國 立 交 通 大 學

電子工程學系電子研究所

碩 士 論 文

6T 靜態隨機存取記憶體的設計與特性分析

Design and characterization of 6T SRAM

研 究 生：林 宜 緯

指導教授：莊 景 德

中 華 民 國 九十九 年 十一 月

6T 靜態隨機存取記憶體的設計與特性分析

Design and characterization of 6T SRAM

研 究 生：林宜緯　　　　Student：Yi-Wei Lin

指導教授 ： 莊景德　　　　Advisor:Prof. Ching-Te Chuang

國 立 交 通 大 學

電子工程學系 電子研究所碩士班

碩 士 論 文

A Thesis

Submitted to Department of Electronics Engineering &

Institute of Electronics

College of Electrical and Computer Engineering

National Chiao Tung University

in Partial Fulfillment of the Requirements

for the Degree of

Master of Science

In

Electronics Engineering

September 2010

Hsinchu, Taiwan, Republic of China

中華民國 九十九 年 十一 月

6T 靜態隨機存取記憶體的設計與特性分析

學生：林宜緯　　　　　　　　　　　　　　指導教授：莊景德教授

國立交通大學電子工程學系電子研究所

摘要

對於幾乎現今所有的電子設備都必須要用到記憶體來當作儲存媒介，於是記憶體的操作效能變掌握了整個系統的操作速度。而因為靜態隨機存取記憶體有著比其他記憶體種類更高的操作速度，靜態隨機存取記憶體通常會被嵌入到系統當中做為儲存媒介或者是快取記憶體。過去 20 年當中，6T 靜態隨機存取記憶體因為有比較緊密的面積以及較高的操作速度而變成靜態隨機存取記憶體設計的主流。然而當製程進行到一百奈米等級之後，製程變異讓 6T 靜態隨機存取記憶體變得很難存活。先進製程中，6T 靜態隨機存取記憶體的讀取和寫入能力都遭受到很大的退化。特別是低壓操作，6T 靜態隨機存取記憶體能夠正常運作的機率就又更小了。

為了要使 6T 靜態隨機存取記憶體能夠正常的在先進製程下工作，我們提出了兩個電路技巧: WL 降壓以及資料依存性寫入幫助電路來幫助讀寫能力。而讀寫的雜訊限度都會得到提升。即使在低壓操作，128k 位元的 6T 靜態隨機存取記憶體測試晶片依然能夠操作。此外，我們另外實現了一個監測電路來對 6T 靜態隨機存取記憶體的變異以及雜訊限度來做特性化。512 位元的測試陣列可以提供我們足夠的資料量來分析其統計上的分佈。而陣列的實現方式可以使得到的資訊能夠很接近真實的 6T 靜態隨機存取記憶體陣列而不是僅僅從模擬上得到。而配合特別設計的的量測方法，可以是我們的解析度達到某個程度，也可以自動化的量測。

Design and characterization of 6T SRAM

Student：Yi-Wei Lin Advisor：Ching-Te Chuang

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

ABSTRACT

Almost the modern electronic devices need memory as its storage media, and the performance of memory always dominant the overall performance of one system. Since SRAM has highest operating speed than other memory family, it is usually embedded into system to storage data or to be a cache. From past decades, standard 6T cell becomes the main stream of SRAM design due to its compact area and high speed. However, as the technology goes beyond 100 nm, variation issue makes 6T SRAM cell hard to survive. The Read/Write ability suffers a serious degradation in advanced technology node. Especially at low voltage, 6T SRAM seem has smaller probability to work.

In order to successfully allow 6T SRAM work at advanced process, we proposed two circuit techniques : Word-Line Under-Drive and Data-Aware Write-Assist to increase the read and write ability. Both Read Static Noise Margin and Write Margin would be improved. Even at low voltage operation, the test chip of one 128kb 6T SRAM still function work. Besides, we implement a monitoring structure to characterize the variation factors and noise margin of 6T cells. The test Array has 512kb cells could provide us a sufficient amount sample to analysis the statistical distribution. The Array Based implementation could allow us to get the information about Noise Margin quite close to real SRAM macro rather than get it just from simulation. And with a special designed measure scheme, the measurement resolution could be guaranteed and the measurement could be automatic.

誌　謝

本論文能順利完成，首先誠摯的感謝指導教授莊景德教授。在這兩年多的研究生涯裡，莊老師以他豐富的學識內容來指導我們，使得我很快就搭上最先端的研究，而省去懵懵懂懂摸索的歷程。莊老師為人和藹可親的態度，讓我在研究過程中遇到問題，都能夠找到一個很好討論以及指點方向的對象。而莊老師的處事不苟且嚴謹的態度，讓我也除了在專業知識以外也學得了許多。

另外感謝楊皓義學長，在研究過程中給了許多的幫助。皓義學長對於研究經驗充足，也在許多方面給了我指導與建議，讓我能夠很快的就對於研究方面上手。至於實驗室的同學以及學弟學妹們，也感謝你們能夠和我一起討論，也讓整個研究室的氣氛相當歡愉，讓我來到實驗室不會有壓力，能夠開心的在這個空間快樂做研究。

最後，謹以此文獻給我摯愛的雙親，感謝你們能夠構將我扶養成人，在我跌跌撞撞的生涯裡給予最適時的幫助與鼓勵。

林宜緯

于新竹交大

2010.09.15

Content

CHAPTER 1 INTRODUCTION **1**

1.1 BACKGROUND 1
1.2 MOTIVATION AND GOALS 1
1.3 THESIS ORGANIZATION 2

CHAPTER 2 OVERVIEW OF THE OPERATION AND THE DESIGN OF 6T SRAM **3**

2.1 MEMORY FAMILY AND SRAM 3
2.2 6T SRAM 4
2.2.1 Introduction of 6T SRAM 4
2.2.2 Read Operation and read disturb voltage of 6T SRAM 5
2.2.3 Write Operation of 6T SRAM 6
2.2.4 Read Static Noise Margin (RSNM) and Write Static Noise Margin (WSNM) 6
2.2.5 Write margin 9
2.2.6 The size of 6T SRAM 10
2.3 SRAM ARRAY ARCHITECTURE 10
2.3.1 Memory Array and half-select problem 10
2.3.2 Decoder 13
2.3.3 Sense scheme and column architecture 15
2.4 VARIATION ISSUE 18
2.4.1 Global variation and Local variation 18
2.4.2 Variation to 6T SRAM 20
2.5 DESIGN METHODOLOGY OF MODERN 6T SRAM 22
2.5.1 Dual Supply Voltage 22
2.5.2 Dynamic Word-Line voltage 24
2.5.3 Negative Bit-Line Level 26
2.6 RELIABILITY ISSUE AND MONITORING STRUCTURE OF SRAM ARRAY 27
2.6.1 BTI effect 27
2.6.2 Monitoring Structures 30

CHAPTER 3 DESIGN OF ONE 128KB TRADITIONAL 6T SRAM WITH WORD-LINE UNDER-DRIVE SKILL AND DATA-AWARE WRITE ASSIST SKILL **32**

3.1 INTRODUCTION 32
3.2 CONCEPT OF WORD-LINE UNDER-DRIVE 35
3.3 PREVIOUS WORK OF WORD-LINE UNDER-DRIVE 35
3.4 PROPOSED WORD-LINE UNDER-DRIVE SKILL 40

3.4.1 RSNM improvement and WM decrease with word-line under-drive ...40
3.4.2 Prior-art comparison ...42
3.4.3 Proposed Word-Line Under Drive (WLUD) Scheme ...44
3.4.4 The chosen of WL voltage ...48
3.5 PREVIOUS WORK OF WRITE ABILITY IMPROVEMENT ...49
3.6 DATA-AWARE WRITE ASSIST CIRCUIT ...52
3.6.1 Proposed Data-Aware Write-Assist Circuits ...53
3.6.2 Write current and Column Half-Selected problems ...54
3.6.3 Asymmetric Hold SNM ...56
3.6.4 Leakage Issue ...57
3.6.5 Keeper Design ...57
3.7 IMPLEMENTATION OF ONE 128KB 6T SRAM ...63
CHAPTER 4 THE DESIGN OF ONE 512KB ARRAY BASED 6T SRAM NOISE MARGIN MEASUREMENT CIRCUIT ...68
4.1 INTRODUCTION ...68
4.2 PREVIOUS WORK ...70
4.2.1 Array Based measurement structure ...70
4.2.2 All Digital measurement scheme ...76
4.3 MODIFIED CELL FOR RSNM MEASUREMENT ...78
4.4 WM MEASUREMENT METHODOLOGY ...81
4.5 ARRAY IMPLEMENTATION ...83
4.6 MEASUREMENT METHODOLOGY AND CIRCUIT BLOCKS ...88
4.7 CALIBRATION ...92
4.8 TEST CHIP IMPLEMENTATION ...93
CHAPTER 5 CONCLUSIONS ...95
REFERENCE OF CHAPTER 2 ...1
REFERENCE OF CHAPTER 3 ...1
REFERENCE OF CHAPTER 4 ...4

List of Figures

Fig. 2-1 Traditional 6T SRAM 5
Fig. 2-2 Voltage Transfer Curve (VTC) of CMOS inverter [2-1] 6
Fig. 2-3 Butterfly curve of HSNM 7
Fig. 2-4 The RSNM butterfly curve 7
Fig. 2-5 The WSNM butterfly curve 7
Fig. 2-6 Write trip voltage and Write Margin (WM)[2-3] 9
Fig. 2-7 Transistor size ratio in 6T SRAM 10
Fig. 2-8 Array architecture of an $2^N x2^M$ memory array[2-1] 11
Fig. 2-9 SRAM critical path [2-5] 12
Fig. 2-10 Half-selected Read Disturb Voltage [2-4] 12
Fig. 2-11 Stability ratio [2-4] 13
Fig. 2-12 CMOS Logic Circuits [2-1] 13
Fig. 2-13 Domino Logics [2-6] 14
Fig. 2-14 Differential sense amplifier [2-1] 16
Fig. 2-15 Large signal sensing scheme of IBM Cell processor [2-7] 17
Fig. 2-16 Global variation and Local variation of threshold voltage [2-6] 19
Fig. 2-17 Static Noise Margin definition and the applied noise or variation that makes the SNM equals zero [9] 20
Fig. 2-18 The effect of the inter-die Vt variation [2-8] 21
Fig. 2-19 The effect of local variation, (a) Write mode worse case, and (b) Read mode worse case 22
Fig. 2-20 Dual VCC Design of [2-10] 23
Fig. 2-21 Negative VSS scheme [2-11] 23
Fig. 2-22 RSNM improvement with lower $V_{WL,}$ (b)WSNM decade with lower V_{WL}[12] 25
Fig. 2-23 Traditional method of word-line under-drive [2-12] 26
Fig. 2-24 Constant Negative Bit-line of [2-13] 27
Fig. 2-25 The generation of interface traps a PMOS transistor is biased in inversion.[2-20] 28
Fig. 2-26 Schematic description of the reaction-diffusion model [2-15] 28
Fig. 2-27 Time dependent effect of stress and relaxation mechanism [2-15] 29
Fig. 2-28 NBTI and PBTI impact to traditional 6T cell 30
Fig. 2-29 RSNM degradation due to NBTI 30
Fig. 3-1 "Level Programmable Word-Line Driver and Dynamic Array Supply Control in Dual-Supply SRAM scheme [3-13]" 35
Fig. 3-2 Adaptive Word-Line Under-Drive (WLUD) Scheme 36
Fig. 3-3 Conventional version (b) resistor added version 37

Fig. 3-4 Simulated word-line voltage of Fig. 3-3 38
Fig. 3-5 The I-V curve of two RACs, and shows the temperature dependence 38
Fig. 3-6 Single-Supply WLUD using Replica Discharging Transistor 39
Fig. 3-7 Standard 6T SRAM in Read Mode 40
Fig. 3-8 RSNM increase with word-line voltage decrease 41
Fig. 3-9 WSNM decrease with word-line voltage decrease 42
Fig. 3-10 Prior art of word-line under-drive circuits 43
Fig. 3-11 Proposed word-line under-drive circuit 44
Fig. 3-12 I-V_{WL} curve of proposed WLUD circuit 46
Fig. 3-13 V_{WL} comparison across process corners with different temperature of the proposed WLUD circuit. 46
Fig. 3-14 Word-line rising time comparison of this work and prior art. 47
Fig. 3-15 RSNM and WM versus WL voltage level. 48
Fig. 3-16 Column based dynamic V_{CC}. 49
Fig. 3-17 Floating Power Line Write 50
Fig. 3-18 Differential Data-Aware V_{DD} scheme 51
Fig. 3-19 Differential Data-Aware V_{SS} scheme 51
Fig. 3-20 Differential Data-Aware Power-supplied (D2AP) 8T cell 52
Fig. 3-21 Proposed word-line under-drive scheme 53
Fig. 3-22 Proposed word-line under-drive circuit 53
Fig. 3-23 Column Arrangement of DAWA and WLUD 55
Fig. 3-24 equivalent circuit of column half-selected cells 56
Fig. 3-25 Hold SNM with Asymmetric supply voltages 56
Fig. 3-26 Leakage drop waveform 57
Fig. 3-27 corner tracing keeper control 58
Fig. 3-28 gate voltage of keeper at different corner 59
Fig. 3-29 Write waveform 59
Fig. 3-30 Read waveform 60
Fig. 3-31 RSNM improvement of constant V_{WL} and WLUD 61
Fig. 3-32 Read margin and read V_{CC_min} improvement 61
Fig. 3-33 WSNM improvement with DAWA 62
Fig. 3-34 Write Trip Voltage (write margin) improvement with DAWA 62
Fig. 3-35 Write Time and write V_{CC_min} Improvement 63
Fig. 3-36 Test Array Architecture 64
Fig. 3-37 Layout view of 128kb 6T SRAM Macro 66
Fig. 3-38 Power Distribution 67
Fig. 4.1 Finding Vwrite structure with LYA (Low Yield Analysis) DFT 70
Fig. 4.2 DBTA circuit of characterization of Bit transistor scheme 71

1896

Fig. 4.3 The Read/Write Margin measurement scheme of [4-3] ... 72
Fig. 4.4 The storage node wired-out SRAM cell layout of [4-3] ... 72
Fig. 4.5 Read SNM, Butterfly curve WNM and I_W measurement of [4-3] ... 73
Fig. 4.6 1-M bit SRAM transistor NBTI test structure of [4-4] ... 74
Fig. 4.7 Modified cells of [4-4] ... 74
Fig. 4.8 Fast stability analysis of large-scale SRAM arrays and the impact of NBTI degradation. 75
Fig. 4.9 measurement of read current and impact of NBTI ... 75
Fig. 4.10 Block diagram of measurement structure ... 76
Fig. 4.11 current sensing scheme ... 77
Fig. 4.12 Shows another frequency difference detection circuit ... 77
Fig. 4.13 traditional standard 6T SRAM cell ... 78
Fig. 4.14 (a) Basic concept to get Vtrip of inverter (b) Concept of Vtrip equivalent circuit in 6T SRAM (c) Vread Disturb Concept ... 79
Fig. 4.15 (a) Vtrip cell schematic (b) Vtrip cell layout (c) Vtrip equivalent circuit ... 79
Fig. 4.16 (a) Vread cell schematic (b) Vread cell layout (c) Vread equivalent circuit ... 79
Fig. 4.17 (a) Vread cell schematic (b) Vread cell layout (c) Vread equivalent circuit ... 81
Fig. 4.18 Write Margin measurement flow ... 82
Fig. 4.19 (a) Vtrip column and trip condition circuit (b) Vread column and read condition circuit (c) WM column and WM condition ... 83
Fig. 4.20 (a) Vtrip column and trip condition circuit (b) Vread column and read condition circuit (c) WM column and WM condition ... 85
Fig. 4.21 (a) Modified Row direction shift register (b) Modified Column direction shift register 86
Fig. 4.22 Measurement Array Architecture ... 87
Fig. 4.23 ADC Measurement Scheme ... 88
Fig. 4.24 VCO + Counter Based scheme ... 89
Fig. 4.25 Building blocks of voltage measurement scheme ... 90
Fig. 4.26 (a) Single stage of VCO (b) 21 stages VCO ... 90
Fig. 4.27 VCO with counter based measurement scheme concept ... 91
Fig. 4.28 Layout of present 512k Test Macro ... 93
Fig. 4.29 Test Macro Location in Test Chip ... 94

Chapter 1

Introduction

1.1 Background

From the past decades, the trend of IC development perfectly fit the Moor's law : the density of the chip capacity is doubled per 18 month. The performance, complexity, and cost to a chip are promoted with the advanced technology. However, the size of device becomes very small at nano-scaled technology, variation is the major obstacle to chip design. The fabricated out transistor may be different to previous designed value and lead to system functional failure. This could result the degradation to the yield during chip manufacture.

From the predicting of ITRS Roadmap, memory area will be the largest portion to SOC. Static Random Access Memory (SRAM) would dominate the performance, power and area. In recent years, multi-core system and clouding computing gradually occupies an important role to the electronic application. These systems usually need high speed and large volume SRAM to do data processing. This makes the design to high speed SRAM an important issue.

1.2 Motivation and Goals

Nowadays, the Static Random Access Memory (SRAM) occupies an important role to current SOC. However, with the shrinking process and reducing supply voltage, regular SRAM design no longer easily survived at advanced technology node. The most crucial issue is the degradation of the read/write ability to 6T SRAM. Since it is

hard to be fixed at the process phase, we would like to sue for the circuit technique solution. For SRAM, the read/write ability are contradictory conditions. So we must use different circuit technique to separately solve the read and write problem. In thesis, we focus on the circuit technique which could improve the functionality to 6T SRAM. Even at low supply voltage, we wish 6T SRAM could also have good yield with these circuit technique. Besides, we also want to investigate the real characteristic to 6T SRAM Array to provide more actual information rather than only from simulation. We also focus on the implementation of measurement circuit. With this implemented measurement structure, we can more precisely characterize several effects to the 6T SRAM cell.

1.3 Thesis Organization

In the following of the thesis, Chapter 2 discuss the basic operation to traditional 6T SRAM and its design issue. The reliability issue would be also mentioned in this chapter. Chapter 3 demonstrates "55nm 6T SRAM with Variation Tolerant Word-Line Under-Drive (WLUD) and Data-Aware Write Assist (DAWA)" design. In this Chapter, Variation Tolerant WLUD is proposed to improve the read stability of 6T SRAM. Data-Aware Write-Assist circuit skill is proposed to improve the 6T cell write ability. The design issue would be also discussed in Chapter 3. Chapter 3 shows an "An array-based all-digital on-chip circuit for 6T SRAM Stability measurement" design to characterize the static noise margin to 6T SRAM. 6T cell is modified to perform trip voltage to 6T cell (V_{trip}) and read-disturb voltage (V_{read}) measurement. Write margin measurement scheme is also implemented in this design. For analog voltage measurement, a VCO based measurement scheme is modified to fit our measurement scheme. Chapter 5 makes a conclusion to this thesis.

Chapter 2

Overview of the operation and the design of 6T SRAM

2.1 Memory Family and SRAM

For current System on Chip (SOC), memory always account for the majority of transistors. As a result, memory always dominates the overall performance of one system. For current CMOS chips, Random Access Memory (RAM) array always be build in the integrated system (IC) as the storage media. RAMs is always associated with volatile memory which the storage data would loss if the power turn off. In contrast, there are several nonvolatile types of memory that are ROM, flash, DISK, etc. In the recent years, flash memory has a great growth at the market share and its application.

RAM is often applied to the Embedded System due to its higher access speed than other memory family. Volatile RAM could roughly divided into two categories : Dynamic RAM (DRAM) and Static RAM (SRAM). DRAM has slightly slower access speed than SRAM, but it has more compact density to save data with much smaller area. DRAM is currently the major storage device of most SOC. SRAM has the highest operation speed which could reach several hundred Mega Hertz or even Giga Hertz, so it is always embedded into system as cache memory. For the past decade, 6-transistor (6T) SRAM is always the main stream to the cache memories.

However, as the technology node goes the nano-scaled, the design of 6T SRAM

array faces several challenges. Process variation and leakage issue may cover the designed margin that may fail the read/write operation. This means we need to use extra circuit technique to counter the process variation and the leakage. In order to properly apply the circuit technique to 6T SRAM, we have to first discuss the basic operation of it.

2.2 6T SRAM

2.2.1Introduction of 6T SRAM

The most common used cell for SRAM is traditional 6T cell (show in Fig.2-1). Basically, this cell contains six transistors →two pull-up PMOS (M1&M4), two pull-down NMOS (M2&M5), and two pass-gate NMOS (M3&M6), so this cell is called 6"T" (Transistor) cell. In this cell, two inverter (M1-M2 and M4-M5) forms a cross couple pair latch that could strongly lock the storage value at logic "0" or "1", because the Voltage Transfer Curve (VTC) of this cell has only two stable point (actually there are 3 cross point, but one is the meta-stable point and hard to be exist) (Fig.2-3). And NMOS M3 and M6 could be seen as port to access the storage data of this cell. And one pair of Bit-lines (BL & BLB) are connected to the source node of pass-gate NMOS. This Bit-Line pair is used to pass in or pass out data. And one signal Word-Line (WL) is the activation signal of this cell.

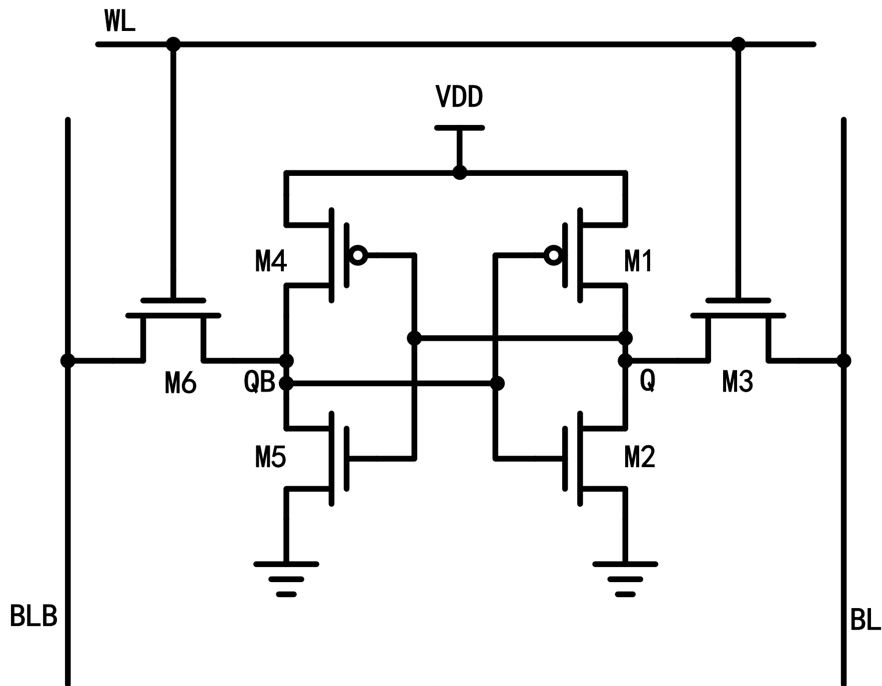


Fig. 2-1 Traditional 6T SRAM

2.2.2Read Operation and read disturb voltage of 6T SRAM

Actually, all of the BL of 6T cell are pre-charged to VDD at standby mode which means there is no operation. Both of BL and BLB are held at high voltage before the read operation. Once the WL is activate, the pass-gate of this cell is turn on for accessing storage value. The BL of storage "0" side will be pulled close to the ground level however the storage "1" side will still keep at high level. Because of this operation, the storage value of this cell is passed to the BL. And we just need extra read circuit to get the value on the BL that we can get exact storage value.

However, for 6T cell, there is a disturb that could destroy the original storage value. From Fig.1, we assume Q is originally storage "0". Once the WL goes high, Q would be a divided voltage of M3 and M2 rather than ground voltage. This non-zero read-disturb voltage at storage node is close to the trip voltage of the other side inverter. If this read-disturb voltage is higher than the trip voltage of the opposite side

inverter, than destructive read operation is performed and the cell storage value would be flipped.

2.2.3 Write Operation of 6T SRAM

Write is an opposite operation to read for a 6T traditional cell. Before WL is selected to perform write operation, the Data-In should be prepared at the BL & BLB. For example, we want to write logic "0" into a cell, we must set the BL at ground voltage and BLB at high voltage before WL goes high. Once the WL goes to the high level, the storage value Q will quickly drop below the trip voltage of the opposite inverter that would flip the original storage value. And the regenerative process of the central latch would again hold the Q at strong ground voltage.

But there is also write fail problem for write operation. If the voltage at Q when the WL turns on is not lower than the trig voltage of opposite inverter, then write fail.

2.2.4 Read Static Noise Margin (RSNM) and Write Static Noise Margin (WSNM)

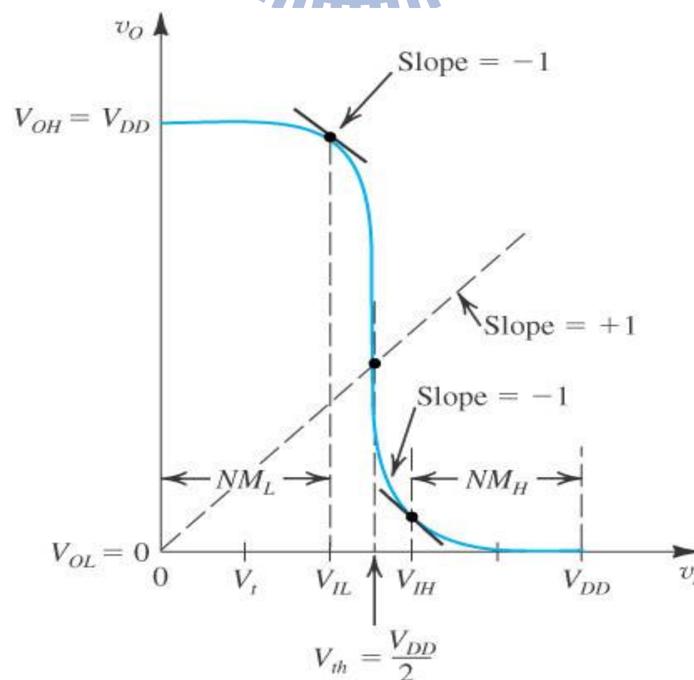


Fig. 2-2 Voltage Transfer Curve (VTC) of CMOS inverter [2-1]

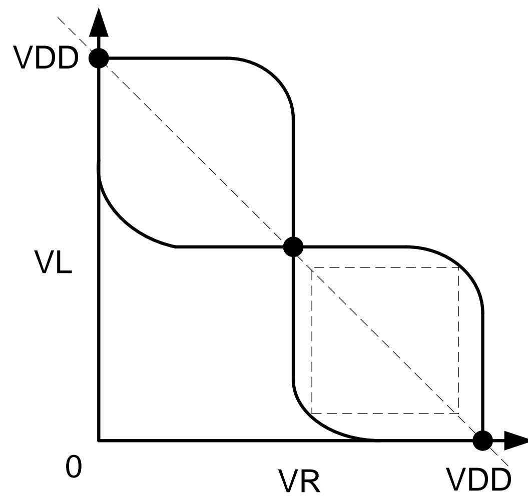


Fig. 2-3 Butterfly curve of HSNM

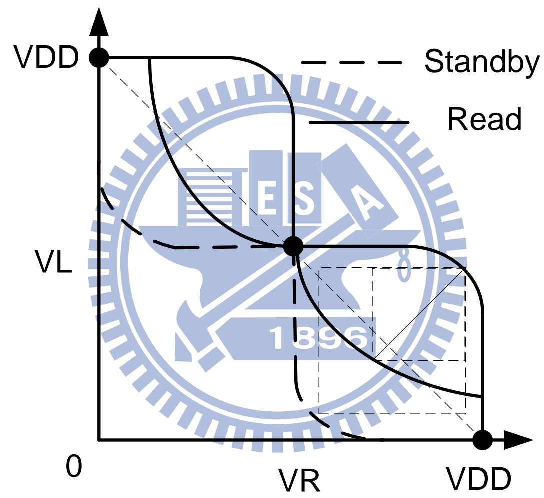


Fig. 2-4 The RSNM butterfly curve

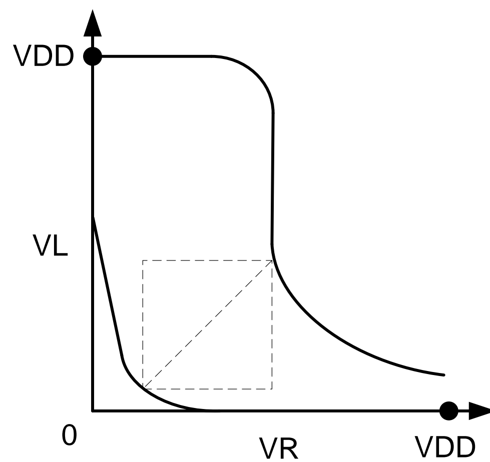


Fig. 2-5 The WSNM butterfly curve

To evaluate the read ability or write ability, the static noise margin is usually a good approach. And static noise margin could be get from the Voltage Transfer Curve (VTC) which we always called as "butterfly curve". For an inverter, as the input voltage arises, the output won't decrease as the increased input voltage (Fig. 2-2). Instead, the output voltage would have a sharp transition at a specific voltage which we always called as trig-voltage of this inverter. (Actually, in definition, the trig-voltage is the input voltage that result the output voltage at half of the supply voltage.). However, to a cross pair latch, the output of one inverter is the input of the other. So we can get another VTC from this relationship but switch the axis. We can overlap these two curves to get a "butterfly-like" curve. Fig2-3 shows the butterfly curve of HSNM or cross coupled pair latch, because the hold mode operation of 6T SRAM is exactly a cross coupled pair latch. In this butterfly curve, we can see it has two "wings". The common used definition of SNM is the largest tolerable square at the either wing and takes the smaller one.

For read operation, both of the pass-gate NMOS are turn-on for accessing data. However, the high voltage level BLs and the turn-on pass-gate will decrease the VTC slope of the inverter of the storage "0" side. But the VTC of the other side remains. As shown in Fig 2-4, the Butterfly curve of read mode is smaller than the read mode, and the largest square in either wing that is the RSNM is smaller than the Hold mode or Standby mode. As mentioned before, there exist the read disturb voltage that decrease the noise margin. So, the minimum RSNM could directly defined as the voltage difference from trig voltage to read disturb voltage. Once the RSNM equals to 0 or less than zero which in other words that there is no any square could locate at each wing, that the destructive read operation occurs and read fail.

Fig 2-5 shows the butterfly curve of write operation. Since one BL would be set at ground and the other BL would be set at high voltage, the butterfly curve would be

wide open and only have one intersect point. This means we can successfully write in a data because we have only one stable point. And these two VTC should be get more farer. By definition, like Read and Hold SNM does, we also fine a largest square in this curve. But rather than find it in wings, we locate a largest square to define how much voltage distance between it which is called WSNM. If the butterfly curve of write mode of 6T cell has no WSNM that means there is more than one interest point, write operation will fail.

2.2.5Write margin

There is another definition which we called as Write Margin (WM) to evaluate the write performance of write ability rather than WSNM. First we both set the BLs are at high level, and sweep down one BL form VDD to ground. At some BL voltage the cell storage value will be suddenly flip, and the BL voltage at this moment is defined as WM. (Fig.2-6)

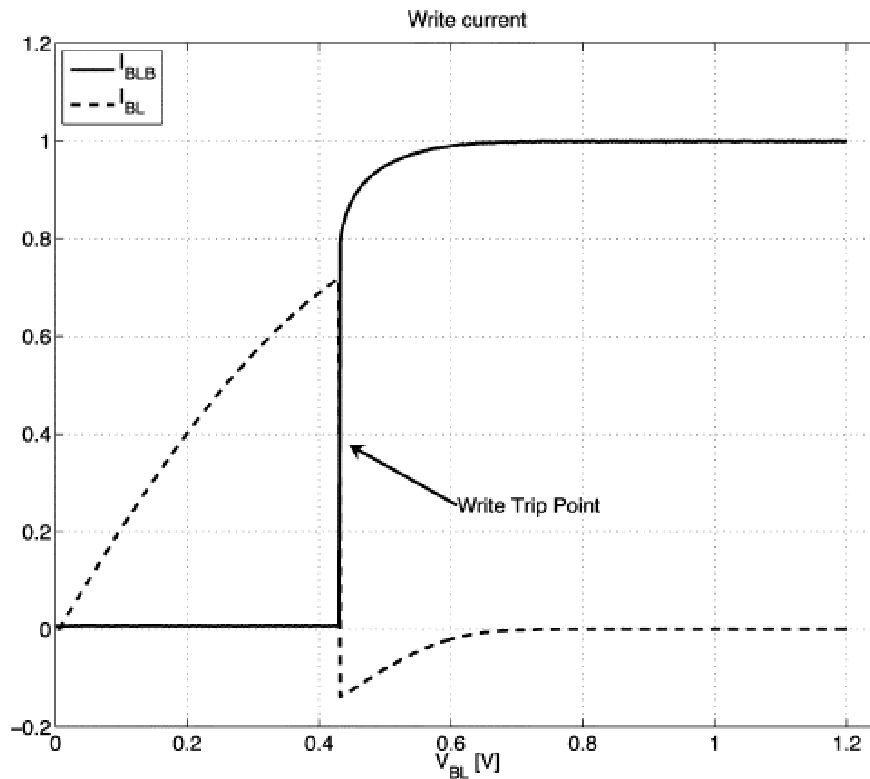


Fig. 2-6 Write trip voltage and Write Margin (WM)[2-3]

2.2.6The size of 6T SRAM

In order to keep the read stability, the pull-down NMOS should be stronger than the pass-gate NMOS. To maintain the write ability, the pass-gate NMOS should be stronger than pull-up PMOS. However to ensure the stability at standby mode, the pull-up PMOS cannot be too weak compares to pull-down NMOS. As a result, the size of each transistor in 6T SRAM cell is specific designed to maximize the read, hold stability and the write ability.

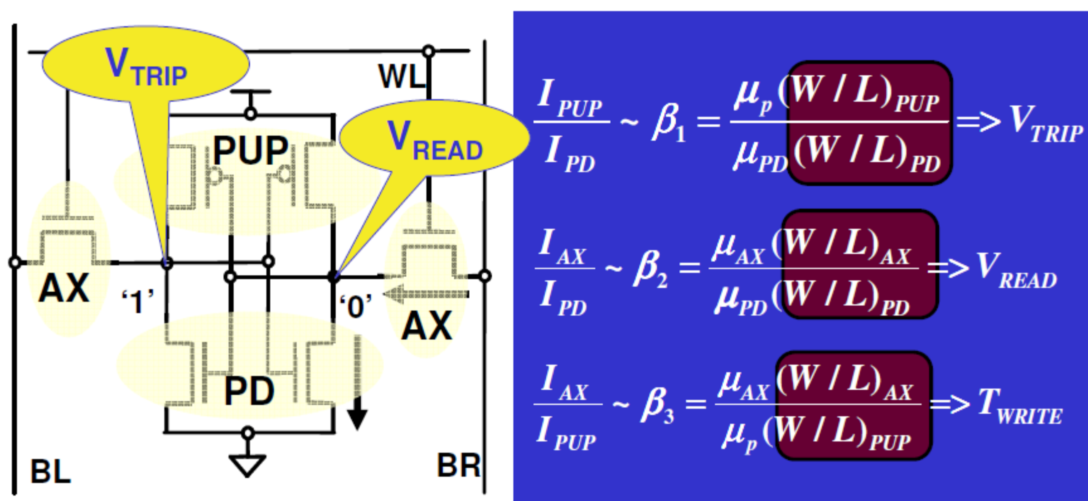


Fig. 2-7 Transistor size ratio in 6T SRAM

2.3 SRAM Array architecture

2.3.1Memory Array and half-select problem

Because we always need a large volume memory, so we have to have a large amount of memory cells. However, these cells are usually formed into an array to save the area and to easy access. Traditionally, all of the SRAM cell are put together which we called SRAM array and the peripheral circuit such as decoder and Sense Amplifier are placed next to the SRAM to control those operation. For a 6T SRAM array, Word-Line (WL) is usually a row direction signal that control pass-gate of all the row direction cells. And Bit-Line (BL) is column direction signal line that could

pass in or out the data. So if we want to select one cell for operation, both the WL and BL must active. The intersect cell is the selected one, and the operation is depend on the BL.

