國 立 交 通 大 學

電子工程學系 電子研究所

碩 士 論 文

40 奈米製程技術操縱在低操縱電壓及管線結構的

512Kb 8T 靜態隨機存取記憶體

40nm Low V_{MIN} Pipeline 512Kb 8T SRAM Design

研 究 生：朱俐瑋

指導教授：莊 景 德 教授

中 華 民 國 一 〇 一 年 九 月

40 奈米製程技術操縱在低操縱電壓及管線結構的 512Kb 8T 靜態隨機存取記憶體

40nm Low V_{MIN} Pipeline 512Kb 8T SRAM Design

研 究 生：朱俐瑋 Student：Li-Wei Chu

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

國 立 交 通 大 學

電 子 工 程 學 系 電 子 研 究 所

碩 士 論 文

A Thesis
Submitted to Department of Electronics Engineering and
Institute of Electronics
College of Electrical Engineering and Computer Engineering
National Chiao Tung University
In Partial Fulfillment of the Requirements
for the Degree of
Master of Science
in
Electronics Engineering
Sep. 2012
Hsinchu, Taiwan, Republic of China

中華民國一〇一年九月

40 奈米製程技術操縱在低操縱電壓及管線結構的 512Kb 8T 靜態隨機存取記憶體

學生：朱俐瑋　　　　　　　　　　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

由於現在的3C產品對整個生活週遭影響越來越大,而這些商品一定都需要記憶體來存放資料。而記憶體的存放或是讀取時間也是會影響整個電子產品的在運作上之效率。在靜態隨機存取記憶體裡，因為他跟其他的記憶體相比有較快的處理速度，一般都是放在CPU附近當作快取記憶體，但相較面積之下，確是相對大一些。所以在先進的SoC晶片設計中，靜態隨機存取記憶體往往都是占整個晶片最大的面積。由於這個理由，我們必需好好設計他的操作速度及能量消耗。在一般靜態隨機存取記憶體結構中，主要是以6個電晶體為目前的趨勢。但隨著現在電子業產品的走向，是希望能在越低的操作電壓中運作，使得整個產品耗電量能越少越好，6個電晶體的靜態隨機存取記憶體架構會在低電壓中難以正常操作。故我們設計了一個8個電晶體的靜態隨機存取記憶體能比傳統6個電晶體的靜態隨機存取記憶體操作在更低的電壓，相對來說，也就比較省電能。為了幫助能在低壓下寫入/讀取成功,我們在此40奈米512kb的記憶體中加入了提升字元線的機制以及幫助寫入的電路設計。雖然8個電晶體的靜態隨機存取記憶體原來就比傳統的靜態隨機存取記憶體操作速度快,但還是能夠比其他的記憶體動作快。我們為了提升此8個電晶體的靜態隨機存取記憶體之操作速度，我們運用了漣波位元線讀取架構及管線結構來增進它。

40nm Low V_{MIN} Pipeline 512Kb 8T SRAM Design

Student: Li-Wei Chu   Advisors: Prof. Ching-Te Chuang

Department of Electronics Engineering & Institute of Electronics

National Chiao-Tung University

ABSTRACT

As 3C products now growing impact on the entire lives goods must have the memory to store data. Memory storage or read time also will affect the operational efficiency of the entire electronic product. Relatively faster processing speed compared with other memory in the SRAM, which are generally on the CPU as a cache near, but compared to the area that indeed larger some. So in advanced SoC chip design, SRAM area often is the largest in total chip. For this reason, we must have a good design performance and energy consumption. In general SRAM, the 6T structure is current trend. However, with the trend of the electronics industry, which is hoping to operate in a lower working voltage, power consumption can be as little as possible in the entire product. But the 6T SRAM architecture in low voltage is difficult to operate normally. Therefore, we have designed an 8T SRAM than 6T SRAM operating at lower voltage. Relatively speaking, there is more power savings for products. In order to help write/read mechanism achievement, we use the boosting WL mechanism as well as to help write in 40nm 512 Kb memories circuit design. 8T SRAM operating speed is slower than conventional 6T SRAM, but still faster than other memories. In order to enhance our 8T SRAM speed, we use a ripple bit-line and pipeline structure to enhance it.

誌 謝

本篇論文能夠完成，首先要感謝的是我的指導教授莊景德教授，很幸運地能夠當他的學生，不僅是個負責、認真，又是個和藹的老師，有困難時也會給我很多的建議，也許我是女孩子吧，老師總是會比較關照及希望不要太給我太大的壓力(但該有的壓力還是有的)，我想應該是上輩子有做好事還有這輩子有積陰德所換來的吧!

再來就是要感謝在這個論文中也給我很多幫助的在職博班學長連南鈞學長。由於學長在業界打滾許久，所以研究上有一些問題的時候總是能馬上抓出重點，及電路上的改良設計也幫忙不少，讓我省了不少力，還有謝謝智原公司的學長們(明賢、Jason、Jerry、Angelo、Paul)大力贊助才能成就此篇論文結果如此的完美。另外還有實驗室的已經畢業的學長們能分享在電路設計上的技巧及設計和能陪我玩桌遊的學弟妹們，讓我有個充實學習又不失快樂的實驗室環境。

最後感謝我的父母，沒有他們也就沒有現在在交大讀書的我。還有感覺像朋友的弟弟，雖然讀的是不同的科系(資工系)，但有時也會給點看法並給我一些幫助。另外，住在交大附近的表姊、表姊夫及他們的好友們讓我在新竹有困難時能給我一些援助，讓我感覺在新竹一點也不孤單。

