Note that the test write operation here is still a write-0. For the front-gate GOS, since the BL/BLB remain floating,

the GOS induced current I_3 will affect the SRAM in the same manner as for the TG-FinFET SRAM. The Proposed_IG can detect the front-gate defect. As to back-gate GOS, the modification of BLB voltage makes the pulling up of QB rely on I_2 only without the driving current from BLB. If the back-gate GOS in Ma decreases the drain current I_2 too much to succeed the pulling up QB, the test write operation would fail. And the back-gate GOS is hence detected.

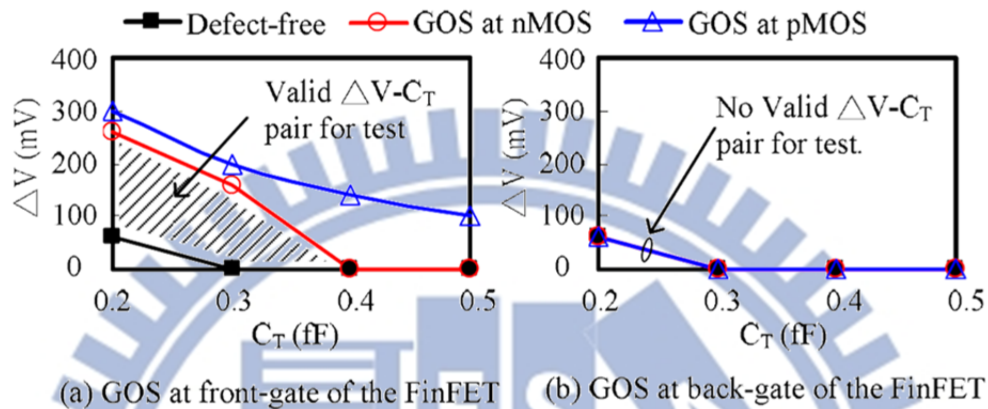


Fig. 5-11. Test results of applying Proposed_TG for detecting GOS in the IG-FinFET SRAM.

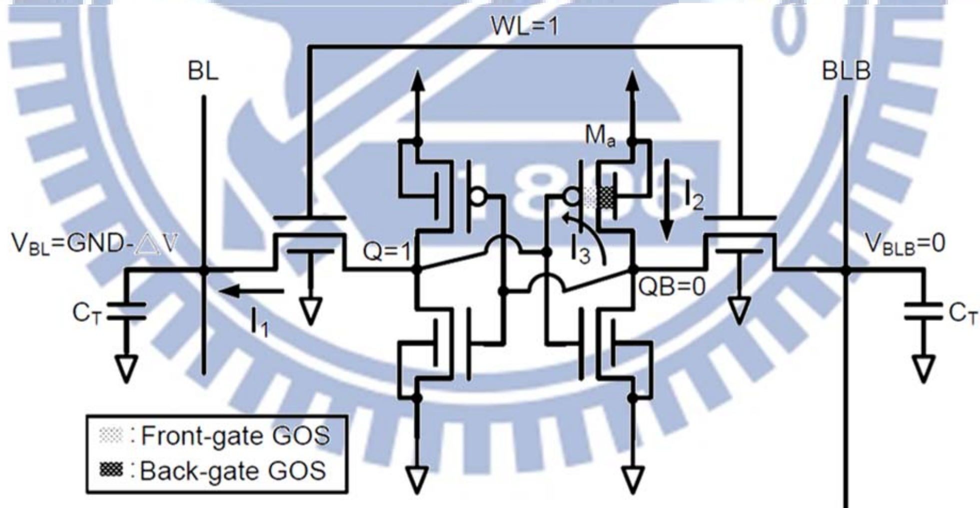


Fig. 5-12. Configuration of the Proposed_IG test method for detecting GOS in IG-FinFET SRAM.

Figure 5-13 shows the simulation details of the Proposed_IG detecting both the front-gate and back-gate GOS at the same time. For the simulations, we apply C_T =0.2fF and ΔV=0.62V as example. In the Figure 5-13, the IG-FinFET SRAM is under three conditions: (a) defect-free, (b) GOS at front gate, and (c) GOS at back gate. For each case, we show the voltage of Q/QB and VBL/VBLB

during the test write-0 period. Firstly, for the defect-free SRAM, Figure 5-13(a) shows the Q and QB successfully flip, and V_{BLB} is always higher than V_{BL}. When GOS occurs at front gate, the Figure 5-13(b) shows the Q is pulled down and reaches QB. But the Q remains higher than QB at the end. The write-0 thus fails. For GOS occurring at back gate, Figure 5-13(c) shows the Q/QB flip at time=2ps. But while V_{BL}/V_{BLB} also flip later on, the driving force of the write-0 disappears. The Q/QB then flip again at time=4ps. The test write operation fails as well, and the back-gate GOS is detected.

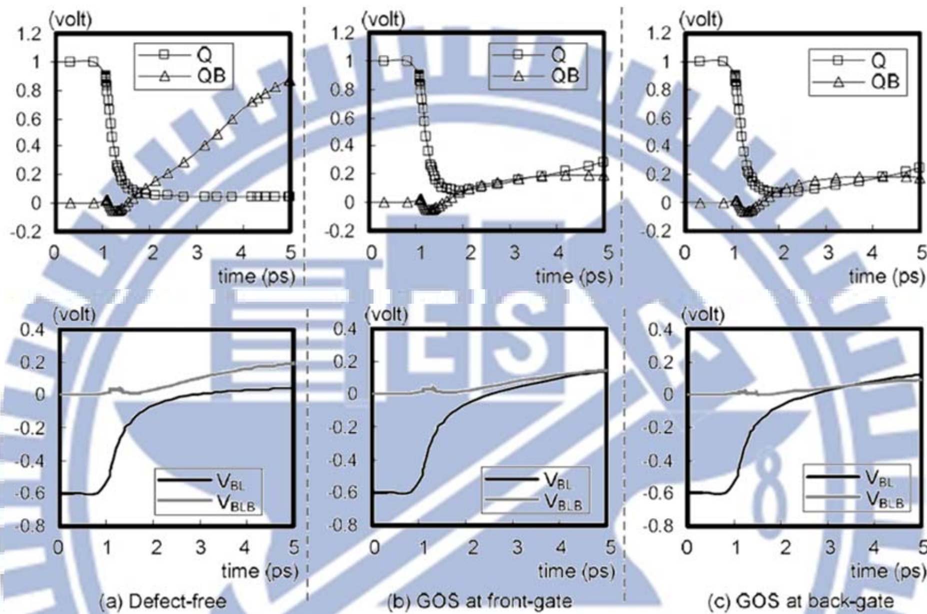


Fig. 5-13. Simulation details of the Proposed_IG detecting front-gate/back-gate GOS at the same time (C_T=0.2fF, ΔV=0.62V).