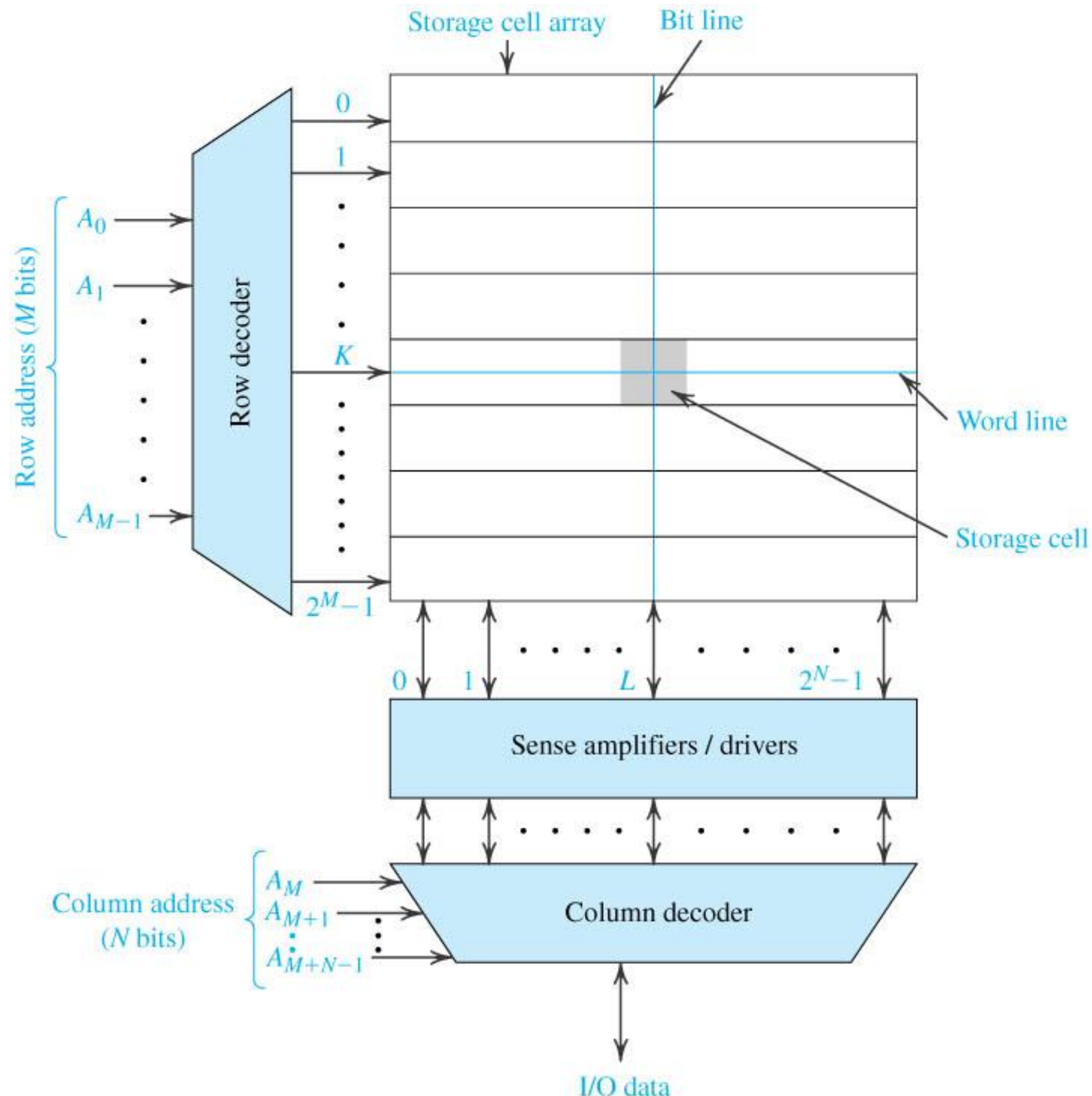


Fig. 2-8 Array architecture of an 2^Nx2^M memory array[2-1]

However, due to the array formation, there would have the half-selected problem. The most common seem problem to 6T SRAM Array is the half-selected read disturb problem. Since the word-line is row-based, all the pass-gate along this direction would be turn-on even those un-selected cells. The BL of these unselected cells are held at high because there are in hold mode, so all of the half-selected cells are performing the half-selected read operation. But this half-selected problem is

unwanted, because this means all of the other standby cells of this row would suffer disturb that would affect the storage value.

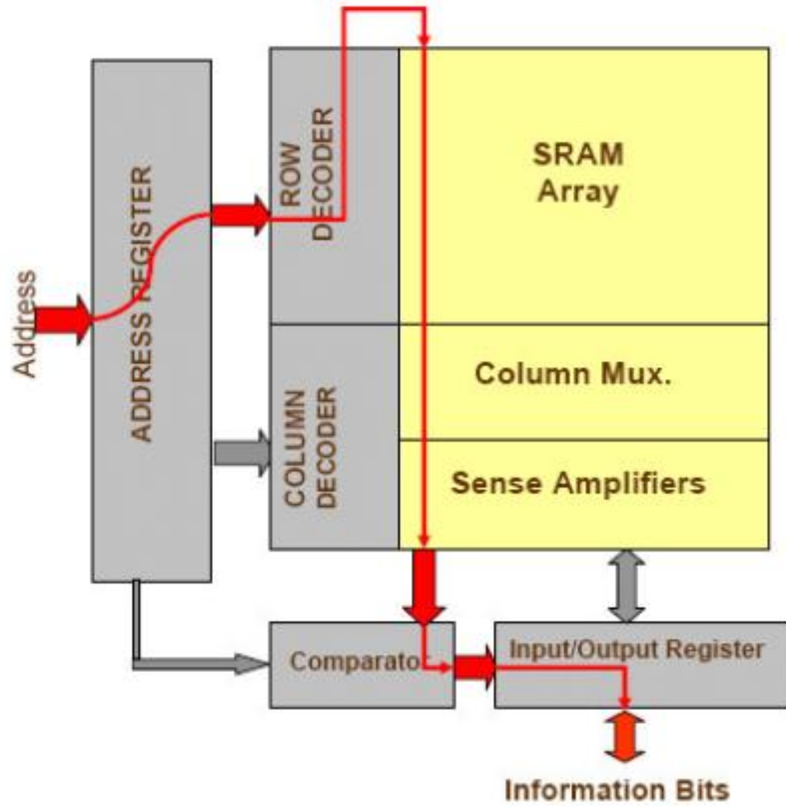


Fig. 2-9 SRAM critical path [2-5]

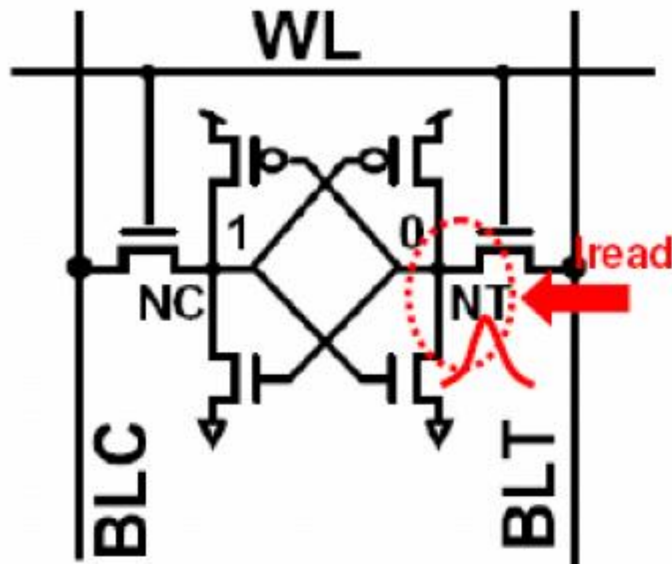


Fig. 2-10 Half-selected Read Disturb Voltage [2-4]

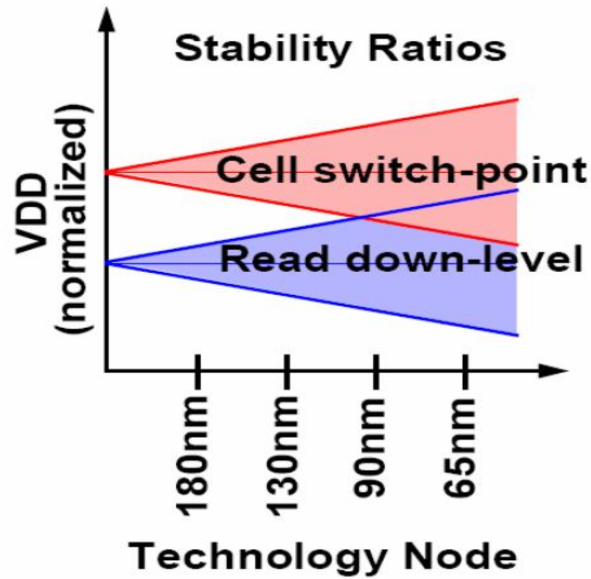


Fig. 2-11 Stability ratio [2-4]

2.3.2Decoder

For a memory, the address pins are usually the input which is used to select the specific cell to read/write data. So the decoder circuit is needed to get the address information. Decoder can be divided into Row Decoder and Column Decoder. Row Decoder (Which is also called X-Decoder or WL Decoder) is used to decode the selected WL. The Column Decoder is used to decode the select column to perform operation.

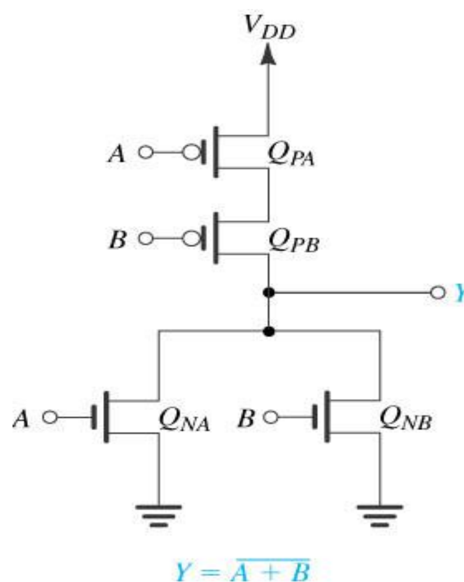


Fig. 2-12 CMOS Logic Circuits [2-1]

The decoder always composed by simple logics. But the decoding time occupies a not small portion of the overall performance of SRAM, so the decoder must as fast as possible. However, decoder always has a large amount branches because we usually has more than hundred BLs and WLs, loadings of parasitic capacitance usually the main factor that slow down the decoding time. So the solution form architecture is to divide the decoder into the Pre-decoder and decoders. The operation Pre-decoders could be overlap at the last operation cycle to save the time. So the result of the Pre-decoder could be get at the early cycle stage. Once the clock triggered, decoder could process to save time.

Another method to boost the decoding time is to use variety logic family. The most common used logic family is CMOS, but the parasitic capacitance always slow down the speed. Domino or Dynamic logic family could reduce the parasitic capacitance to increase operation speed. And the pre-charge operation of one-side could allow one signal transmission path is almost 0. But the design of these dynamic circuits must be careful to their margin and secondary effect issue such as charge sharing, etc. Pass-Transistor Logic family is also favored in decoder design.

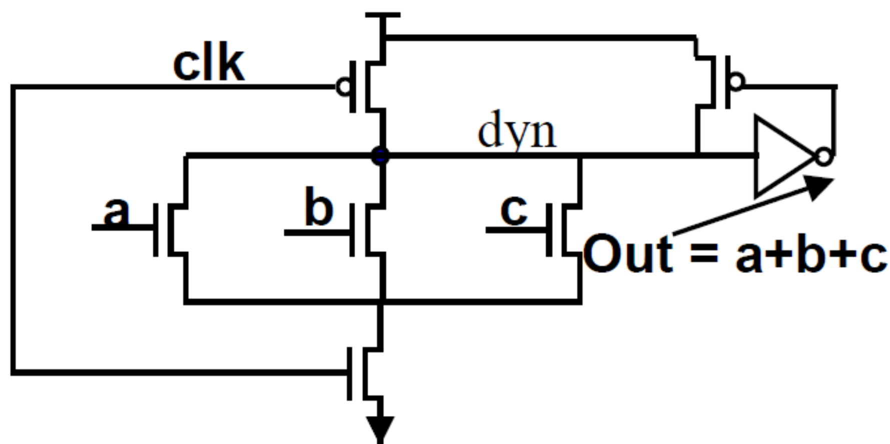


Fig. 2-13 Domino Logics [2-6]

2.3.3Sense scheme and column architecture

Since we have pass the storage value on the BL at read operation, we need to extra circuit to sense whether the storage value is “1” or “0”. Basically, there are two kind of sensing scheme : one is differential sensing and the other one is large signal sensing.

Differential sensing or small signal sensing is the traditional sensing scheme. The basic idea of small signal is to sense the voltage difference between BLs than amplify it, so that we can get the logic “0” or “1” from the amplified signal of differential sense amplifier. The traditional differential sense amplifier is a cross-coupled latch with two access transistor that is quite like a 6T SRAM cell. But the size tuning and design issue is different from 6T cell. During read operation, WL is active and the storage value start to pass to the Bit-Line, either BL or BLB will start to go low. Once the voltage difference of BLs reaches a value that differential sense amplifier could operate, then a sense amplifier enable (SAE) signal would be activated to turn on the sense amplifier. Then the low going Bit-Line would be amplified to fully ground voltage and the logic “0” or “1” could be get.

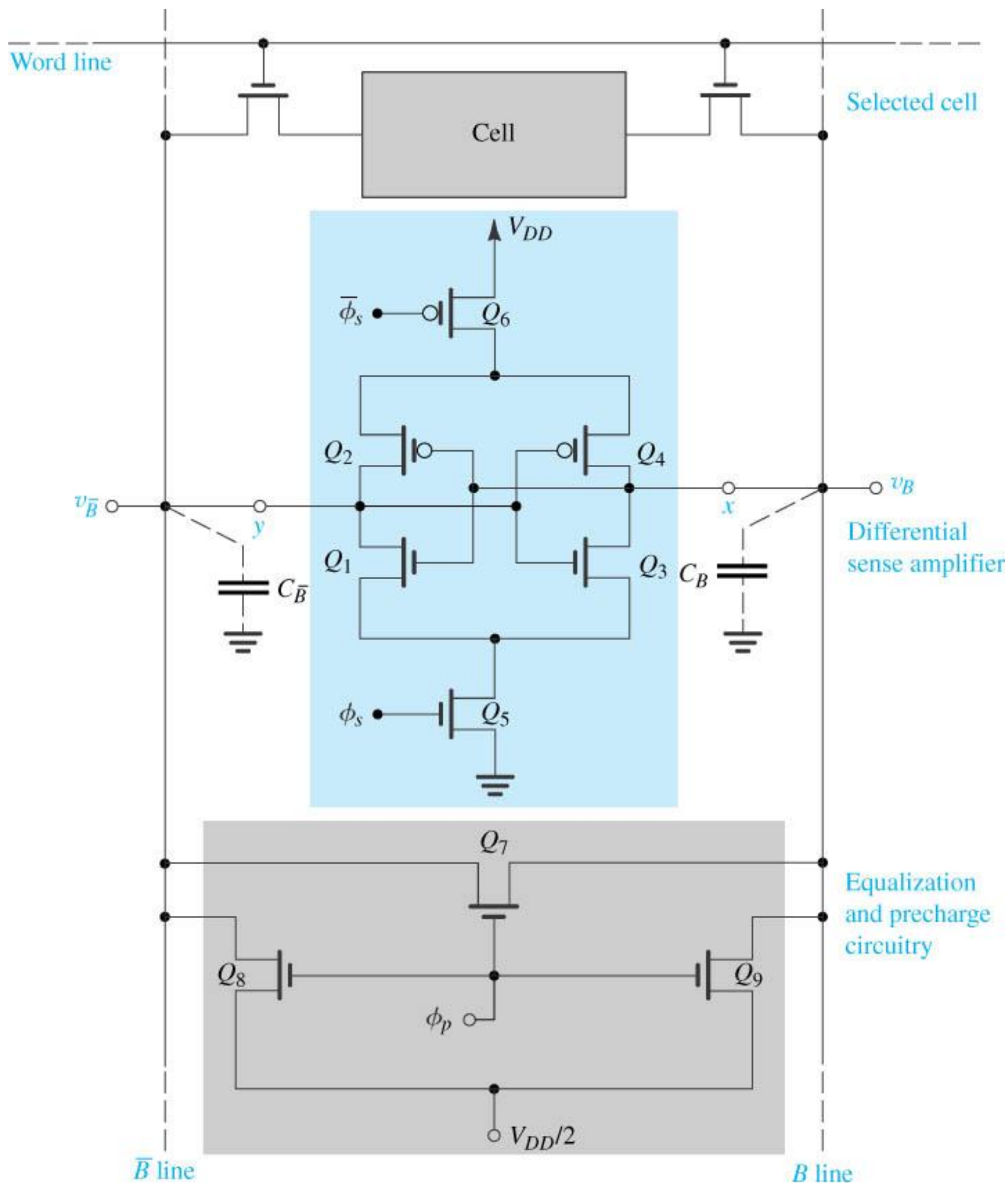


Fig. 2-14 Differential sense amplifier [2-1]

The differential sensing scheme usually adopt with long bit-line structure which means there are many cells along the Bit-line (usually more than hundred cells). Because large number of cells result heavy loading, the storage value fully pass to BLs is a slow process in the long Bit-line structure. The Differential Sense Amplifier is to sense the small difference of BLs, so the use small signal scheme could arise the operation speed greatly.

As the process scales down, however, leakage becomes a crucial issue. Long Bit-line structure first faces the problems because there are large amount devices

along Bit-line. Even the WL is turn off, the large amount leakage current may flip the standby cells that would result retention fail. And the small signal sensing scheme also may fail to sense the data at Bit-line due to the un-neglectable leakage current. So the Bit-Line length must be decreased or adopt the short Bit-line Structure. For the Short Bit-Line scheme, the Bit-Line length is about 8, 16, or 32 cells which is quite smaller than long Bit-Line scheme. Because the lesser cell, the total leakage current could be reduced.

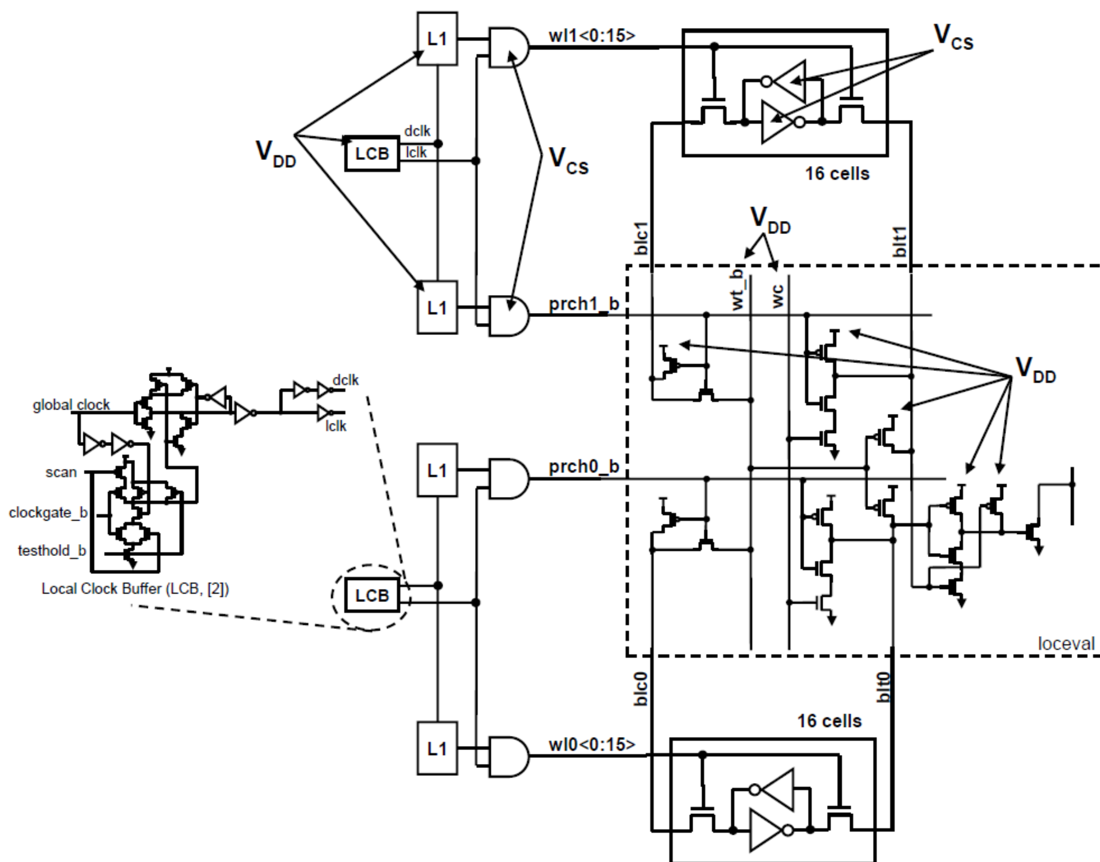


Fig. 2-15 Large signal sensing scheme of IBM Cell processor [2-7]

Rather than adopt the differential sensing scheme, large signal sensing scheme is usually used with short Bit-Line structure. Large signal sensing is directly use one transistor or one inverter to detect the logic signal on the BLs. Once the Bit-line goes lower than the trig voltage of the sensing transistor or sensing inverter, the logic signal would be passed out to the other circuit. And the shorter Bit-Line length reduces the

BL loading, so the signal pass to the BL is faster. Since the large sensing is almost single ended sensing, so the leakage impaction is almost half to the differential sensing scheme where both the Bit-line are connect to the Sense Amplifier. And the implementation of single transistor or inverter is much simple than the Sense Amplifier which is always exhausted time to find the optimized gain. But the short Bit-line structure may have much more area overhead than the long Bit-line structure. Because the long Bit-line structure may only need one Sense Amplifier per each column, short Bit-line structure needs several sensing devices in each column.

2.4 Variation Issue

2.4.1 Global variation and Local variation

For the real chip implementation, there must exist the variation factor that makes the real outcome device is different from that we the original designed. For CMOS process, these variations can be totally reflected on the threshold voltage V_T. Since the VT is different from the designed value, the current drive ability of transistors would deviate and the leakage may be larger. These could even result the functional fail or more power consumption. The existence of variation is various. The most common reason are the lithography variation at each process, the dopant number fluctuation, and the line edge roughness, etc. In a deeper sight to view the variation formation, we can separate the variation into the Global and Local. Global variation which is also called the inter-die variation means the variation between die to die. And the Local variation that also called intra-die variation means the variation of transistors at the same die. The total variation to one transistor could be the superposition of Global variation and Local variation. 2-16 shows the case for threshold voltage variation (δV_t), the variation can be expressed as [8]：

$$\delta V_t = \Delta V_{t-GLOBAL} + \delta V_{t-LOCAL}$$

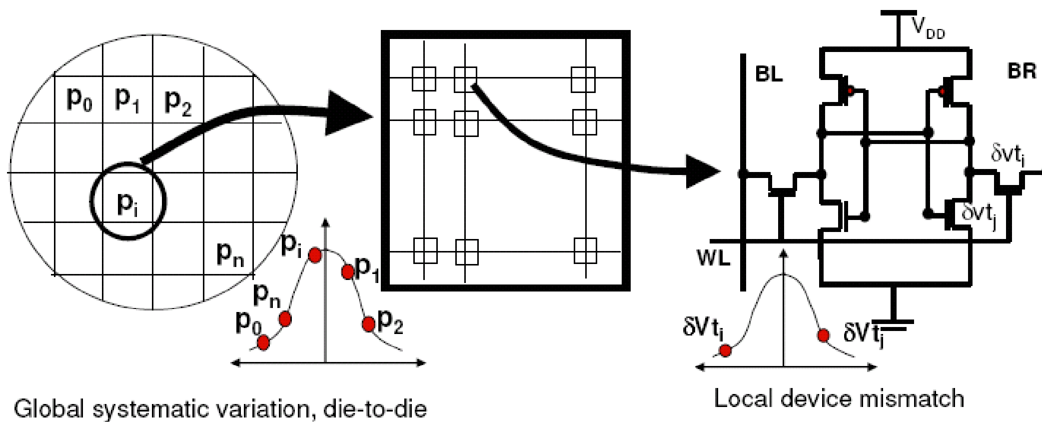


Fig. 2-16 Global variation and Local variation of threshold voltage [2-6]

The Global variation is also the corner variation. At each time we make the real chip, the lithography, machine settings, temperature, doping concentration would be different. Different die may suffer the different process condition, so the characteristic would vary inter dies. And due to the formation reason, the Global variation is systematic and predictable.

But for the two same transistors in the same die but different location, the dopant number may fluctuate. This would result the VT of transistor with the same size in the same die not the same. This we called local variation. However, the dopant number is random and follows the statistical Gaussian distribution, Local variation is unpredictable and hard to be controlled.

For the current nanometer technology, the channel length is short and the dopant number is a few. Variation becomes a most critical issue at the design phase. Shorter channel length makes the Line Edge Roughness variation become more serious. And a few number dopant let one missing dopant result a large percent dopant concentration change. Random Dopant Fluctuation is almost the main factor of Local variation and become more serious at modern process.

2.4.2 Variation to 6T SRAM

The variation issue also affects the characteristic of 6T SRAM. Due to the condense and compact environment, 6T SRAM cell may suffer much serious variation than logic devices. Both of the Global and Local variation would greatly affect cells

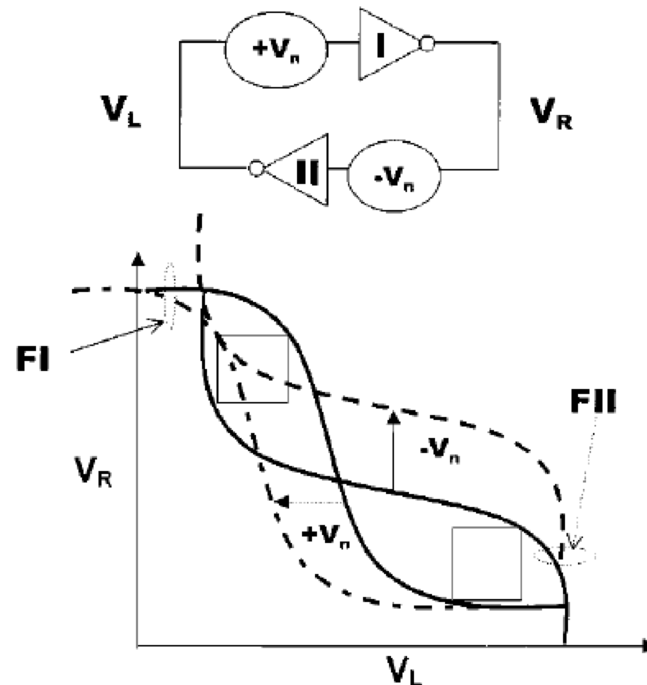


Fig. 2-17 Static Noise Margin definition and the applied noise or variation that makes the SNM equals zero [2-9]

We first discuss the read operation. The Global variation locate at PSNF corner would make the 6T SRAM has the smallest RSNM. At PSNF, the trig voltage of inverter would be lower than normal, so that the noise margin between trig voltage and read disturb voltage would be smaller. However, there would be also the Local variation that is random. The variation of each transistor would be random and independent. From the Fig. we can get that the worst case cell of read operation. High VT M4 and Low VT M5 forms higher read-disturb voltage and High VT M1 and Low VT M2 forms smaller trig voltage of inverter. The existence of Local variation could result a worst case for read.

And for the write operation, the worst case of Global corner is at PFNS corner. Since the pass-gate NMOS is weak, the write trip voltage is much higher than nominal. Consider the Local variation, the worst case cell of write is shown in Fig. Low VT M3 and High VT M5 result higher write trip voltage and High VT M1 and Low VT M2 make a lower trig voltage. So that the WM would be lesser if we both consider these variation issues.

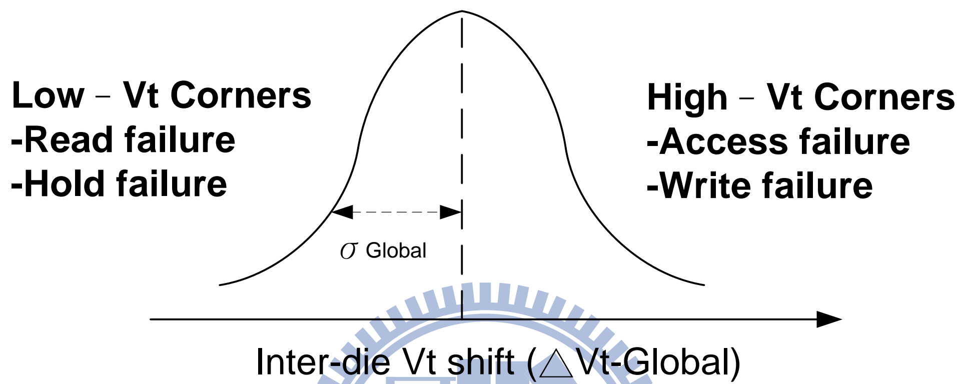
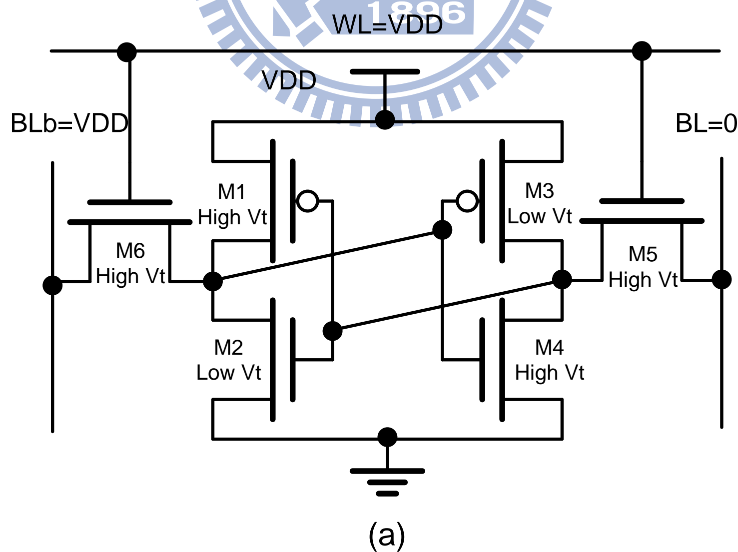


Fig. 2-18 The effect of the inter-die Vt variation [2-8]



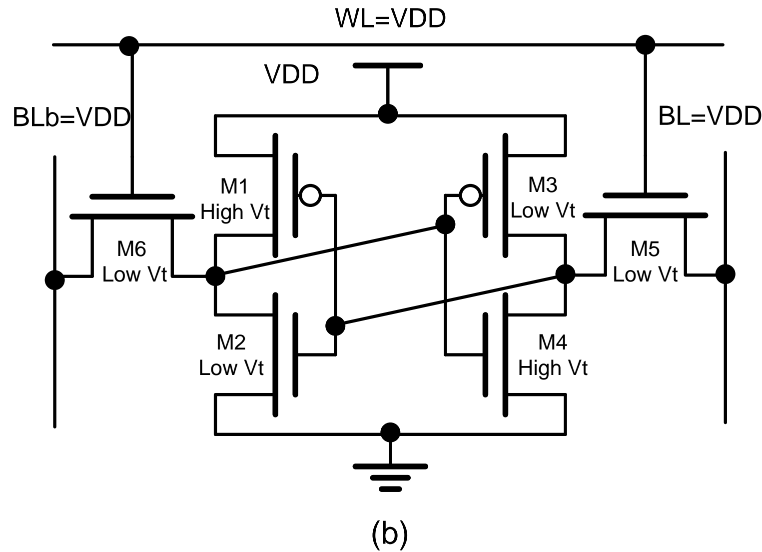


Fig. 2-19 The effect of local variation, (a) Write mode worse case, and (b) Read mode worse case

2.5 Design methodology of modern 6T SRAM

So far we have discuss several critical issue of 6T SRAM such as read-disturb voltage, half selected problem, leakage, variation, and SNM issue, etc. Unfortunately, these critical issues become more serious than before, and the read/write ability of 6T SRAM has greatly degraded. In order to increase the read/write ability, we need extra circuits to improve. In fact, the most common methodology to improve the read/write ability is three: 1) Dual Supply voltage 2) Lower Bit-Line level 3) Negative Bit-Lin. These issues would be discussed below.

2.5.1 Dual Supply Voltage

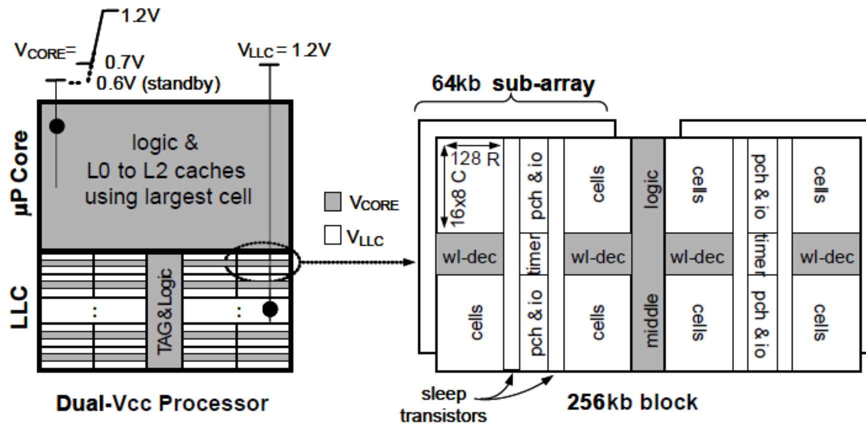


Fig. 2-20 Dual VCC Design of [2-10]

In order to improve the read/write ability, one straight forward concept is to change the supply voltage. If we want to increase the read ability or the noise margin, we could raise the supply voltage of the cell supply or reduce the cell ground voltage to a negative voltage. Or if we want to increase the write ability, we can reduce the supply of the cell. In order to achieve these, the simplest method is to use the second supply voltage directly, so that the supply voltage of cell and logic circuit are separated. However, the implementation of second power supply is expensive. For a general embedded system, the second supply voltage means we have to use another set of power converter, power regulator, and battery, etc. This will greatly increase the cost of chip. So the much practical method is still use single supply, but generates second supply voltage with circuits.

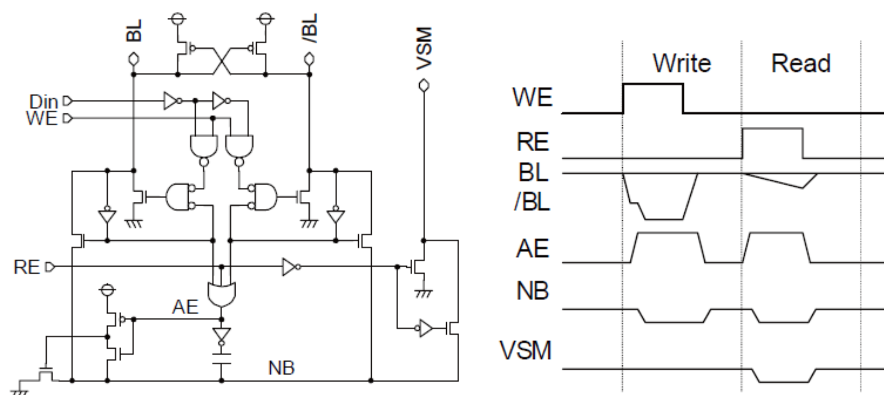


Fig. 2-21 Negative VSS scheme [2-11]

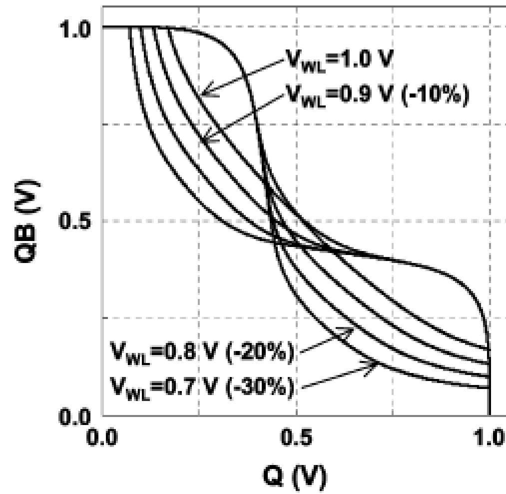
The most common circuit technique of adjusting the voltage is to use the boost

capacitance. By using the capacitance which is previously precharged, we can "add or subtract" amount of voltage on the original voltage. So that, this "pseudo second supply" could achieve almost the same effect of using second supply. For example, we could only boost the cell supply voltage higher during read cycle or boost the GND to a negative voltage to increase the read ability. This implementation is just need to realize capacitance and has precise timing control.