Contents

Chapter 1 Introduction 1

1.1 Background 1

1.2 Motivation 1

1.3 Thesis Organization.. 2

Chapter 2 Overview of Low-voltage SRAM Design in Recent Years .4

2.1 Introduction 4

2.2 Memory Family... 4

2.2.1 Flash 4

2.2.2 DRAM...6

2.2.3 FinFET SRAM 8

2.3 SRAM.. 11

2.3.1 6T SRAM 11

2.3.2 Conventional 8T SRAM...14

2.4 SRAM Static Noise Margin (SNM) 15

2.4.1 Hold Static Noise Margin (HSNM) 16

2.4.2 Read Static Noise Margin (RSNM)...17

2.4.3 Write Static Noise Margin (WSNM)..19

2.5 SRAM Write Margin (WM).. 20

2.6 SRAM Array Structure... 21

2.7 Variation Issue.. 23

2.7.1 Global and Local Variation 23

2.7.2 SRAM Cell Variation...26

2.8 Modern SRAM Design Methodology 29

2.8.1 Dual Supply Voltage 29

2.8.2 Negative Bit-Line 32

2.8.3 Dynamic Word-Line Voltage 35

2.9 Power Consumption 38

2.9.1 Dynamic Power Dissipation..38

2.9.2 Short-Circuit Power Dissipation 39

2.9.3 Static Power Dissipation 40

Chapter 3 40m 512Kb Pipeline Low VDD 8T SRAM.........................44

3.1 Introduction 44

3.2 8T SRAM Operation 45

3.2.1 Conventional Single-Ended 8T SRAM..45

3.2.2 Disturb-free DAWA Single-Ended 8T SRAM 46

3.3 Data-Aware Write-Assist (DAWA) and Interleaving 49

3.4 Pipeline Structure and Clock Distribution 51

3.5 Word-Line Booster..54

3.5.1 Word-Line Booster Circuit...54

3.5.2 Booster Capacitance Design...56

3.5.3 Simulation Result...57

3.6 Voltage Detector 58

3.6.1 Voltage Detector Design 58

3.6.2 Simulation Result 60

3.7 Data-Aware Write-Assist Tracking Circuit...61

3.7.1 DAWA Tracking Circuit Design...61

3.7.2 Simulation Result 63

3.8 Ripple Bit-Lin (BL) and Multiplexer 64

3.8.1 Ripple BL and Multiplexer Designs..64

3.8.2 Simulation Result 67

3.9 The Keeper Design of LBL 68

3.9.1 The Keeper Design..68

3.9.2 Simulation Result 74

3.10 Data-In Data-Out (DIDO) Design 75

3.10.1 GBL Latch 75

3.10.2 Write Through Circuit 77

Chapter 4 The Chip Structure Result 79

4.1 Chip Sepc. 79

4.2 Chip Per-Simulation vs. Post-Simulation 81

4.3 Test Flow 83

4.4 Testing Result 86

Chapter 5 Conclusion 90

5.1 Conclusion. 90

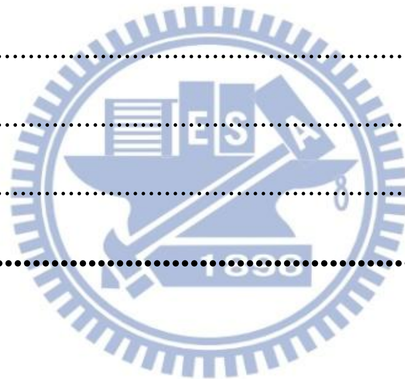
Reference 91

Chapter 1 91

Chapter 2 91

Chapter 3 94

Vita 98



List of Figures

Fig. 1.1 (a) Conventional 8T SRAM cell (b) single-ended 8T SRAM cell 2

Fig. 2.1 Schematic cross section of the Flash cell [2.1] 5

Fig. 2.2 NOR Flash writing mechanism [2.1] 5

Fig. 2.3 Floating-gate MOSFET reading operation [2.1]... 6

Fig. 2.4 DRAM structure.. 7

Fig. 2.5 Multi-fin FinFET [2.3] 8

Fig. 2.6 TG-FinFET 6TSRAM cells schematic [2.4] 9

Fig. 2.7 TG-FinFET 6T SRAM cells layout [2.4] 10

Fig. 2.8 IG-FinFET 6T SRAM cells schematic [2.4] 10

Fig. 2.9 IG-FinFET 6T SRAM cells layout [2.4] 11

Fig. 2.10 6T SRAM cell structure 11

Fig. 2.11 6T SRAM read current flow 12

Fig. 2.12 6T SRAM writing current flow 13

Fig. 2.13 6T SRAM transistor ratios 13

Fig. 2.14 Conventional 8T SRAM 14

Fig. 2.15 Conventional 8T SRAM layout [2.5] 15

Fig. 2.16 Voltage transfer curve of inverter [2.6] 16

Fig. 2.17 HSNM detected circuit 16

Fig. 2.18 The butterfly curve in hold mode 17

Fig. 2.19 RSNM detected circuit 18

Fig. 2.20 The butterfly curve in read mode (6T) 18

Fig. 2.21 The 8T butterfly curve in read mode 19

Fig. 2.22 WSNM detected circuit 19

Fig. 2.23 The write-1 VTC 20

Fig. 2.24 Diagram of new write margin [2.8] 20

Fig. 2.25 Basic SRAM organization 21

Fig. 2.26 SRAM critical path [2.9] 22

Fig. 2.27 Global variation and Local variation of V_T [2.10] 23

Fig. 2.28 (a) Number of dopant atoms in the channel as function of effective channel length (b) "sigma" of V_T variation of function of technology node 24

Fig. 2.29 Plot of randomly generated dopant positions in a 25nm MOSFET, viewed from the side. [2.12] 25

Fig. 2.30 Plot of threshold voltage uncertainty (1σ) versus the vertical depth parameter, d, of the source /drain doping for 25nm n-channel MOSFETs, comparing discrete donor effects to discrete acceptor effects. [2.12] 26

Fig. 2.31 Scaling and cell stability margin of 6T [2.13] 27

Fig. 2.32 6T RSNM with V_T variation [2.11] 28

Fig. 2.33 The effect of local variation in write-mode with worse case 29

Fig. 2.34 The effect of local variation in read-mode with worse case 29

Fig. 2.35 Single-Vcc and Dual-Vcc processors [2.14] 30

Fig. 2.36 (a) Cell stability in the Vmin (b) Vnim in different bit desnsity [2.14] .. 30

Fig. 2.37 The cross-point 8T SRAM schematic, layout, read/write negative-biased circuit, and waveform [2.15] 31

Fig. 2.38 Constant-negative-level write buffer. [2.16].. 32

Fig. 2.39 Simulated negative bit-line level. [2.16].. 33

Fig. 2.40 Write driver with boost control [2.17] 34

Fig. 2.41 Write cycle simulation waveforms and results l [2.17] 34

Fig. 2.42 SNM improvement by lowering WL voltage in 6T SRAM in case of (a) with and (b) without local V_T variations. [2.18] 35

Fig. 2.43 Improving read stability depending on process and temperature variations. [2.18-19].. 36

Fig. 2.44 Write assist circuit (WAC) improving write stability [2.18-19]............... 37

Fig. 2.45 (a) Simulated waveform of the ary-VDM and dmy-VDM in the write status. (b) Comparison of the write ability by DC simulation result of the write-trip-point. [2.19].. 37

Fig. 2.46 The inverter schematic [2.20] 38

Fig. 2.47 Current behavior of an inverter without load. [2.21] 39

Fig. 2.48 Leakage components in a scaled transistor [2.22] 40

Fig. 2.49 Injection of hot electrons from substrate to oxide [2.23] 41

Fig. 2.50 As n+ region is depleted or inverted with high negative gate bias, condition of the depletion region near the drain-gate overlap region of an MOS transistor [2.23] 42

Fig. 2.51 Variation of minority carrier concentration in the channel of a MOSFET biased in the weak inversion. [2.23] 42

Fig. 3.1 (a) A conventional single-ended 8T SRAM cell.[3.11] (b) A disturb-free DAWA single-ended 8T SRAM cell. [3.12] 45

Fig. 3.2 (a) The operation for stand-by mode. (b) The read-mode operation. (c) The operation for wirte-1 mode. (d) The operation for write-0 mode. 47

Fig. 3.3 Cell layout for disturb-free DAWA single-ended 8T SRAM. 48

Fig. 3.4 Write-1 mode and write-0 mode with half-select 50

Fig. 3.5 Interleaving array structure 51

Fig. 3.6 The CLK signal use H-tree way. The black line represents the farthest path of CLK and the black cycle stand for the farthest cell. And the red line represents the nearest path of CLK and the red cycle stand for the nearest cell. 52

Fig. 3.7 The master-slave latches. The master latch part enclosed by the red wireframe; then slave latch part enclosed by the green wireframe. The master latch capture data in the positive clock, we called the latch of L1. Also, the slave latch to do is launching the data, we called the latch of L2. 53

Fig. 3.8 The pipeline SRAM operation in whole chip 54

Fig. 3.9 (a) Write margin of writing 0 with boosting. (b) Write margin of writing 1 with boosting. 55

Fig. 3.10 The wordline booster which BST_EN come from voltage detector. If we want to boost WL, then the BST_EN signal goes high, vice versa. 55

Fig. 3.11 The WL capacitance (C_{VVDD}) estimation 56

Fig. 3.12 Boosting Δv steps. 57

Fig. 3.13 The simulation of WL, farthest cell array write/read, and data-out (DO) 58

Fig. 3.14 (a) The delay chain by the CSB controlled. (b) The section by an external voltage to decide whether or not as boosting WL. (c) Three different corner (SS, TT, FF) be controlled by OSD<0>~<2>. voltage at VDD is 0.9 59

Fig. 3.15 The simulation waveform of voltage detector. 60

Fig. 3.16 DAWA tracking circuit. 61

Fig. 3.17 1-bit dummy cell and tied 0/1 schematic. 62

Fig. 3.18 DAWA tracking circuit waveform @ TT, 25℃, VDD=1.1V, OSD<0>: on. 63

Fig. 3.19 The ripple path structure and the LEV_TOP and LEV_DN read part circuit. 64

Fig. 3.20 The ripple multiplexer circuit 65

Fig. 3.21 The waveform for the ripple multiplexer 67

Fig. 3.22 (a) The keeper of ripple multiplexer, (b) the keeper of LEV_TOP, and (c) the keeper of LEV_DN in red part 69

Fig. 3.23 LCR keeper dynamic gate topology [3.22] 69

Fig. 3.24 The waveform by replica keeper @ ff, 125℃, VDD=1.1V. 70

Fig. 3.25 Read local bit-line with contention free shared (CFS) keeper and global keeper delay element. [3.23] 71

Fig. 3.26 The waveform for GBL by CFS keeper @ ff, 125℃, VDD=1.1V. 72

Fig. 3.27 The local bit-line keeper design in detail 72

Fig. 3.28 The local bit-line keeper design is chose from the LBL_KP<0>. 73

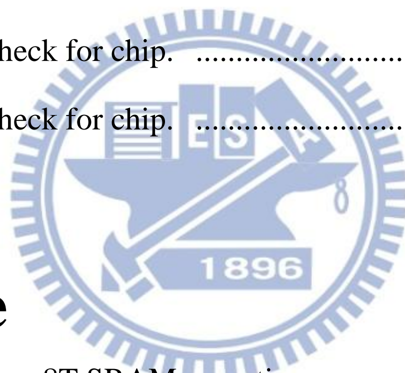
Fig. 3.29 The local bit-line keeper design is chose from the LBL_KP<1>. 74

Fig. 3.30 The waveform for local bit-line keeper design @SS, 125℃, VDD=0.99V 75

Fig. 3.31 The GBL latch circuit 76

Fig. 3.32 The waveform of GBL latch circuit 76

Fig. 3.33 Writer through design 77
Fig. 3.34 Writer through design waveform 78
Fig. 4.1 The simple diagram of 512Kb 8T SRAM 79
Fig. 4.2 Cycle time diagram for read/write 80
Fig. 4.3 Signal line distribution in chip 81
Fig. 4.4 Po-sim vs. pre-sim @ TT, 25℃, VDD=1.1V 82
Fig. 4.5 Po-sim vs. pre-sim @ FF, -40℃, VDD=1.21V 82
Fig. 4.6 Po-sim vs. pre-sim @ SS, 125℃, VDD=0.99V 83
Fig. 4.7 The design places on half chip. 83
Fig. 4.8 The chip testing flow. 85
Fig. 4.9 The layout of chip. 86
Fig. 4.10 The function check for chip. 86
Fig. 4.11 The function check for chip. 86



List of Table

Table 3.1 Disturb-free new 8T SRAM operation 48
Table 3.2 GBL_KP voltage on different corner 70
Table 4.1 Pin descriptions for the chip 80
Table 4.2 Three options for the chip. 87
Table 4.3 The pass rate for TT, 25℃ 87
Table 4.4 The pass rate for FF, 25℃ 87
Table 4.5 The pass rate for SS, 25℃ 87

Chapter 1
Introduction

1.1 Background

In the IC's world, we are following the Moore's Law today. In another word, the density of the chip will become double each 18 months. So we can design more complex circuit and improve performance at the same area in the past nano-technology. For this reason, the most of market's 3C products are becoming smaller, light and portable. Cause of the size of the transistor trend is getting smaller; we wish the power consumption reduce, too. Beside, the leakage current and the process pass into more important and critical. These problems should research and study in today IC products.