Table 5-IV summarizes the test efficacy of the methods for detecting GOS in the IG-FinFET SRAM. Firstly, Proposed_TG detects the front-gate GOSs for pull-down and pull-up IG-FinFETs. But for back-gate GOSs, the test method cannot detect the defects. Proposed_IG on the other hand detects all the GOSs including front-gate/back-gate GOSs in both pull-down and the pull-up IG-FinFETs. For comparison, we also include the severe write test method [5-34] in experiments. The method is included because the decreased $I_{D(SAT)}$ caused by back- gate GOSs is similar to the fault behavior of open defects. And the severe write method has been shown useful for detecting the open defects in SRAMs. According to the results, the severe write method does not detect any back-gate GOSs. After improvement, the modified severe write can detect the back-gate GOS, but only the one in the pull-down IG FinFET.

TABLE 5-IV
Test Efficacy Comparison of The Methods To Detect The GOSs in IG-FinFET SRAM

Test method	Detecting the GOS in IG-FinFET SRAM			
	pull-down n-FinFET		pull-up p-FinFET	
	FG	BG	FG	BG
Proposed-TG	yes	no	yes	no
Proposed-IG	yes	yes	yes	yes
Severe write [34]	no	no	no	no
Modified severe write	no	yes	no	no

FG: front gate BG: back gate

5.5 Conclusion

In this paper, we investigate the gate oxide short (GOS) in FinFETs by using the TCAD mixed-mode simulation. According to the results, we discover the GOS fault behaviors in FinFETs are more complex than those in planar bulk MOSFETs. For tied-gate FinFETs, the fault behavior of saturation drain current decreasing becomes much less obvious. This leads the test method, which intends to detect the reducing of device's drive ability, to be unable to detect the defect. For independent-gate FinFETs, the fault behaviors of the GOS at front gate and back gate completely differ from each other. When locating at the front gate, the GOS increases the saturation drain current and induces the gate leakage current. But when locating at the back gate, the GOS decreases the saturation drain current on the contrary and causes no fault behavior related to gate leakage current. Based on the investigation for single FinFETs, we also discuss the corresponding SRAM testing. For detecting GOS in the SRAMs, we propose two new test methods. One is for tied-gate FinFET based SRAM, and the other is for independent-gate FinFET based one. Both the methods can detect the undetectable GOSs of traditional methods and are verified by using the TCAD transient simulations.

Chapter 6
Summary

In this thesis, testing of subthreshold SRAMs is discussed. For the Sub-V_{th} SRAMs with extra-read paths, severe write and LVW-HVR test methods are verified to cover the stability faults at the internal pull-down nMOSs which cannot be detected by traditional tests. As to the single-ended read/write type of Sub-V_{th} SRAM cell, Read-equivalent stress would be the appropriated test method for the corresponding stability faults. For the data-aware write-assist SRAM, a new self-loop attacking method is proposed to detect all the open defects. In addition, detectable open resistances can be further lowered with test time reduced by more than 55% when being compared to the conventional March C- algorithm.

The discussion of the specific gate-oxide short defect comes with a newly proposed circuit-level description model. The model considers both the DC and the transient behaviors which have not been investigated by previous works. To verify the model, device-level simulation with TCAD tool and HSPICE BSIM model extraction were proceeded. More than 20% of fitting error between SPICE and TCASD curves can be reduced by using the newly proposed model. Based on the device-level analysis, the GOS fault behaviors at traditional planar bulk CMOS and advanced FinFETs are also studied. The distinguishing of the different fault behaviors between technologies helps the development and modification for an effective SRAM test method.

Chapter 7
Future Work

Based on the categorization, analysis, and development of the test methods for Sub-V_{th} SRAMs in this thesis, the next step would be the corresponding implementation. To embed the proposed test methods into the chip, integration with the original BIST circuitry would be the best choice. Hence, my first future work is to collect the necessary March algorithms as well as the proposed tests in this thesis. Then, by deleting repeated or redundant operations, a succinct and efficient test process can be acquired. Finally, the memory BIST compiler would be used to generate the BIST circuitry with minimized area overhead.

When parsing the SPICE models applied in the experiments of this thesis, I have discovered the BSIM model always describes the device characteristics in the manner of "region by region". As a result, device operated at the region near threshold is often interpreted by the two inconsistent regions: sub-threshold and super-threshold, which leads to the imprecise description. Hence, the second future work is to develop the appropriate analytical model for characterizing the device behavior near threshold. Then, the optimization of near-threshold logic/SRAM design could be further improved.

The third future work would be the extension of the discussion for complicated manufacture defects/faults. For example, in the modern VLSI, the impact of random telegraph noise becomes more evident. However, the traditional tests which often aim at hard defects can hardly capture the soft errors. Hence, how to evaluate the interface quality between oxide and Si channel and to stress the chip in the early stage for triggering the fault become the next importance topic.

Bibliography

[2-1] B. Zhai, D. Blaauw, D. Sylvester, and K. Flautner, ”Theoretical and practical limits of dynamic voltage scaling,” P. 868-873, ACM/IEEE Design Automation Conference, 2004.

[2-2] B.H. Calhoun and A. Chandrakasan, ”Characterizing and modeling minimum energy operation for subthreshold circuits,” P. 90–95, IEEE International Symposium on Low Power Electronics and Design, 2004.

[2-3] R.G. Dreslinski, B. Zhai, T. Mudge, D. Blaauw, and D. Sylvester, ”An Energy Efficient Parallel Architecture Using Near Threshold Operation,” P. 175-188, International Conference on Parallel Architecture and Compilation Techniques, 2007.

[2-4] R.G. Dreslinski, M. Wieckowski, D. Blaauw, D. Sylvester, and T. Mudge, “Near-Threshold Computing: Reclaiming Moore’s Law Through Energy Efficient Integrated Circuits,” VOL. 98, ISSUE 2, P. 253–266, Proceedings of the IEEE, 2010.

[2-5] A. Wang, B. H. Calhoun, and A. P. Chandrakasan ”Sub-threshold Design for Ultra Low-Power Systems,” Springer Science+Business Media, LLC, 2006.