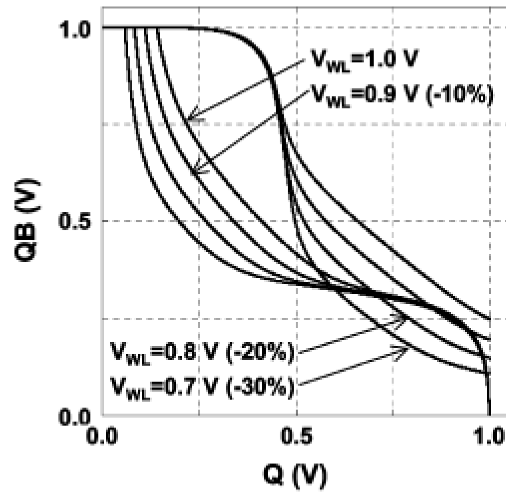
In order to increase the write ability, we would like to reduce the cell supply voltage. But it is difficult to reduce the cell supply voltage with the boost skill. The alternative method is to float cell supply voltage during write. Since the cell power is floating during write cycle, the write ability would increase with the decreasing floating power. Basically the write ability would be greatly improved. However, the floating cell supply would affect the stability of cell and reduce the regenerative process speed of the cross- coupled latch. One compromised method is only floating once side power during write. This could greatly improve the stability and strengthen the regenerative process speed. Another approach is to use keeper. Rather than directly floating supply node, a weak keeper is added to replace the strong power supply. The write ability could also be improved because of the weaker supply. But the keeper size has to be fine tuned to maximize the write ability improvement.

2.5.2 Dynamic Word-Line voltage

Another method is to control the WL voltage. If we raise the WL voltage, the pass-gate would be strengthened. The write ability and read speed would be improved. However, higher WL voltage would increase the read disturb voltage of 6T cell which would decrease the RSNM. More important, all of the cells along the selected WL except of the selected one are performing half-selected read operation. Insufficient RSNM could result a serious retention fail to those half-selected cells.



(a)



(b)

Fig. 2-22 RSNM improvement with lower V_{WL}, (b)WSNM decade with lower V_{WL}[2-12]

Another direction is to reduce the WL voltage or also called word-line under drive. Although the lower WL voltage weakened the pass-gate and read speed and write ability are weakened. But with this word-line under drive skill, RSNM of all cells are improved. This could at least prevent the destructive read problem of insufficient RSNM. Moreover, we can mix the skill of the dynamic supply skill to improve the read speed and write ability. The mix use of word-line under drive and dynamic supply skill has no contradiction and could improve RSNM, read speed, write ability at the same time. So many benefit could be get by the mix use, we the

word-line under drive has more and more become the main stream.

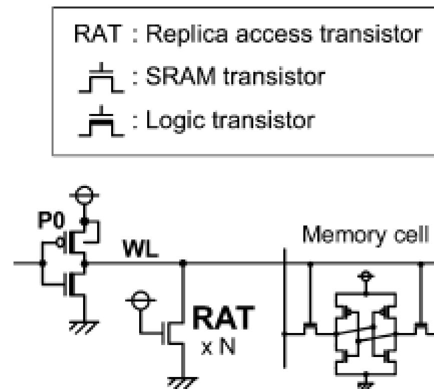
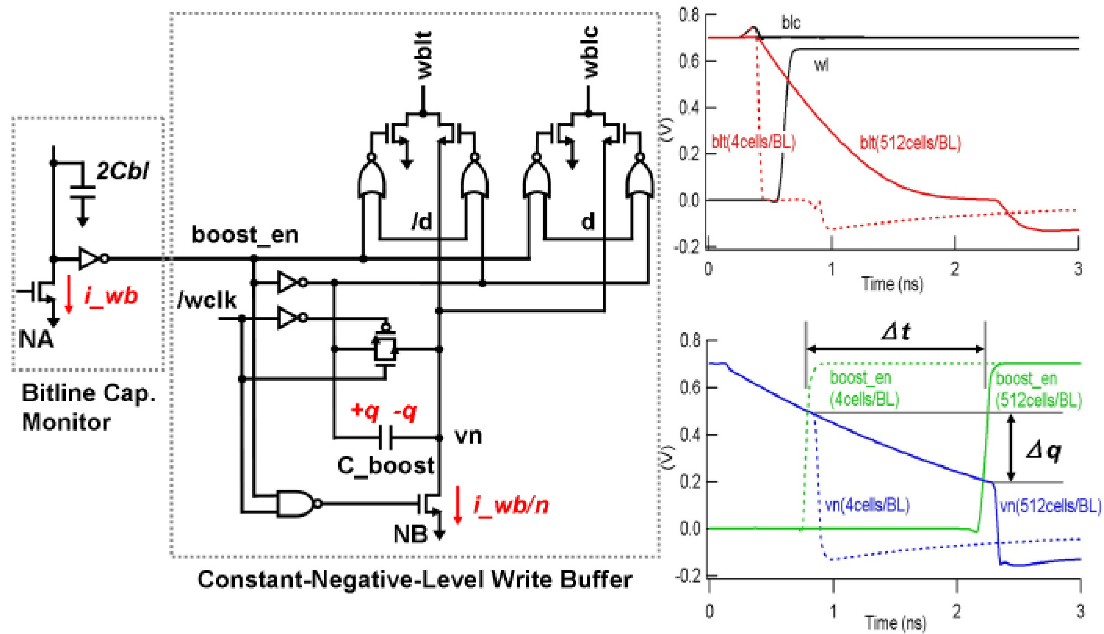


Fig. 2-23 Traditional method of word-line under-drive [2-12]

2.5.3 Negative Bit-Line Level

Negative Bit-Line (NBL) is a skill to improve the write ability. The basic idea is to generate a negative voltage on Bit-Line with the boosting skill we mention previously pass into the cell storage node to improve write performance. Much lower voltage could help trigger the cross coupled pair latch to flip its value of cells which have insufficient write margin. And the stability could also maintain if the NBL is used. From many previous works, NBL could greatly improve the write ability and write margin. However, it is necessary to implementation a large capacitance if we want to use NBL skill. A large and good performance capacitance is difficult to get if we use the MOS capacitance. The large area overhead would be a serious problem of NBL. Besides, NBL circuit needs a quite precise timing control to maximize the improvement of write. Basically, the negative voltage should be generated before WL turned on. But the timing is quite hard to be controlled and to be a crucial issue to NBL.



$$\Delta q = \Delta t * i_wb/n = (\Delta Cbl * Vdd / i_wb) * (i_wb/n) \propto \Delta Cbl$$

Fig. 2-24 Constant Negative Bit-line of [2-13]

2.6 Reliability issue and monitoring structure of SRAM Array

2.6.1 BTI effect

As the supply voltage decrease with the scaled down technology, the Hot Carrier Injection (HCI) has become a minor effect to MOSFET degradation. However, the Bias Temperature Instability (BTI, especially the Negative Bias Temperature Instability, NBTI) rules the major effect that degrades the MOSFET reliability. Several works has done deep investigation to NBTI mechanism and its impact to digital circuits [15]-[22].

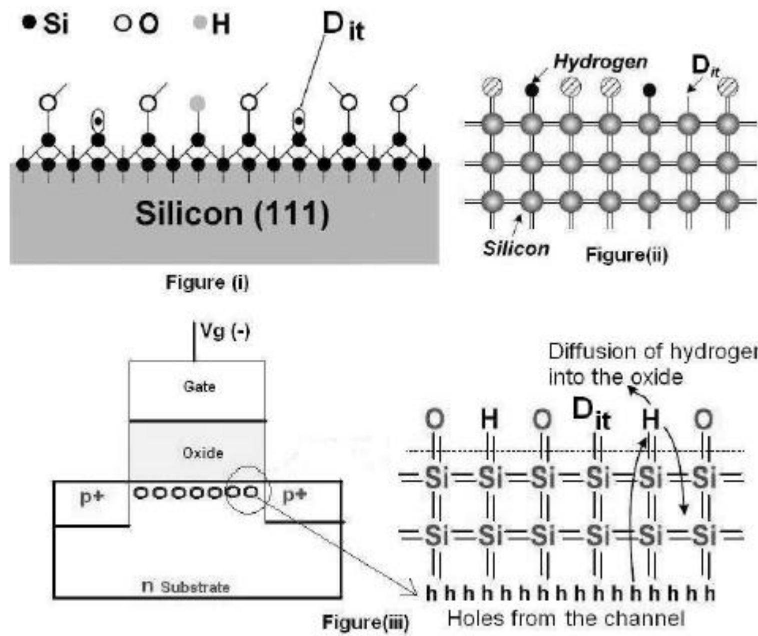


Fig. 2-25 The generation of interface traps a PMOS transistor is biased in inversion.[2-20]

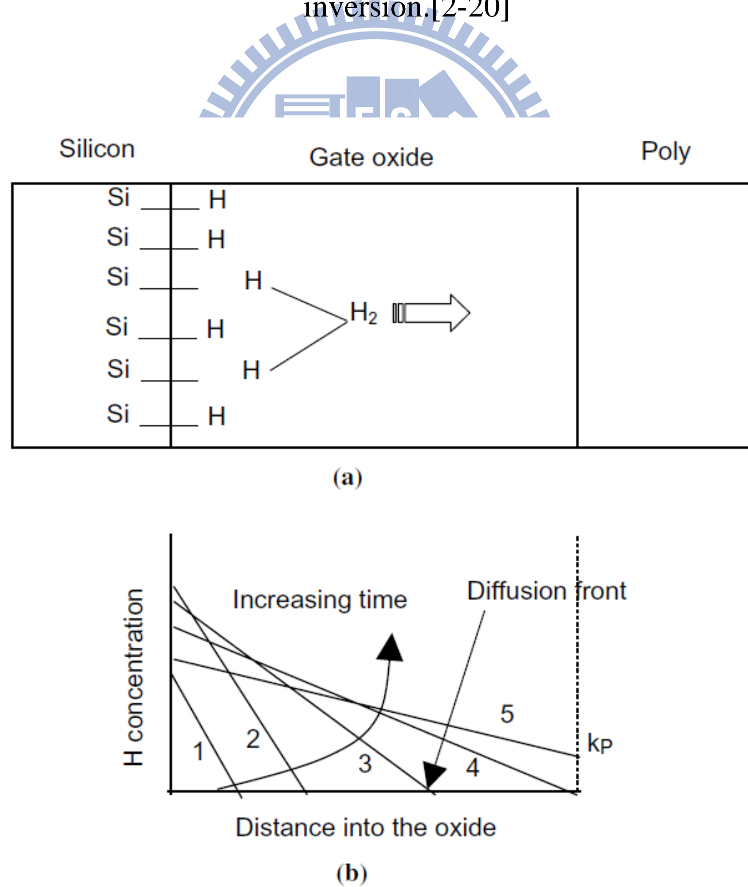


Fig. 2-26 Schematic description of the reaction-diffusion model [2-15]

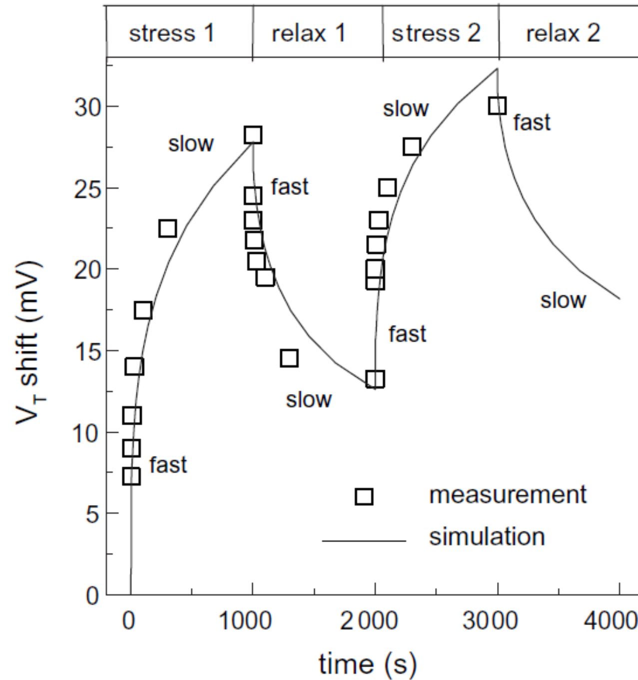


Fig. 2-27 Time dependent effect of stress and relaxation mechanism [2-15]

Several formation mechanism of NBTI is proposed, and the most acceptable one is the Reaction-Diffusion (R-D) model (Fig.2-25 & Fig.2-26). For standard silicon MOSFET process, especially PMOS, there would generate the Si-H bond at the interface between bulk silicon and oxide. Once the PMOS suffers the Negative-Bias, the interface Si-H bond would break and format traps and the formation H_2, D_{it}, etc would diffusion away from the interface of bulk silicon and oxide. As a result, the effective threshold voltage (V_T) would degrade according this process. This NBTI effect is an time-dependent effect and getting worse especially at high temperature, so we called NB"T"I. Due to the time-dependent characteristic to NBTI, it would result the CMOS Digital Circuit has the finite lifetime. Fortunately, there is a recovery mechanism to NBTI effect only if we induce the positive-bias. Besides, the recovery mechanism is also time-dependent (Fig. 2-27).

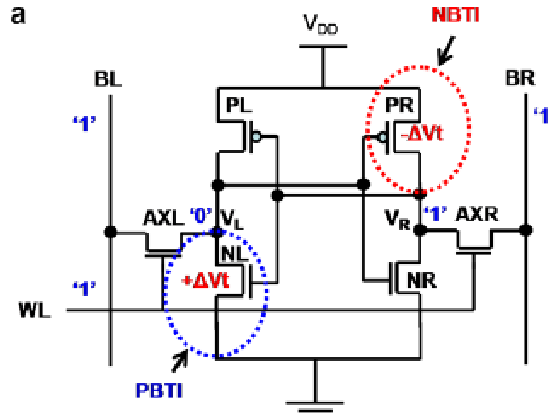


Fig. 2-28 NBTI and PBTI impact to traditional 6T cell

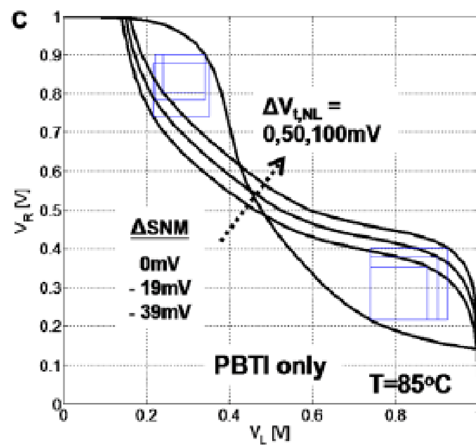


Fig. 2-29 RSNM degradation due to NBTI

NBTI would be also happened to the traditional cell (PBTI would be also happened to High-K metal gate NMOS) (Fig.2-28). Due to NBTI, the absolute V_T to PMOS becomes larger than nominal value. This could result a lower V_{trip} voltage to cell inverter. As a result, the RSNM would be getting smaller with time that degrades the read stability to traditional 6T SRAM cell. Due to the half select read problem, the NBTI degradation would be a critical issue that affects the reliability and lifetime to 6T SRAM.

2.6.2 Monitoring Structures

Reliability is an important issue as the speed, and power. For electronic devices, the characteristic of transistor would change with time due to some specific physical phenomenon such HCI, NBTI, etc. This would result the performance degradation or

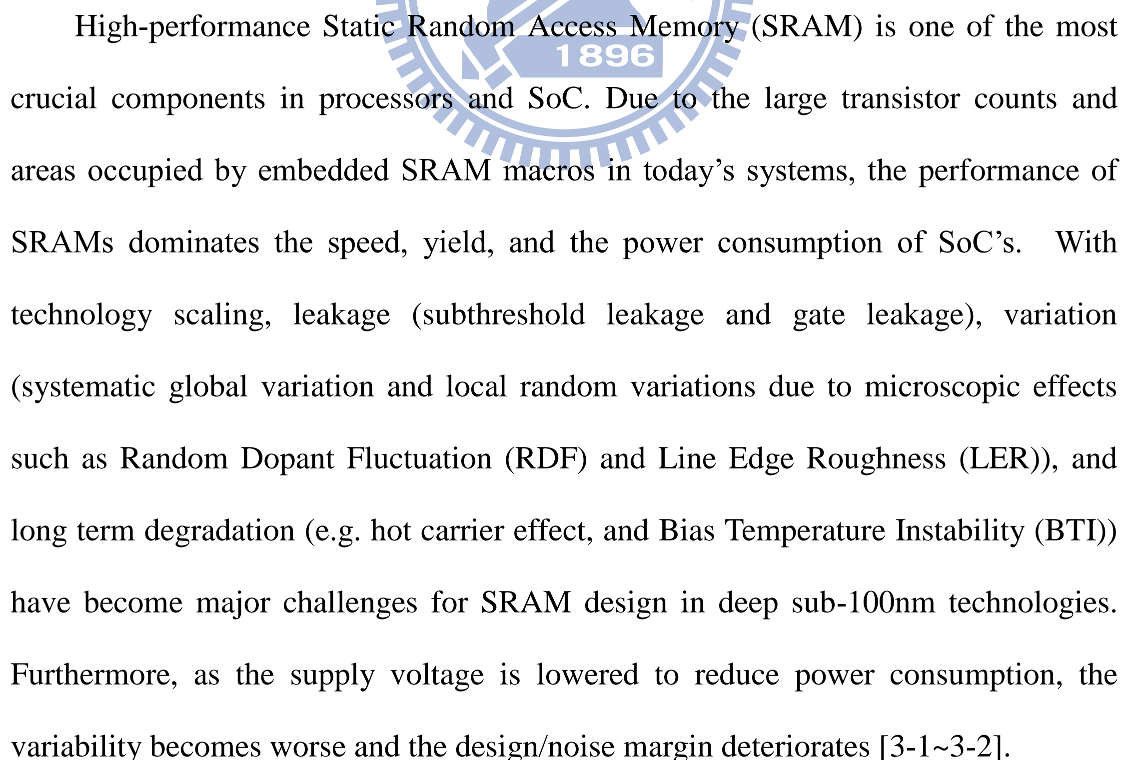
even function fail, and every chips has a finite life time. From a view point of circuit designer, to understand and characterize the formation of degradation and realize the quality and quantity of physical phenomenon is necessary. If designer can get the information about the transistor characteristic and the degradation from the real chip, it would be a great help to understand real physical phenomenon and add some reliability improvement circuit at early design stage.

The simplest and most straight forward way is to build a monitoring structure on chip to get the information what we want. The monitoring circuit could be designed for measure variety characteristics, like statistical distribution and degradation of transistor VT, the NBTI and PBTI degradation to transistor, HCI of transistor, etc. For a SRAM, we could build circuit to measure the SNM or the PBI impaction to SRAM cell. By analyzing the collected data, we can find out the statistical distribution of these parameters. The means value, standard deviation of it could tell us the real information about the real chip.

Several types monitoring structure has been proposed to measure different parameters. Some previous work would be noticed in Chapter 4. Besides, An Array Based monitoring structure to measure the static noise margin is proposed and would be discussed in Chapter 4 in this essay. An “Embedded SRAM Ring Oscillator for BTI Measurement” is also proposed at the work [2-23].

Chapter 3

Design of one 128kb traditional 6T SRAM with Word-line Under-Drive skill and Data-Aware Write Assist skill



3.1 Introduction

High-performance Static Random Access Memory (SRAM) is one of the most crucial components in processors and SoC. Due to the large transistor counts and areas occupied by embedded SRAM macros in today's systems, the performance of SRAMs dominates the speed, yield, and the power consumption of SoC's. With technology scaling, leakage (subthreshold leakage and gate leakage), variation (systematic global variation and local random variations due to microscopic effects such as Random Dopant Fluctuation (RDF) and Line Edge Roughness (LER)), and long term degradation (e.g. hot carrier effect, and Bias Temperature Instability (BTI)) have become major challenges for SRAM design in deep sub-100nm technologies. Furthermore, as the supply voltage is lowered to reduce power consumption, the variability becomes worse and the design/noise margin deteriorates [3-1~3-2].

For conventional 6T SRAM, The cell stability, Static Noise Margin (SNM), and V_{min} are limited by leakage, variation, and supply voltage in the physical domain; and

by the conflicting Read/Write requirements and cell disturbs in the design domain. In the design and optimization of 6T SRAM, the procedure typically starts by reducing Read-disturb to maintain adequate RSNM to make sure the cell won't flip during Read operation. Circuit techniques such as "unclamped" or "weakly-clamped" Local Bit-Lines (LBL) where all LBL's are released ("unclamped" or "wealy clamped" by a diode drop) in active mode have been widely adopted to reduce Half-Select disturb. In high-end systems which can afford dual supply, higher supply voltage (VCS) can be used for cell array to enhance cell stability and Read performance, while BL is precharged to the lower (VDD) supply to reduce Read/Half-Select disturb. For cost-performance applications and portable/ handheld devices where the cost, form factor, and weight considerations dictate the use of single supply, Read-assist scheme such as Word-Line Under-Drive (WLUD) (also called Suppressed Word-Line (SWL)) can be employed to mitigate Read disturb and Half-Select disturb[3-4]~[3-10]. Notice that the WLUD scheme is essentially a "Poor Man's Dual Supply" scheme to generate a lower supply level. With WLUD, the Read performance and Write-ability degrade. The Read performance can be recovered by circuit techniques such as transient negative cell VGND [3-12]. The Write-ability, Write Margin (WM), and Write performance are typically recovered/enhanced by Floating power-line Write (where the cell supply is left floating during Write) [3-15]~[3-20], dynamic Read/Write supply (where the cell supply is lowered during Write), or Negative Bit-Line (NBL, where the low-going BL is transiently pulled down below GND during Write) techniques [3-11]~[3-14].

In this work, we proposed a variation-tolerant Word-Line Under-Drive (WLUD) circuit technique to improve the RSNM. Each word-line has one WLUD circuit and the level of the selected word-line is lowered. Our WLUD circuit comprises of one PMOS which tracks the WL driver pull-up PMOS, and one NMOS which tracks the

cell access NMOS. This resistor-less implementation reduces the area overhead, and eliminates the variation due to the sheet resistance variation of the resistor which can easily be ±30% in sub-100nm technologies [3-11]~[3-16].

The proposed WLUD scheme provides excellent tracking capability across process corners. The Write-ability and Write performance are enhanced by a novel variation-tolerant Data-Aware Write-Assist (DAWA) scheme where the virtual cell supply nodes for the left- and right-half cells on a column are control by separate power-switch/keeper pairs. The power-switch on the low-going BL side for the selected column is turned off to allow the corresponding virtual supply node voltage to go down to enhance Write-ability, Write Margin, and Write performance. The keeper remains "On" to maintain a proper (low) virtual cell supply level, thus ensuring adequate Hold SNM for half-selected cells on the selected column. Switching only the half-cell virtual supply improves the supply switching speed and reduces the switching power/noise by half. The scheme also offers fast WL rise time as the WL drivers is driven by full VDD [3-4], and die-to-die programming capability can be easily implemented if desired. The power-switch control logic is simple, and thus requires much smaller area overhead compared with Negative Bit-Line (NBL) Write-assist scheme which requires large boosting capacitor and more complicated timing control [3-4][3-17]~[3-20]. The half-cell virtual supply switching is controlled/initiated by the low-going BL during Write operation, thus tolerant to PVT variation and V_T scatter. The DAWA scheme adds minimum loading to the BL (only 1 C_{gate} per BL), hence no degradation to the BL sensing speed. Implemented in a 55nm 128Kb 6T SRAM, the WLUD and DAWA schemes together require 18% area overhead compared with the base 6T SRAM design. The chip can operate across wide voltage range from 1.2V to 0.6V, with operating frequency of 940MHz@1.0V and 25°C.

3.2 Concept of word-line under-drive

Just as said previously, word-line under-drive could increase the RSNM of 6T SRAM cells. Due to every half-selected cells along row direction are performing a read operation once some word-line is activated, the increase of read stability could do great help to resist the variations which could result function fail. The detail of how the word-line under-drive skill increase the RSNM we will be discussed later. Many previous works has proposed to achieve the requirement of word-line under-drive, however, they always has some drawbacks when implementation. We propose a simple word-line under-drive circuit which uses full CMOS transistor. And the proposed circuit has better corner tracking ability and previous works.

3.3 Previous work of word-line under-drive

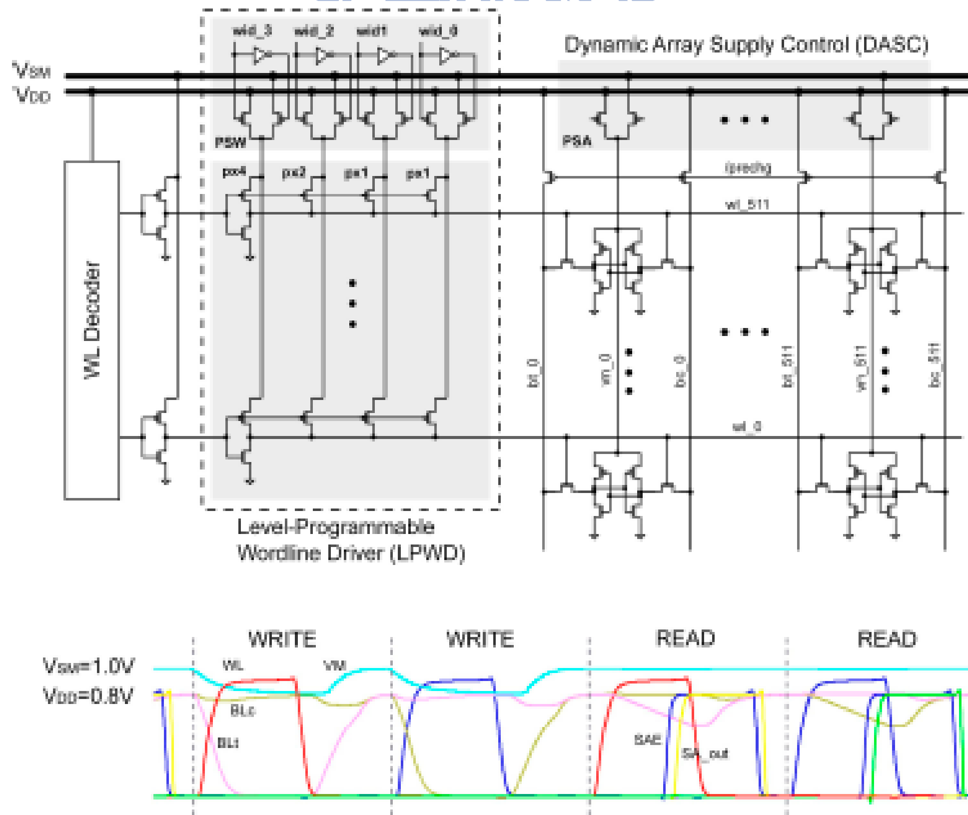


Fig. 3-1 "Level Programmable Word-Line Driver and Dynamic Array Supply Control in Dual-Supply SRAM scheme [3-13]"

Fig. 3-1 shows previous work of “Level Programmable Word-Line Driver and Dynamic Array Supply Control in Dual-Supply SRAM scheme”. In this scheme, VSM is higher than VDD, so the programmable word-line level would determined by the ratio of PMOS of VSM and PMOS VDD. And we can either switch cell supply higher to improve read or switch cell supply lower to improve write. But this method needs dual supply which is an expensive cost when realization. And there is nothing mechanism to prevent the process variation.

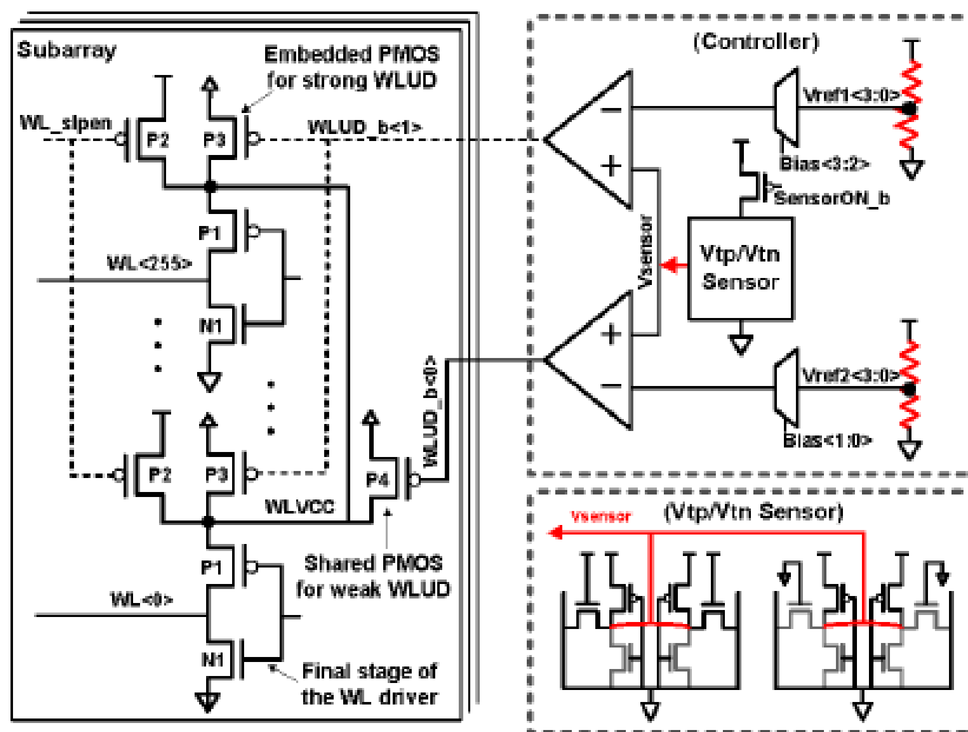


Fig. 3-2 Adaptive Word-Line Under-Drive (WLUD) Scheme

Fig. 3-2 shows the “Adaptive Word-Line Under-Drive Scheme” method. This scheme eliminates the high cost of dual supply. By controlling the gate voltage of PMOS, word-line level would be separated into two levels. But this implementation is quite complicate and needs a lot of extra component such as operational amplifier (OPA). The area cost of this scheme is considerable. The much practical implementation method to under drive the word-line level is the Read-Assist Circuit

(RAC) in [3-14].

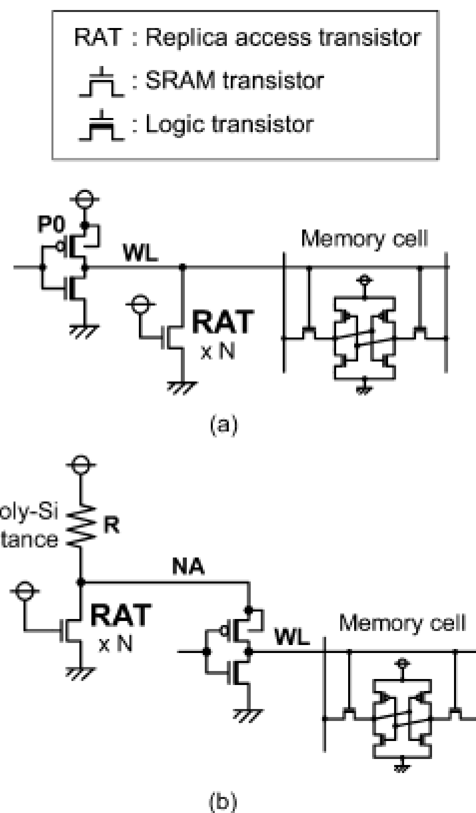


Fig. 3-3 Conventional version (b) resistor added version

Fig. 3-3 [3-16] shows two version of this simple word-line under drive circuits. Both of these two are single supply design. Fig.3-3 (a) is the conventional version and only use single transistor. One NMOS is used with drain connect to the word-line and gate connect to the supply. This allows the word-line voltage to be lower when word-line is active by the forming divide voltage between P0 and RAT NMOS. This implementation is quite simple and even without other control signal. The RAT NMOS can track the pass-gate NMOS to adjust the word-line level around corners.

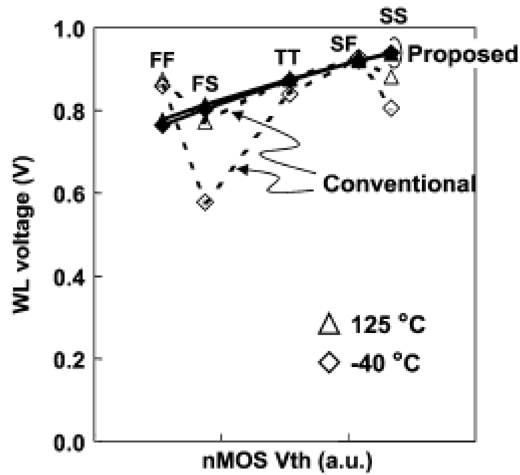


Fig. 3-4 Simulated word-line voltage of Fig. 3-3

But as shows in Fig. 3-4, this single NMOS implementation encounters large temperate variations. Especially at NFPS corner, we can see that from Fig. 3-4 that the word-line voltage drops to a quite low level. In order to fix this, previous work of Fig. 3-3 (a) is proposed to fix temperature variation. Fig.3-3 (b) shows the circuits of the resistor added implementation. Rather than directly connect to the word-line, this method controls the supply by using resistor and NMOS to form divide voltage.

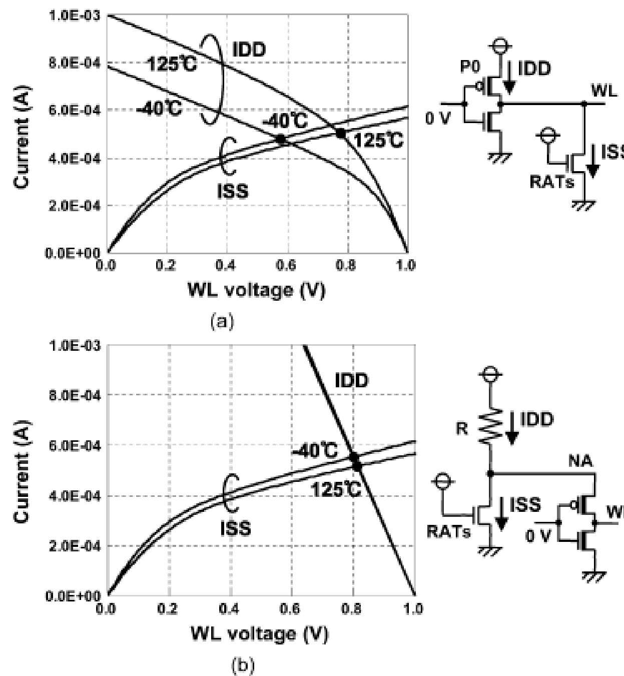


Fig. 3-5 The I-V curve of two RACs, and shows the temperature dependence

From the I-V curve shows in Fig. 3-5, we can get that this implementation could reduce the temperature dependence due to the current-voltage relationship of resistor is stable to temperature. Fig. 3-4 also shows this method could achieve stable word-line level around temperature. However there exists the resistor variation that would lead the variation to the divided voltage. And the result word-line level would also suffer this fluctuation. Besides, the decrease of the word-line driver supply would decrease the rising time of word-line activation. This could lead the performance degradation.