The memory area will occupy around 90% of the system on SOC base on the International Technology Roadmap for Semiconductors (ITRS). The prediction tells us the Static Random Access Memory (SRAM) will affect the total power dissipation and area seriously. The modern smart phone and note book has begun to keep watch for power, because they should use the battery to work. And they internal circuits both have the SRAM, so we will reduce the SRAM power is the major and important work.

1.2 Motivation

Governments around the world advocate use power effective and saving energy. Moreover, the IC companies also want to reduce the support external power for their produces. That show us the power is an important role, not only hope use it available but also how to decrease the power in recent years. As a result of the subject, I choose the SRAM in order to cut down the total power as best as possible.

For reading a lot papers and listening memory classes, I realize the 8T SRAM is the best choice. In spite of 8T SRAM's speed is lower than conventions 6T SRAM, but it can work at lower power than 6T SRAM. Well-known manufacturers like IBM publish paper that their SRAM also use 8T SRAM due to diminish power dissipation. Besides, 8T SRAM has a better ability of read-disturb than 6T SRAM which the read disturb is free that allows a robust operation in lower voltage supply. And I show the conventional 6T cell and the single-ended 8T SRAM cell below the article (fig. 1.1) [1.1].

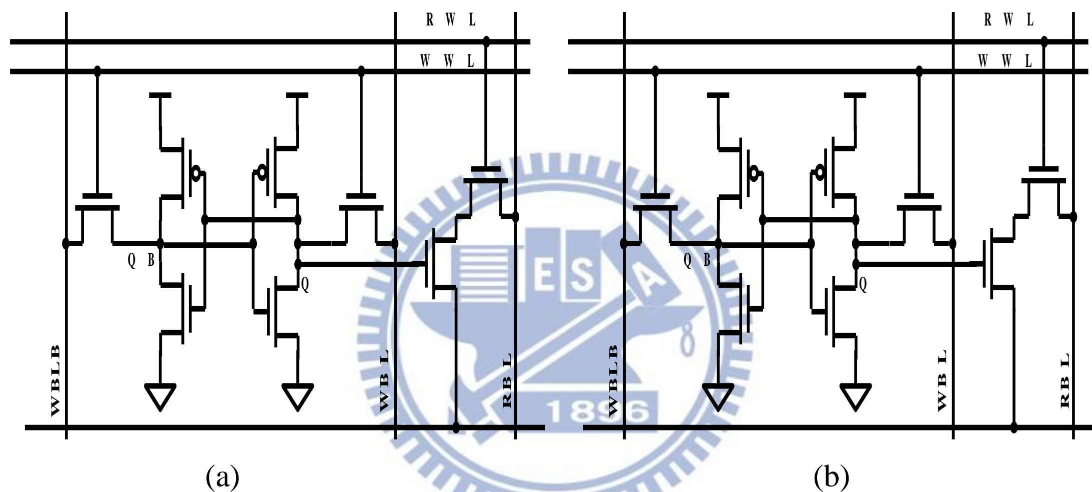


Fig1.1 (a) Conventional 8T SRAM cell (b) single-ended 8T SRAM cell [1.1]

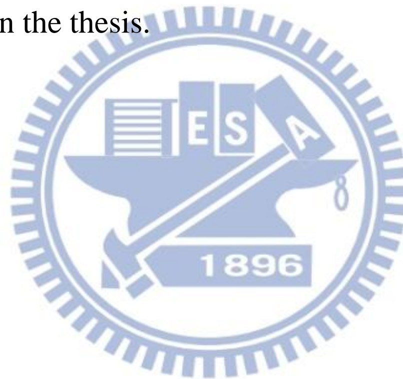
1.3 Thesis Organization

The thesis will introduce you the overview Low-Voltage SRAM design in chapter 2. I will discuss and analyze the low-voltage SRAM development. Afterwards, how to produce the power consumption and have to take notice of the current leakage from the transistor that also writes in chapter 2.

In chapter 3, I will tell you my 8T SRAM design structure in detail. I want to do the design better and use some circuit to improve the performance in the 512Kb memory chip. The 8T SRAM chip use the pipeline structure that wishes the performance can get better and working faster. Besides, we choose the H-tree clock distribution to

transmit the external clock that is able to make the pipeline working preferable. Then, the 8T SRAM made Data-Aware Write-Assist (DAWA) to enhance its write ability and reduce power. Then, we use a new scheme of voltage controller that control whether boosting world-line (WL) or not to write/read ability better. Doing DAWA operation, we can’t always turn on the DAWA for long time that the cell data is flipped probably. For this reason, we design the DAWA tracking circuit to control the DAWA switch.

Finally, chapter 4 is displaying the test flow how to test the design and showing the chip measure results. We will discuss the data by real measurement and analyze the design where should modify that let the chip performance better. Chapter 5 is making a conclusion of reference in the thesis.



Chapter 2

Overview of Low-Voltage SRAM Design in Recent Years

2.1 Introduction

The chapter first would introduce the family of memory development and discuss their identity. To realize read-disturb, read static noise margin (RSNM), write static noise margin (WSNM) definitions, and so on. We just choose major memory to realize how they operation in present IC industry.

Next, I will tell you the conventional 6T SRAM, the conventional single-ended 8T SRAM, and the new 8T disturb-free single-ended 8T SRAM basic operation. Even though two single-ended 8T structures are different, but the RSNM and WSNM is the same. Besides, I show the simulations of the 6T/8T SRAM RSNM and WSNM. Since the thesis is talking about power, we discuss the power consumption where produce in the transistor in this chapter ultimately.

2.2 Memory Family

2.2.1 Flash

Flash has developed for 20 years since in 1986 product. The main structure are NOR and NAND flash memories. Its belong Non-Volatile Memory (NVM) and play an important role on disk cashes. It also viable subsystem on PC computing and uses other applications as 3G/UMNT mobile phones. The NOR composition main make use of cellular phones or embedded application and the NAND composition is for memory cards. The NOR flash for code and storage application and the NAND flash only for data storage. [2.1][2.2]

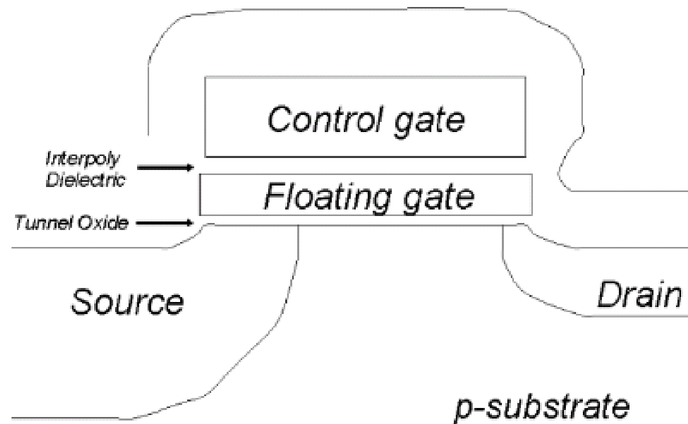
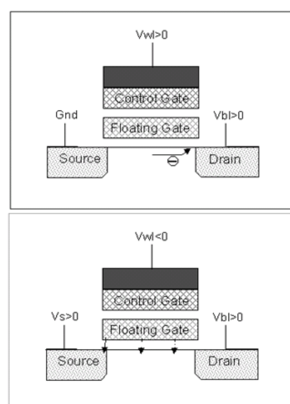


Fig. 2.1 Schematic cross section of the Flash cell. [2.1]

We simply explain the flash operation. We see the Fig. 2.1, this is a Flash cell which has two gates surrounding by dielectrics. The couple capacitor is produce between control gate (CG) and floating gate (FG). In normal state (or positively charged), the data is logic “1”. And the negatively charged state stand for the data value is logic “0” as electrics store in the FG.

For writing operation, a NOR Flash is programming by channel hot electron (CHE) injection in the floating gate (FG) at the drain side; it is eared by the flower-nordheim (FN) electron tunneling oxide from the FG to the silicon surface. (Fig. 2.2) [2.1].