[2-6] M. Yamaoka, N. Maeda, Y. Shinozaki, Y. Shimazaki, K. Nii, S. Shimada, K. Yanagisawa, and T. Kawahara, ”Low-power embedded SRAM modules with expanded margins for writing,” VOL. 1, P. 480–611, IEEE International Solid-State Circuits Conference, 2005.

[2-7] M. Yamaoka, K. Osada, R. Tsuchiya, M. Horiuchi, S. Kimura, and T. Kawahara, ”Low power SRAM menu for SOC application using Yin-Yang-feedback memory cell technology,” P. 288–291, Symposium on VLSI Circuits, 2004.

[2-8] K. Zhang, U. Bhattacharya, Z. Chen, F. Hamzaoglu, D. Murray, N. Vallepalli, Y. Wang, B. Zheng, and M. Bohr, ”SRAM design on 65nm CMOS technology with integrated leakage reduction scheme,” P. 294–295, Symposium on VLSI Circuits, 2004.

[2-9] B. H. Calhoun and A. Chandrakasan, ”A 256kb Sub-threshold SRAM in 65nm CMOS,” IEEE International Solid-State Circuits Conference, 2006.

[2-10] T. H. Kim, J. Liu, J. Keane, and C. H. Kim, ”A 0.2 V, 480 kb Subthreshold SRAM With 1 k Cells Per Bitline for Ultra-Low-Voltage Computing,” P. 518–529, IEEE Journal of Solid-State Circuits, 2008.

[2-11] N. Verma and A. P. Chandrakasan, ”A 256 kb 65 nm 8T Subthreshold

SRAM Employing Sense-Amplifier Redundancy," P. 141–149, IEEE Journal of Solid-State Circuits, 2008.

[2-12] B. Zhai, S. Hanson, D. Blaauw, and D. Sylvester, "A Variation-Tolerant Sub-200 mV 6-T Subthreshold SRAM," P. 2338–2348, IEEE Journal of Solid-State Circuits, 2008.

[2-13] J. Singh, J. Mathew, D. K. Pradhan, and S. P. Mohanty, "A Subthreshold Single Ended I/O SRAM Cell Design for Nanometer CMOS Technologies," IEEE International SOC Conference, 2008.

[2-14] I. J. Chang, J. J. Kim, S. P. Park, and K. Roy, "A 32 kb 10T Sub-Threshold SRAM Array With Bit-Interleaving and Differential Read Scheme in 90 nm CMOS," P. 650–658, IEEE Journal of Solid-State Circuits, 2009.

[2-15] M. T. Chang and W. Hwang, "A Fully-Differential Subthreshold SRAM cell with Auto-Compensation," IEEE Asia Pacific Conference on Circuits and Systems, 2008.

[2-16] T. H. Kim, J. Liu, and C. H. Kim, "An 8T Subthreshold SRAM Cell Utilizing Reverse Short Channel Effect for Write Margin and Read Performance Improvement," P. 241–244, IEEE Custom Integrated Circuits Conference, 2007.

[2-17] L. Dilillo, P. Girard, S. Pravossoudovitch, A. Virazel, and H. H. Magali Bastian, "Data Retention Fault in SRAM Memories: Analysis and Detection Procedures," IEEE VLSI Test Symposium, 2005.

[2-18] A. Pavlov, M. Azimane, J. P. de Gyvez, and M. Sachdev, "Word Line Pulsing Technique for Stability Fault Detection in SRAM Cells," IEEE International Test Conference, 2005.

[2-19] A. Ney, L. Dilillo, P. Girard, S. Pravossoudovitch, A. Virazel, M. Bastian, and V. Gouin, "A New Design-for-Test Technique for SRAM Core-Cell Stability Faults," Design, Automation & Test in Europe Conference & Exhibition, 2009.

[2-20] J. Yang, B. Wang, Y. Wu, and A. Ivanov, "Fast Detection of Data Retention Faults and Other SRAM Cell Open Defects," P. 167–180, IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems, 2006.

[2-21] L. Dilillo, P. Girard, S. Pravossoudovitch, A. Virazel, S. Borri, and H. H. Magali, "Resistive-Open Defects in Embedded-SRAM core cells: Analysis and March Test Solution," Asian Test Symposium, 2004.

[2-22] A. J. van de Goor and G. N. Gaydadjiev, "An Analysis of (Linked) Address Decoder Faults," International Workshop on Memory Technology,

Design and Testing, 1997.

[2-23] J. Otterstedt, D. Niggemeyer, and T. W. Williams, ”Detection of CMOS Address Decoder Open Faults with March and Pseudo Random Memory Tests,” International Test Conference, 1998.

[2-24] A. J. van de Goor, ”Testing Semiconductor Memories,” John Wiley & Sons Ltd, 1991.

[2-25] H. Geib, W. Weber, E. Wohlrab, and L. Risch, ”Experimental Investigation of the Minimum Signal for Reliable Operation of DRAM Sense Amplifiers,” VOL. 27, ISSUE 7, IEEE Journal of Solid-State Circuits, 1992.

[2-26] E. J. Sprogris, ”A Technique for Measuring Threshold Mismatch in DRAM Sense Amplifier Devices,” P. 103–106, Microelectronic Test Structures, 1991.

[2-27] Neil H. E. Weste and D. Harris, ”CMOS VLSI Design,” Pearson Education, Inc, 2005.

[2-28] A. Golda and A. Kos, ”Temperature Influence on Power Consumption and Time Delay,” P. 378–382, Euromicro Symposium on Digital System Design, 2003.

[2-29] B. Datta and W. Burleson, ”Temperature Effects on Energy Optimization in Sub-Threshold Circuit Design,” P. 680–685, Quality of Electronic Design, 2009.

[2-30] R. Kumar and V. Kursun, ”Temperature-Adaptive Energy Reduction for Ultra-Low Power-Supply-Voltage Subthreshold Logic Circuits,” P. 1280–1283, IEEE International Conference on Electronics, Circuits and Systems, 2007.

[2-31] M. Sachdev, ”Test and testability techniques for open defects in RAM address decoders,” P. 428–434, European Design and Test Conference, 1996.

[2-32] M. Sachdev, ”Open defects in CMOS RAM address decoders,” VOL. 14, ISSUE 2, P. 26–33, IEEE Design & Test of Computers, 1997.

[2-33] K. M. Thompson, ”Intel and the myths of test,” IEEE Des. Test. Comput., VOL. 13, NO. 1, pp. 79–81, 1996.