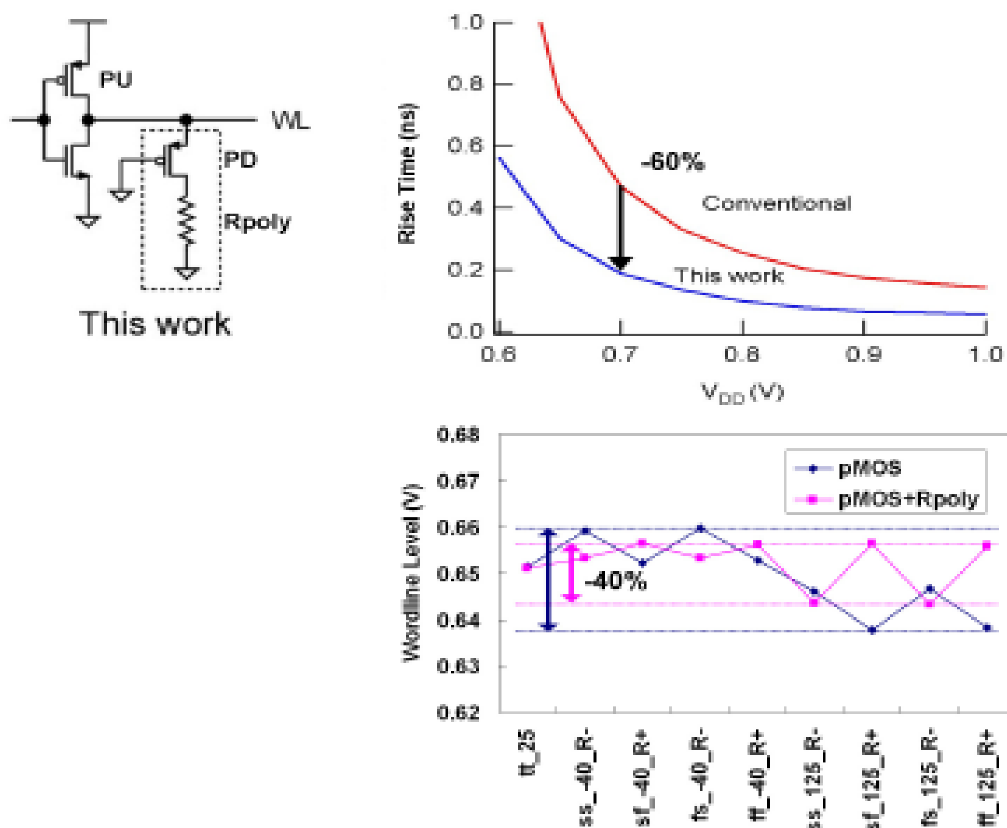


Fig. 3-6 Single-Supply WLUD using Replica Discharging Transistor

Fig. 3-6 [3-4] shows another word-line under-drive scheme proposed from Toshiba. A simple PMOS and resistor is series connected to the WL. And the PMOS is used to track the PU and the POLY resistor is used to suppress the temperature variation. From the simulation result of this work, we can see that the WL rising time

has great improvement than the previous works. And the WL level variation could be suppressed at certain level. However, the resistor may have variation that will cause extra deviation to the WL voltage.

3.4 Proposed word-line under-drive skill

3.4.1RSNM improvement and WM decrease with word-line under-drive

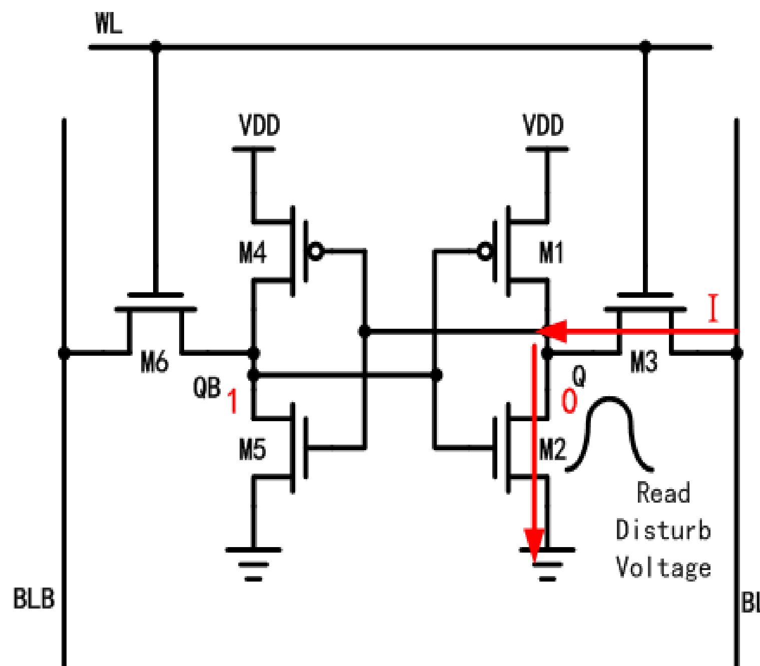


Fig. 3-7　Standard 6T SRAM in Read Mode

For a 6T SRAM cell shown in Fig.3-7, a read disturb voltage would be generated at the storage node once the word-line is turn-on for read. This read disturb voltage would decrease the noise margin of cell during read and even result a destructive read failure if the read static noise margin is smaller than zero. According to the architecture of memory array, if one column is selected to do read/write operation, the other half selected cells in the row direction is performing a read operation. Those half-selected cells may fail to retain its storage value due to this disturbed read. With the unsystematic local variation, the RSNM of cells may degrade more, and the half-selected problem would become more seriously. As a result, we have to have

higher priority to fix this read failure than write failure if we want to achieve higher yield chip.

The simplest method to alleviate this read failure is to reduce the read disturb voltage. This read disturb voltage can be seem as the divided voltage between pass-gate (M3) and pull-down NMOS (M2), and we can decrease the read disturb voltage by decreasing the current drive ability of pass-gate (M3). As a result, the lower current drive ability is similar to a larger resistance which will force the divided voltage smaller. Fig.3-8 shows the RSNM increase with the low-going WL voltage.

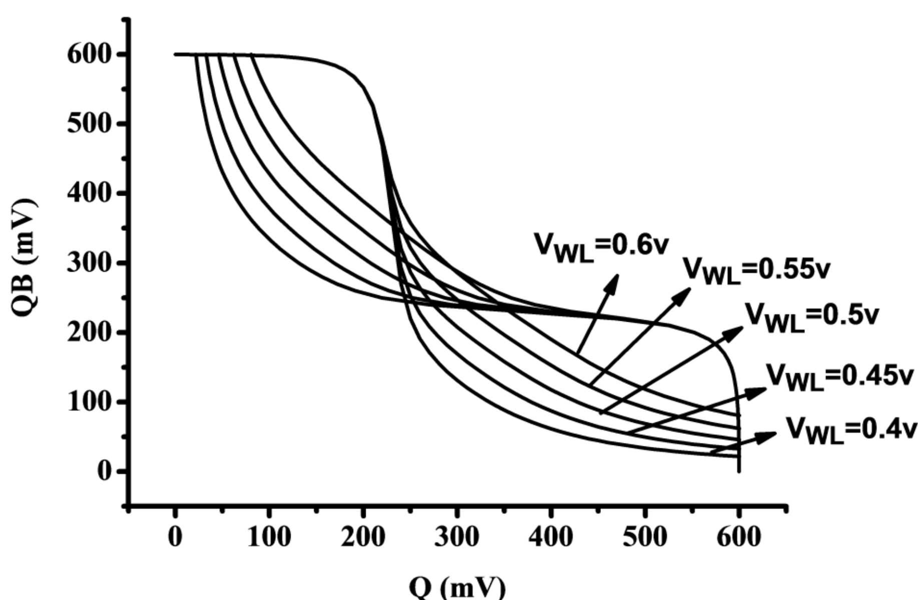


Fig. 3-8 RSNM increase with word-line voltage decrease

However, word-line under-drive increases the RSNM but reduce the write ability. Because the current drive ability of pass-gate is weakened, cell will become harder to pull the storage node into another state. As shown in Fig.3-9, WSNM reduces with WL decrease. Once the WL drop to a certain value, the butterfly curve of cell no longer has only one root. More than one root of butterfly curve means the cell will write fail. So, we must keep certain write ability when we increase RSNM with word-line under-drive.

Besides, the weaker strength of pass-gate will also degrade the read speed. The

chosen of WL voltage is an important issue.

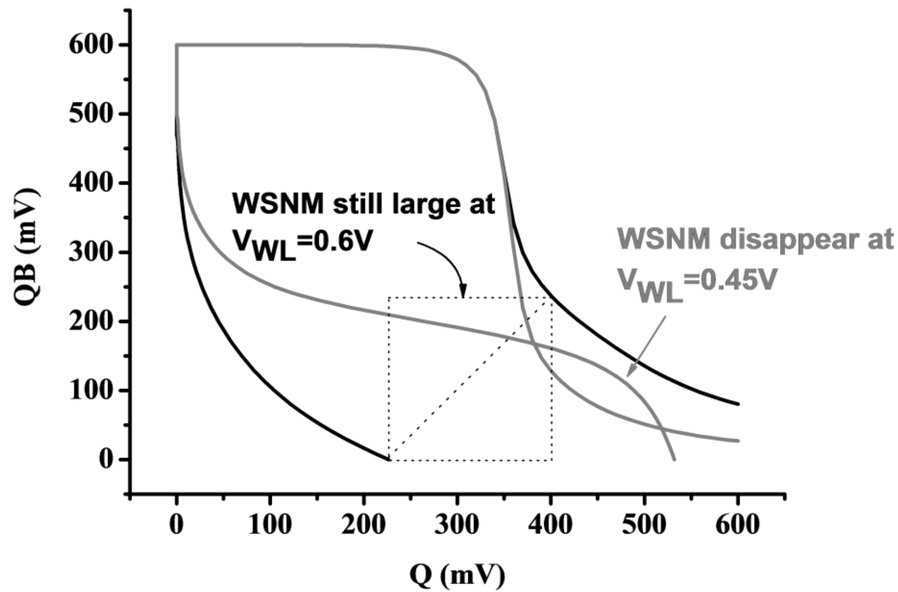


Fig. 3-9 WSNM decrease with word-line voltage decrease

3.4.2Prior-art comparison

Fig. 3-10(a) [3-17] shows a further improved version utilizes the tracking NMOS and series feedback resistor to control the virtual supply node of the WL driver. Notice that in Fig. 3-10(a), the virtual supply node of the WL driver taps the poly-resistor somewhere in the middle to minimum the sensitivity of WL level to poly sheet-resistance variation, which can be easily ±30% in deep sub-100nm technology. However, due to the presence of resistor in the virtual supply path and the lowered virtual supply for the WL driver, the scheme suffers severe degradation of WL rise time. Another scheme, depicted in Fig. 3-10(b), uses PMOS and series feedback resistor to lower the WL level [4]. This method could prevent the degradation of WL rise time and provide a stable WL level across process corners and temperature, but still suffers from the poly sheet-resistance variation and the large area needed to implement the resistor.

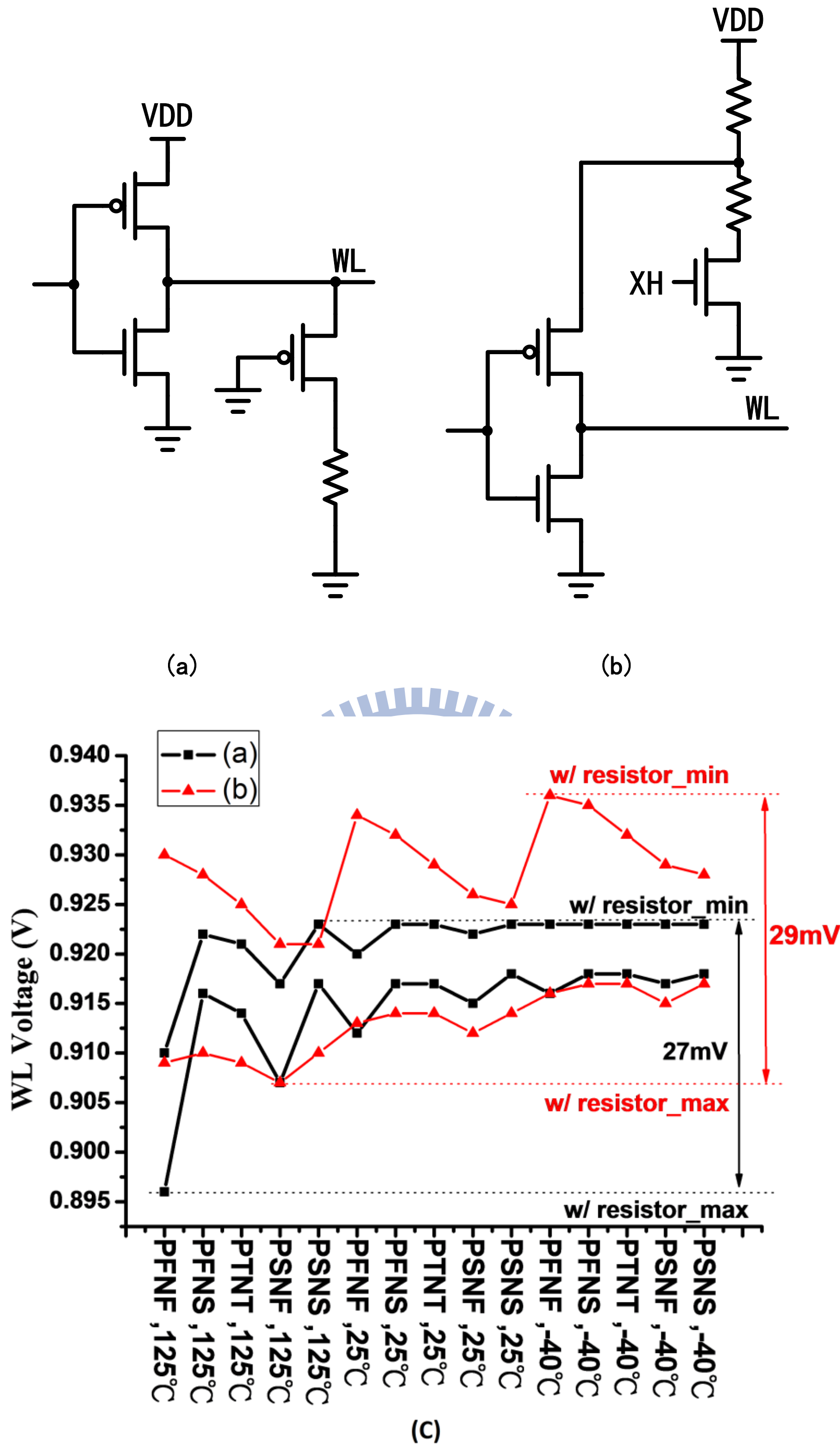


Fig. 3-10 Prior art of word-line under-drive circuits

Fig. 3-10(c) shows the simulated WL levels of the schemes in Fig. 3-10(a) and 3-10(b). The scheme of Fig. 3-10(a) exhibits larger temperature dependence, especially at high temperature. At high-T and PFNF corner, the WL voltage drops significantly more than other cases due to the large temperature dependence of NMOS.

The scheme of Fig. 3-10(b) has smaller temperature dependence, yet exhibits large corner-to-corner variations. Moreover, the level control deteriorates significantly when poly sheet-resistance variation is taken into account. Finally, the scheme in Fig. 3-10(a) only tracks the cell access NMOS, and that in Fig. 3-10(b) only tracks the WL driver pull-up PMOS. As such, they do not fully track the corner variations, and the WL level is not optimally adjusted across process corners.

3.4.3 Proposed Word-Line Under Drive (WLUD) Scheme

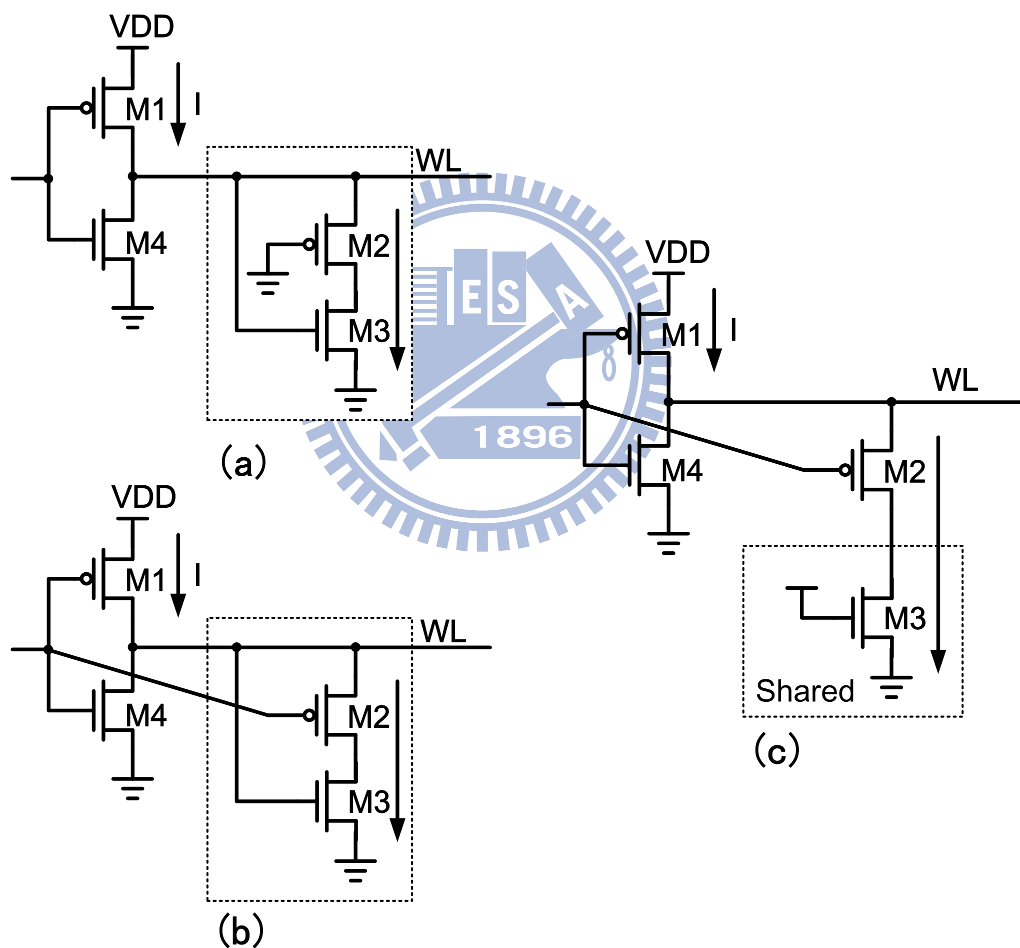


Fig. 3-11 Proposed word-line under-drive circuit

Fig. 3-11 (a) shows the schematic of the proposed word-line under-drive (WLUD) circuit. One NMOS (M3) and one PMOS (M2) form a series connection. The gate of M3 is connected to the word-line and the gate of M2 is connected to ground. M3 is

turn on when the word-line voltage goes high, so the WLUD circuit is activated only when the word-line is selected. PMOS M2 tracks WL driver pull-up PMOS M1, while NMOS M3 tracks the cell access NMOS (M3/M6 in Fig. 3-7). In essence, PMOS M2 serves as the feedback "resistor" for NMOS M3, and NMOS M3 serves as the feedback "resistor" for PMOS M2. This resistor-less feedback for WL discharging circuit mitigates the WL level sensitivity to process corners and temperature, eliminates sensitivity to poly sheet-resistance variation, and can be easily implemented with minimum device/area overhead. The scheme also offers fast WL rise time, as the WL driver is driven by full VDD, and the WL is pulled up by PMOS M1 with no resistor in the pull-up path. Notice that as the WL voltage goes low, the channel resistance of M3 increases and the current through M3 decreases. This helps to stabilize the WL voltage at a certain level. The tracking of both cell access NMOS and WL driver pull-up PMOS offers the capability to optimally place/adjust WL level across process corners. For example, at PSNF corner, if there is no M2 to track M1 (i.e. if M2 is replaced by a poly-resistor), the WL level would be pulled-down too low by the strong NMOS. With M2, to track M1, the channel resistance of M2 increases and its current decreases, thus limiting the WL pull-down current to maintain a proper WL level, Similarly, at PFNS corner, the slow NMOS M3 allows the WL level to stay higher, thus compensating for the slow (weak) cell access NMOS. Notice that the WL driver pull-up PMOS M1 works at triode (linear) region, and M2-M3 forms a diode like I-V characteristics (Fig. 3-12). The diode like I-V curve of M2-M3 intersects the PMOS M1 I-V curve in triode (linear) region where the transistors behave like resistors and have small temperature sensitivity. As a result, the resulting WL level has small sensitivity to temperature.

As illustrated in Fig. 3-11(b), the gate of PMOS M2 can be connected to the input of the WL driver. Functionally, it is the same as that in Fig. 3-11(a), with the

difference in the layout and possibly area overhead. Fig. 3-11(c) depicts another scheme where the discharging NMOS M3 is shared among WL drivers to further reduce device and area overhead. Notice that the gate of the shared discharging NMOS can be connected to a "Control Voltage" to offer Die-to-Die programming capability.

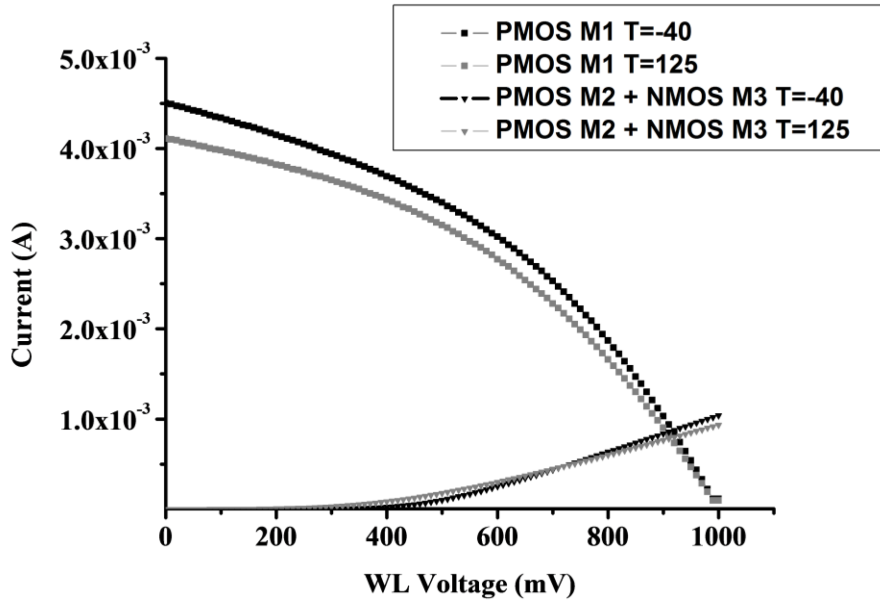


Fig. 3-12 I-V_{WL} curve of proposed WLUD circuit

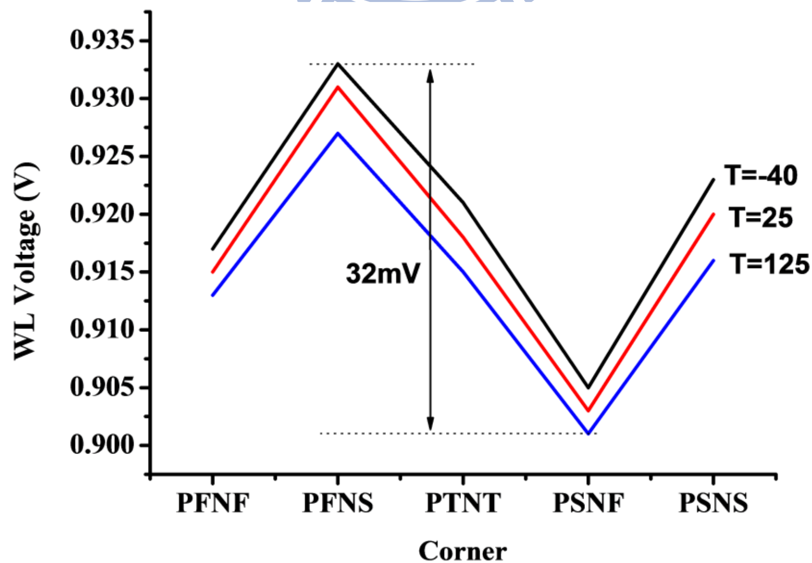


Fig. 3-13 V_{WL} comparison across process corners with different temperature of the proposed WLUD circuit.

Fig. 3-13 shows the simulated WL voltage level for the proposed scheme. We can see that the proposed WLUD circuit results in about 32mV variation across process corners and temperature. The temperature sensitivity is very small due to the dual-tracking capability and the triode-operated resistor-like device behavior discussed above. Furthermore, the WL level is optimally adjusted/placed across process corners. For example, the WL level is the highest at PFNS corner (which would compensate for the slow cell access NMOS), and the lowest at PSNF corner (which would compensate for the fast cell access NMOS). The scheme also reduces/confines the WL level difference between PFNF and PSNF corners. Fig. 3-14 compares the WL rise time for the schemes in Fig. 3-10(a),(b), and the proposed scheme in Fig. 3-11(a) across various process corners at different temperature. The scheme in Fig. 3-10(a) can be seen to have slow WL rise time, as the WL driver is driven by a lower voltage, and poly-resistor is present in the WL virtual supply path. Our proposed scheme and the scheme in Fig. 3-11(b) have significantly better WL rise times than that in Fig. 3-11(a) as the WL driver is driven by full VDD and there is no resistor in the WL pull-up path.

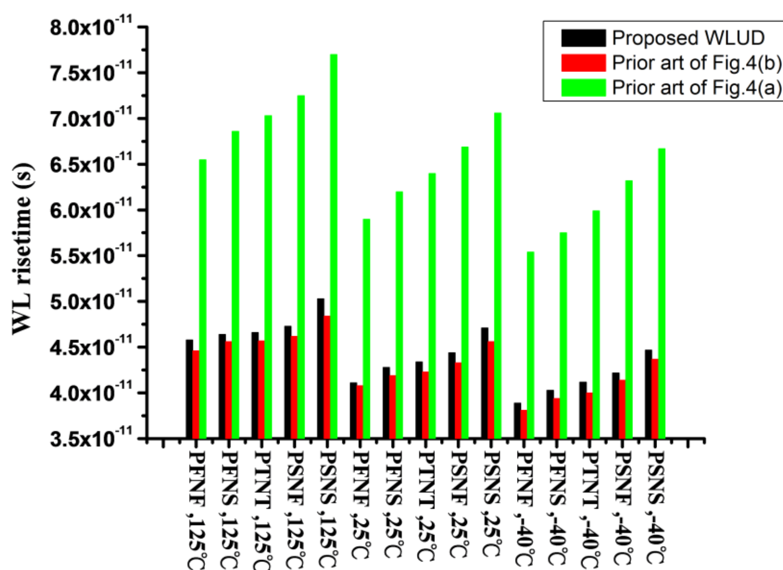


Fig. 3-14 Word-line rising time comparison of this work and prior art.

3.4.4 The chosen of WL voltage

With WLUD, the RSNM improves while WSNM degrades as shown in Fig. 9. If we choose the $V_{WL_WLUD'}$ which makes WSNM equals RSNM, we can strike to a balance point that read/write margin are equal. However, this $V_{WL_WLUD'}$ is lower and the read performance and write ability suffers more. In our methodology, we choose the WL voltage level V_{WL_WLUD2} where the RSNM equals to WM as Write-assist can be applied to improve the Write-ability, WSNM, WM, and Write performance. This V_{WL_WLUD} is higher than $V_{WL_WLUD'}$ to prevent read performance and write-ability degrades too much. The WL voltage level V_{WL_WLUD} is chosen based on PSNF corner, which needs the most RSNM improvement and thus the lowest WL level among all corners. The WL levels of other corners would be higher to prevent too much Write-ability degradation, especially at PFNS corner which has the worst Write-ability.

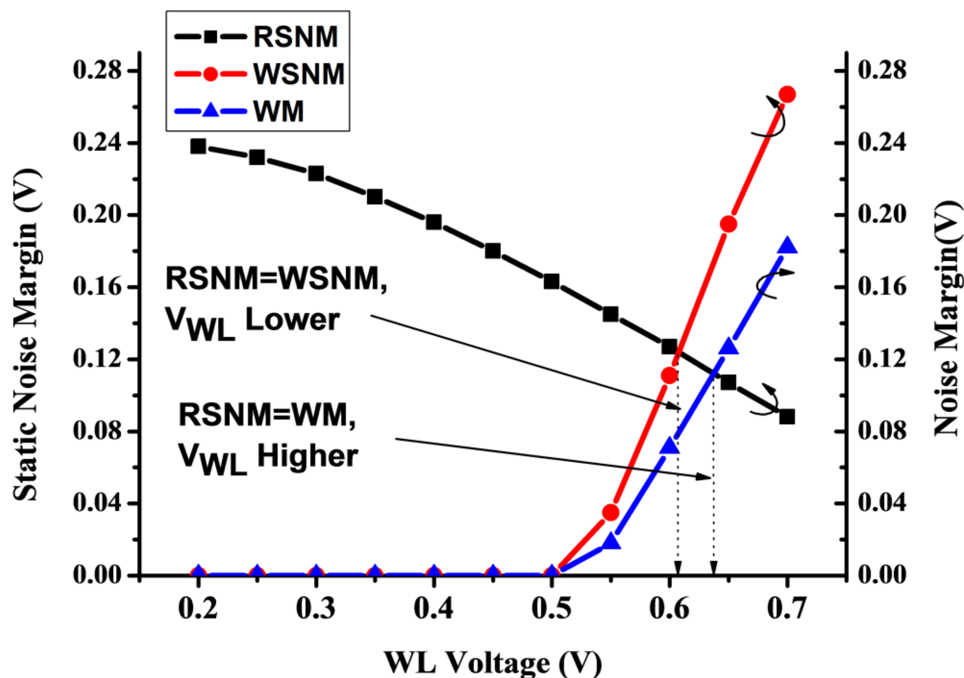


Fig. 3-15 RSNM and WM versus WL voltage level.

3.5 Previous work of write ability improvement

As the technology node goes to nanometer scale, write performance has become a critical issue. Especially with the supply voltage is reduced, write ability and write margin (WM) of traditional 6T-SRAM face a severely degradation. There are several ways to improve write ability: 1) Higher WL voltage, 2) Negative Bit Line (NBL), 3) Lower cell supply. Higher WL voltage could help write ability because the word-line voltage is arisen higher that makes the pass-gate stronger. But this would let the half-select disturb problem more serious because other cells in the row direction are performing half-selected read operation. Since the RSNM is reduced due to the higher word-line voltage, these half-selected cells may loss their storage value. [3-3] Negative Bit Line (NBL) skill is to pass a negative voltage in to cell rather than a ground voltage during write operation. This could improve write ability without losing noise margin. However, negative voltage is hard to pass into the cell and a large capacitor is always needed to maximize the coupling ability. But the implementation of large capacitor cost lots of area. [3-10]~[3-14] Besides, the timing of capacitor boosting mechanism should be precisely controlled. This timing control issue makes the NBL design much harder. Another good method to improve write ability is the lower cell supply write.

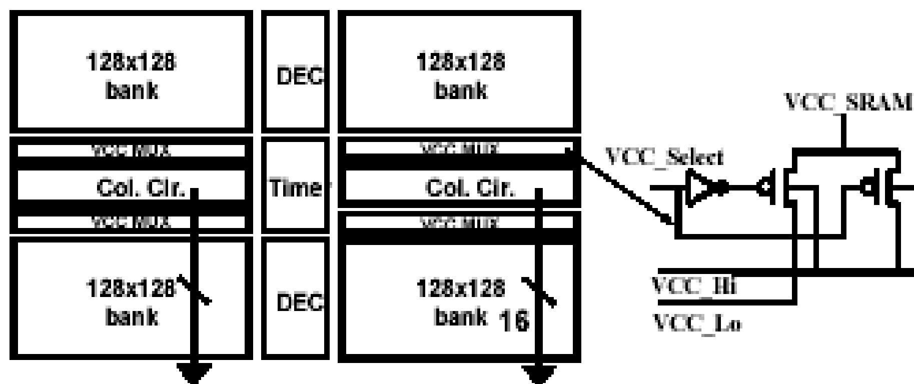


Fig. 3-16 Column based dynamic V_{CC}.

Lower cell supply is another direction to improve write ability. Fig. 3-16 shows a traditional method with column based dynamic V_{CC}. The V_{CC} would be switch to the lower one during read cycle. However, two supplies are needed in this implementation, and the power switching of whole chip is a slow process.

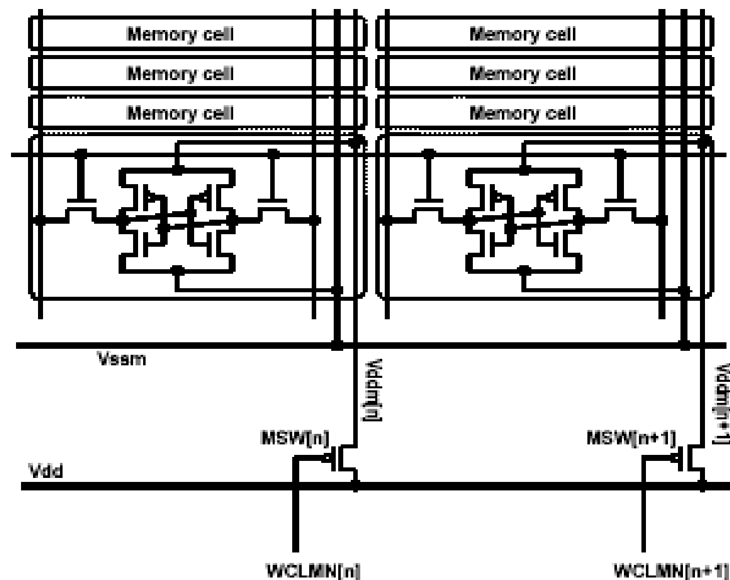


Fig. 3-17 Floating Power Line Write

Rather than directly use the second power, Floating Power Line Write demonstrates another idea to improve the write ability. Because the power line is floating during the write cycle, the floating power line will continuous go low due to the write current. This means the write margin would increase with the low going supply voltage. And column direction switch could prevent large carrier movement that cause the slow switching time. But this scheme floats the both side power line of cell. This would degrade the latch regenerate process which could result slow settle time. And the floating drop too low would degrade the SNM, cause the stability problem of other column direction half-selected cells.

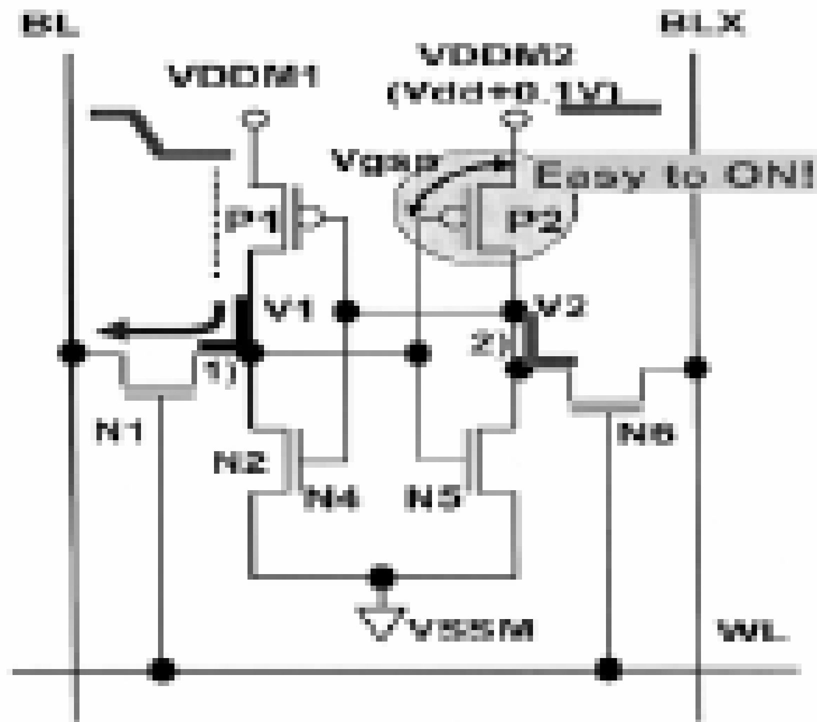


Fig. 3-18 Differential Data-Aware V_{DD} scheme

Fig. 3-18 shows the differential Data-Aware V_{DD} circuit. This scheme only weaken the power of write "0" side with a diode drop. The power line of the opposite half cell is kept to enhance the latch regenerative effect. And the SNM has a better improvement than previous work. But the diode drop limit the granularity of supply tuning with poor level-setting capability. The SNM of column direction half-selected cells is still degraded.