- **Programming**: channel hot electron (CHE) injection in the floating gate at the drain side
- **Erasing**: Fowler-Nordheim (FN) electron tunneling current through the tunnel oxide from the floating gate to the silicon surface

Fig. 2.2 NOR Flash writing mechanism. [2.1]

For reading operation, we can measure the FG MOS transistor voltage to decide the cell data whether is logic “1” or “0”. We see the Fig.2.3 [2.1] that realize the Flash cell store logic “1” that transconductance is the same with logic “0” in current-voltage characteristic curve. Only difference is the threshold voltage shift ΔV_T that is proportional to store the electron charge Q in fixed gate bias. While the current is very high we measure that means the stored data is logic “1” in fixed bias voltage. In other case, as we measure the current is 0 that represent the Flash cell storage logic “0”. [2.1]

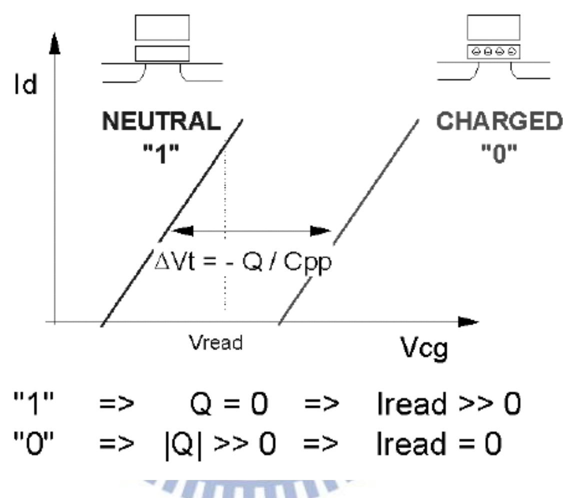


Fig. 2.3 Floating-gate MOSFET reading operation [2.1]

2.2.2 DRAM

Dynamic Random Access Memory (DRAM) is very important memory for PC that main application is used to most of memory for disk in this day. DARM is the Volatile Memory (VM). If you do not support the fixed voltage for it, the DRAM cell storage data can’t remain original data in last time. The DRAM structure is combined into one transistor and one capacitor as Fig. 2.4. Due to the simple composition (1T+1C) that density is very well. With SRAM comparison, the cost is cheaper than SRAM, but accesses data slower and consumes more power.

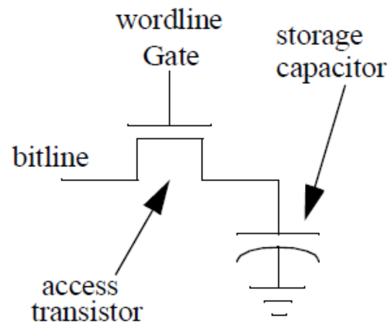


Fig. 2.4 DRAM structure

Next, the DRAM cell is introduced into its operation of reading and writing. In write mode, turns on the access transistor is controlled over word-line at first. If we want to transmit the logic “0” data in the DRAM cell, first step the bit-line (Fig. 2.4) discharge to GND. Then second step, original charges in the storage capacitor are produced the discharged path passing through the access transistor to bit-line. Else if we wish to write logic “1” into the DRAM cell, we assume the storage capacitor has no charge into in beginning. The bit-line charges to the support voltage (VDD) and turns on the access transistor naturally. Then it comes into a charge path from bit-line through the access transistor and finishes the working.

In read mode, of course the cell must turn on the access transistor. The bit-line voltage always is VDD that fixed supporting voltage in stand-by mode. When in read mode, the bit-line voltage changes from VDD to 1/2VDD. Besides, the cell storage logic “1” let the bit-line voltage larger than 1/2VDD. On the contrary, the cell stores logic “0” make the bit-line voltage less than 1/2VDD. This bit-line voltage wills read the result by sense amplifier with 1/2VDD as a diving line.

But cause of the DRAM structure has a capacitor that discharges to ground for period of time. This problem makes the write “1” data in cell for long time, the capacitor voltage change floating “1” to floating “0” and finally read the cell data

result is wrong. To solve the problem, there is a mechanism for each fixed time will be recharged once.

2.2.3 FinFET SRAM

According to Moore's law, today the process gets smaller and smaller but cause to the physical properties discover we will meet the bottleneck below 10nm process. Therefore, some people use the same process however the system place the way of 3D that can improve the performance. Another way is changing the MOS structure, and afresh is issued with properties, and then does modify and modeling. The structure is FinFET that is a trending in the future!

Now we begin to introduce the FinFET composition. There are three FinFETs in Fig. 2.5 [2.3]. This composition unlike planar single- and double-gate devices in plane, the channel width is placed perpendicular to the semiconductor plane. It is not only increase the drive current due to raise per unit planar area by fin-height but also reduces the delay time cause to the equation: C_{load}/I_{drive}. [2.3]

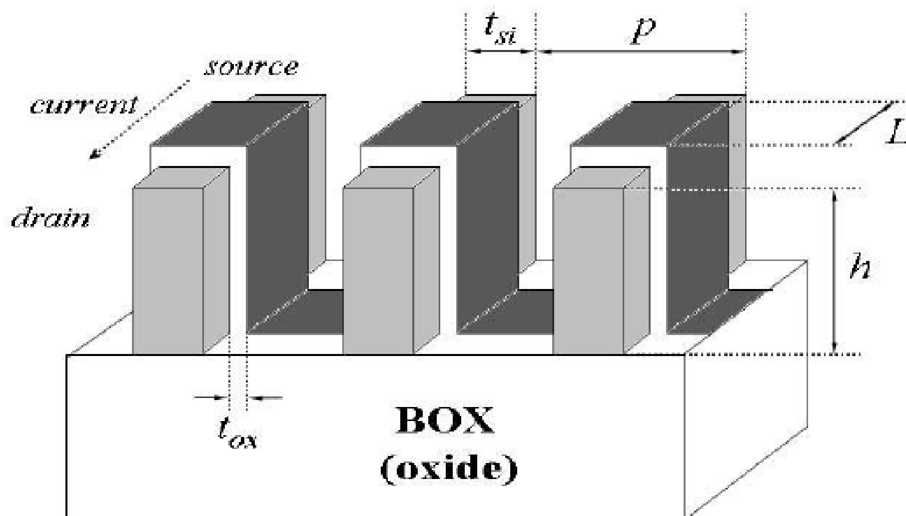


Fig. 2.5 Multi-fin FinFET structure [2.3]

In the Fig. 2.5, the silicon fin of thickness t_{si} is located on an SOI wafer. The t_{si} is the body-thickness of the resulting double-gate structure where both gates are tied together. Current flow is parallel to the wafer plane while channel width is perpendicular to the plane. The effective channel width is equal to *2h* because the height of SOI thickness also is *h*. Higher widths are achieved by drawing multiple fins in parallel and wrapping the gate around them. The effective channel width for a multi-fin FinFET on a given planar area of silicon is determined by h and fin-pitch *p*. The minimum *h* required to achieve equivalent planar are efficiency is thus *p/2*. In other words, increasing *h* beyond *p/2* increases area efficiency. The upper bound on *h* is set by the maximum fin aspect ratio (a_{max}=H_{max}/ t_{si}) allowed by the process. [2.3]

The design considerations for the reliable operation of the 6T FinFET SRAM circuits are provided in this section. The standard tied-gate (TG) FinFET SRAM cell is shown at Fig. 2.6 that all six transistors are sized minimum in 32nm process and Fig. 2.7 is that layout. [2.4]

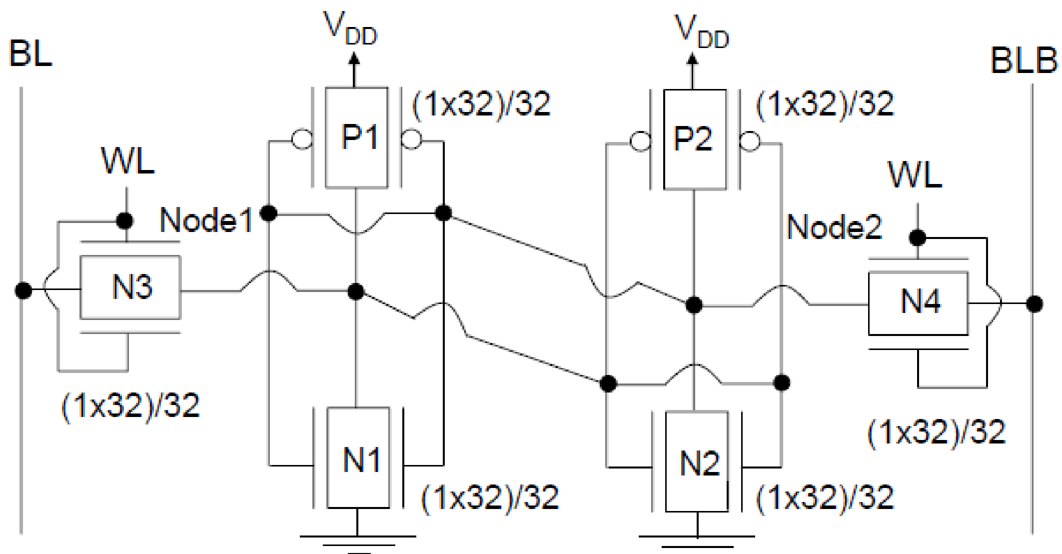


Fig. 2.6 TG-FinFET 6T SRAM cells schematic [2.4]