[2-34] W. Needham, C. Prunty, and E. H. Yeoh, ”High volume microprocessor test escapes, an analysis of defects our tests are missing,” IEEE International Test Conference, 1998.

[2-35] A. K. Majhi, M. Azimane, G. Gronthoud, M. Lousberg, S. Eichenberger, and F. Bowen, ”Memory testing under different stress conditions: an industrial evaluation,” IEEE Design, Automation and Test in Europe, 2005.

[3-1] B. Zhai, D. Blaauw, D. Sylvester, K. Flautner, "Theoretical and practical limits of dynamic voltage scaling," Design Automation Conference, pp. 868-873, 2004.
[3-2] B.H. Calhoun, A. Chandrakasan, "Characterizing and modeling minimum energy operation for subthreshold circuits," International Symposium on Low Power Electronics and Design, pp. 90-95, 2004.
[3-3] A. Wang, B. H. Calhoun, A. P. Chandrakasan, "Sub-threshold Design for Ultra Low-Power Systems," Springer Science+Business Media, LLC, 2006.
[3-4] M. Yamaoka et al., "Low-power embedded SRAM modules with expanded margins for writing," IEEE International Solid-State Circuit Conference, pp. 480-611, 2005.
[3-5] K. Zhang et al., "SRAM design on 65nm CMOS technology with integrated leakage reduction scheme," Symposium on VLSI Circuits, pp. 294-295, 2004.
[3-6] M. Yamaoka et al., "Low power SRAM menu for SOC application using Yin-Yang-feedback memory cell technology," Symposium on VLSI Circuits, pp. 288-291, 2004.
[3-7] Y. W. Lin, "Analysis and Design of Data-Aware Dynamic Supply Write-Assist Scheme for Cross-Point 8T SRAM," Unpublished thesis, National Chiao Tung University, Taiwan, 2010.
[3-8] H. I. Yang et al., "A high-performance low VMIN 55nm 512Kb disturb-free 8T SRAM with adaptive VVSS control," IEEE International SOC Conference, 2011.
[3-9] C. T. Chuang et al., "Data-aware dynamic supply random access memory," US 2012/0044779 A1, U.S. Patent Application Publication, 2012.
[3-10] B.H. Calhoun, A. Chandrakasan, "A 256kb Sub-threshold SRAM in 65nmCMOS," International Solid-State Circuits Conference, 2006.
[3-11] T.H. Kim, J. Liu, J. Keane, C.H. Kim, "A 0.2 V, 480 kb Subthreshold SRAM With 1 k Cells Per Bitline for Ultra-Low-Voltage Computing," Journal of Solid-State Circuits, pp. 518-529, 2008.
[3-12] N. Verma, A.P. Chandrakasan, "A 256 kb 65 nm 8T Subthreshold SRAM Employing Sense-Amplifier Redundancy," Journal of Solid-State Circuits, pp. 141-149, 2008.
[3-13] I.J. Chang, J.J. Kim, S.P. Park, K. Roy, "A 32 kb 10T Sub-Threshold SRAM Array With Bit-Interleaving and Differential Read Scheme in 90 nm CMOS," Journal of Solid-State Circuits, pp. 650-658, 2009.
[3-14] M.T. Chang, W. Hwang, "A Fully-Differential Subthreshold SRAM cell

with Auto-Compensation," Asia Pacific Conference on Circuits and Systems, 2008.

[3-15] Q. Li, T.T. Kim, "A 9T subthreshold SRAM bitcell with data-independent bitline leakage for improved bitline swing and variation tolerance," IEEE Asia Pacific Conference on Circuits and Systems (APCCAS), pp. 260-263, 2010.

[3-16] A.R. Ramani, K. Choi, "A novel 9T SRAM design in sub-threshold region," IEEE International Conference on Electro/Information Technology (EIT), pp. 1-6, 2011.

[3-17] M.-H. Chang, Y.-T. Chiu, S.-L. Lai, W. Hwang, "A 1kb 9T subthreshold SRAM with bit-interleaving scheme in 65nm CMOS," International Symposium on Low Power Electronics and Design (ISLPED), pp. 291-296, 2011.

[3-18] T.H. Kim, J. Liu, C.H. Kim, "An 8T Subthreshold SRAM Cell Utilizing Reverse Short Channel Effect for Write Margin and Read Performance Improvement," Custom Integrated Circuits Conference, pp. 241-244, 2007.

[3-19] J. Singh, J. Mathew, D.K. Pradhan, S.P. Mohanty, "A Subthreshold Single Ended I/O SRAM Cell Design for Nanometer CMOS Technologies," International SOC Conference, 2008.

[3-20] C.-W. Lin et al., "Fault Models and Test Methods for Subthreshold SRAMs," IEEE Transactions on Computers, 2011.

[3-21] A. Ney et al., "A New Design-for-Test Technique for SRAM Core-Cell Stability Faults," Design, Automation & Test in Europe Conference & Exhibition, 2009.

[3-22] J. Yang et al., "Fast Detection of Data Retention Faults and Other SRAM Cell Open Defects," P. 167–180, IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems, 2006.

[3-23] Y. W. Chiu et al., "8T Single-ended sub-threshold SRAM with crosspoint data-aware write operation," International Symposium on Low Power Electronics and Design, 2011.

[3-24] M. H. Tu et al., "A Single-Ended Disturb-Free 9T Subthreshold SRAM With Cross-Point Data-Aware Write Word-Line Structure, Negative Bit-Line, and Adaptive Read Operation Timing Tracing," Vol. 47, Issue 6, IEEE Journal of Solid-State Circuits, 2012.

[3-25] M. F. Chang et al., "A Differential Data Aware Power-supplied (D2AP) 8T SRAM Cell with Expanded Write/Read Stabilities for Lower VDDmin Applications," Vol. 45, No. 6, IEEE Journal of Solid-State Circuits, 2010.

[3-26] Y. W. Lin et al., "A 55nm 0.55V 6T SRAM with Variation-Tolerant

Dual-Tracking Word-Line Under-Drive and Data-Aware Write-Assist," IEEE International Symposium on Low Power Electronics and Design (ISLPED), 2012.

[4-1] Kaushik Roy, S. Mukhopadhyay, H. Mahmood-Meimand, "Leakage Current Mechanisms and Leakage Reduction Techniques in Deep-Submicrometer CMOS Circuits," Proceedings of the IEEE, Vol. 91, No. 2, Feb. 2003.