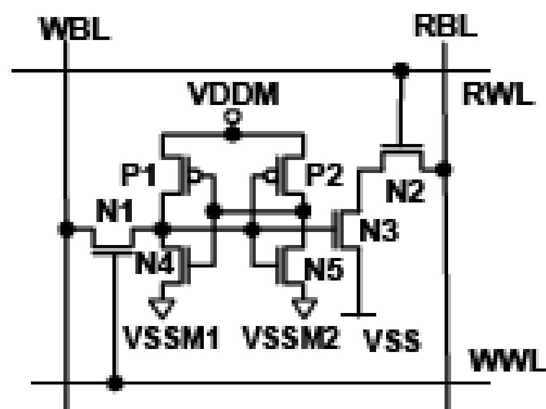


Fig. 3-19 Differential Data-Aware V_{SS} scheme

Besides, we can also control the V_{SS} with the data-aware scheme rather than control the $V_{DD.}$ (As shown in Fig. 3-19)

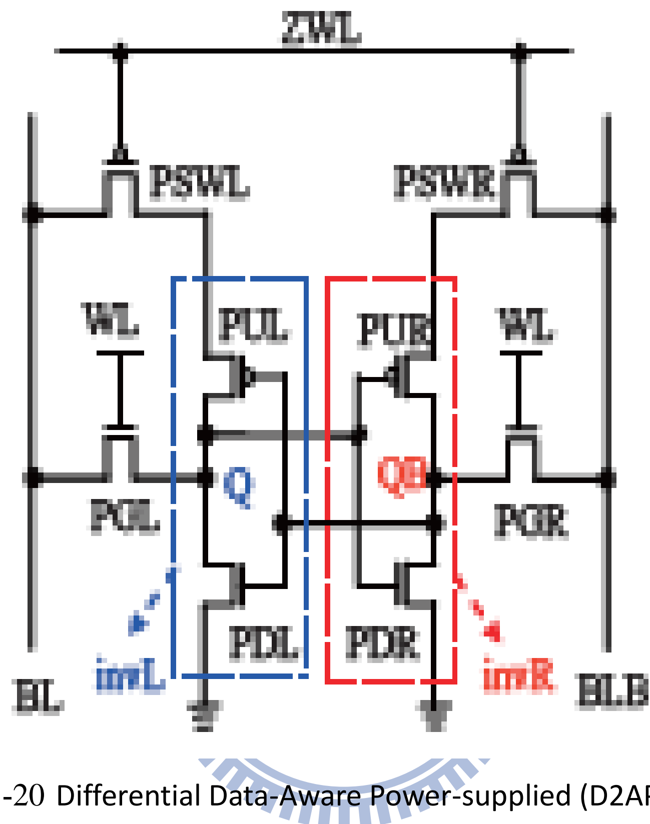


Fig. 3-20 Differential Data-Aware Power-supplied (D2AP) 8T cell

The column control of power line could result the SNM problem of those column half-selected cells, the Differential Data-Aware Power-supplied (D2AP) 8T cell directly add 2 transistors into original 6T cell to control the power with the Data-Aware scheme. But extra 2 transistors of each cell add extra area and performance overhead. Cell supply for unselected rows cut-off during read/write still degrades the SNM. BL suffer 2 time pass-gate loading that would degrade the sensing margin, performance, and density.

3.6 Data-Aware Write Assist circuit

3.6.1 Proposed Data-Aware Write-Assist Circuits

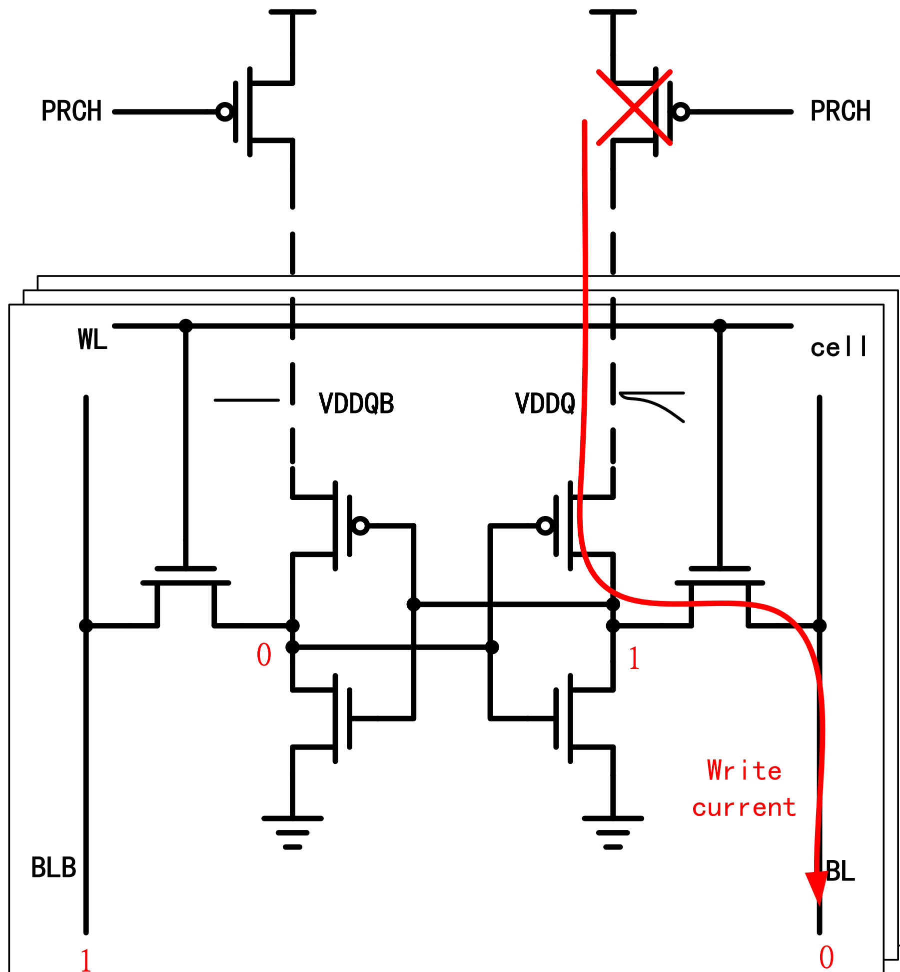


Fig. 3-21 Proposed word-line under-drive scheme

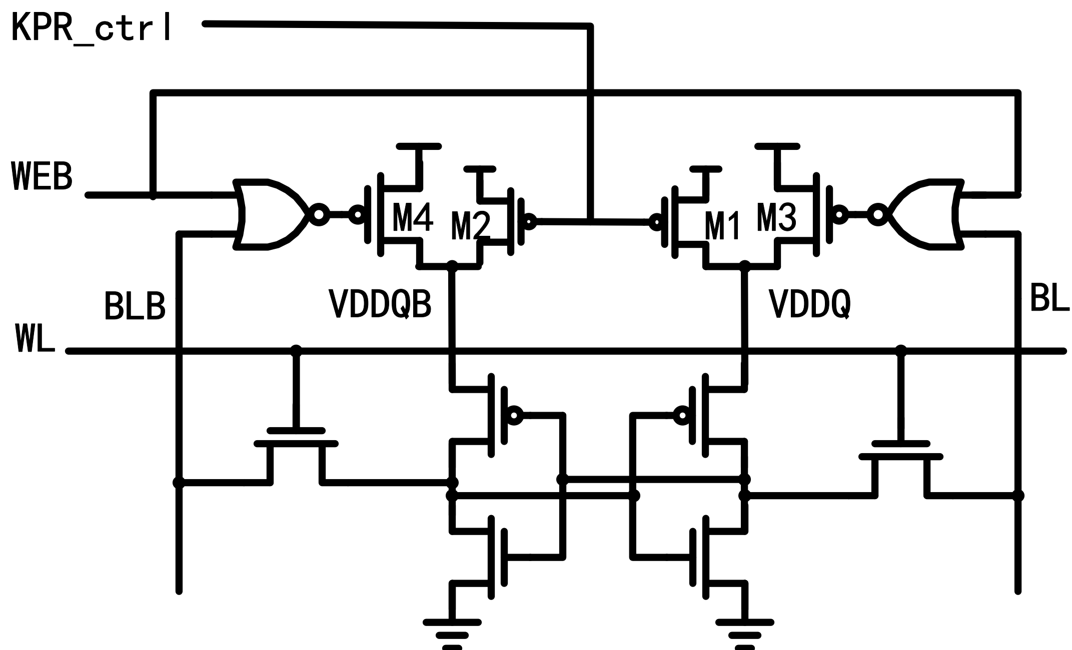


Fig. 3-22 Proposed word-line under-drive circuit

Fig. 3-22 shows the proposed DAWA circuit. Each side of virtual power node (VDDQ and VDDQB) is connected to one power switch PMOS rather than directly to the power supply. One NOR logic gate is connected to the gate of power switch PMOS, and the NOR gate only activate when WEB is at "0" (WEB means write enable signal and low-active during write cycle) and BL is pulled to "0", too. So that, the virtual power node VDD of write "0" side (we assume VDDQ is the power of write 0 side) will be floating. Once word-line is selected to perform the write operation, there will generate a write current from power node to the ground voltage bit-line. The VDDQB will go low because it is floating now and this write current reduce the charges of this floating node. VDDQ continuous go low until the cell is written a value. Because the storage value is flipped when write success, the pull-up PMOS of floating side will be turned-off to stop the VDDQ to decrease. With this low going VDDQ, write ability can be greatly improved because the WM widen a large value. The cell pull-up ability and latch feedback effect won't be weakened because of that the power switch PMOS of opposite side remains on and VDDQ of that side is a full VDD voltage level. Besides, the control is directly from the bit-line without other logics and delay chain that makes circuit is quite simple and has better timing control.

3.6.2 Write current and Column Half-Selected problems

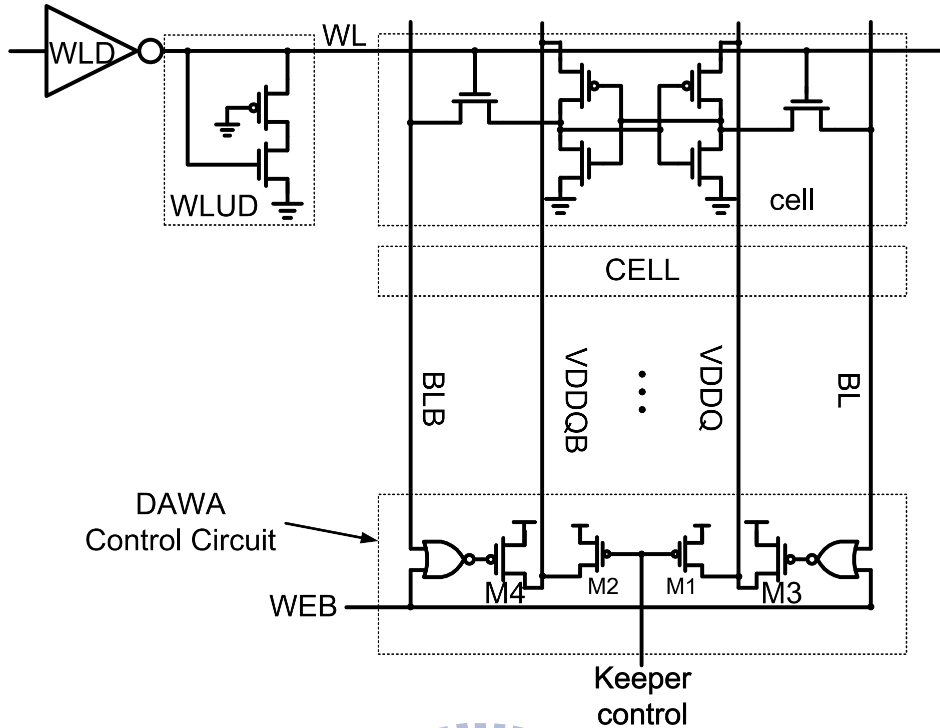


Fig. 3-23 Column Arrangement of DAWA and WLUD

Due to the architecture arrangement, all of the cells in the same column share the same DAWA control logic circuit (Fig. 3-23). If one cell is writing a value, all of the other cells in this column will face this floating VDDQ which is called column-half-selected effect. This lower voltage VDDQ will reduce the HOLD Stability Noise Margin (HSNM) of these standby column-half-selected cells and even result in retention fail to these cells if VDDQ is too low to keep its HSNM (Fig. 3-25). To prevent this, we add a keeper (M1 & M2 in Fig. 3-23) on floating node VCSBL that its voltage would be above a certain value (VDDQ_min) that won't flip those standby cells. The VDDQ_min should be define as the asymmetry supply voltage that just make the minimum of asymmetric Hold SNM of all corner with local variation equal or slightly above zero. So that once the designed keeper let the VDDQ won't below this value, the other column half-selected cells won't retention fail even with variation.

3.6.3 Asymmetric Hold SNM

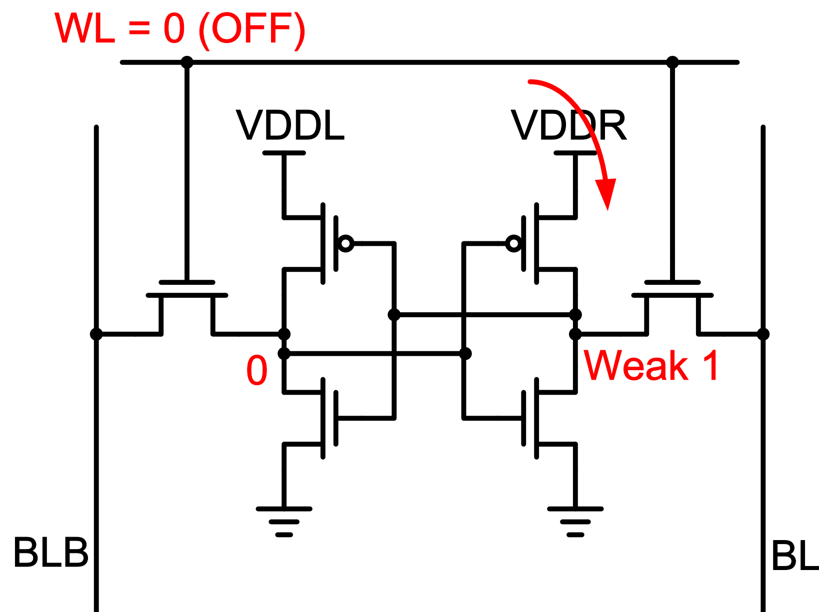


Fig. 3-24 equivalent circuit of column half-selected cells

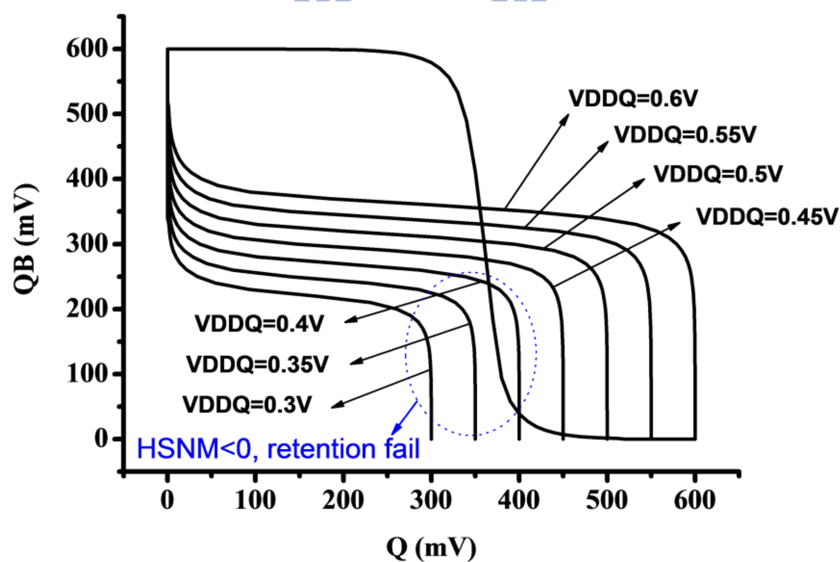


Fig. 3-25 Hold SNM with Asymmetric supply voltages

In order to guarantee the stability to those column selected cells, we must exam the Hold SNM of them. For a single cell, the HOLD SNM would decrease due to the voltage drop of one side power. Fig. 3-24 shows the condition of equivalent circuit of the column half-selected cells. And Fig. 3-25 shows that one wing of the butterfly curve would decrease with the VCSR decreasing. Once one wing disappears, in other

words, Hold SNM equals or less than 0, cell would flip.

3.6.4 Leakage Issue

Another source that would also drop VDD is the cell leakage. Since there are many cells in a column and the cell power node of those cells is connected to the power switch PMOS which is a weaker power source than power supply, the total leakage current would be an un-neglectable value that would also drop VDD voltage as write current does. Especially at PFNF corner, this leakage induced VDD drop could be more than the drop by using DAWA and also let those column-half-selected cells fail to keep its value. So we must both consider the write current of DAWA and total leakage current when designing the keeper size.

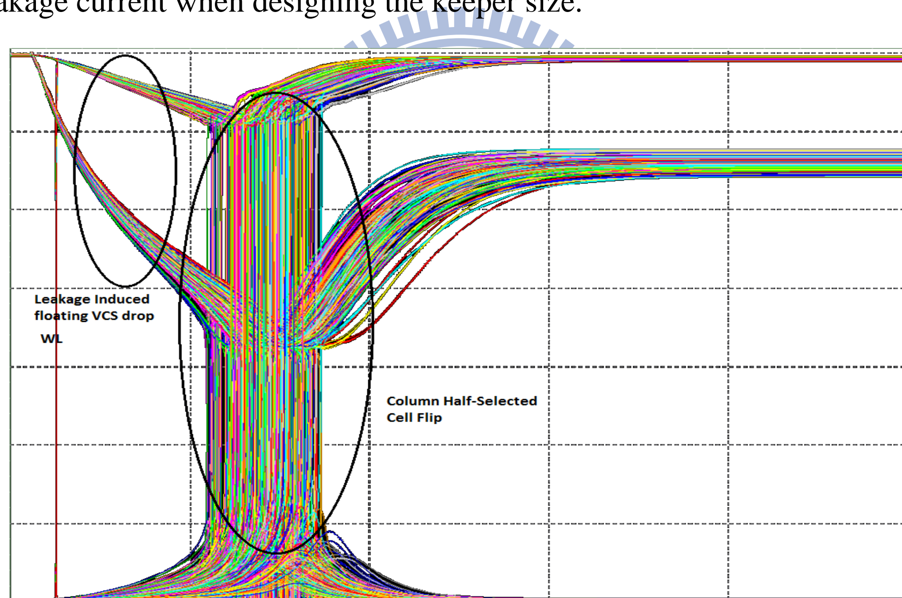


Fig. 3-26 Leakage drop waveform

3.6.5 Keeper Design

Because we add the keeper which could be seen as another power source, VDD could not drop as low as possible. The write ability to be saved would be lesser if the keeper is stronger. So it is critical to find the keeper size that can save the write ability as much as possible but won't result in any retention fail. However, there exist a paradox: if we design a keeper size to prevent retention fail at PFNF where in this

corner that VDD would drop to the lowest value due to large leakage in this corner, the keeper size would be too strong to fix the write ability at PFNS which is the corner has the worst write ability but has smaller leakage. The effect of DAWA would be greatly limited by this too-strong keeper. So we add a keeper control circuit that let the current drive ability of keeper would track with corner. (Fig. 3-27) shows the keeper control circuits. We use one PMOS and one NMOS to divide a voltage Vkpr_ctrl and connect it to the gate voltage. At PFNF, the keeper size and Vkpr_ctrl is tuned that VDDQ would above VVDD_min. However at PFNS, Vkpr_ctrl would be higher than PFNF. Keeper would be weakened to improve the write ability of using DAWA. At PSNF that has the best write ability, Vkpr_ctrl would be lower than others. Keeper is very strong to keep the stability of those column-half-selected cells. Fig. 3-28 shows the gate voltage of keeper with the corner tracing circuit.

For Read mode and Standby mode, the power switch always turns on and VDD is held at power supply voltage. The RSNM, stability and read performance would not be disturbed.

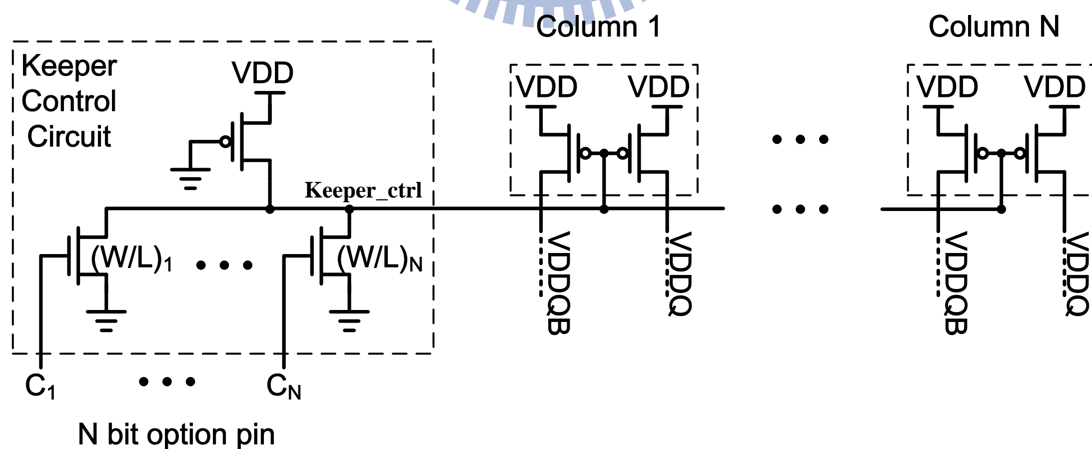


Fig. 3-27 corner tracing keeper control

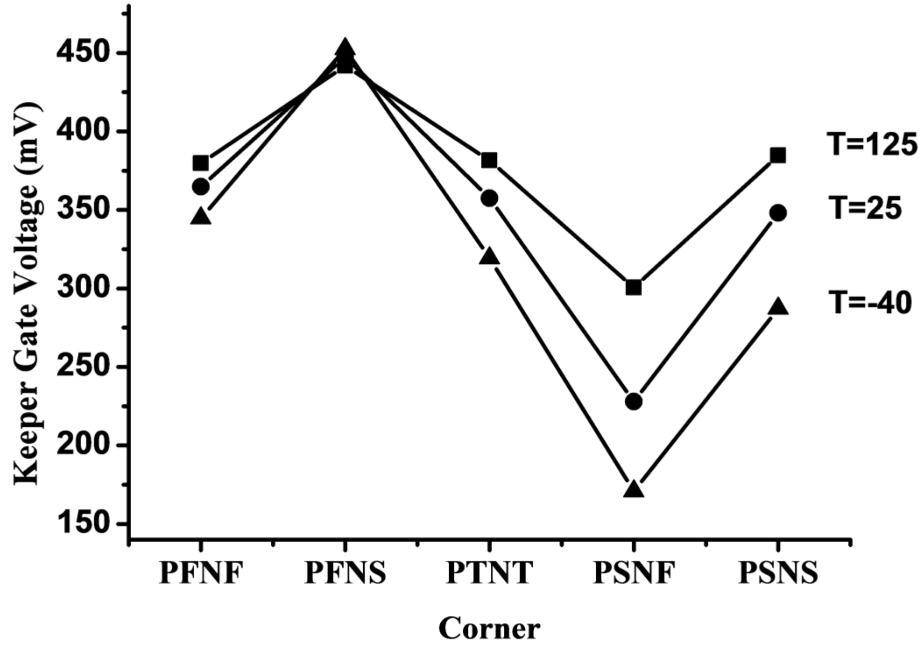


Fig. 3-28 gate voltage of keeper at different corner

3.6 Simulation result

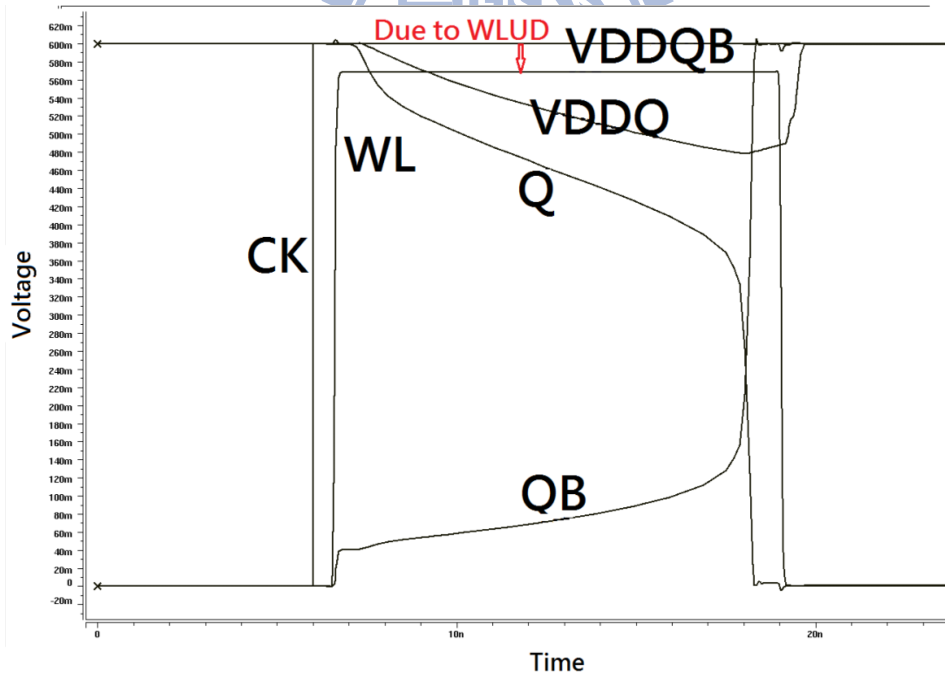


Fig. 3-29 Write waveform

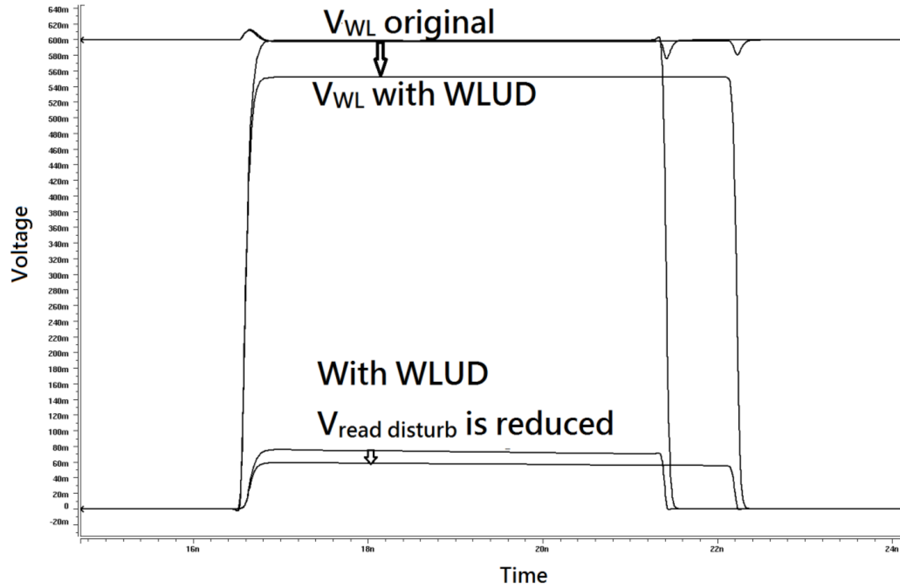


Fig. 3-30 Read waveform

Fig. 3-29 shows the simulated write waveform. We can see that the floating VDDQ decrease due to the write current until the cell is written a value. We also can see that the word-line voltage decrease a value from supply voltage due to WLUD circuit. Fig.3-30 shows the read waveform comparison w/ and w/o WLUD circuit. The read disturb voltage is decreased with the lower V_{WL}. Fig. 3-31 shows the RSNM improvement around corners and Fig. 3-32 shows RSNM improvement at different supply voltage. In Fig. 3-31, if we adopt the constant WL level scheme, we could only have 21% improvement RSNM improvement at PSNF corner, which is the read worst corner. But with the proposed WLUD circuit scheme, the RSNM improvement can be increased to 29%, which is 8% larger than constant WL level. As shown in Fig. 3-32, the original read V_{CC_min} is about 0.4V. But the read V_{CC_min} could be further improved to 0.3V if we adopt the WLUD. And at supply 1.2V operation, the RSNM with WLUD has 39mV improvement.

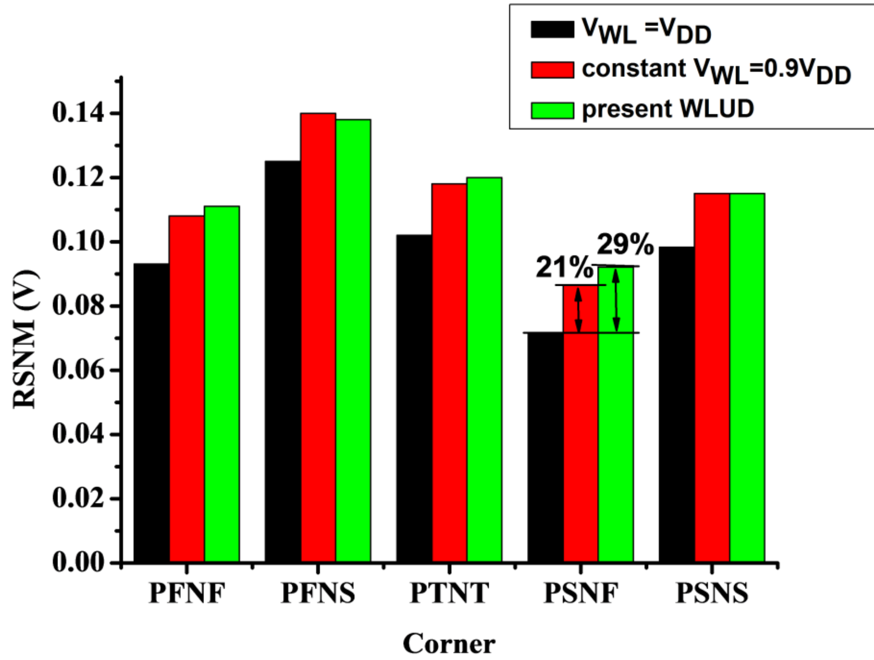


Fig. 3-31 RSNM improvement of constant V_{WL} and WLUD

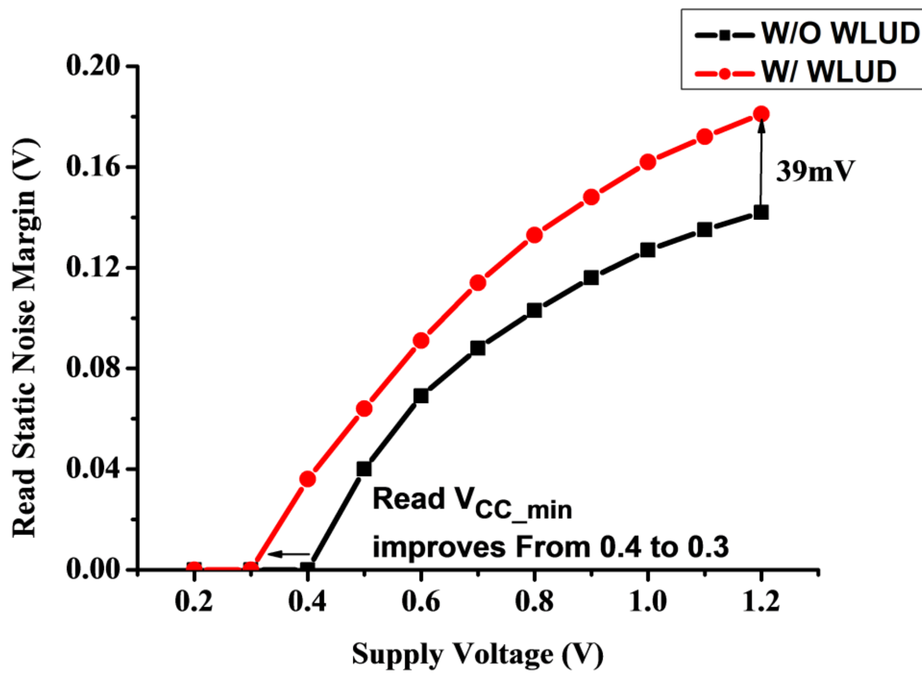


Fig. 3-32 Read margin and read V_{CC_min} improvement

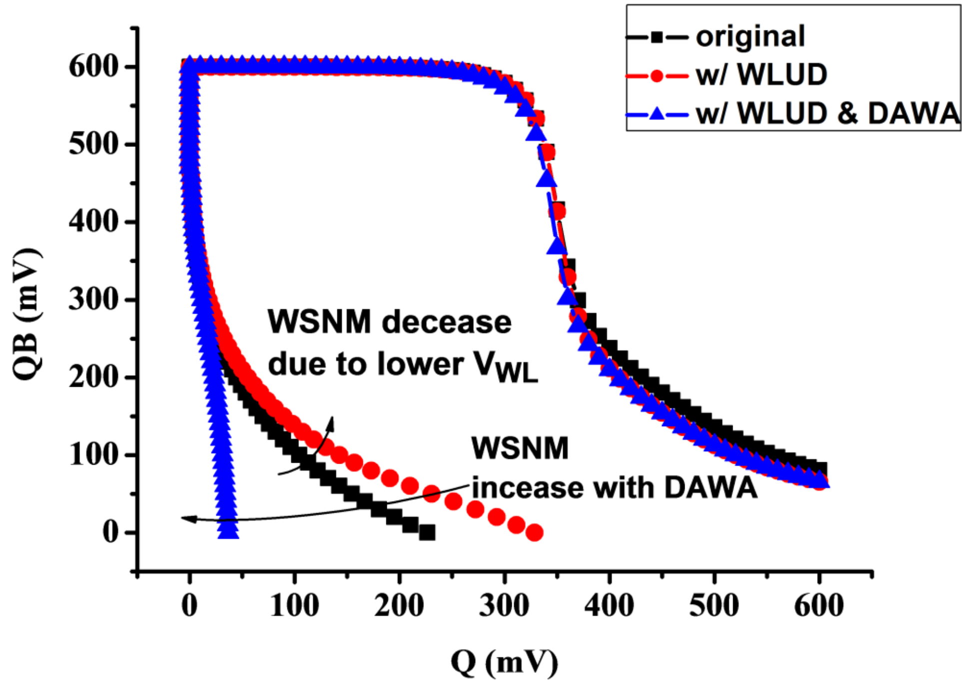


Fig. 3-33 WSNM improvement with DAWA

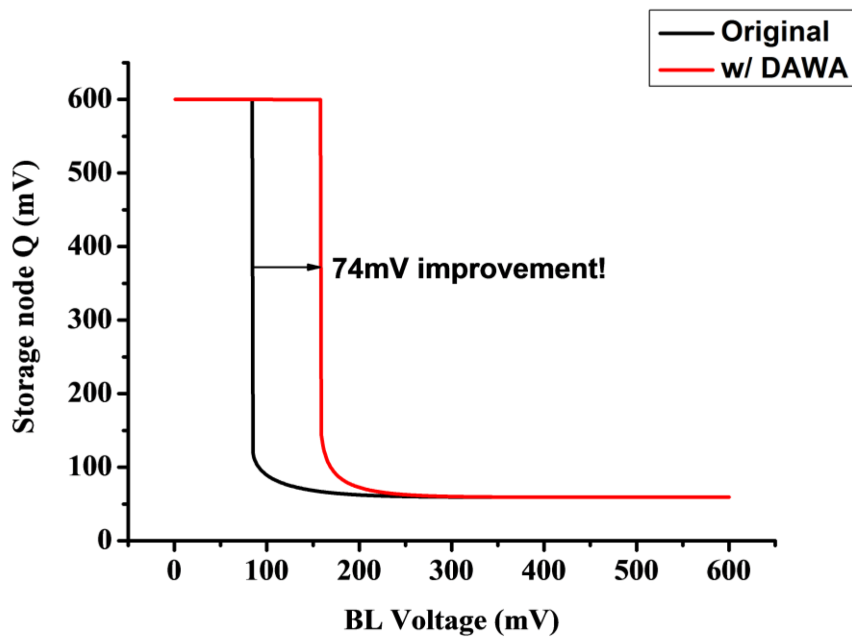


Fig. 3-34 Write Trip Voltage (write margin) improvement with DAWA

Since we use the WLUD scheme, the WSNM would decrease. As shown in Fig. 3-33, the butterfly curve would get closer that makes the WSNM decrease. But we add the DAWA with keeper, the WSNM can get a large improvement. From the Fig. 3-33, butterfly curve is widen that the WSNM increase a large amount. Fig. 3-34

shows the Write trip voltage improvement with DAWA. The original write trip voltage with WLUD scheme is about 84 mV. However, the write trip voltage can improve to 158mV with DAWA circuit with corner tracing keeper control. Fig.3-35 shows the write time. With DAWA, the write time could save about 10 % at low supply (VDD=0.4). At other supply voltage, the write time seem almost the same with the original design.