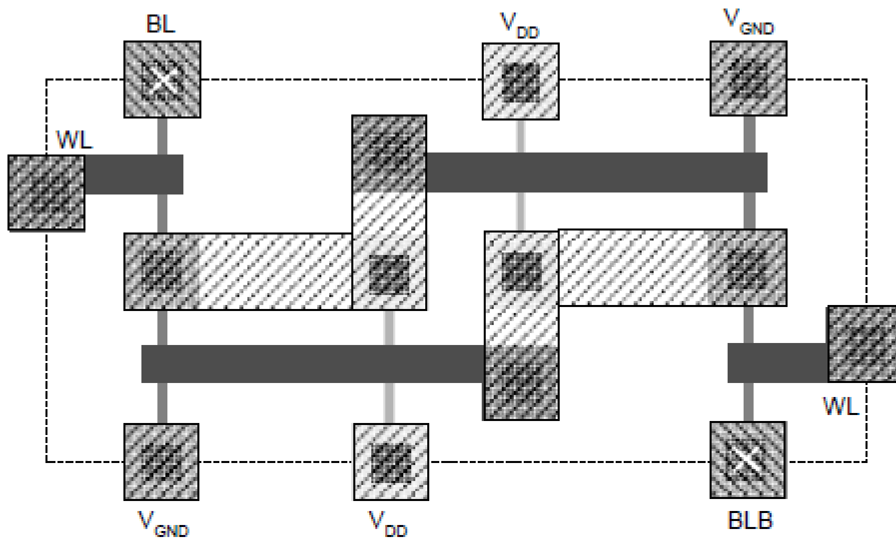


Fig. 2.7 TG-FinFET 6T SRAM cells layout [2.4]

The idle mode leakage power consumption is reduced with the Independent-Gate (IG) FinFET 6T SRAM that can enhance the data stability and the integration density as compare to the TG-FinFET SRAM circuits. The schematic is in the Fig. 2.8 and the layout is shown at Fig. 2.9 that the six transistors also use the minimum size.

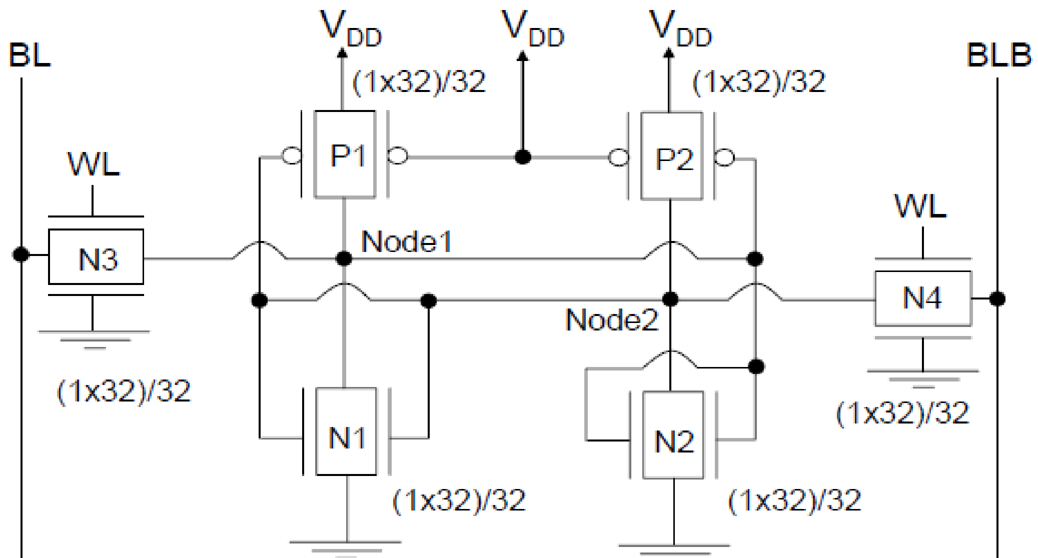


Fig. 2.8 IG-FinFET 6T SRAM cells schematic [2.4]

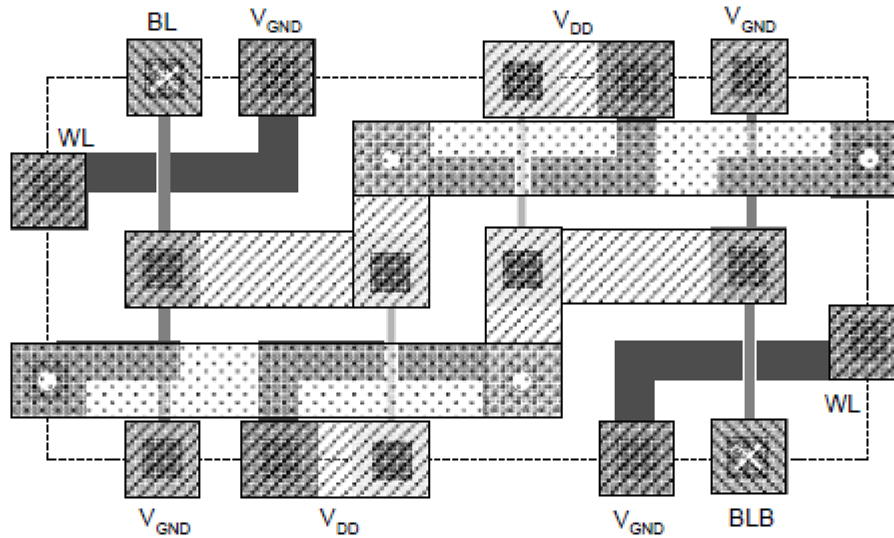


Fig. 2.9 IG-FinFET 6T SRAM cells layout [2.4]

2.3 SRAM

2.3.1 6T SRAM

The 6T SRAM is the most companies using to produce SRAM memory in recent years. The construction is shown in Fig. 2.10. It is composed of two back-to-back inverters and adds one NMOS in inverter input side respectively that accomplishes the 6T SRAM structure. We usually say MP1 and MP2 are pull-up gates; MN1 and MN2 are called pull-down gates; M1 and M2 are called pass-gates in the field. In standby, the 6T SRAM setting is bit-line (BL) and bit-line-bar (BLB) is charging to VDD, but the word-line (WL) is not turning on.

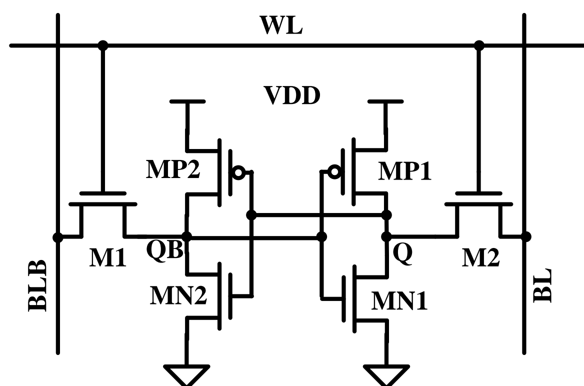


Fig. 2.10 6T SRAM cell structure

First, we start to introduce the read operation in 6T SRAM. Above the description, the standby mode, the BL and BLB will pre-charge to VDD at beginning. As reading the cell, the selected cell WL is going high (logic “1”) that turns on the M1 and M2. Then, the BL and BLB voltage starts to float. If we assume the Q data is logic “1” and QB is logic “0”, cause to switch on M1 the BLB voltage is going to low gradually until the voltage to GND. The cell right side BL voltage is floating high according to the Q is logic “1”. We can see the Fig. 2.11 to realize the current flows in the 6T cell.

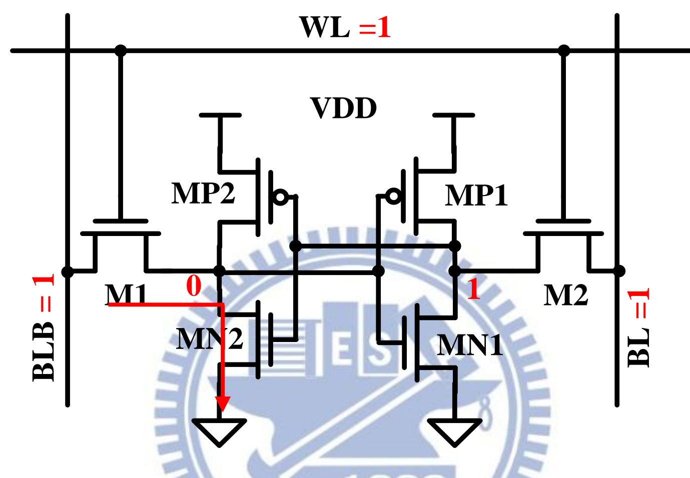


Fig. 2.11 6T SRAM read current flow

Second, we explain the 6T SRAM write operation. Consider the Q point wish to write logic “0”, so the Q point initial value is logic “1”. Turning on the WL and BL discharge to GND but BLB still keep on VDD. Q is going to write “1” operation and the other side, QB is going to do writing “0” working. In write mode, the BL and BLB are two opposite signals. To see Fig. 2.12 that displays the write operation and current flows. The red numbers are initial values in the figure and the data will be written to the 6T cell in period.