[4-2] J. Segura, C. De Benito, A. Rubio, C.F. Hawkins, "A detailed analysis of GOS defects in MOS transistors: testing implications at circuit level," International Test Conference, 1995.

[4-3] Abu Khari bin, Kian Sin Sim, Cheow Kwee Siong, "The effects of gate oxide short in 6-transistors SRAM cell," IEEE International Semiconductor Electronics, 2004.

[4-4] T.W. Chen, K. Kim, Y.M. Kim, S. Mitra, "Gate-Oxide Early Life Failure Prediction," VLSI Test Symposium, 2008.

[4-5] J.M. Galliere, F. Azais, M. Renovell, L. Dilillo, "Influence of gate oxide short defects on the stability of minimal sized SRAM core-cell by applying non-split models," Design & Technology of Integrated Systems in Nanoscal Era, 2009.

[4-6] A. Ille, W. Stadler, T. Pompl, H. Gossner, T. Brodbeck, K. Esmark, P. Riess, D. Alvarez, K. Chatty, R. Gauthier, A. Bravaix, "Reliability aspects of gateoxide under ESD pulse stress," Vol. 49, Issue 12, Microelectronics Reliability, 2009.

[4-7] L. Lakshmanan, L. Herlinger, and K. Miller, "A Sample Preparation Technique to Localize Gate Oxide Defects in Memory Arrays Using Conducting Atomic Force Microscopy," International Symposium for Testing and Failure Analysis, 2011.

[4-8] Hao Hong, E.J. McCluskey, "Analysis of gate oxide shorts in CMOS circuits," IEEE Transactions on Computers, 1993.

[4-9] Hao Hong, E.J. McCluskey, "On the modeling and testing of gate oxide shorts in CMOS logic gate," International Workshop on Defect and Fault Tolerance on VLSI Systems, 1991.

[4-10] Hao Hong, E.J. McCluskey, "Very-low-voltage testing for weak CMOS logic ICs," International Test Conference, 1993.

[4-11] M. Sytrzycki, "Modeling of gate oxide shorts in MOS transistors," Vol 8, Issue 3, IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems, 1989.

[4-12] J. Segura, A. Rubio, J. Figueras, "Analysis and modeling of MOS

devices with gate oxide short failures," IEEE International Symposium on Circuits and Systems, 1991.

[4-13] S.I. Syed, D.M. Wu, "Test generation for gate oxide short in CMOS ICs," IEEE Southeastcon, 1990.

[4-14] R. Rodriguez-Montanes, J.A. Segura, V.H. Champac, J. Figueras, J.A Rubio, "CURRENT VS. LOGIC TESTING OF GATE OXIDE SHORT, FLOATING GATE AND BRIDGING FAILURES IN CMOS," International Test Conference, 1991.

[4-15] D.M. Wu, "Detecting gate-oxide shorts and delay defects in CMOS IC's," IEEE Southeastcon, 1991.

[4-16] J.A. Segura, A. Rubio, "GOS defects in SRAM: fault modeling and testing possibilities," IEEE International Workshop on Memory Technology, 1994.

[4-17] J.M. Soden, C.F. Hawkins, "Test Considerations for Gate Oxide Shorts in CMOS ICs," Vol 3, Issue 4, IEEE Design & Test of Computers, 1986.

[4-18] S.I. Syed, D.M. Wu, "Defect analysis and test generation for gate oxide shorts in CMOS ICs," IEEE Custom Integrated Circuits Conference, 1990.

[4-19] S.I. Syed, D.M.Wu, "Defect analysis, test generation and fault simulation for gate oxide shorts in CMOS ICs," IEEE International Symposium on Circuits and Systems, 1990.

[4-20] M. Renovell, J.M. Galliere, F. Azais, Y. Bertrand, "Modeling gate oxide short defects in CMOS minimum transistors," IEEE European Test Workshop, 2002.

[4-21] M. Renovell, J.M. Galliere, F. Azais and Y. Bertrand, "Modeling the Random Parameters Effects in a Non-Split Model of Gate Oxide Short," Vol 19, Journal of Electronic Testing, 2003.

[4-22] S. Chehade, A. Chehab, A. Kayssi, "Modeling of gate oxide short defects in MOSFETS," International Design and Test Workshop , 2009.

[4-23] M. Renovell, J.M. Galliere, F. Azais, Y. Bertrand, "Delay testing of MOS transistor with gate oxide short," Asian Test Symposium, 2003.

[4-24] Xiang Lu, Zhuo Li, Wangqi Qiu, D.M.H. Walker, Weiping Shi, "A circuit level fault model for resistive shorts of MOS gate oxide," Fibres and Optical Passive Components, 2005.

[4-25] J. M. Galliere, M. Renovell, F. Azais, Y. Bertrand, "Delay testing viability of gate oxide short defects," Vol 20, Issue 2, Journal of Computer Science and Technology, 2005.

[4-26] "Technology computer aided design (TCAD) : Process and device simulation tools," http://www.synopsys.com.

[4-27] ”UTMOST IV : Optimization Module for Compact/Macro-Modeling,” http://www.silvaco.com.

[4-28] ”BSIM : Berkeley Short-channel IGFET Model,” http://wwwdevice.eecs.berkeley.edu/~bsim.

[4-29] J. Segura, A. Rubio, ”A detailed analysis of CMOS SRAM’s with gate oxide short defects,” Vol 32, Issue 10, IEEE Journal of Solid-State Circuits, 1997.

[4-30] S. Bhunia, Li Hai, K. Roy, ”A high performance IDDQ testable cache for scaled CMOS technologies,” Asian Test Symposium, 2002.

[4-31] M. Yabuuchi, K. Nii, Y. Tsukamoto, S. Ohbayashi, Y. Nakase, H. Shinohara, “A 45nm 0.6V cross-point 8T SRAM with negative biased read/write assist,” Symposium on VLSI Circuits, Page 158–159, 2009.

[4-32] S. Mukhopadhyay, R.-M. Rao, J.-J. Kim, and Ching-Te 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, 2011.

[4-33] A. Karandikar, K. K. Parhi, ”Low power SRAM design using hierarchical divided bit-line approach,” International Conference on Computer Design: VLSI in Computers and Processors, 1998.

[4-34] V. Dabholkar, S. Chakravarty, ”Computing stress tests for gate-oxide shorts,” International Conference on VLSI Design, 1998.

[4-35] E. Isern, M. Roca, J. Segura, ”Analyzing the need for ATPG targeting GOS defects,” IEEE VLSI Test Symposium, 1999.