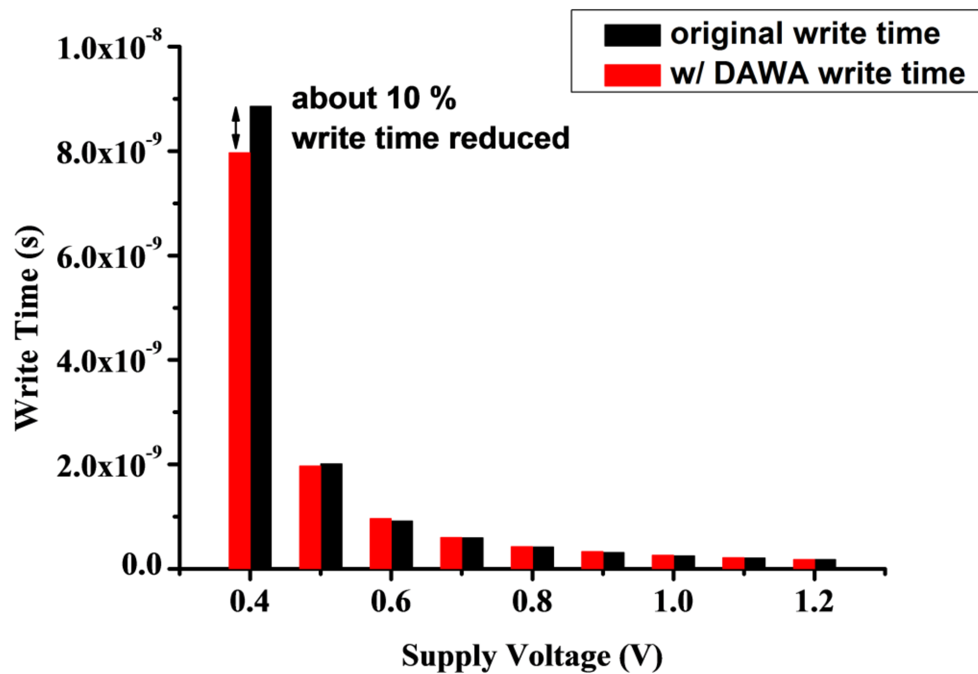


Fig. 3-35 Write Time and write V_{CC_min} Improvement

3.7 Implementation of one 128kb 6T SRAM

This test macro is modified from a 128kbit standard 6T-SRAM array in 55nm standard regular V_T process without changing its architecture. The test chip is directly modified from a standard 128kb SRAM macro. In this array, there are 1024 Word-Lines and 128 Column with inter-leaving 8. Total data input number is 16 and data output is 16. At each column, local bit line partition is 128 so that 1024 rows will be divided into 8 local blocks. We add a set of DAWA control logics and keeper to

control the cell floating power line to this macro. The keeper gate voltage control circuit has 4 pin to external control, and we place it with decoders. Every word-line is connected with a WLUD circuit to suppress Word-Line level. Fig. 3-36 shows the modified architecture of this test macro.

This area of this test chip is 691.33um*215.76um which is about 18% larger than previous design. In order to provide options and level programmable ability, extra 7 pin is added in this design. Fig. 3-37 shows the layout view.

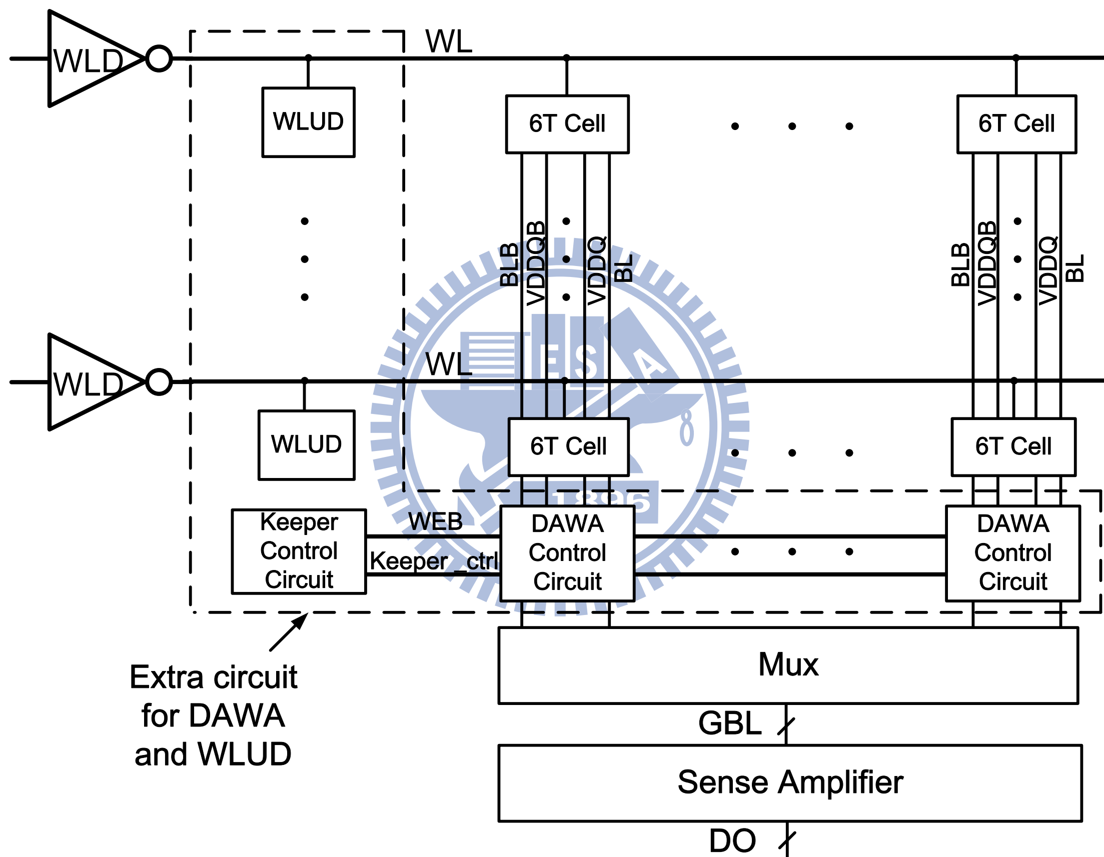


Fig. 3-36 Test Array Architecture

Capacity	1024*128 bits SRAM (Total 128k bits)
Process	55nm 1P9M SPRVT
Max frequency @ 1.0V FF,T=125	941MHz
Average Power @ 1.0V FF,T=125	2.0828mW
Max frequency @ 0.6V FF,T=125	362MHz
Average Power @ 0.6V FF,T=125	1.475mW
Core Area	714um*238um
VCC min	0.4 (TT)

Table. 3-1 Post-layout simulation summary

Corner and Temperature	Original Macro	Modified Macro w/ WLUD and DAWA
SS, 0.9V, -40℃	1.920 ns	1.976 ns

Table. 3-2 Access Time comparison at High Operation Voltage

Corner and Temperature	Original Macro	Modified Macro w/ WLUD and DAWA
FF, 1.0V, 125℃	2.0297 uW	6.052 uW

Table. 3-3 Power Consumption comparison

	Original	w/ WLUD	w/ DAWA
Power	2.08mW	2.254mW	2.1322mW
%	1	1+**8%**	1+**2.5%**
Area	0.126mm2	0.132mm2	0.146mm2
%	1	1+**4%**	1+**14%**

Table. 3-4 Power and Area comparison

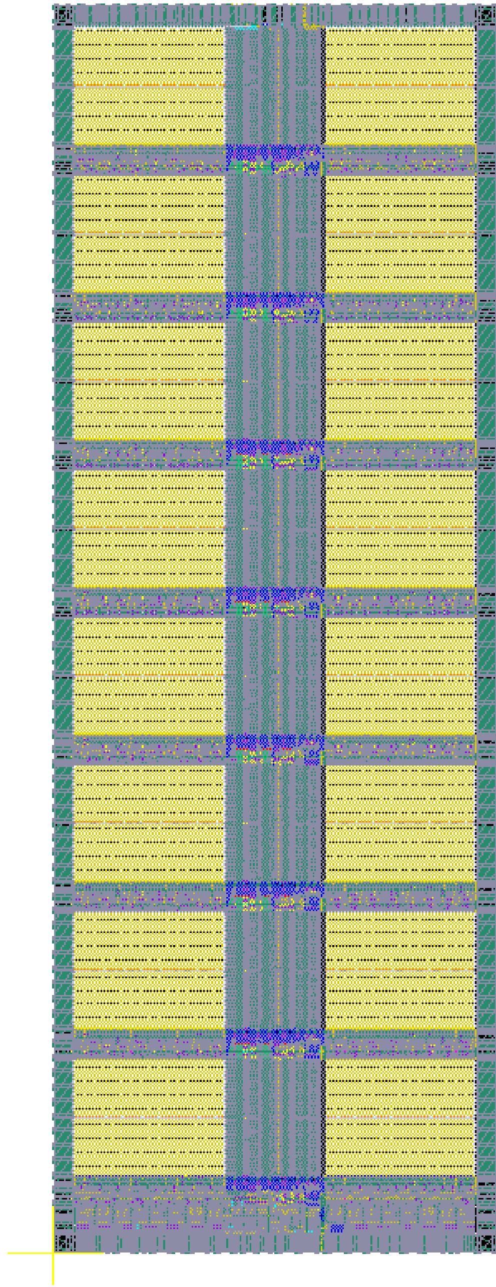


Fig. 3-37 Layout view of 128kb 6T SRAM Macro

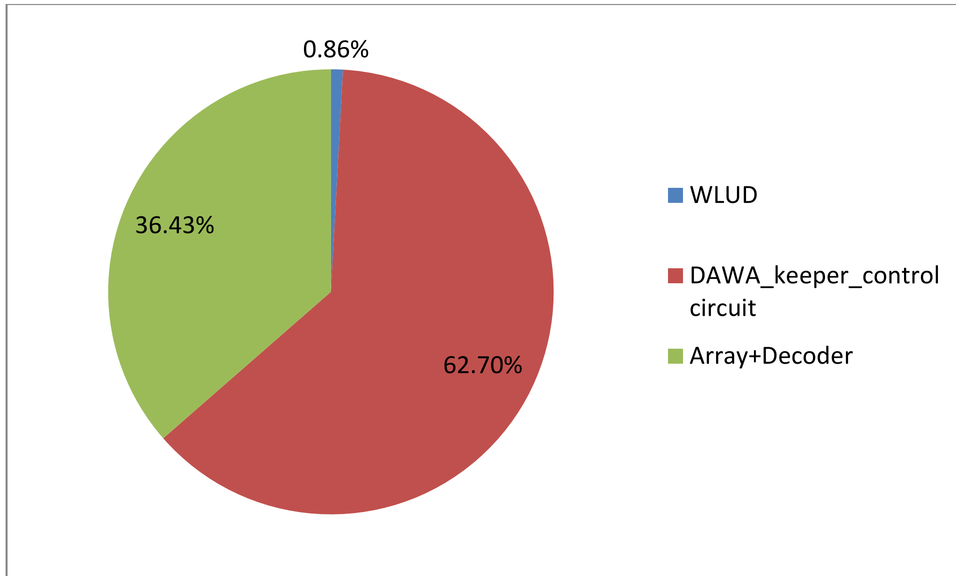


Fig. 3-38 Power Distribution

From Fig. 3-38, we can see that the power of test macro increased about 2X than original design. The main factor is the DC path of corner tracing keeper gate voltage control circuit. This contributes a quite large amount power.

	DAWA (Height*Width)	SHAA55_SH (Height*Width)	Increased Area
Core	691.33um*215.76um	609.87um*207.23um	81.46um*8.53um
LBLMux	16.895um*9.92um	7.8um*9.92um	9.095um*1
GAP	1.8um*9.92um	0.8um*9.92um	1.0um*1
LBLSEL	16.895um*50.5um	7.8um*41.97um	9.095um*1

Area Overhead = ((DAWA / SHAA55_SH)-1)*100% = 18.02%

Table. 3-7 Area Overhead

Chapter 4

The Design of one 512kb Array Based 6T SRAM Noise Margin measurement circuit

4.1 Introduction

In the modern nano-scaled devices, one major design issue is the random variation that always severely influences the stability and reliability of Integrated Chips (ICs). Many previous works shows that local variation such as random dopant fluctuation (RDF), etc is random and a statistical Gaussian distribution due to its physical characteristics.[4-1]~[4-6] Because of that there exists these variations, parameter of transistor such as threshold voltage (V_{th}), current (I) will spread over a wide region and deviate from nominal value. This makes the transistors will not work at the designed operating point when the chip is fabricated out. In Synchronous Random Access Memory (SRAM), transistor is much vulnerable to local variations due to its compact and dense environment. The read/write noise margin would suffer great degradation due to local random variation and result a fault read/write operation. For a real implemented SRAM macro, we could not fail any one cell even the variation exists. As a result, we have to characterize the variation impact to SRAM noise margin.

In order to get the large amount statistical data fast, the simplest method is to build

a testing macro to measure the required parameter. Many previous works have proposed various measurement circuits to measure the noise margin, transistor VT in SRAM cell, and BTI degradation. [4-10][4-12][4-16] demonstrate measurement circuits to measure the noise margin, read current and write current in SRAM which could help to evaluate the cell stability. [4-11][4-13][4-14][4-15] discuss the V_T distribution and the degradation due to NBTI. [13][14] even modify the standard 6T cell layout to help V_T measurement. [4-17]~[4-21] proposed several all-digital measurement scheme to measure the V_T , Gate Dielectric Breakdown, HCI, BTI, TDDB, etc. These full digital measurement could allow the measurement process be fully on-chip and has good resolution for resulted value.

In this work, we developed a test structure to measure the trip voltage of cell inverter, read-disturb voltage and write margin (WM) of tradition 6T SRAM cell. By using the Static definition of Read Static Noise Margin (RSNM), we separate RSNM into trip voltage of cell inverter and read-disturb voltage and subtract them to get RSNM. To do so, we modify standard 6T-SRAM cell to get trip voltage of cell inverter and read-disturb voltage from circuit structure directly. For WM measurement, we continuously write a cell with BL voltage from V_{DD} to ground till the value is written. The bit line of the first time the cell is written value is defined as WM. Equivalent 1-M 6T-SRAM cells array is implemented in this structure to provide us enough amounts of data to analysis the read static noise margin (RSNM) and WM.

In order to measure the voltage from each cell, we add an extra measurement circuits to measure analog voltage without losing its precision [4-17] [4-18]. In order to sweep bit-line voltage, a resistor-voltage divider is designed to provide different voltage level from (0v~640mV) with 10mV a step. These can be controlled by all digital signals and test automatically.

4.2 Previous work

4.2.1 Array Based measurement structure

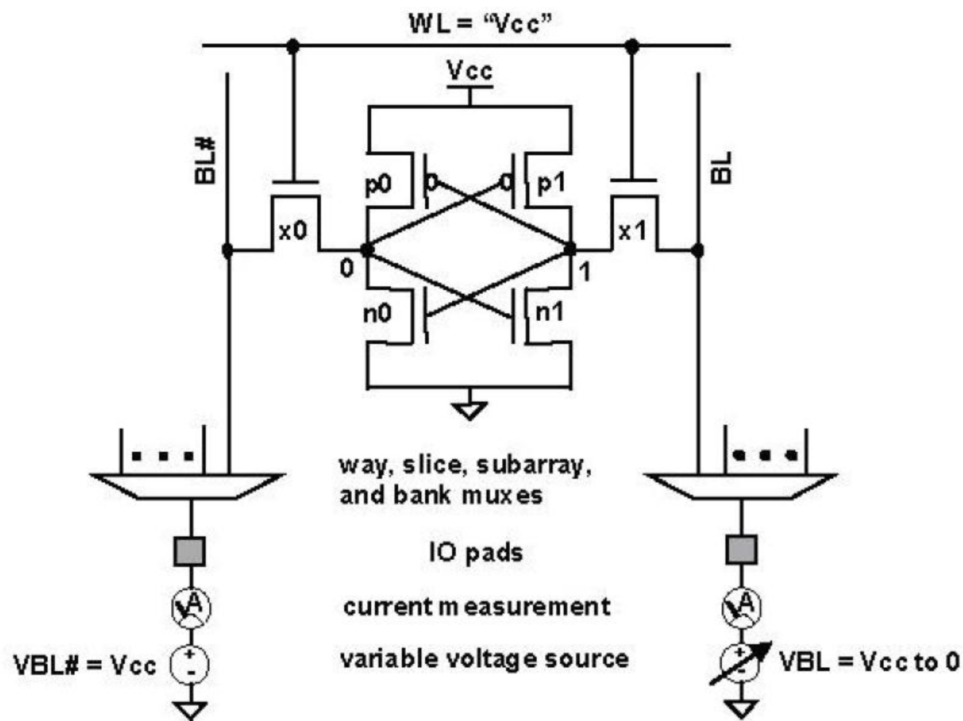


Fig. 4.1 Finding Vwrite structure with LYA (Low Yield Analysis) DFT

Fig. 4-1 shows the LYA (Low Yield Analysis) DFT structure that could find the Vwrite during the supply is suffering the noise [4-1]. The bit-line voltage is control by off-chip equipment, and sweep from VDD to the ground. In this bit-line gradually decreasing process, the selected cell would turn failure write into successful write. The highest bit-line voltage that could successful write in data is the DC write margin. In this structure, the SRAM cell is remain the original one and forms an array which is similar to the regular SRAM array structure which could allow us to get the DC write margin from the real cell array environment. And all of the bit-lines are connected to several Mux, and these mux is directly connected to the I/O. This structure allows us to measurement DC write margin, but we cannot get RSNM from

it.

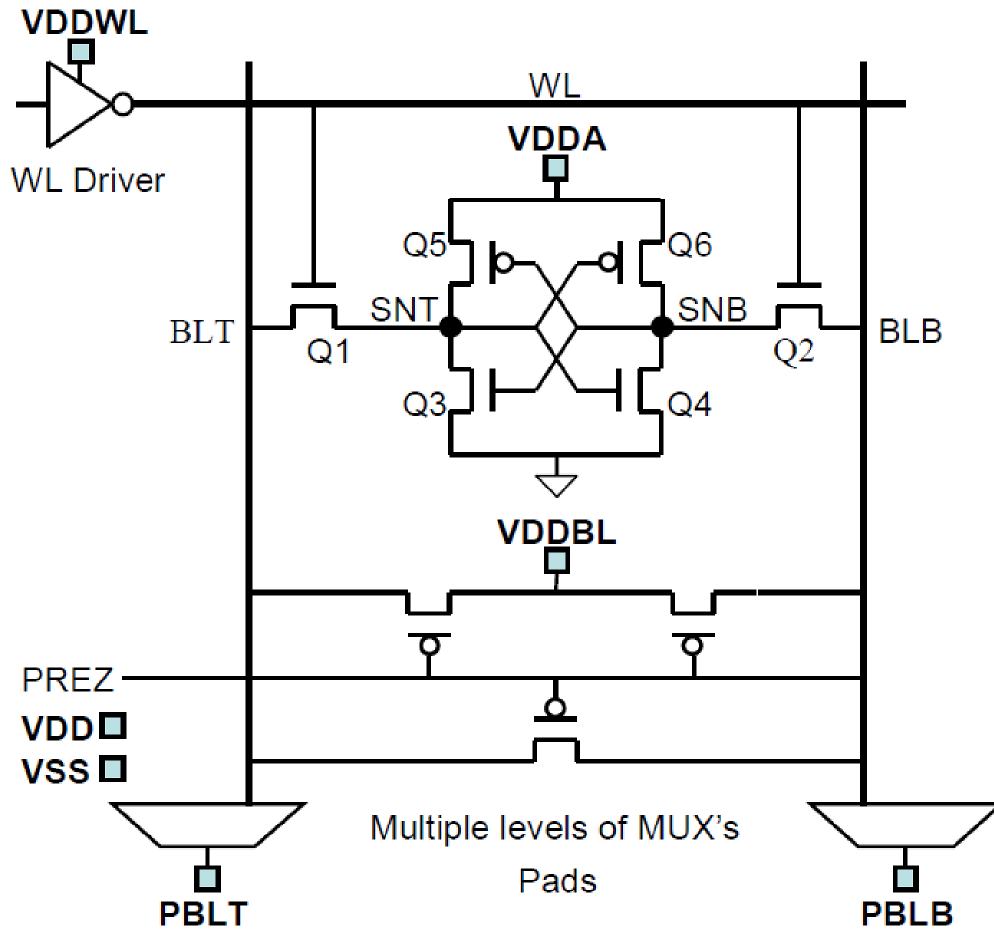


Fig. 4.2 DBTA circuit of characterization of Bit transistor scheme

The method in Fig. 4-2 is modified the LYA structure from previous one to get the Ig-Vg curves [4-2]. The test structure is also formed as real array. From the measured Ig-Vg curves, we can get the Vt from it, so that we can make a characterization from it. For this structure, all the cell are remain and the VT of each transistor (Pull-up, Pull-down, and Pass-Gate) could be separately gotten. However, the current measurement scheme consist many variation factor such as the leakage that would deviate the measured current. The current measurement scheme needs extra equipment and the measurement could exhaust time and not efficient.

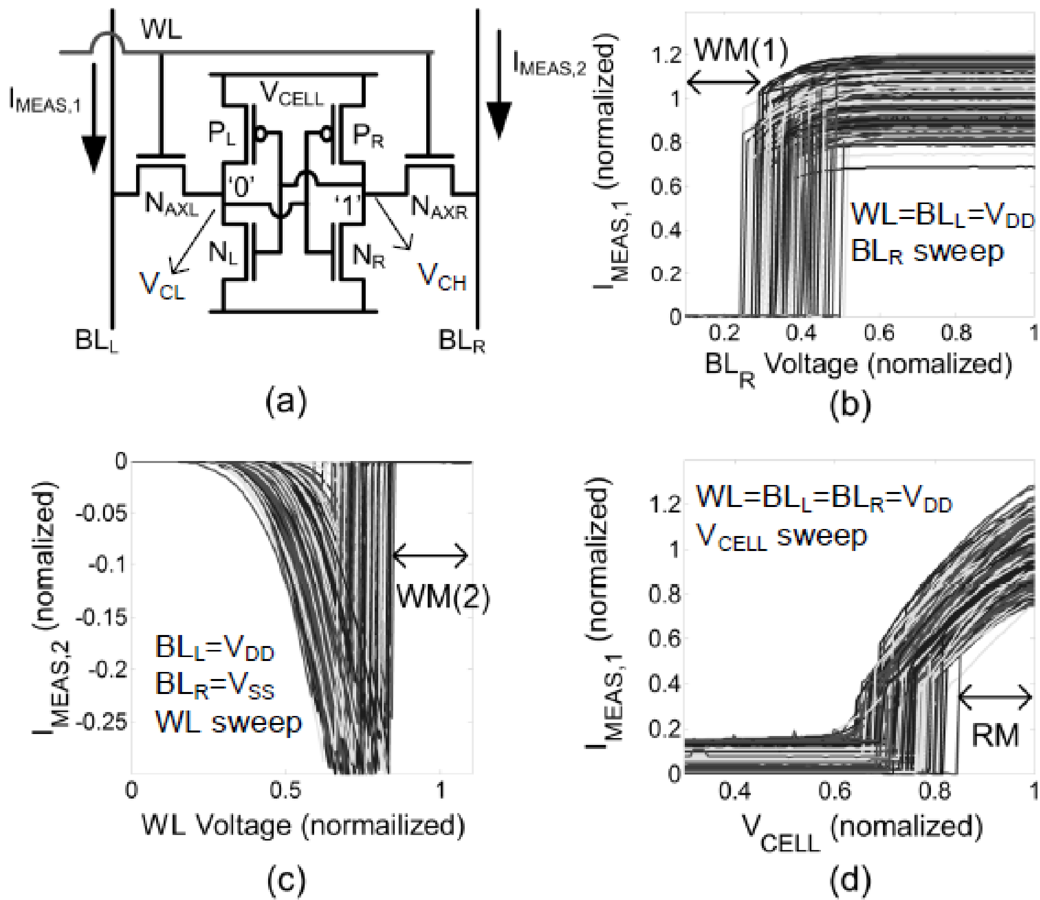


Fig. 4.3 The Read/Write Margin measurement scheme of [4-3]

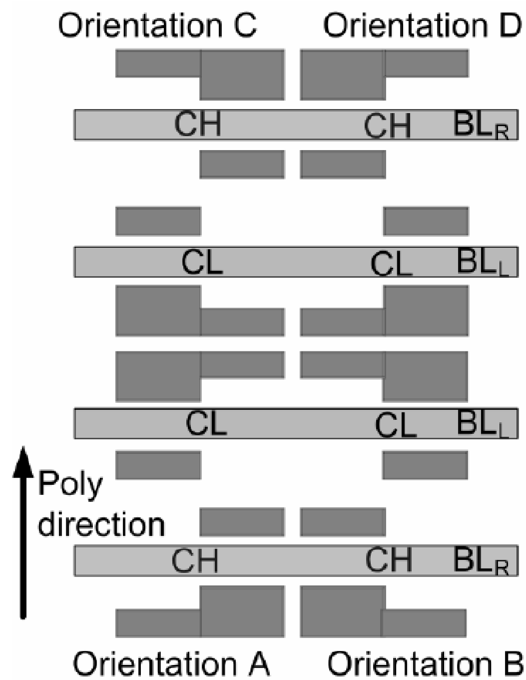


Fig. 4.4 The storage node wired-out SRAM cell layout of [4-3]

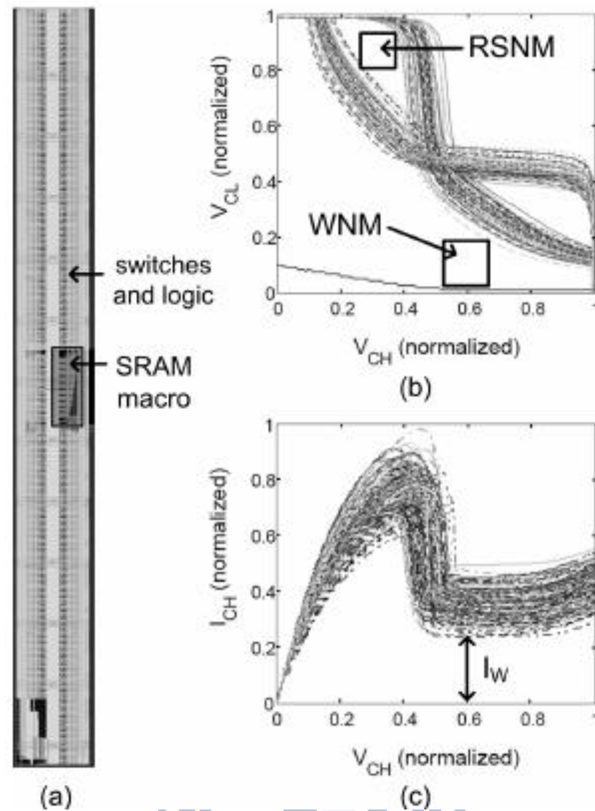


Fig. 4.5 Read SNM, Butterfly curve WNM and I_W measurement of [4-3]

This structure forms array with several banks, and all the storage node of cell is wired out for access [4-3]. It could allow us to directly sweep voltage to get the butterfly curve of RSNM and WSNM. Besides directly to form the SNM, this structure could also measure the write margin and read margin. For the write margin scheme, it adopts two schemes. One is the BL write margin, which is to sweep Bit-line and find the write trip point and we mentioned before. Another one is the WL write margin, which is to decreasing WL voltage to find the write trip voltage. Unlike the BL write margin, WL write margin is maximum value of V_{DD}-V_{WL} that could perform successful write. As to the read margin measurement, it decreases the supply voltage and monitor the bit-line current to get the read margin.

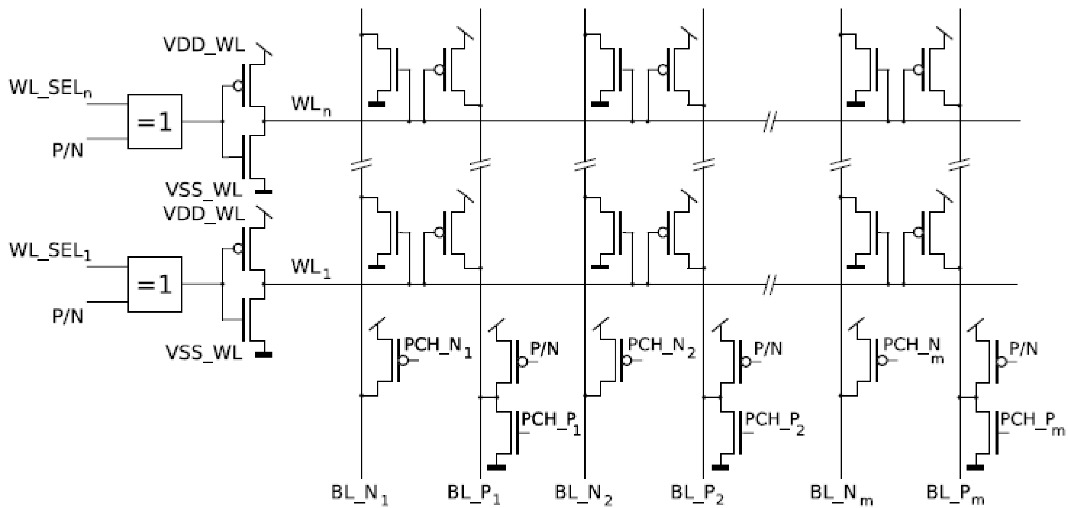
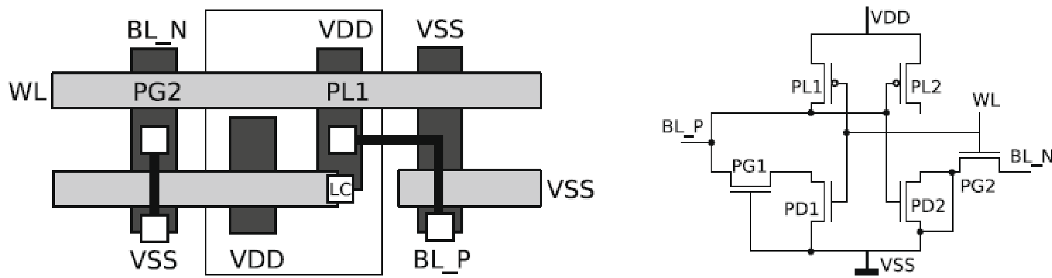


Fig. 4.6 1-M bit SRAM transistor NBTI test structure of [4-4]



1896

Fig. 4.7 Modified cells of [4-4]

Another array based test structure to analysis the statistical variation in SRAM transistor is shown in Fig.4-6 and Fig. 4-7[4-4][4-5]. With the modified cell, we can access each transistor to measure the I-V curve. So that we can measure the Vth fluctuation and even the impact of NBTI. And the cell is form into array to keep the density as real SRAM.

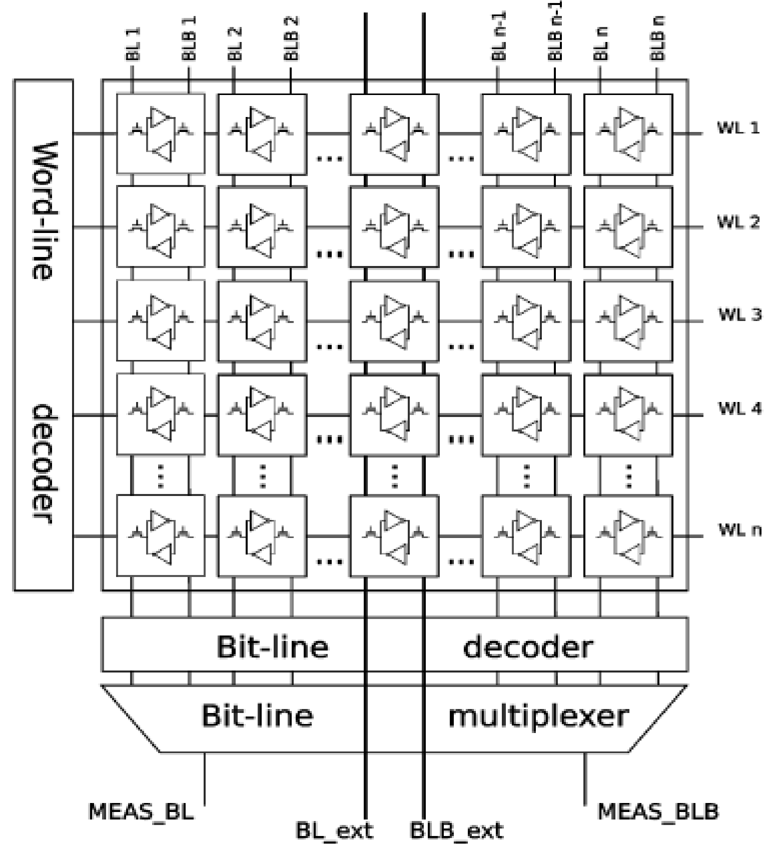


Fig. 4.8 Fast stability analysis of large-scale SRAM arrays and the impact of NBTI degradation.

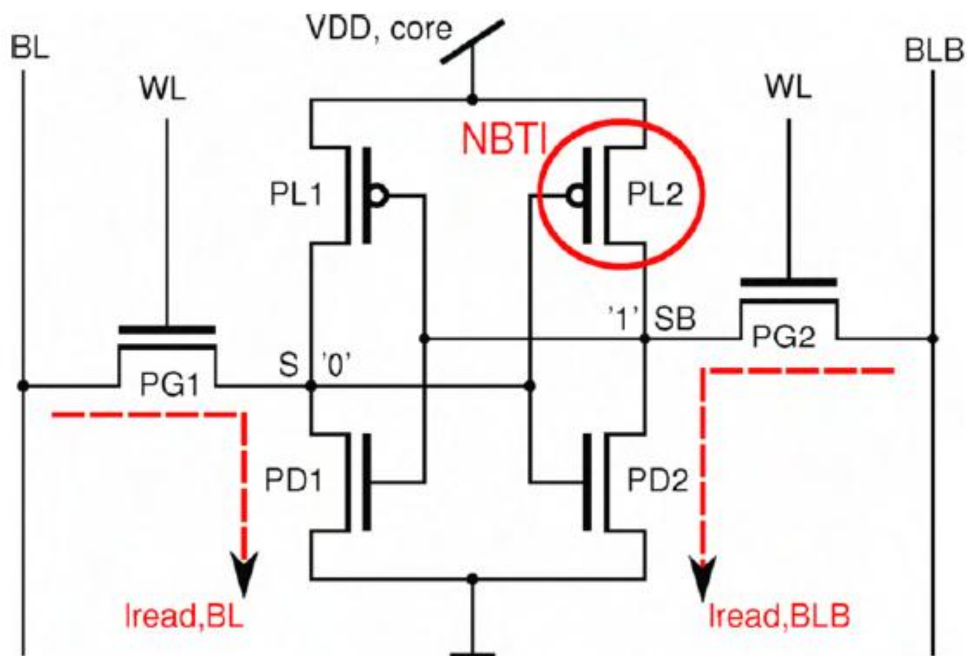


Fig. 4.9 measurement of read current and impact of NBTI

The test scheme adopts the read margin and write margin definition which we mentioned in the scheme of Fig. 4-3 [4-6][4-7]. It also measure sweep the core supply and measure the read current to find the read margin. But the decoder of this structure is modified with shift register, so we can perform fast switching to shorten the measurement time. Write margin could also be got by sweep bit-line. In this structure, the supply of the core could be raised high to stress cell PMOS, so that we also could get the impact of NBTI.

4.2.2 All Digital measurement scheme

All digital on chip measurement could let us perform more efficient measurement process. Here introduce some previous work that could allow us to get the absolute value of voltage in chip without the implementation of ADC.