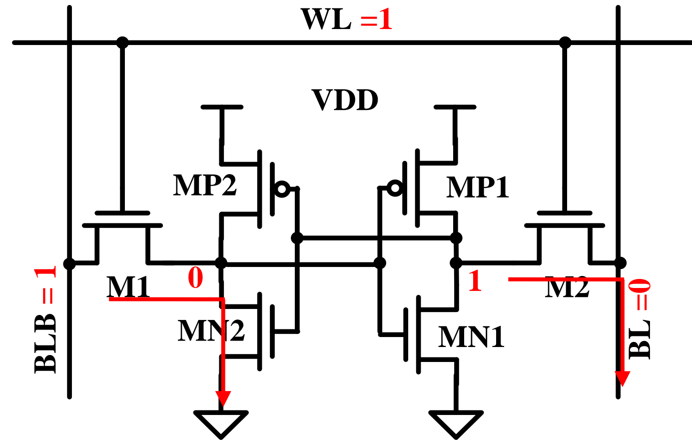


Fig. 2.12 6T SRAM writing current flow

After understanding the read/write operation in 6T SRAM, we always want to read and write data easily but these two running conflict each other. If we consider read ability well, the pull-down NMOS must be stronger than the pass-gate NMOS. Else if we wish write ability good, the pass-gate NMOS have to stronger than pull-up PMOS. Unfortunately, we also thick about the stability in standby mode and the pull-up PMOS do not be weaker overly than pull-down NMOS. In Fig. 2.13, there are three ratios β_1, β_2, andβ_3 in blue background that are represented the proportion of standby, read, and write ability. In simple terms,β_1 value cannot much smaller; β_2 value is more smaller more better; β_3 values is more larger more better.

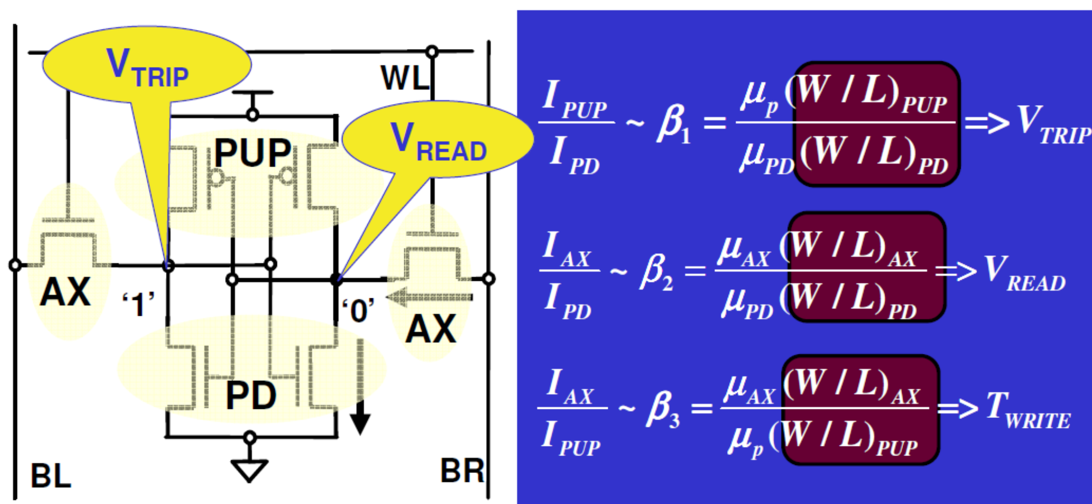


Fig. 2.13 6T SRAM transistor ratios

2.3.2 Conventional 8T SRAM

Although 6T SRAM performance is very well and simplify the SRAM, but the trend for all 3C products with the external voltage requirements are low. The 6T SRAM external power cannot too low that the reason will explain in below selection. For working in low VDD, we try to design another SRAM cell and the 8T SRAM is produced directly. In Fig. 2.14 is shown the conventional 8T SRAM. [2.5]

Fig. 2.14 Conventional 8T SRAM

This 8T cell (Fig. 2.14) adds two-transistor read stack in the conventional 6T cell. The below transistor gate of the stacking-transistor is connected to the node of Q. While write operation, we just turn on the WWL directly and RWL not turn on. Then the working way is the same by the conventional 6T SRAM cell. Afterwards, the read operation is turning on RWL and turning off WWL that the current only through two stack transistors if the Q value is logic “1”, then RBL is pulled down to logic “0”. If Q is logic “0” that RBL voltage is a floating “1”. In addition, the two cross-coupled inverters of 8T SRAM condition is alike standby state so that don’t have read-disturb problem.

The Fig. 2.15 [2.5] is the conventional single-ended 8T SRAM layout. The compact layout is the 6T cells at the left side that just take on the RWL and RBL at the right side, and then the layout is accomplished. The WWL is 6T cells’ WL. Similarly, the WBL and WBLB are 6T cells’ BL and BLB separately.

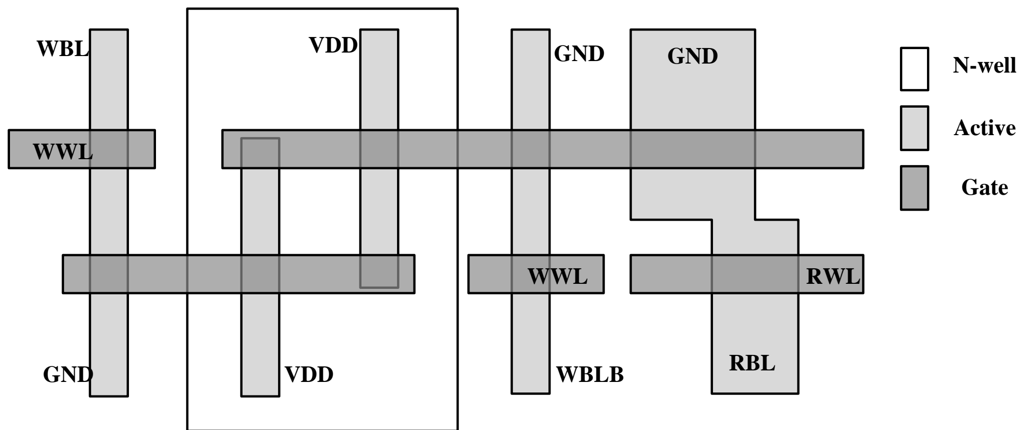


Fig. 2.15 Conventional 8T SRAM layout [2.5]

However, the thesis 8T SRAM cell schematic is not the above description. We use the different design way and have the same benefits, for example: Read Static Noise Margin (RSNM) is better than 6T SRAM and can work in low VDD. The 8T SRAM we use will introduce in chapter 3.

2.4 SRAM Static Noise Margin (SNM)

Static Noise Margin(SNM) means that the maximum DC noise voltage. In simple words to say that is a bit-cell can be tolerated by the moaximum noise value. If exceed the value, then the storage in the cell will be filed and be incorroected. Measure the SNM is getting form Voltage Transfer Curve (VTC) that usually is called as "butterfly curve" and SRAM cell must have two inverters that place back-to-back. We show Fig. 2.16 that is the VTC of inverter. Besides, we use the example for 6T cell explain SNM for below article.

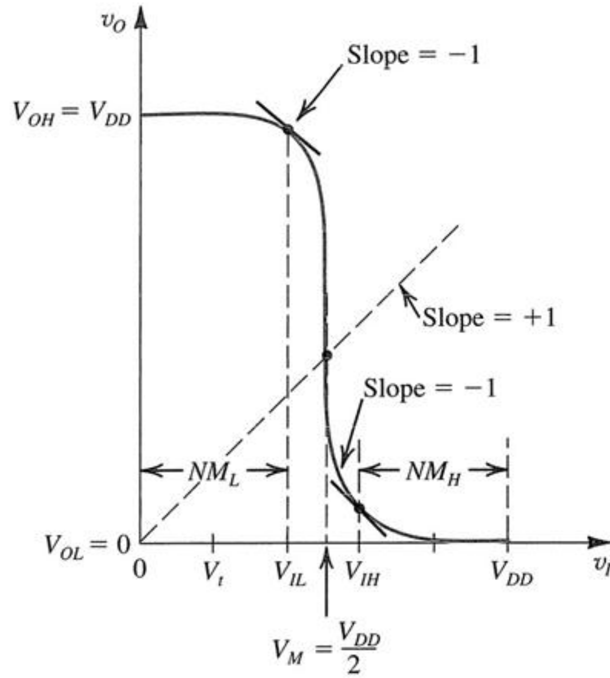


Fig. 2.16 Voltage transfer curve of inverter [2.6]

2.4.1 Hold Static Noise Margin (HSNM)

HSNM is means when the SRAM cell in the stand-by mode. The WL signal is logic "0" and do not turn on the cell pass-gate. Next, BL and BL signals are pre-charged logic "1". Cause of SNM have to use DC voltage to measure, we use the way [2.7] to test the value that shown in Fig. 2.17.