[4-36] A. Yassine, K. Wieczorek, K. Olasupo, V. Heinig, ”A novel electrical test to differentiate gate-to-source/drain silicide short from gate oxide short,” IEEE International Integrated Reliability Workshop Final Report, 2000.

[4-37] M. Renovell, J.M. Galliere, F. Azais, Y. Bertrand, ”Boolean and current detection of MOS transistor with gate oxide short,” International Test Conference, 2001.

[4-38] S.A. Kumar, R.Z. Makki, D.M. Binkley, ”IDDT testing of embedded CMOS SRAMs,” Design, Automation and Test in Europe Conference and Exhibition, 2002.

[4-39] Chua Yong Moh, A’ain, A.Kb., ”A study on characterization of gate oxide shorts using non-split model,” Student Conference on Research and Development, 2003.

[4-40] A. Chehab, A. Kayssi, A. Ghandour, ”Transient Current Testing of Gate-Oxide Shorts in CMOS,” International Design and Test Workshop, 2007.

[4-41] A. Ghandour, K. Fawaz, A. Chehab, A. Kayssi, ”Wavelet transform-based transient current analysis for detection of gate-oxide shorts in CMOS,” International Design and Test Workshop , 2009.
[4-42] STMicroelectronics, ”Deep Sub-Micron Processes: 120nm, 90nm, 65 nm, 45nm CMOS,” CMP annual users meeting, PARIS, January 2008.
[4-43] F. Arnaud et. al., ”A Functional 0.69μm2 Embedded 6T-SRAM bit cell for 65nm CMOS platform,” in the Digest of Technical Papers of the Symposium on VLSI Technology, 10-12 June 2003, pp. 65-66.
[4-44] Teong-San Yeoh, N. R. Kamat, R. S. Nair, and Shze-Jer Hu, ”Gate oxide breakdown model in MOS transistors,” IEEE International Reliability Physics Symposium, 1995.
[4-45] A. Meixner, and J. Banik, ”Weak Write Test Mode: an SRAM cell stability design for test technique,” International Test Conference, 1997.
[4-46] Vikas Chandra, Cezary Pietrzyk, and Robert C. Aitken, ”On the efficacy of write-assist techniques in low voltage nanoscale SRAMs,” Design, Automation & Test in Europe Conference & Exhibition (DATE), 2010.
[4-47] Saibal Mukhopadhyay, Rahul M.Rao, Jae-Joon Joon Kim, and Ching-Te T. Chuang, ”Capacitive coupling based transient negative bit-line voltage (Tran-NBL) scheme for improving write-ability of SRAM design in nanometer technologies,” IEEE International Symposium on Circuits and Systems, 2008.
[4-48] Daeyeon Kim, Gregory K. Chen, Matthew Fojtik, Mingoo Seok, David T. Blaauw, and Dennis M. Sylvester, ”A 1.85fW/bit ultra low leakage 10T SRAM with speed compensation scheme,” IEEE International Symposium on Circuits and Systems (ISCAS), 2011
[4-49] Lawrence F. Childs, Craig D. Gunderson, Olga R. Lu, and James D. Burnett, ”Memory having self-timed bit line boost circuit and method therefor,” U.S. Patent application number: 20100074032, 2010.
[4-50] Igor Arsovski, Harold Pilo, Vinod Ramadurai, and Vinod Ramadurai, “Static random access memory (SRAM) write assist circuit with leakage suppression and level control,” U.S. Patent application number: 20120140551, 2012.
[4-51] Vibhu Sharma, Francky Catthoor, and Wim Dehaene, ”SRAM Design for Wireless Sensor Networks,” Page 31–64, ISBN 978-1-4614-4039-0, Springer, 2013.
[5-1] Vita P. H. Hu et al., ”FinFET SRAM Cell Optimization Considering Temporal Variability Due to NBTI/PBTI, Surface Orientation and Various Gate Dielectrics,” Vol. 58, Issue 3, IEEE Transactions on Electron Devices,

2011.
[5-2] B. Raj et al., "Nanoscale FinFET Based SRAM Cell Design: Analysis of Performance Metric, Process Variation, Underlapped FinFET, and Temperature Effect," Vol. 11, Issue 3, IEEE Circuits and Systems Magazine, 2011.
[5-3] M. Kang et al., "FinFET SRAM Optimization With Fin Thickness and Surface Orientation," Vol. 57, Issue 11, IEEE Transactions on Electron Devices, 2010.
[5-4] Y. W. Li et al., "A reconfigurable distributed all-digital clock generator core with SSC and skew correction in 22nm high-k tri-gate LP CMOS," IEEE International Solid-State Circuits Conference, 2012.
[5-5] E. Karl et al., "A 4.6GHz 162Mb SRAM design in 22nm tri-gate CMOS technology with integrated active VMIN-enhancing assist circuitry," IEEE International Solid-State Circuits Conference, 2012.
[5-6] E. Baravelli et al., "Impact of Line-Edge Roughness on FinFET Matching Performance," Vol. 54, No. 9, IEEE Transactions on Electron Devices, 2007.
[5-7] S. Yu et al., "The impact of line edge roughness on the stability of a FinFET SRAM," Vol. 24, No. 2, Semiconductor Science and Technology, 2009.
[5-8] P. Majhi, et al., "CMOS Scaling Beyond High-k and Metal Gates," Issue 22, Future Fab Intl., 2007.
[5-9] K. Shin et al., "Dual stress capping layer enhancement study for hybrid orientation FinFET CMOS technology," IEDM Tech. Dig., pp. 998–991, 2005.
[5-10] E. A. Cartier et al., "FinFET with thin gate dielectric layer," US 2011/0175163 A1, U.S Patent Application Publication, 2011.
[5-11] T. G. Dziura et al., "Measurement of high-k and metal film thickness on FinFET sidewalls using scatterometry," Vol. 6922, pp. 69220V-69220V-8, Proceedings of the SPIE, 2008.
[5-12] B. Foran et al., "Tomography of High-k Dielectrics on Fin-FET Sidewalls," Vol. 14, pp. 444–445, Microscopy and Microanalysis, 2008.
[5-13] J. Segura et al., "A detailed analysis of GOS defects in MOS transistors: testing implications at circuit level," International Test Conference, 1995.
[5-14] Abu Khari bin et al., "The effects of gate oxide short in 6-transistors SRAM cell," IEEE International Semiconductor Electronics, 2004.
[5-15] Kaushik Roy et al., "Leakage Current Mechanisms and Leakage Reduction Techniques in Deep-Submicrometer CMOS Circuits,"