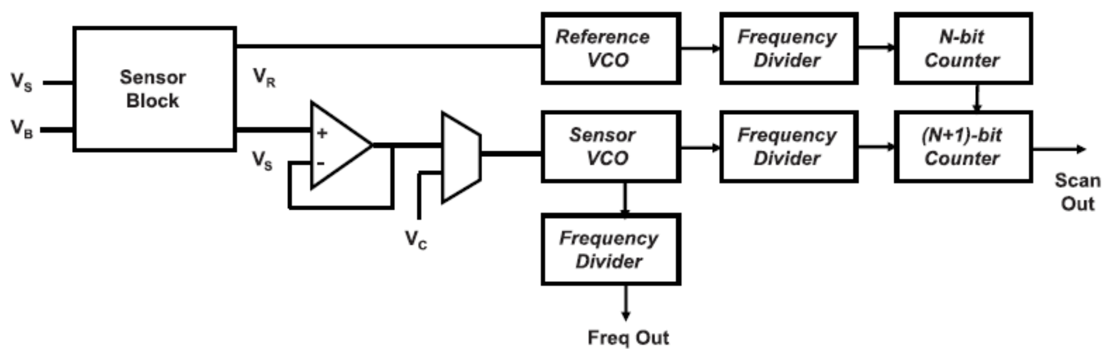


Fig. 4.10 Block diagram of measurement structure.

Fig. 4-10 shows the previous work that use the VCO based measurement scheme to measure the V_T [4-8][4-9]. For this structure, an array of normal transistor is build for VT measurement. Since an analog voltage would be gate, this measure scheme transform voltage signal into frequency and count out with digital counter. All digital control I/O could allow us easily control and get data simply by using a computer rather expensive equipment.

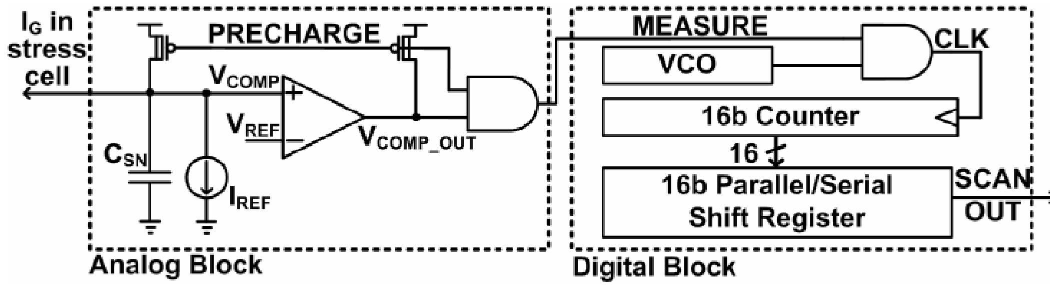


Fig. 4.11 current sensing scheme

Fig. 4-11 demonstrates a current sensing scheme. This scheme also based on the VCO, but the input the voltage which is transformed by the under-measurement current [4-10]. With the aid of OPA based comparator and the VCO scheme, it can check the current exceed the reference value or not by checking the output frequency difference.

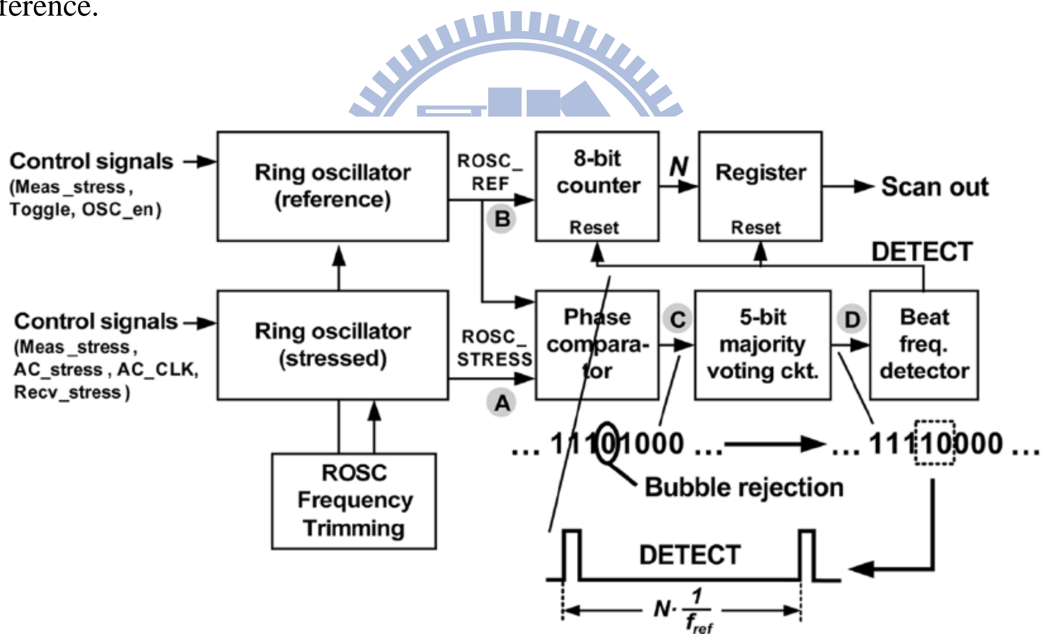


Fig. 4.12 Shows another frequency difference detection circuit

Since we might want to know the voltage difference of two, we can form two VCO and check their frequency difference [4-11]. To do so, one phase detector is added to detect the frequency difference by detecting their phase difference. As a result, the reference count number could be get and the difference number would also be got. And we still can get the real absolute value by adding them.

4.3 Modified cell for RSNM measurement

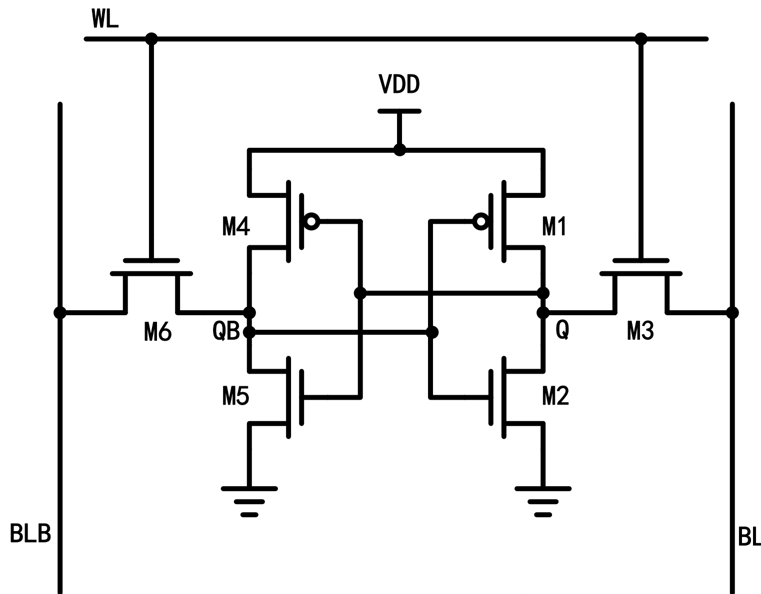


Fig. 4.13 traditional standard 6T SRAM cell

Fig.4.13 shows a traditional standard 6T SRAM cell. By definition, the RSNM could be the voltage difference of the trig voltage V_{trip} and read-disturb voltage V_{read} (RSNM = V_{trip} - V_{read}). During read process, pass-gate M3 is turned on and forms a divided voltage with M2 which is called read-disturb voltage V_{read} at storage node. V_{read} is an unwanted voltage and should be lower than the trig voltage of opposite inverter M4-M5 to keep the storage value. Once the V_{read} larger than V_{trip}, cell would lose its storage value. So the difference could be seen as the margin to prevent the cell flip its value and also be called as RSNM. To understand the characteristic of RSNM, we can further investigate the V_{trip} and V_{read}.

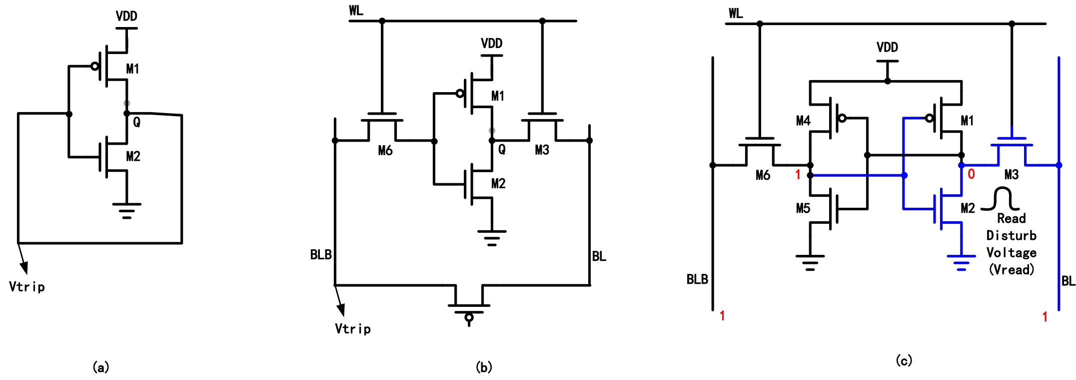


Fig. 4.14 (a) Basic concept to get Vtrip of inverter (b) Concept of Vtrip equivalent circuit in 6T SRAM (c) Vread Disturb Concept

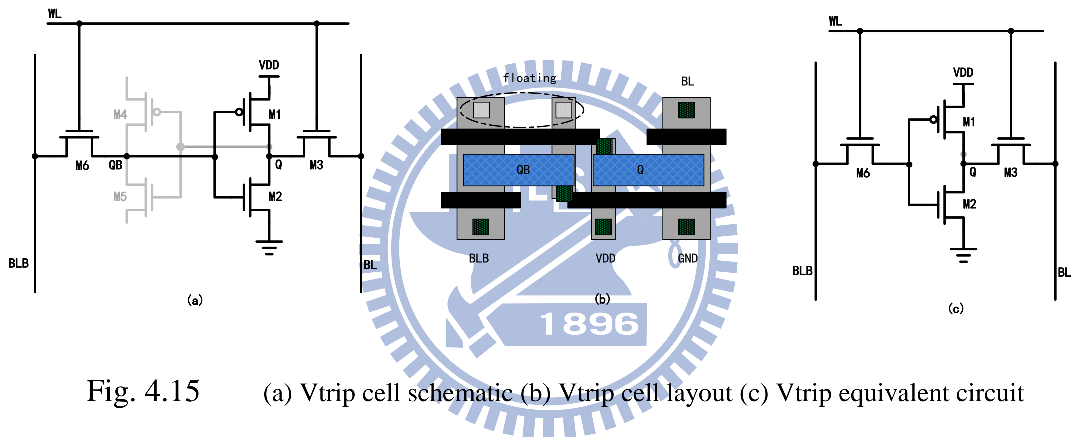


Fig. 4.15 (a) Vtrip cell schematic (b) Vtrip cell layout (c) Vtrip equivalent circuit

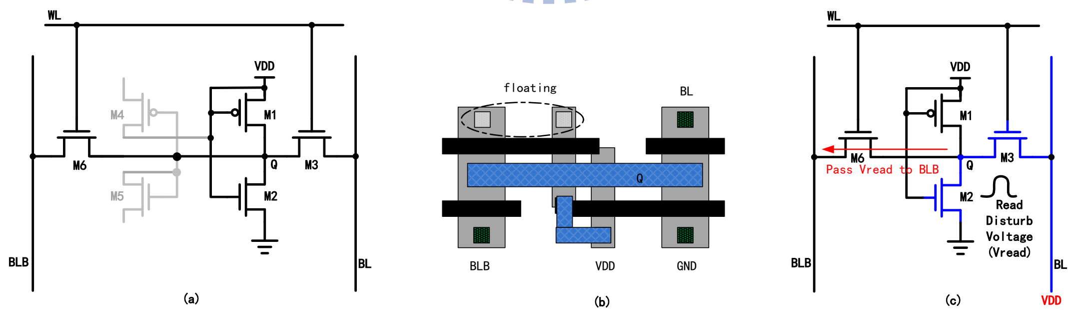
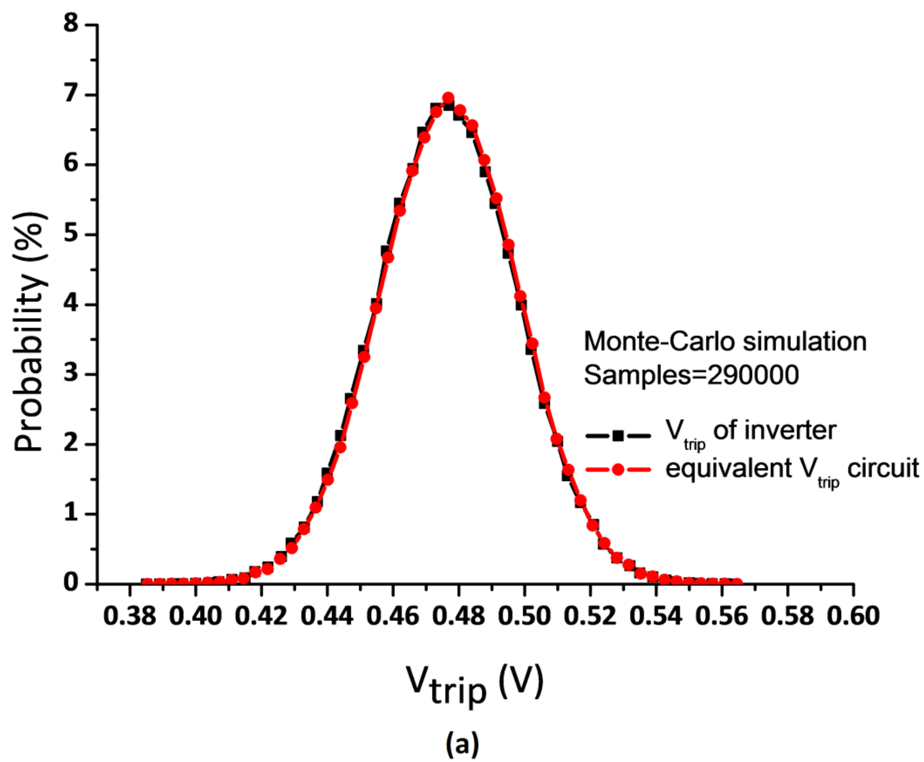


Fig. 4.16 (a) Vread cell schematic (b) Vread cell layout (c) Vread equivalent circuit

In order to directly get the V_{trip} and V_{read} from the cell array, we modified the connection of standard 6T cell. The simplest method to get the trig voltage of an inverter is to short the input and output(Fig.4-14 (a)(b)). So we isolate the power and ground of half part of inverter (Fig.4-15 (a)) and both pass-gate M3 and M6 is kept as

port to short the input and output. Fig.4-15 (b) is the layout of the modified trig voltage measurement cell (Vtrip cell). To get Vread, we must form a divide voltage from pass-gate M3 and pull-down NMOS M2 voltage (Fig.4-14 (c)). So we also isolate the power and ground of M4 and M5 and connect the gate of inverter M1-M2 to power to force this cell storage "0" (Fig.4-16 (a)). The pass-gate M6 is modified to connect the storage node Q to access Vread. Once BL is raised to high and word-line is selected, there will generate Vread at storage node Q and pass out to BLB through pass-gate M6. Fig.4-16 (b) shows the modified layout of read-disturb measurement cell (Vread cell).



(a)

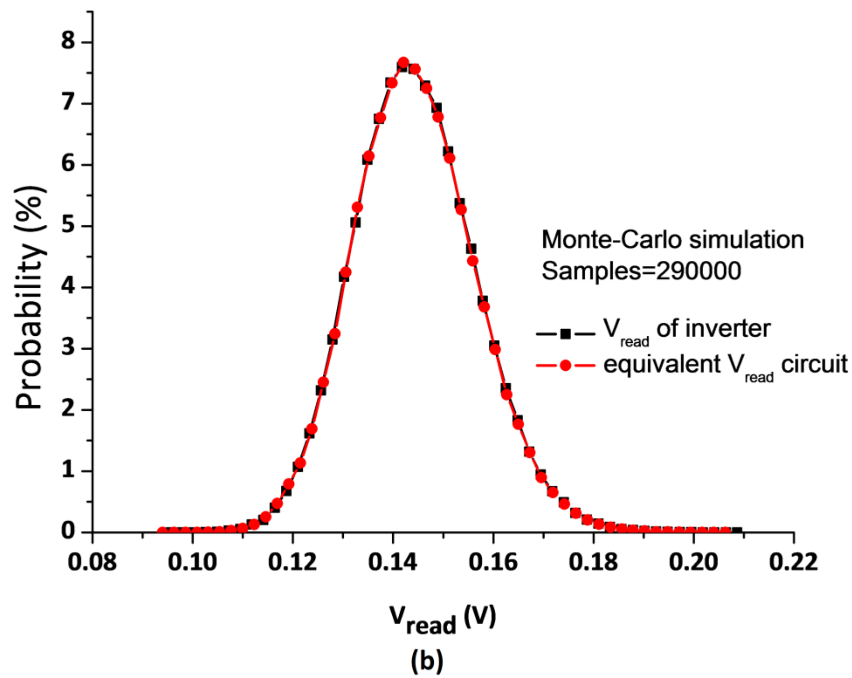


Fig. 4.17 (a) Vread cell schematic (b) Vread cell layout (c) Vread equivalent circuit

Fig. 4-17(a) and 4-17(b) shows the Monte Carlo simulation comparison to the modified cell of Vtrip (Vread) between its equivalent circuit. From the simulation result of Fig.5, the modified cell of Vtrip (Vread) has a good approach to its equivalent circuit. This mean we can use the modified cell to perform the proper measurement.

To maintain the compact environment and density, we only modified layers above Metal 1. Poly, Diffusion, and contact is unmodified and the same as a traditional 6T SRAM cell.

4.4 WM measurement methodology

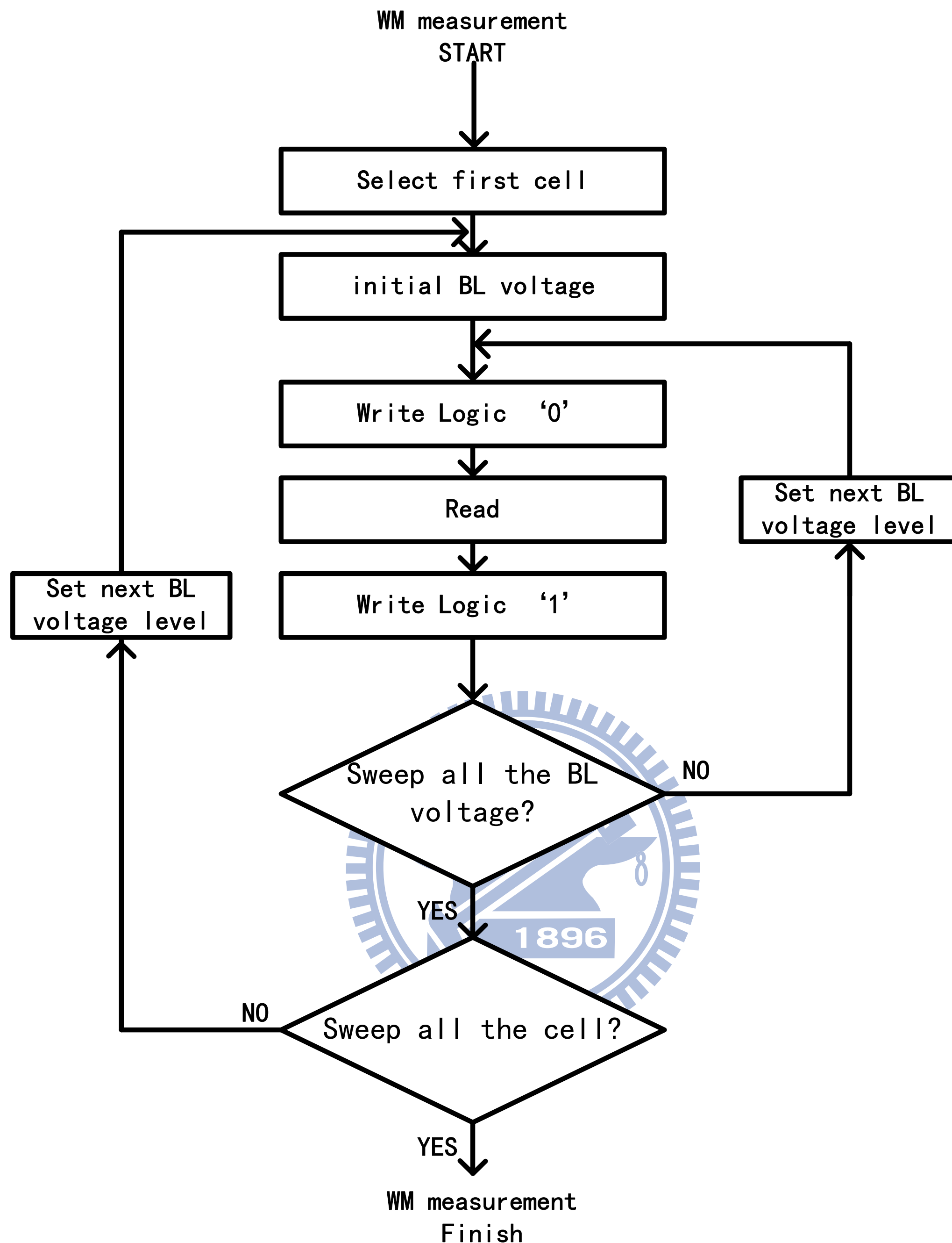


Fig. 4.18 Write Margin measurement flow

In order to get the write margin of a 6T SRAM cell, we can sweep the bit-line voltage from supply voltage VDD to ground voltage gnd to check that at which BL voltage the cell is first time written in a value. And this BL voltage of this time could be the write margin WM of this cell. For our write margin test, we use unmodified 6T cell. Our write margin methodology is that: 1) first set BL voltage level 2) turn on the word-line to write 3) read out to check write success or write fail. This progress would

continue until the BL voltage is set to ground level. So that we just to check the read record of this cell and find the first time write success BL voltage to get the write margin.

4.5 Array Implementation

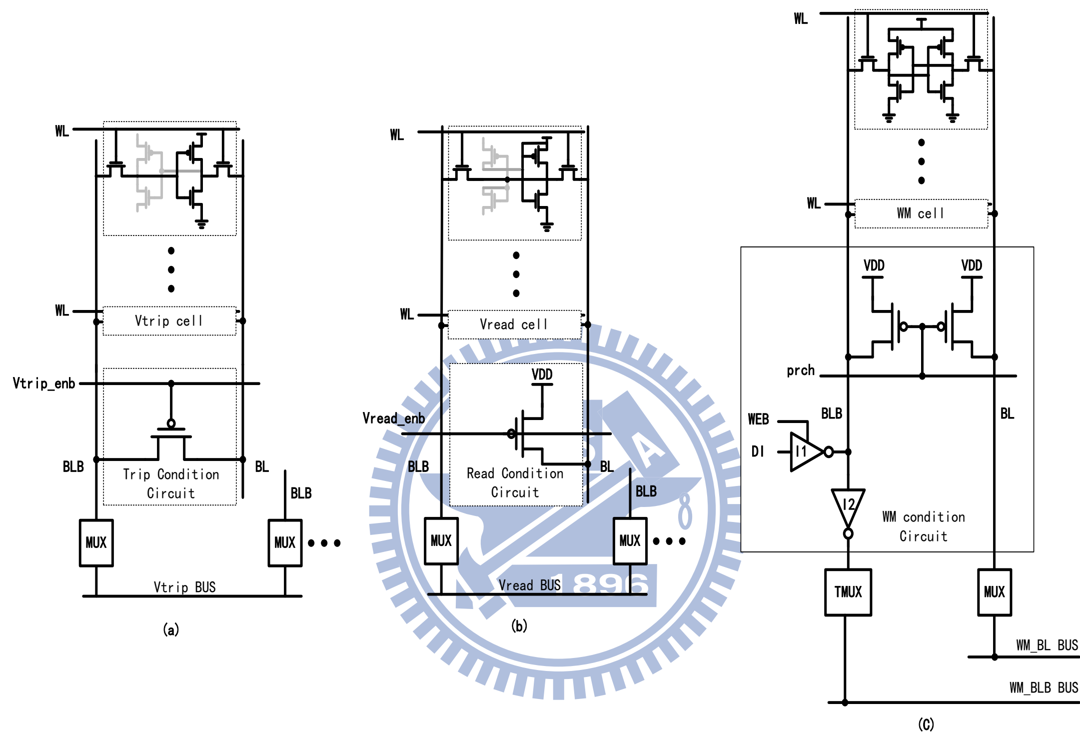


Fig. 4.19 (a) Vtrip column and trip condition circuit (b) Vread column and read condition circuit (c) WM column and WM condition

Since we modified the cell for each V_{trip} and V_{read} measurement, we still need extra conditional circuit to help to get those voltage information. For V_{trip} cell, we have tapped out the input and output through two pass-gates and connect to the BL and BLB. Fig.4-19 (a) shows the conditional circuit of V_{trip} measurement that could short BL and BLB together if the V_{trip} measurement enable signal is active. This V_{trip} conditional circuit contains only one PMOS and is just like the equalizer PMOS that

is connected to BL and BLB in the real SRAM. Once the word-line is selected, the corresponding V_{trip} of the selected cell will pass to the BL and then pass to the bus through the MUX.

Fig.4-19 (b) shows the condition circuit of V_{read} measurement. Because we have to clamp the BL at VDD to form a divide voltage V_{read}, the conditional circuit of V_{trip} is a PMOS which is like the pre-charge PMOS in the regular SRAM design. Once V_{read} measurement enable signal is active, the PMOS will turn and BL will keep high. The selected word-line turns on the pass-gate of V_{read} cell to generate V_{read}, and V_{read} is passed through the opposite side pass-gate to BLB and transmit to bus through the MUX.

Fig.4-19 (c) demonstrates the condition circuit of WM test. For WM measurement, we use unmodified tradition 6T standard cells. So the WM condition is the same as the simplified read/write of regular SRAM design. For each set of test, we must write the cell into a logic "0" with different BL voltage. Different BL voltage is generated from the voltage divide circuit outside of array (we will discuss at part.) and pass through the MUX and bus to the target BL. The opposite side BLB will clamp at VDD through device I1 to perform write '0' operation. The second step is read operation, so we adopt single ended read method to prevent the implementation of latch-type sense amplifier. Fig.4-19 (c) shows the I2 is the read inverter and the M3-M4 is the pre-charge PMOS. Just as the regular SRAM operation, first we generate a pulse to pre-charge the BL and BLB to VDD before the word-line is active. If word-line is selected, the cell start to discharge BLB if QB is '0' or keep at high if QB is '1'. And the I2 will sense the voltage level to read out logic '0' or '1' of BLB and pass to the MUX. The last step of WM test is Write logic '1' back to the cell. Contradict to the Write '0' process, the BL is at voltage VDD now, and the BLB is set to ground voltage by device I1 to perform write '1' process.

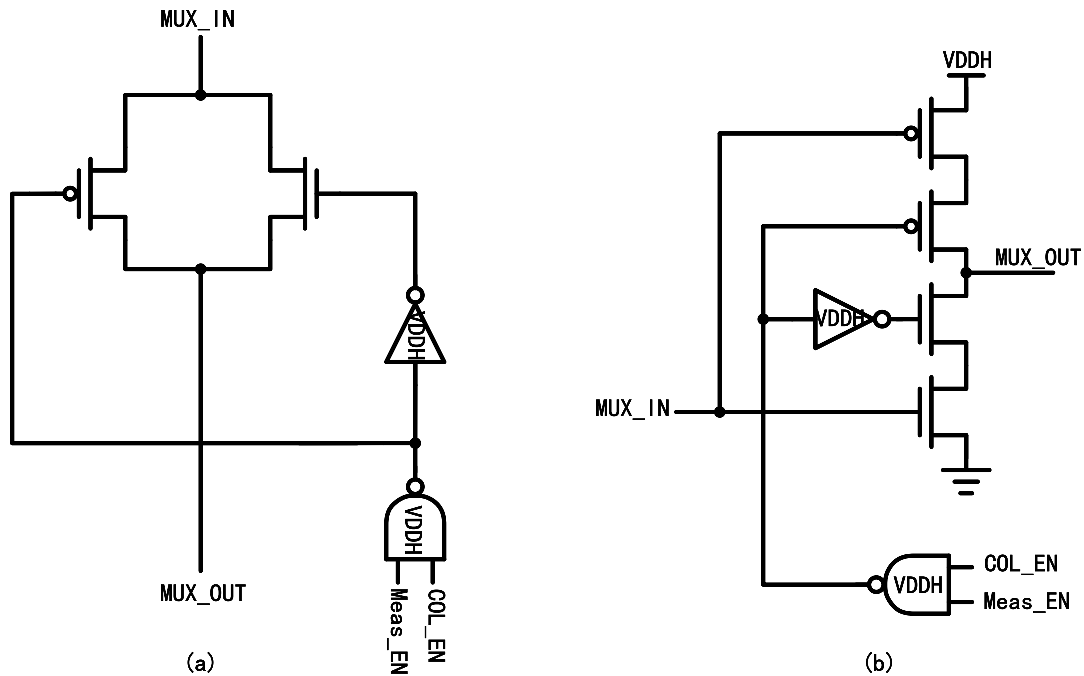
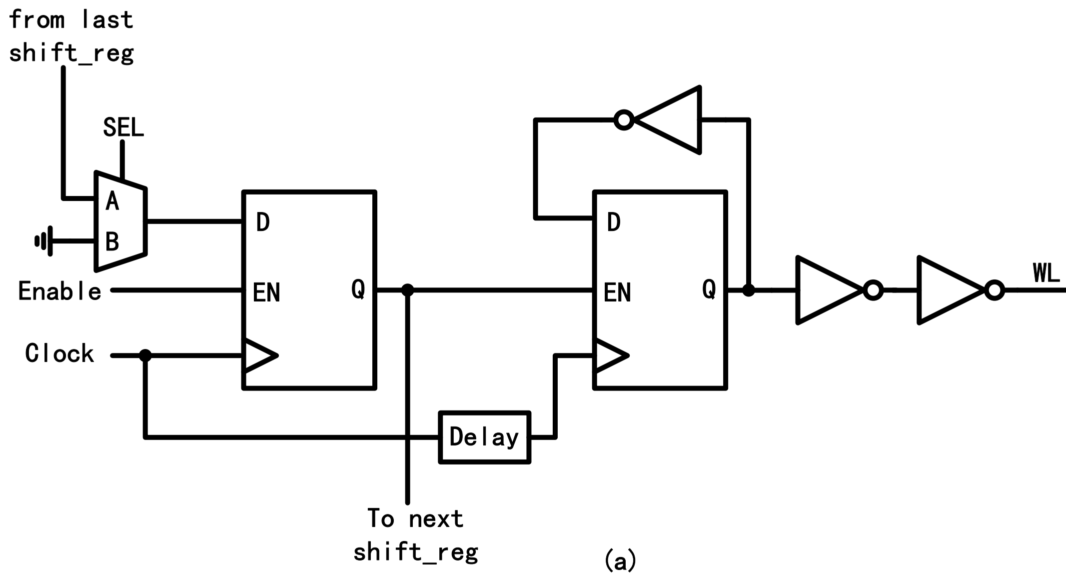


Fig. 4.20 (a) Vtrip column and trip condition circuit (b) Vread column and read condition circuit (c) WM column and WM condition

Fig.4-20 show the Mux design of Vtrip, Vread, and WM test. Because we must pass in/out a analog voltage, so we use a transmission gate type Mux to transmit precise voltage (Fig.4-20 (a)). To prevent the voltage level degradation through the transmission gate type Mux, we raise the control voltge to 1.2V to achieve precise voltage. But all the voltage of cells supply and voltage transmission path is at voltage 1.0V. For our implementation, every column of Vtrip column are connect to the Vtrip bus through the MUX of each column and Vread column does too. Unlike the Vtrip and Vread column which we only MUX the BLB, both of BL and BLB of WM column have to connect to each bus through MUXes. As a result we have 4 BUS : Vtrip bus, Vread bus, WM_BL bus, and WM_BLB bus.



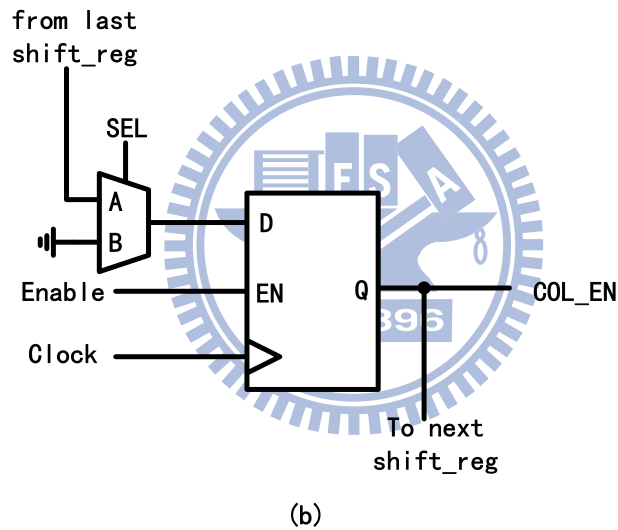


Fig. 4.21 (a) Modified Row direction shift register (b) Modified Column direction shift register

To reduce the control signal and input address pin, we simply the row/column decoder with shift register. (Fig.4-21) Both row and column will shift to next by the clock signal. This could allow us to sweep all the word-line and column only by two signals (each for row/column) rather a huge number of address pin. However for the WM measurement, we have to continuously turn on one row for the write-read-write process with different BL voltage, so we further modify the row shift register to

provide a continuous switch on/off mechanism. Fig. Two set of shift register is used for a row shift register. Set A controls that which row should be selected and Set B controls the regular switch mode or continuously switch mode. With proper control signal, the word-line could perform two modes. Fig. demonstrates the row decoder. For our implementation, each Vtrip, Vread, and WM column owns independent set of column shift register.

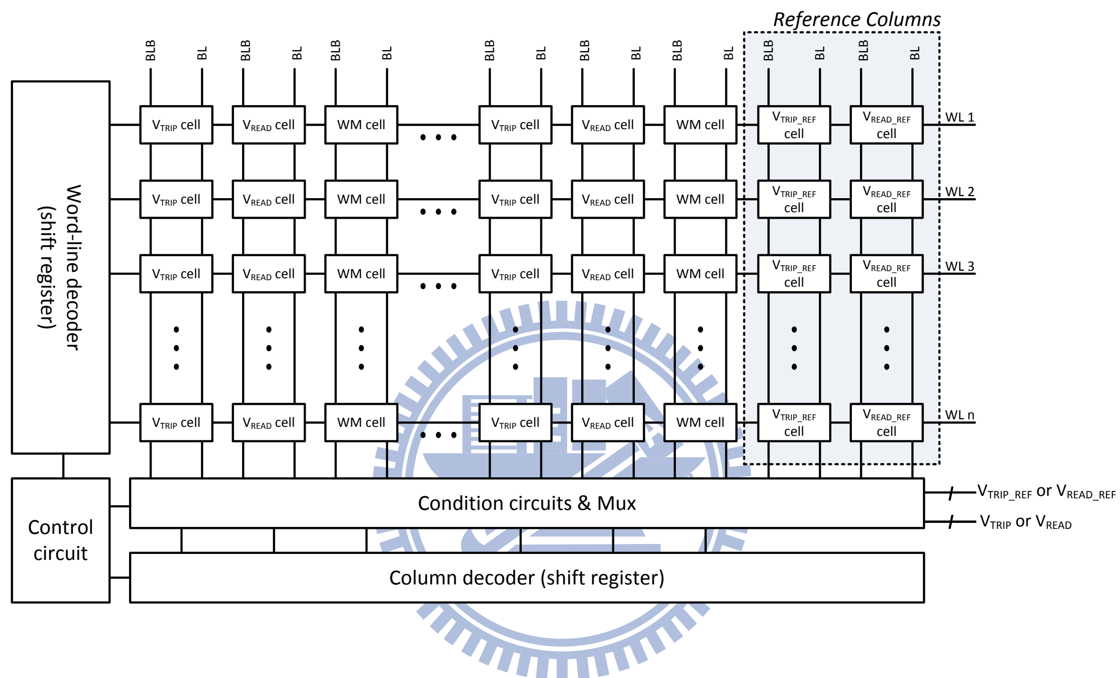


Fig. 4.22 Measurement Array Architecture

Fig.4-22 shows the placement of these cell columns. Each measurement cell occupies its own column and 128 cells along a column share one pair of BL and BLB. And we interleaving place the Vtrip, Vread, and WM column rather than place all the cells together. These could allow us to test cells at different location of a die to exam the local variation distribution. Because local variation is spatial independent and a normal distribution, so the statistical distribution of measurement result of Vtrip, Vread, WM should form a bell shape of Gaussian distribution. The condition circuit and MUX are placed at below and close to these cell columns. And the row/column circuit is placed next to these circuits. One set of a cell test contains one Vtrip, one

Vread, and one WM cell. We have 85 set which means we have 3*85 columns and 128 word-lines in a bank. So we totally have about 128*85*3~32K cell in a bank. And we have 16 banks, so the total cell numbers are 16*32K ~ 512kb. For each Vtrip, Vread, or WM test, we can get about 128*85*16~166.4K samples. Beside we add a second level Mux to connect the bus of each bank together.