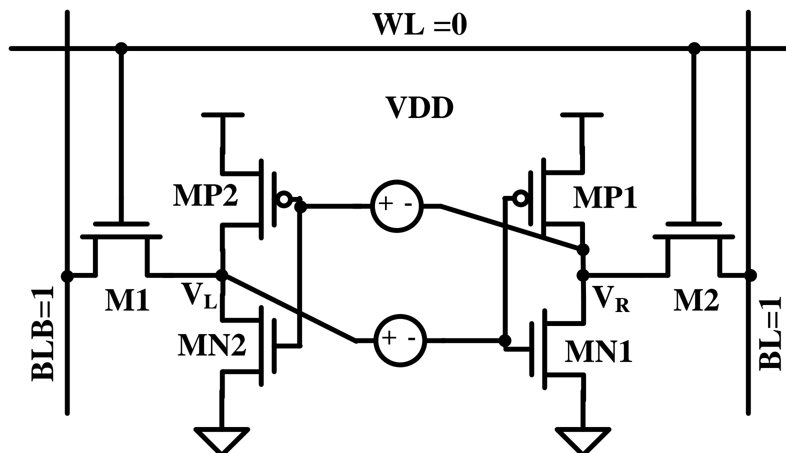


Fig. 2.17 HSNM detected circuit

The method of measuring HSNM is connected to two DC noise source with 6T cell in Fig. 2.17. We sweep the DC noise source form high voltage (ex: 1V) to low voltage (ex: 0V), then observing the result of V_R and V_L voltage. These results individually use the VDD and V_R/V_L to do two axes that accomplish the voltage transfer curves. VTCs overlapping outcome that the shape like a butterfly, so the mapping also say "butterfly curve" as shown in Fig. 2.18. There are two SNM in the figure, we choose the smaller SNM as the cell SNM.

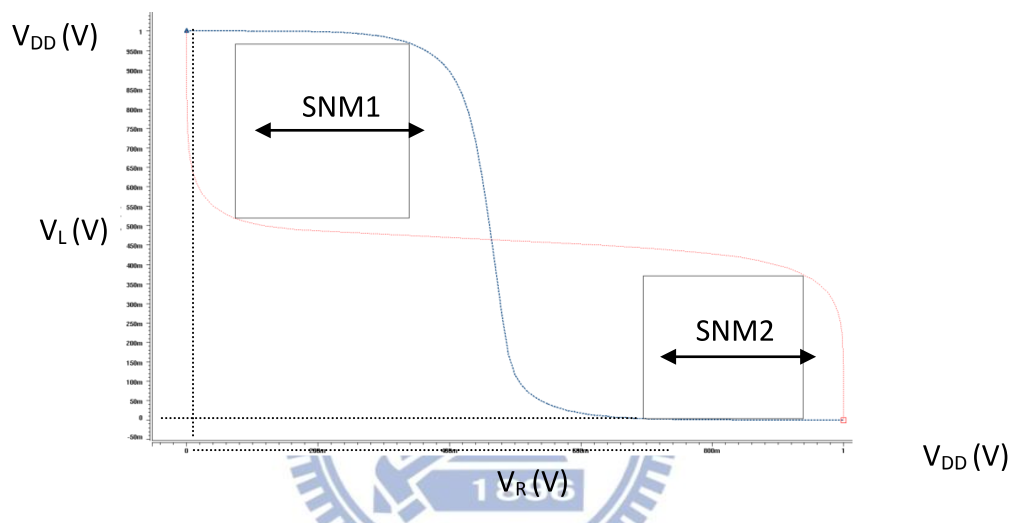


Fig. 2.18 The butterfly curve in hold mode

2.4.2 Read Static Noise Margin (RSNM)

The method is the same as the above description that difference is initial condition. We turn on the WL in read mode and per-charge BL and BLB to VDD are changeless in 6T cell that show in Fig. 2.19. Besides, the butterfly curve is shown in Fig. 2.20. Expressly, the RSNM is worse than HSNM in 6T cell. The reason is that WL turns on the M1 and M2, so produced the current path pass through the pull-down NMOS: MN1 and MN2. Pull-down NMOSs have an equivalent resistor due to the V_R/V_L voltage cannot low to GND. This voltage difference from the minimum voltage to GND we are called read-disturb.

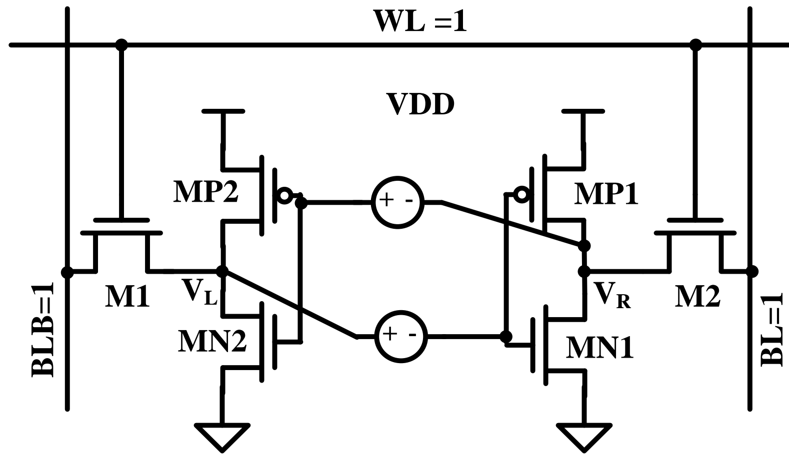


Fig. 2.19 RSNM detected circuit

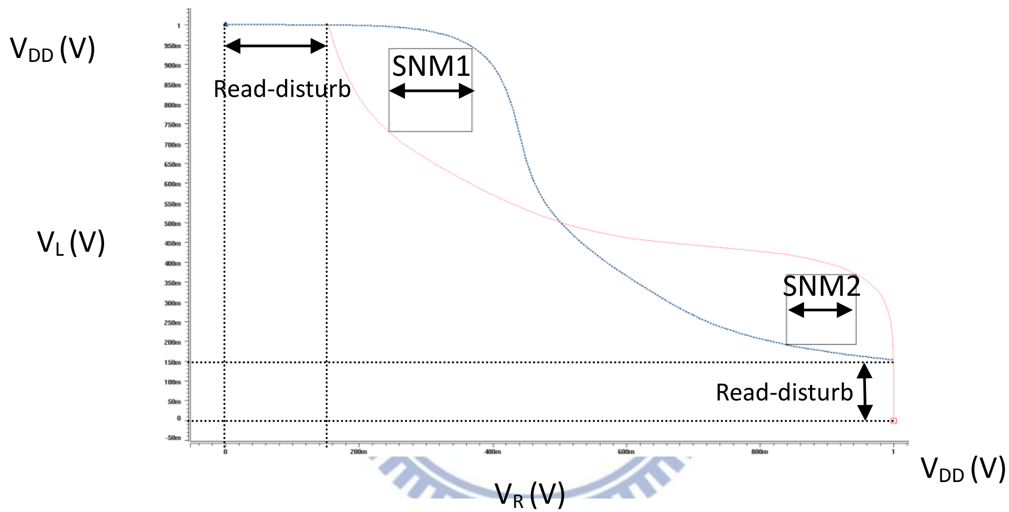


Fig. 2.20 The butterfly curve in read mode (6T)

In read mode, the 6T cell have the read-disturb problem that the RSHM is worse but the 8T cell will solve the read-disturb problem due to the initial condition is the same in hold mode. The 8T cell in read mode just turns on the RWL and the inside WWL don't turn on at Fig. 2.14. The 8T RSNM is shown in Fig. 2.21 and the thesis 8T cell also have the same conclusion.