Proceedings of the IEEE, Vol. 91, No. 2, Feb. 2003.
[5-16] M. O.Simsir et al., ”Fault modeling for FinFET circuits,” International Symposium on Nanoscale Architectures, 2010.
[5-17] A. N. Bhoj et al., ”Fault Models for Logic Circuits in the Multigate Era,” Vol. 11, Issue 1, IEEE Transactions on Nanotechnology, 2012.
[5-18] M. Renovell et al., ”Modeling the Random Parameters Effects in a Non-Split Model of Gate Oxide Short,” Vol 19, Journal of Electronic Testing, 2003.
[5-19] S. Chehade et al., ”Modeling of gate oxide short defects in MOSFETS,” International Design and Test Workshop , 2009.
[5-20] J. A. Segura et al., ”GOS defects in SRAM: fault modeling and testing possibilities,” IEEE International Workshop on Memory Technology, 1994.
[5-21] J. Segura et al., ”A detailed analysis of CMOS SRAM's with gate oxide short defects,” Vol 32, Issue 10, IEEE Journal of Solid-State Circuits, 1997.
[5-22] J. M. Soden et al., ”Test Considerations for Gate Oxide Shorts in CMOS ICs,” Vol 3, Issue 4, IEEE Design & Test of Computers, 1986.
[5-23] M. L. Fan et al., ”Investigation of Cell Stability and Write Ability of FinFET Subthreshold SRAM Using Analytical SNM Model,” Vol. 57, Issue 6, IEEE Transactions on Electron Devices, 2010.
[5-24] M. Sytrzycki, ”Modeling of gate oxide shorts in MOS transistors,” Vol 8, Issue 3, IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems, 1989.
[5-25] J. Segura et al., ”Analysis and modeling of MOS devices with gate oxide short failures,” IEEE International Symposium on Circuits and Systems, 1991.
[5-26] Z. Liu et al., ”An independent-gate FinFET SRAM cell for high data stability and enhanced integration density,” International SOC Conference, 2007.
[5-27] O. Thomas et al., ”Sub-1V, Robust and Compact 6T SRAM cell in Double Gate MOS technology,” IEEE International Symposium on Circuits and Systems, 2007.
[5-28] M. M. Pelella et al., ”On the Performance Advantage of PD/SOI CMOS With Floating Bodies,” Vol. 49, No. 1, IEEE TRANSACTIONS ON ELECTRON DEVICES, 2002.
[5-29] P. G. Der Agopian et al., ”Study of the linear kink effect in PD SOI nMOSFETs,” Vol. 38, Issue 1, pp. 114–119, Microelectronics Journal, 2007.
[5-30] S. A. Tawfik et al., ”Low-power and robust six-FinFET memory cell using selective gate-drain/source overlap engineering,” International

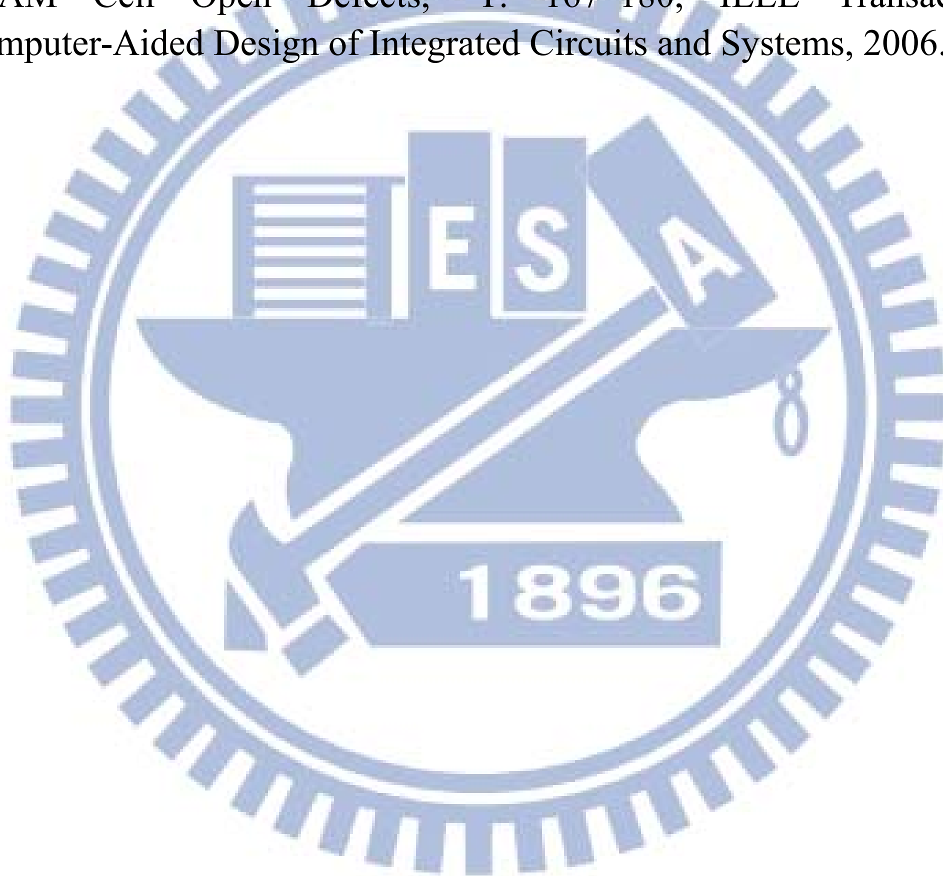
Symposium on Integrated Circuits, 2009.

[5-31] Z. Liu et al., ”Statistical Data Stability and Leakage Evaluation of FinFET SRAM Cells with Dynamic Threshold Voltage Tuning under Process Parameter Fluctuations,” International Symposium on Quality Electronic Design, 2008.

[5-32] F. Sheikh et al., ”The impact of device-width quantization on digital circuit design using FinFET structures,” Proc. EE241 Spring, pp.1–6, 2004.

[5-33] B. Wicht et al., ”A yield-optimized latch-type SRAM sense amplifier,” European Solid-State Circuits Conference, 2003.

[5-34] J. Yang et al., ”Fast Detection of Data Retention Faults and Other SRAM Cell Open Defects,” P. 167–180, IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems, 2006.