4.6 Measurement methodology and circuit blocks

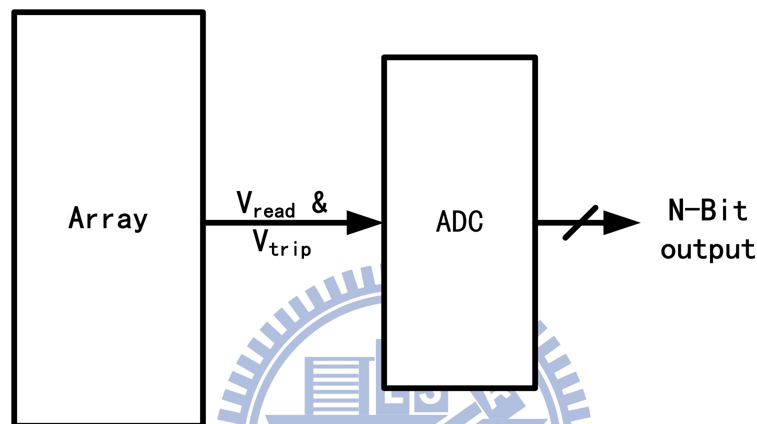


Fig. 4.23 ADC Measurement Scheme

As discuss before, we have implement a 512kb modified cell array, so we can get voltage value of Vtrip and Vread from the bus of each bank. However, it is difficult to directly get precise a voltage value from the packaged chip. Traditional way to measure a chip is to directly probe the nacked die. But probe method is hard to be automation and the measurement equipment always expensive that makes the measurement cost lots of money and time. In order to save the cost, we adopt the digital measurement scheme to measure our target voltage value. Since the digital measurement scheme is used, all the input and output will be digital pin which could be automatic control and easily retrieve voltage value from digital output. One method is to implement the Analog-to-Digital-Converter (ADC) to realize the digital measurement scheme. However, ADC is quite difficult to design and ADC always

occupies a large portion area and power in the whole system. In this work, we adopt another measurement scheme of VCO which is proposed from previous work [17][18]. We only implement the operational amplifier, VCO, and digital counter to prevent the complex design of ADC.

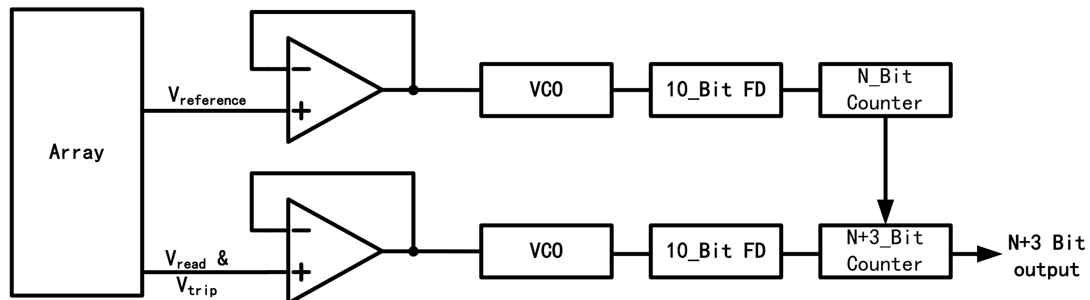


Fig. 4.24 VCO + Counter Based scheme

The basic idea of this measurement scheme which is proposed by [17][18] is to transform the analog voltage to frequency by VCO, and set a specific time period T to count the cycle number M. Then we divide M by T, we can get the frequency period of VCO. If we have the mapping table of frequency to input voltage, we would retrieve the measured voltage such as Vtrip and Vread by this relation without using ADC. In this work, we build two path : reference path and measurement path. Both reference path and measurement path have the same circuit component. However the difference of these two path is that the input voltage of reference path is a specific voltage from the cell in dummy column we build in each bank but the input voltage of measurement path can be the Vtrip (or Vread) of any other cell in the array. And the reference path is to provide a time base T to allow the measurement path to count the frequency cycle numbers which reflect the absolute value of Vread and Vtrip. If there exists voltage difference between two pathes, the counting number M would change and we still could get the value by mapping table. So the counting bit decides the

resolution of this scheme and we should set it a proper number to fit the resolution specification.

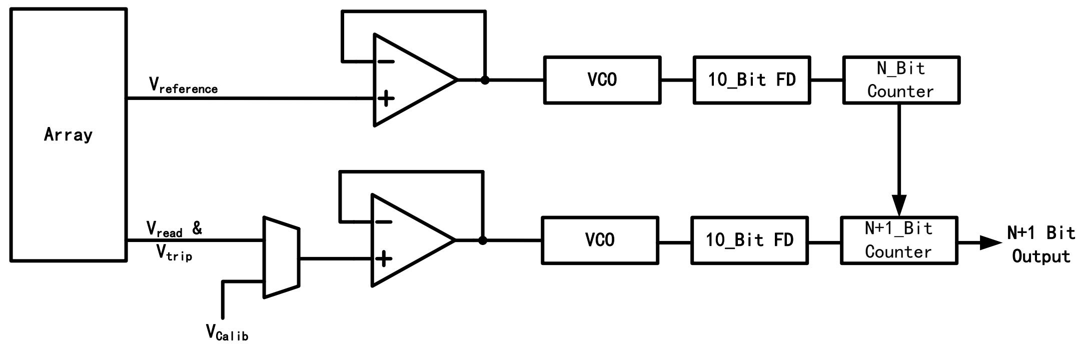


Fig. 4.25 Building blocks of voltage measurement scheme

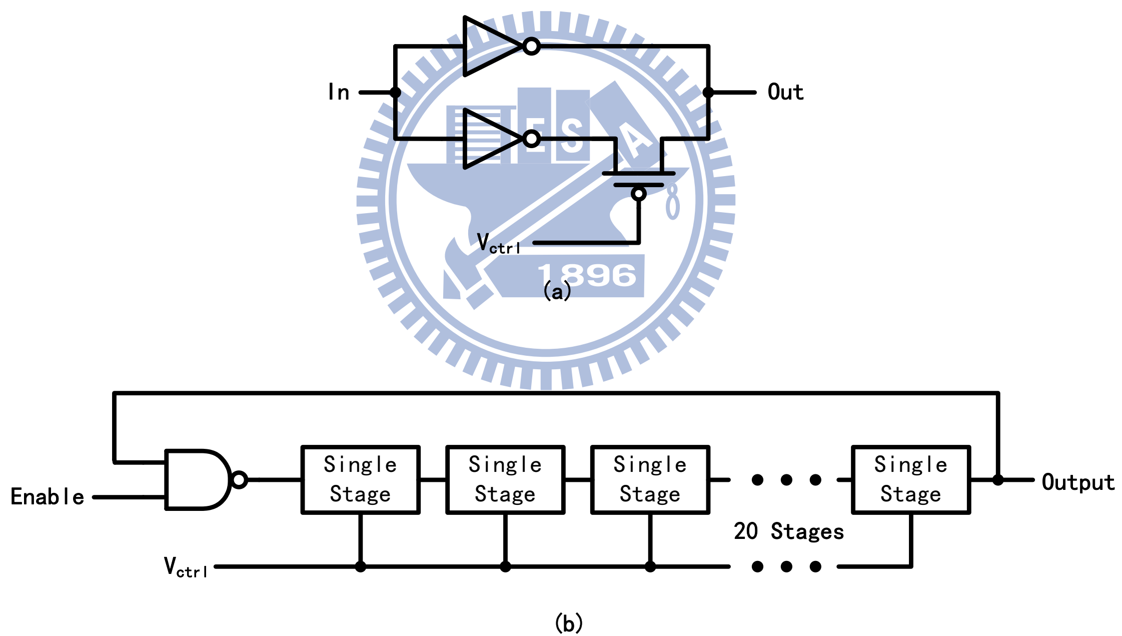


Fig. 4.26 (a) Single stage of VCO (b) 21 stages VCO

Fig. 4-25 shows the overall component of this measurement scheme. And Fig. 4-26(a) shows the circuit schematic of single stage of the VCO. We connect measured voltage to the gate of MOSFET to control the current drive ability, so that the delay of each stage will change with input voltage. However, the measured Vtrip and Vread

always be a lower level from VDD and we hope the output frequency has higher sensitivity to control voltage, we use PMOS as the control of the delay chain. As the input voltage rise, the delay and the output frequency period increase. Fig shows the voltage to frequency plot of this VCO. We can achieve approximately 480 MHz/mV (Fig.4-27) sensitivity to this VCO implementation.

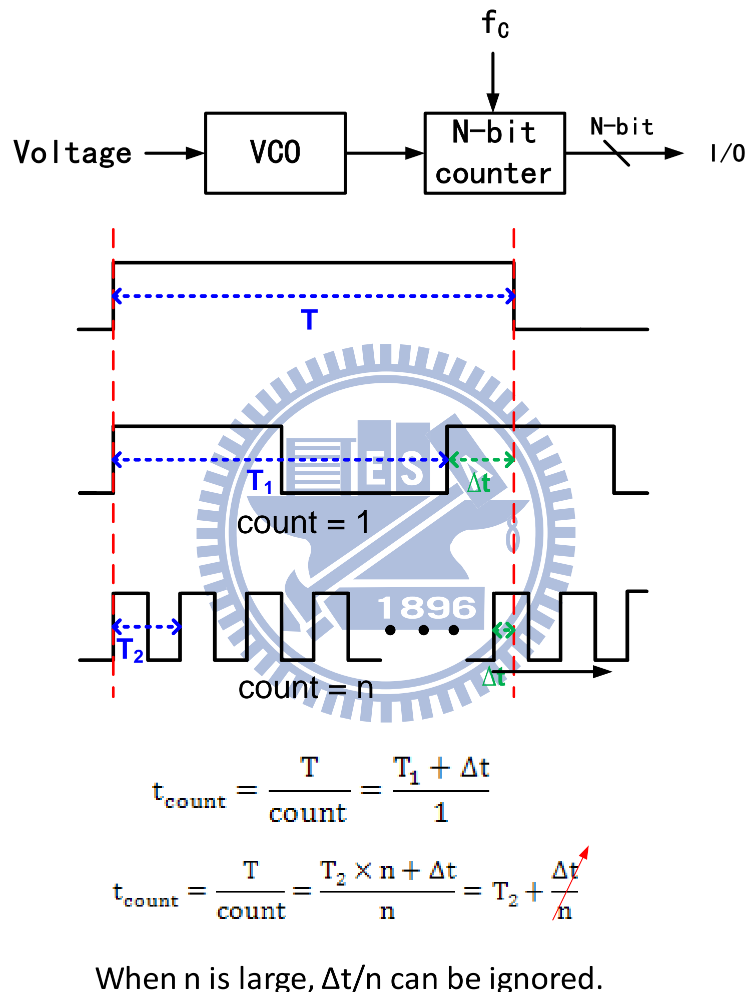


Fig. 4.27 VCO with counter based measurement scheme concept

Every bus of each bank would be mux together before we connect it to the voltage input of VCO. In order to reduce the noise and the couple effect, we add a unit-gain buffer between them. Unit gain buffer is composed by a feedback connected

OPA and shows in the Fig. This OPA should have larger gain to reduce the voltage difference between the input and output. And the input voltage range should prefer lower voltage level because Vtrip and Vread usually no more than 400~500mV. As for the frequency bandwidth, we don't have specific request. Besides, the output frequency of VCO reaches several hundred MHz, so we divide the frequency about 1000 by a 10 bit frequency divider (FD) before we connect it to counter.

We set counter of the reference path as a 13 bit counter, and the counter of measurement path is a 14 bit counter. This could allow us to have a resolution about 0.167mV to cover the voltage-frequency resolution.

For WM measurement, we realize one set of resistor voltage divider (Fig. 14) which has 64 voltage levels with 10mV of each step from 0mV to 640mV for sweeping BL level. Because the WM usually not a high value, so we choose a narrower range to increase resolution. Besides, we simply use the poly resistor to form a resistor voltage divider so that we can prevent the resistor variations. One NMOS is connected to the voltage divider as a port to access the each voltage level. Since we have 64 voltage level, a 6 bit decoder is implemented for selection. So that we can control these 6 bit input to set BL voltage level by the outside equipment. Furthermore the output of these resistor voltage dividers is connected to a unit-gain buffer to decouple noise. And this voltage would be directly connected to the wm_BL bus.

4.7 Calibration

Generally, the total measurement can partition into four parts: (1) calibration of VCO, (2) V_{TRIP} measurement, (3) V_{READ} measurement, and (4) WM measurement. The most important of all is calibration of VCO. Since that we don't know the characteristic curve of VCO's V_C and frequency in real chip, we need to do some

operations to caliber the characteristics of VCO. During calibration mode, we sweep V_C voltage, and observe the corresponding digital-bit result on the output. We record the VC and the mapping output into a mapping table.

After finishing measurement, we just need to collect the output results of V_{TRIP} and V_{READ}, then check the mapping table and find out the respective voltage level before VCO. With this approach, we can get the V_{TRIP} and V_{READ} easily without need to calculate the frequency from output results and revert it back to voltage. Most important of all, even if the curve of frequency to V_C is non-linear, we still can get the correct data. Hence, we don't need to have a lot of efforts to design the linearity of VCO.

4.8 Test Chip Implementation

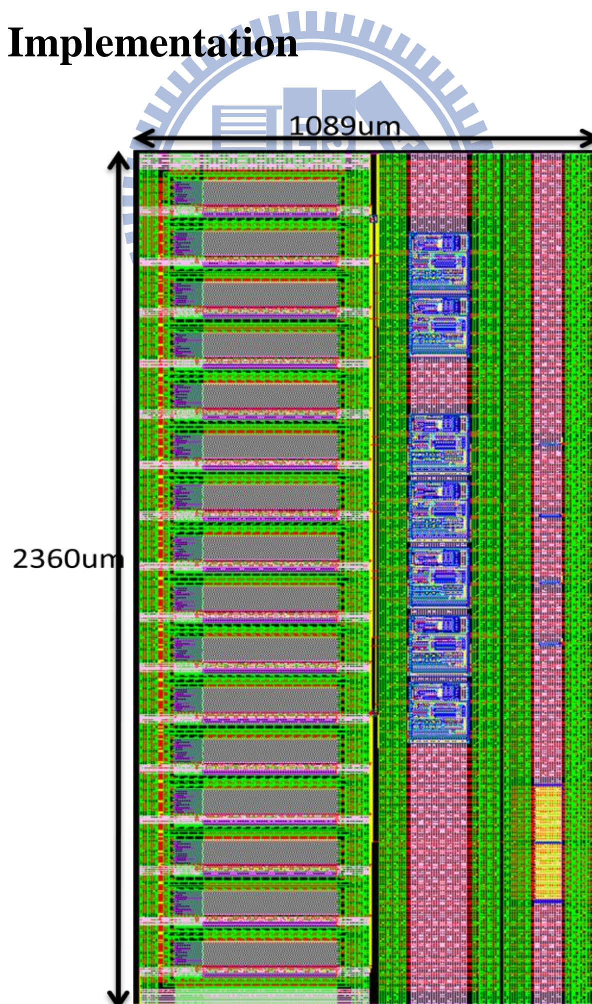


Fig. 4.28 Layout of present 512k Test Macro

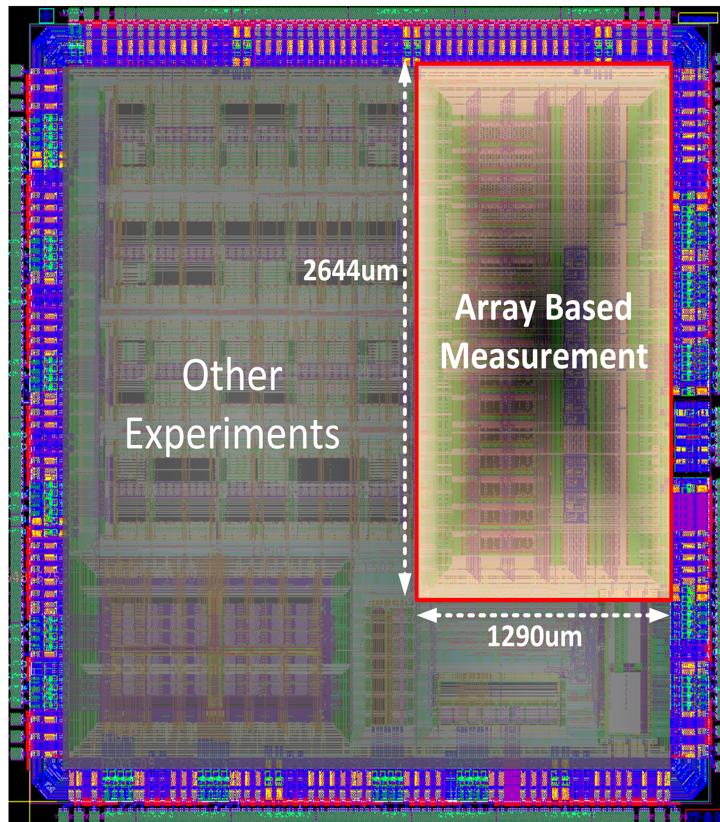


Fig. 4.29 Test Macro Location in Test Chip

Fig. 4-28 shows the overall layout view and Fig.4-29 shows the test chip location. We can see that we partition the 512kb cells array into 16 banks and place it at left half of this die. Array has two power domain : 1.0V for cell array and 1.2V for control logic. The array occupies the largest part of area. Since we have to get the Vread and Vtrip separately, we implement two set of measure circuits. So total number of OPA is seven (4 for two set voltage circuit, one for voltage divider, and two for calibration) . All the OPA is locate at the middle strip of the chip with a 3.3 V power supply. Four VCO and resistor voltage divider are implemented at the right side of the chip and owns their individual power supply. The overall die area is 2555umx1333um with 27 input pin and 7 power domain.

Chapter 5

Conclusions

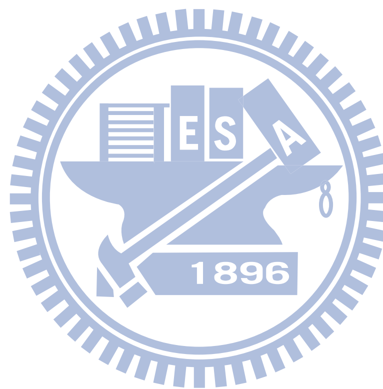
For the past decades, the Moore's law has a good match to the CMOS technology development trend. The process has been pushed to the 20nm in the industry. Although the single core Base Band Digital system has bound to a limit due to the several issue as leakage, etc, new approach such as multi-core applicant has emerging at recent years. With network technology spread wide, new coming clouding computing is also an important issue. As a result, the high speed application has not stopped its progress. However, multi-core system and clouding system always need large amount data access process which has to assist with memories, especially SRAM. This result memory always dominant the overall speed of one system.

With the beyond 100nm technology, memory (SRAM) design suffers variation issue. SRAM no longer easy survived at novel technology due to several issue such as local random variation, leakage, etc. To allow SRAM fully functionally operation at novel technology and without loss the operation speed, or cause too much power consumption is the main stream of current SRAM design.

In this thesis we discuss two adaptive read/write ability improvement circuit techniques. As shown in simulation result, we could successfully push the read/write VCC_min, and allow this design to keep higher stability at nominal voltage operation. Both the WLUD and DAWA are implemented in full CMOS technology that allow lesser cost. And the corner tracking ability could resist the die-to die variation that could maximize the read/write improvement. With Both implemented two circuit

techniques, operation speed has just slightly degraded. This means we still could keep the original operation speed.

Another part of this thesis discuss one RSNM/WM testing macro in 6T SRAM Array. To build in a large array macro could let us collect enough amount data to analyze the statistical distribution of RSNM/WM. With modified RSNM cell and proposed Vtrip and Vread measurement scheme, we could can easily to get the extremely case of read disturb and write margin. With the aid of voltage measurement scheme, all the control, In/output are fully digital. This means we could all use simple computer control without expensive measurement equipment. All the measurement could be digital automation which could save the testing time and money cost.



Reference of Chapter 2

[2-1] Adel S. Sedra, Kenneth C. Smith, “Microelectronic Circuits” 5rd ed. Oxford University Press, 2003.

[2-2] Neil H.E. Weste, and David Harris, “CMOS VLSI DESIGN – A Circuit and System Perspective” 3rd ed. Addison Wesley, 2004.

[2-3] T. Fischer, E. Amirante, P. Huber, T. Nirschl, A. Olbrich and M. Ostermayr et al., Analysis of read current and write trip voltage variability from a 1-mb sram test structure, IEEE Trans Semiconduct Manuf 21 (4) ,2008, pp. 534–541.

[2-4] Ching-Te Chuang et al., “High-Performance SRAM in Nanoscale CMOS: Design Challenges and Techniques,” IEEE International Workshop on Memory Technology, Design and Testing, 2007, pp.4-12.

[2-5] Sridhar Ramalingam, Elakkumanan Praveen, Natarajan Sreedhar, “Tutorial 6: Design Challenges and Solutions for Nanoscale Memories”, IEEE International Symposium on Circuits and Systems, 2007, nil28 - nil29

[2-6] Jayakumaran Sivagnaname, Hung C. Ngo, Kevin J. Nowka, Robert K. Montoye and Richard B. Brown,” Study of Wide LSDL Circuit Implementations”, 19th International Conference onVLSI Design, 2006. Held jointly with 5th International Conference on Embedded Systems and Design., 2006, pp.6 .

[2-7] Pille, J. et al., “Implementation of the CELL Broadband Engine in a 65nm SOI Technology Featuring Dual-Supply SRAM Arrays Supporting 6GHz at 1.3V”, IEEE International Solid-State Circuits Conference, 2007, pp. 322 – 606

[2-8] A. Bhavanagarwala, X. Tang, and J. Meindl, "The impact of intrinsic device fluctuations on CMOS SRAM cell stability," IEEE J. Solid State Circuits, Apr. 2001, vol. 36, no. 4, pp. 658-665.

[2-9] M. Khellah et al., “A 4.2GHz 0.3mm2 256kb Dual-Vcc SRAM Building Block in 65nm CMOS,” Digest of Tech. Papers, ISSCC, 2006, pp. 624-625.

[2-10] Yabuuchi M., Nii K., Tsukamoto Y., Ohbayashi S, Nakase Y, Shinohara H., “A 45nm 0.6V cross-point 8T SRAM with negative biased read/write assist”, Symposium on VLSI Circuits, 2009, pp. 158 – 159

[2-11] K. Nii et al., “A 45-nm Bulk CMOS Embedded SRAM With Improved Immunity Against Process and Temperature Variations,” IEEE JOURNAL OF SOLIDSTATE CIRCUITS, JANUARY 2008 , VOL. 43, NO. 1.

[2-12] M. Alam and S. Mahapatra, “A comprehensive model of PMOS NBTI degradation,” Microelectron. Reliab., vol. 45, no. 1, pp. 71–81, 2005.

[2-13] Fujimura, Y. et at. “A Configurable SRAM with Constant-Negative-Level Write Buffer for Low-Voltage Operation with 0.149µm2 Cell in 32nm High-Metal-Gate CMOS”, ISSCC 2010, pp. 348 – 349

[2-14] K. Kang, H. Kufluoglu, K. Roy and M. Alam, "Impact of negative-bias temperature instability in nanoscale SRAM array: modeling and analysis", computer-aided design of integrated circuits and systems, IEEE Trans 26 (10) (2007), pp. 1770–1781.

[2-15] K. Kang, S.P. Park, K. Roy and M.A. Alam, “Estimation of statistical variation in temporal NBTI degradation and its impact on lifetime circuit performance”, IEEE ICCAD (2007), pp. 730–734.

[2-16] Stathis JH, Zafar S. “The negative bias temperature instability in MOS devices: a review”. Microelectron Rel 2006; pp. 46:270–86.

[2-17] N. Kimizuka, T. Yamamoto, T. Mogami, K. Yamaguchi, K. Imai, and T.

Horiuchi, "The impact of bias temperature instability for direct-tunneling ultra-thin gate oxide on MOSFET scaling," IEEE Symposium on VLSI Technology Digest of Technical Papers, pp. 73-74, 1999.

[2-18] Alam M. A critical examination of the mechanics of dynamic NBTI for p-MOSFETs. In: Proc Int Electron Device Meet, 2003. p. 346–9.

[2-19] S. V. Kumar, K. H. Kim, and S. S. Sapatnekar, "Impact of NBTI on SRAM read stability and design for reliability", in Internal Symposium for Quality Electronic Design (ISQED), pp. 27-29, March 2006.

[2-20] B. C. Paul, K. Kang, H. Kufluoglu, M. A. Alam, and K. Roy, "Impact of NBTI on the temporal performance degration of digital circuits," IEEE Electron Device Letters, vol. 26, n. 8, pp. 560--562, Aug., 2005.

[2-21] A. Bansal et. al., "Impact of NBTI and PBTI on SRAM static/dynamic noise margins and cell failure probability," Microelectronics Reliability, Vol 49, pp. 642-649, 2009.

[2-22] Ming-Chien Tsai, "NBTI/PBTI Degradation and Noise Margin Measurement Circuit of Nano-scale CMOS SRAM", master thesis of Department of Electronics Engineering of National Chiao Tung University, 2010.

Reference of Chapter 3

[3-1] International Technology Roadmap for Semiconductors, ITRS, http://public.itrs.net

[3-2] S. Mukhopadhyay et al., Modeling of Failure Probability and Statistical Design of SRAM Array for Yield Enhancement in Nanoscaled CMOS, TCAD, pp1859-1880, Dec. 2005.

[3-3] C.-T. Chuang et al., “High-Performance SRAM in Nanoscale CMOS: Design Challenges and Techniques,” IEEE International Workshop on Memory Technology, Design and Testing, 2007, pp.4-12.

[3-4] Fujimura, Y., “A configurable SRAM with constant-negative-level write buffer for low-voltage operation with 0.149μm2 cell in 32nm high-k metal-gate CMOS”, Digest of Tech. Papers, ISSCC, 2010, pp. 348-349.

[3-5] K. Zhang et al., “A 3-GHz 70Mb SRAM in 65nm CMOS Technology with Integrated Column-Based Dynamic Power Supply,” Digest of Tech. Papers, ISSCC, 2005, pp. 474-475.

[3-6] M. Yamaoka et al., “Low-Power Embedded SRAM Modules with Expanded Margins for Writing,” Digest of Tech. Papers, ISSCC, 2005, pp. 480-481.

[3-7] Yamauchi Hiroyuki, et al., "A Differential Cell Terminal Biasing Scheme Enabling a Stable Write Operation against a Large Random Threshold Voltage (Vth) Variation," IEICE - Transactions on Electronics, 2006, pp:1526-1534..

[3-8] Toshikazu Suzuki, et al., "A Stable SRAM Cell Design Against Simultaneously R/W Disturbed Accesses," Digest of Tech. Papers, Symp. VLSI Circuits, 2006, pp. 11-12.

[3-9] Meng-Fan Chang, et al., "A Differential Data Aware Power-supplied (D2AP) 8T SRAM Cell with Expanded Write/Read Stabilities for Lower VDDmin Applications," Dig. Tech. Papers, Symp. VLSI Circuits, 2009, pp. 156-157

[3-10] Ajay Bhatia, "Memory Cells with Power Switch Circuit for Improved Low Voltage Operation," U. S. Patent US2009/0016138 A1, Pub. Date: Jan. 15, 2009.

[3-11] Pilo, H. et al., "An 833 MHz 1.5 W 18 Mb CMOS SRAM with 1.67 Gb/s/pin", Digest of Tech. Papers, ISSCC, 2005, pp. 266-267.

[3-12] Pille, J., "A 32kB 2R/1W L1 data cache in 45nm SOI technology for the POWER7TM processor", Digest of Tech. Papers, ISSCC, 2010, pp. 344-345.

[3-13] Hirabayashi, O., "A process-variation-tolerant dual-power-supply SRAM with 0.179µm2 Cell in 40nm CMOS using level-programmable wordline driver", Digest of Tech. Papers, ISSCC, 2009, pp. 458-459.

[3-14] Hyunwoo Nho et al., "A 32nm High-k metal gate SRAM with adaptive dynamic stability enhancement for low-voltage operation", Digest of Tech. Papers, ISSCC, 2010, pp. 346-347.

[3-15] S. Ohbayashi, et al., "A 65-nm SoC Embedded 6T-SRAM Designed for Manufacturability with Read and Write Operation Stabilizing Circuits", JSSC, vol 42, No. 4, pp.820-829, April 2007.

[3-16] Nii, K. , "A 45-nm Bulk CMOS Embedded SRAM With Improved Immunity Against Process and Temperature Variations", IEEE JSSC vol.43, no.1, pp. 180-191.

[3-17] Nii, K., "A 45-nm single-port and dual-port SRAM family with robust read/write stabilizing circuitry under DVFS environment", Digest of Tech. Papers, Symp. VLSI Circuits, 2008, pp. 212-213.

[3-18] Yabuuchi M., Nii K., Tsukamoto Y., Ohbayashi S., Nakase Y., Shinohara H., "A 45nm 0.6V Cross-Point 8T SRAM with Negative Biased Read/Write Assist", Symposium on VLSI Circuits, 16-18 June 2009, pp. 158 - 159.

[3-19] S. Mukhopadhyay, R. Rao, J. J. Kim, and C. T. Chuang, "Capacitive Coupling Based Transient Negative Bit-line Voltage (Tran-NBL) Scheme for Improving Write-ability of SRAM Design in Nanoscale Technologies," Proc. IEEE International Symposium on Circuits and Systems (ISCAS), Seattle, Washington, May 18-21, 2008, pp. 384-387.

[3-20] D. P. Wang, H. J. Liao, H. Yamauchi, Y. H. Chen, Y. L. Lin, S. H. Lin, D. C. Liu, H. C. Chang, and W. Hwang, " A 45nm Dual-Port SRAM with Write and Read Capability Enhancement at Low Voltage," Proc. International SoC Conf., 2007, pp. 211-214.

Reference of Chapter 4

[4-1] Neil H.E. Weste, and David Harris, "CMOS VLSI DESIGN – A Circuit and System Perspective" 3rd ed. Addison Wesley, 2004.

[4-2] K. Takeda et al. Redefinition of write-margin for next generation SRAM and write-margin monitoring circuit. International Solid-State-Circuit Conference, 6-9 Feb. 2006,pp. 2602-2603.

[4-3] E. Seevinck, et al., "Static-noise margin analysis of MOS SRAM cells", IEEE J. Solid-State Circuits, vol. SC-22, no. 2, pp. 748 - 754, 1987.

[4-4] Khellah M., Khalil D.E., Somasekhar D., Ismail Y., Karnik T., De, V., "Effect of Power Supply Noise on SRAM Dynamic Stability," IEEE Symposium on VLSI Circuits, 14-16 June 2007, pp. 76 - 77.

[4-5] Xiaowei Deng, Wah Kit Loh, Pious B., Houston T.W., Liu L., Bashar Khan, Corum D., "Characterization of bit transistors in a functional SRAM," IEEE Symposium on VLSI Circuits, 18-20 June 2008, pp. 44 - 45.

[4-6] Zheng Guo, Carlson A., Liang-Teck Pang, Duong K., Tsu-Jae King Liu, Nikolic B., "Large-scale read/write margin measurement in 45nm CMOS SRAM arrays," IEEE Symposium on VLSI Circuits, 18-20 June 2008, pp. 42 - 43.

[4-7] Fischer T., Amirante E., Hofmann K., Ostermayr M., Huber P., Schmitt-Landsiedel D., "A 65nm test structure for the analysis of NBTI induced statistical variation in SRAM transistors," ESSDERC Solid-State Device Research Conference, 38th European, 15-19 Sept. 2008, pp. 51 - 54.

[4-8] Thomas Fischera, Ettore Amiranteb, Peter Huberb, Karl Hofmannb, Martin Ostermayrb and Doris Schmitt-Landsiedela, "A 65 nm test structure

for SRAM device variability and NBTI statistics," Solid-State Electronics, Volume 53, Issue 7, July 2009, pp. 773-77.

[4-9] Drapatz S., Fischer T., Hofmann K., Amirante E., Huber P., Ostermayr M., Georgakos G., Schmitt-Landsiedel D., "Fast stability analysis of large-scale SRAM arrays and the impact of NBTI degradation," Proceedings of ESSCIRC, 14-18 Sept. 2009, pp. 92 – 95.

[4-10] Fischer T., Amirante E., Huber P., Nirschl T., Olbrich A., Ostermayr M., Schmitt-Landsiedel D.," Analysis of Read Current and Write Trip Voltage Variability From a 1-MB SRAM Test Structure," IEEE Transactions on Semiconductor Manufacturing, Nov. 2008, Volume 21, Issue 4, 6-10 Feb. 2005, pp. 534 - 541.

[4-11] Rao R., Jenkins K.A., Jae-Joon Kim, "A Completely Digital On-Chip Circuit for Local-Random-Variability Measurement," IEEE Solid-State Circuits Conference, 3-7 Feb. 2008, pp. 412 - 623

[4-12] Rao R., Jenkins K.A., Jae-Joon Kim, "A Local Random Variability Detector With Complete Digital On-Chip Measurement Circuitry," IEEE Journal of Solid-State Circuits, Sept. 2009, Volume 44, Issue 9, pp. 2616 – 2623.

[4-13] Keane J., Venkatraman S., Butzen P., Kim C.H., "An array-based test circuit for fully automated gate dielectric breakdown characterization," IEEE Custom Integrated Circuits Conference, 21-24 Sept. 2008, pp. 121 - 124

[4-14] Tae-Hyoung Kim, Persaud R., Kim C.H., "Silicon Odometer: An On-Chip Reliability Monitor for Measuring Frequency Degradation of Digital Circuits," IEEE Solid-State Circuits, April 2008 , pp. 874 - 880

[4-15] Keane John, Persaud Devin, Kim Chris H., "An all-in-one silicon Odometer for separately monitoring HCI, BTI, and TDDB," Symposium on

VLSI Circuits, 16-18 June 2009, pp. 108 - 109.

[4-16] S. R. Nassif, "Modeling and analysis of manufacturing variations,"in Proc. Custom Integrated Circuit Conf., San Diego, CA, 2001,pp. 223–228.

[4-17] S. Borkar, T. Karnik, S. Narendra, J. Tschanz, A. Keshavarzi, and V. De,"Parameter variation and impact on circuits and microarchitecture," in Proc. Design Automation Conf., Anaheim, CA, 2003, pp. 338–342.

[4-18] A. Bhavnagarwala, X. Tang, and J. D. Meindl, "The impact of intrinsic device fluctuations on CMOS SRAM cell stability," IEEE J. Solid-State Circuits, vol. 36, no. 4, pp. 658–665, Apr. 2001.

[4-19] X. Tang, V. De, and J. D. Meindl, "Intrinsic MOSFET parameter fluctuations due to random dopant placement," IEEE Trans. Very Large Scale Integr. (VLSI) Syst., vol. 5, no. 4, pp. 369–376, Dec. 1997.

[4-20] S. Mukhopadhyay, H. Mahmoodi, and K. Roy, "Modeling and estimation of failure probability due to parameter variation in nano-scale SRAMs for yield enhancement," in Dig. Tech. Papers VLSI Circuit Symp., Honolulu, HI, Jun. 2004, pp. 64–67.

[4-21] S. Mukhopadhyay, "Modeling of failure probability and statistical design of SRAM array for yield enhancement in nanoscaled CMOS," IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems, Dec. 2005, vol. 24, no. 12, pp. 1859-1880.