Publication list

Journal Papers

- **Chen-Wei Lin**, Mango Chao, Chih-Chieh Hsu, "A Novel Circuit-level Model for Gate Oxide Short and its Testing Method in SRAMs," *IEEE Transactions on Very Large Scale Integration Systems (TVLSI)*, **accepted**, 2013.
- **Chen-Wei Lin**, Hung-Hsin Chen, Hao-Yu Yang, Chin-Yuan Huang, Mango C.-T. Chao, and Rei-Fu Huang, "Fault Models and Test Methods for Subthreshold SRAMs," *IEEE Transactions on Computers (TC)*, **accepted**, 2012.
- **Chen-Wei Lin**, Mango C.-T. Chao, and Yen-Shih Huang, "A Novel Pixel Design for AM-OLED Displays Using Nanocrystalline Silicon TFTs," *IEEE Transactions on Very Large Scale Integration Systems (TVLSI)*, 2011.
- **Chen-Wei Lin** and Jiun-Lang Huang, "A Charge-Sensing-Capable Source Driver for TFT Array Testing in System-on-Panel Displays," *Journal of Computers (JCP)*, 2009.

Conference Papers

- **Chen-Wei Lin**, Chih-Hsiang Ho, Chao Lu, Mango C.-T. Chao, and Kaushik Roy, "A Process/Device/Circuit/System Compatible Simulation Framework for Poly-Si TFT Based SRAM Design," *IEEE International Conference on Simulation of Semiconductor Processes and Devices (SISPAD)*, 2013.
- **Chen-Wei Lin**, Chin-Yuan Huang, Mango Chao, "Testing of a Low-V_{MIN} Data-Aware Dynamic-Supply 8T SRAM," *IEEE VLSI Test Symposium (VTS)*, 2013.
- **Chen-Wei Lin**, Mango Chao, Chih-Chieh Hsu, "Investigation of Gate Oxide Short in FinFETs and the Test Methods for FinFET SRAMs," *IEEE VLSI Test Symposium (VTS)*, 2013.
- Hao-Yu Yang, **Chen-Wei Lin**, Hung-Hsin Chen, Mango C.-T. Chao, Ming-Hsien Tu, Shyh-Jye Jou, Ching-Te Chuang, "Testing Strategies for a 9T Sub-threshold SRAM," *International Test Conference (ITC)*, 2012.
- **Chen-Wei Lin**, Hao-Yu Yang, Chin-Yuan Huang, Hung-Hsin Chen, and Mango C.-T. Chao, "Detecting Stability Faults in Sub-threshold SRAMs," *International Conference on Computer-Aided Design (ICCAD)*, 2011.
- **Chen-Wei Lin**, Hung-Hsin Chen, Hao-Yu Yang, Mango C.-T. Chao, and Rei-Fu Huang, "Fault Models and Test Methods for Subthreshold SRAMs," *International Test Conference (ITC)*, 2010.
- Lin Hung-Chien, King-Yuan Ho, Han-Sheng Dai, Yen-Shih Huang, Jing-Yi Yan,

Po-Feng Lee, Shih-Yung Peng, Chih-Hung Cheng, **Chen-Wei Lin**, Hua-Chi Cheng, "High Contrast Active Matrix Organic Light Emitting Display With Top Gate Amorphous Silicon Thin Film Transistors," *International Meeting on Information Display (IMID)*, 2010.

- Mango C.-T. Chao, Ching-Yu Chin, and **Chen-Wei Lin**, "Mathematical Yield Estimation for Two-Dimensional-Redundancy Memory Arrays," *International Conference on Computer-Aided Design (ICCAD)*, 2010.
- Hung-Chien Lin, Han-Sheng Dai, King-Yuan Ho, Chih-Chieh Hsu, Shih-Yung Peng, **Chen-Wei Lin**, Po-Feng Lee, Chih-Hung Cheng, Bo-Cheng Kung, Wen-Ya Chao, Jing-Yi Yan, Hua-Chi Cheng, "High Performance Nanocrystalline Silicon Thin Film Transistors: Mechanical Device Reliability," *International Display Workshops (IDW)*, 2010.
- King-Yuan Ho, Hung-Chien Lin, Han-Sheng Dai, Po-Feng Lee, Chih-Chieh Hsu, Shih-Yung Peng, **Chen-Wei Lin**, Chih-Hung Cheng, Bo-Cheng Kung, Wen-Ya Chao, Jing-Yi Yan, Hua-Chi Cheng, "4.1-inch Full Color AMOLED Driving by Top Gate Nanocrystalline Silicon Thin Film Transistor Array," *International Display Workshops (IDW)*, 2010.
- Jung-Jie Huang, Yung-Pei Chen, Yeh-Shih Huang, Guo-Ren Hu, **Chen-Wei Lin**, Yu-Jen Chen, Po-Feng Lee, Cheng-Ju Tsai, Chan-Jui Liu, Hsiao-Chiang Yao, King-Yuan Ho, Bo-Cheng Kung, Shih-Yung Peng, Chyi-Ming Leu, Jing-Yi Yan, Shu-Tang Yeh, Heng-Lin Pan, Hua-Chi Cheng and Cheng-Chung Lee, "A 4.1-inch Flexible QVGA AMOLED using a Microcrystalline-Si:H TFT on a Polyimide Substrate," *Society for Information Display (SID)*, 2009.
- Hua-Chi Cheng, Yen Shih Huang, Chan-Jui Liu, **Chen-Wei Lin**, King-Yuan Ho, Chih-Hung Cheng, Shih Yung Peng, Yung-Pei Chen, Hung-Chien Lin, Bo-Cheng Kung, Po-Feng Lee, Jung-Jie Huang, Liang-You Jiang, and Cheng-Chung Lee, "Plastic Substrate and Backplane for Flexible AMOLED by Sheet to Sheet Process," *International Display Workshops (IDW)*, 2009.
- **Chen-Wei Lin** and Jiun-Lang Huang, "A Built-In TFT Array Charge-Sensing Technique for System-on-Panel Displays," *IEEE VLSI Test Symposium (VTS)*, 2008.
- Yen Shih Huanga, Hua Chi Cheng, Chan-Jui Liu, **Chen-Wei Lin**, King-Yuan Ho, Chih-Hung Cheng, Shih Yung Peng, Pi Hsien Wang, Yung-Pei Chen, Min-Hung Lee, Hung-Chien Lin, Bo-Cheng Kung, Cheng-Ju Tsai, Yu-Ten Chen, Po-Feng Lee, Guo-Ren Hu, Jung-Jie Huang, and Cheng-Chung Lee, "High Flexibility of AMOLED Displays on Colorless PI Substrate," *International Display Workshops (IDW)*, 2008.