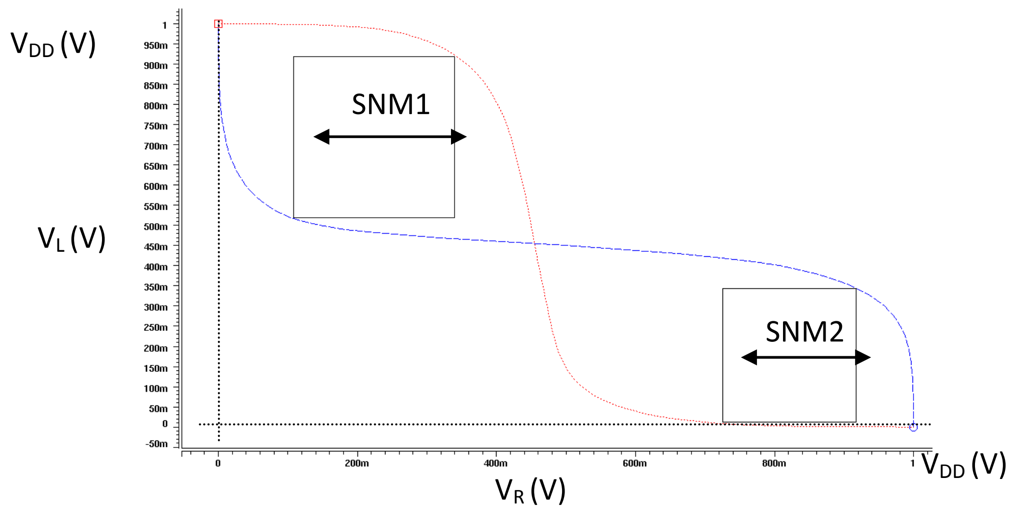


Fig. 2.21 The 8T butterfly curve in read mode

2.4.3 Write Static Noise Margin (WSNM)

In write mode, we assume V_R voltage is logic "0", then V_L is logic "1". Let the WL voltage go high and switch on M1, M2. BL is logic "1" according to V_R voltage, for the same reason that LBL is setting logic "0". The writing operation is ready and use again DC source to input two inverters gate detecting V_R/V_L variation as shown in Fig. 2.22. As a result of the right-side inverter setting is the same with RSNM, so the one of Voltage Transfer Curve (VTC) is equal in Fig. 2.23.

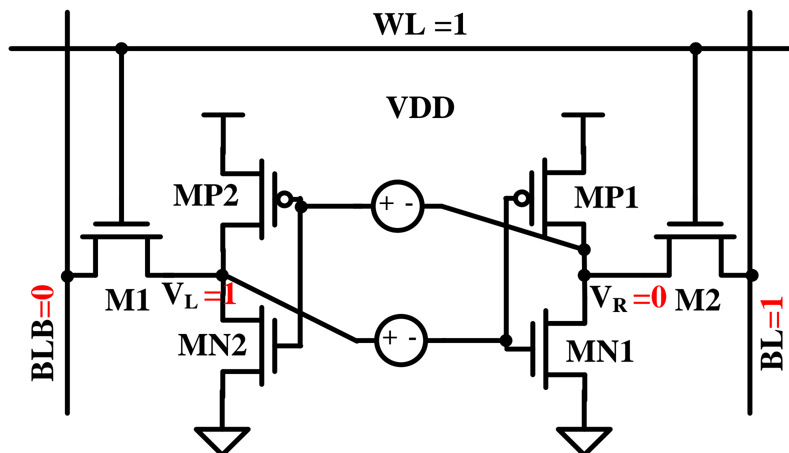


Fig. 2.22 WSNM detected circuit

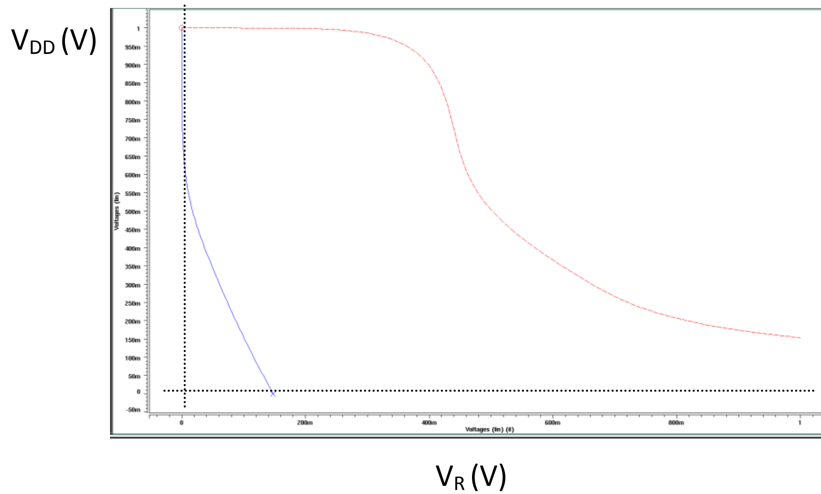


Fig. 2.23 The write-1 VTC

2.5 SRAM Write Margin (WM)

Write ability to use other methods this way is to Write Margin (WM). How to survey the value in a cell? We read on. First, we let the BLs to high level (VDD), and sweep down BL form VDD to GND. When the cell storage value is flipped all at once, now the BL voltage is defined as WM. Or change another way to say, we test the BL voltage start from GND to rise gradually as we detect the cell write fail that the voltage is WM. We see the Fig. 2.24 [2.8] to realize the definition.

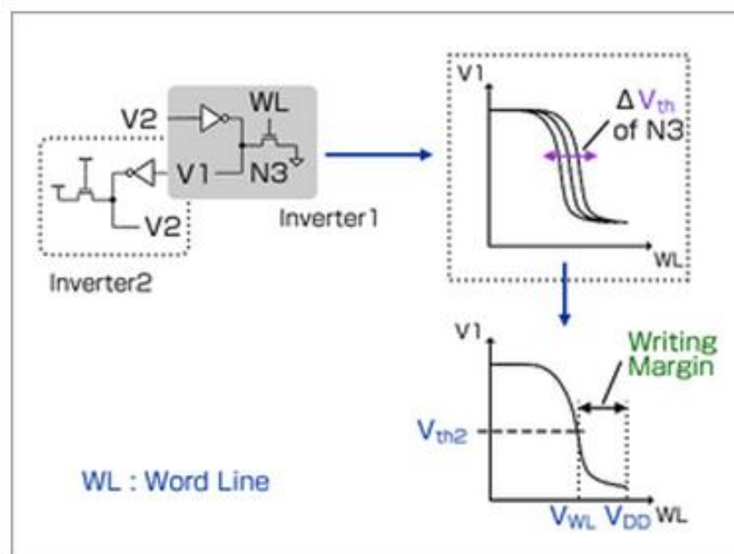


Fig. 2.24 Diagram of new write margin [2.8]

2.6 SRAM Array Structure

Cause of the memory in 3D products is very important that has to save or read many data in the period time, so the designed memory system need to a lot of areas for memory cells. Nevertheless, there is the way to design the memory array structure already and follows it can make well easily. The memory array basic structure is in Fig. 2.25. We always put many memory cells together that is called SRAM array if the memory is by SRAM. The SRAM array left side places the row decoder that working is decoded one signal for address to select a cell word line. Then, the SRAM array below puts a column decoder that running one signal by address to choose a cell bit line. To select one word line and one bit line their across a cell is called storage cell that will is going to write or read operation. If the working is read that pass through the sense amplifiers and comes out after some gating computations.

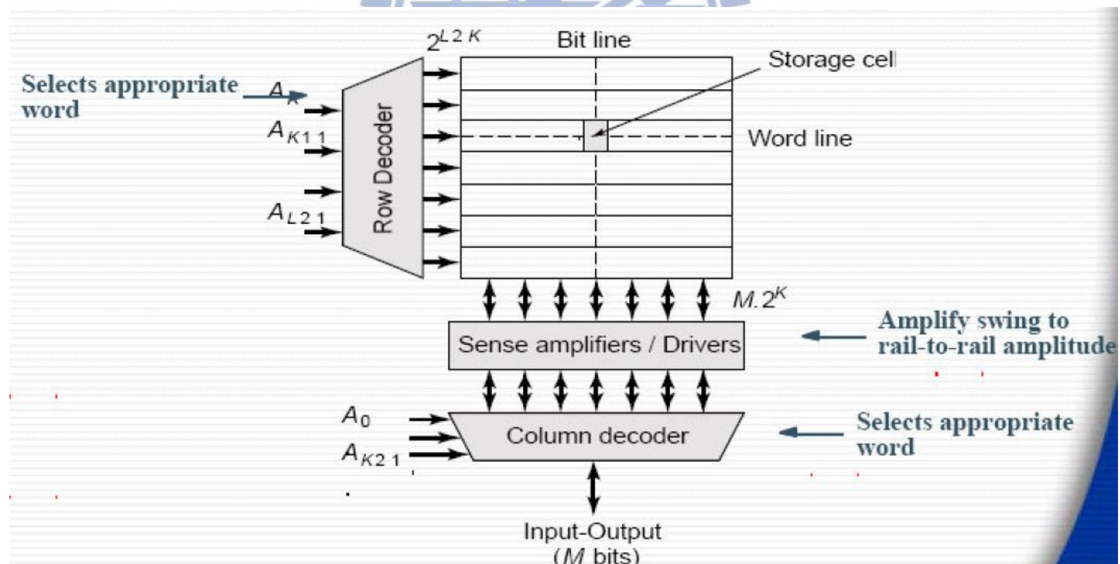


Fig. 2.25 Basic SRAM organization

To design a memory system we have many tests to test. Usually one circuit has one critical path to use tracing a lot of problems. Of course, SRAM organization also has a critical path as in Fig. 2.26. Most of critical paths are the longest of total system by designer. The reason is that the longer path will go through more gates in general. The

critical path starts from address and then sequence passes through address register, row decoder, column mux., sense amplifiers, comparator, and finally input/output register in Fig. 2.26. Most of delay is though the SRAM column, column mux., and sense amplifier. If want to reduce the delay time for this case, we should consider to discuss and solve how to cut down the gate delay that the path can be shorter than before.

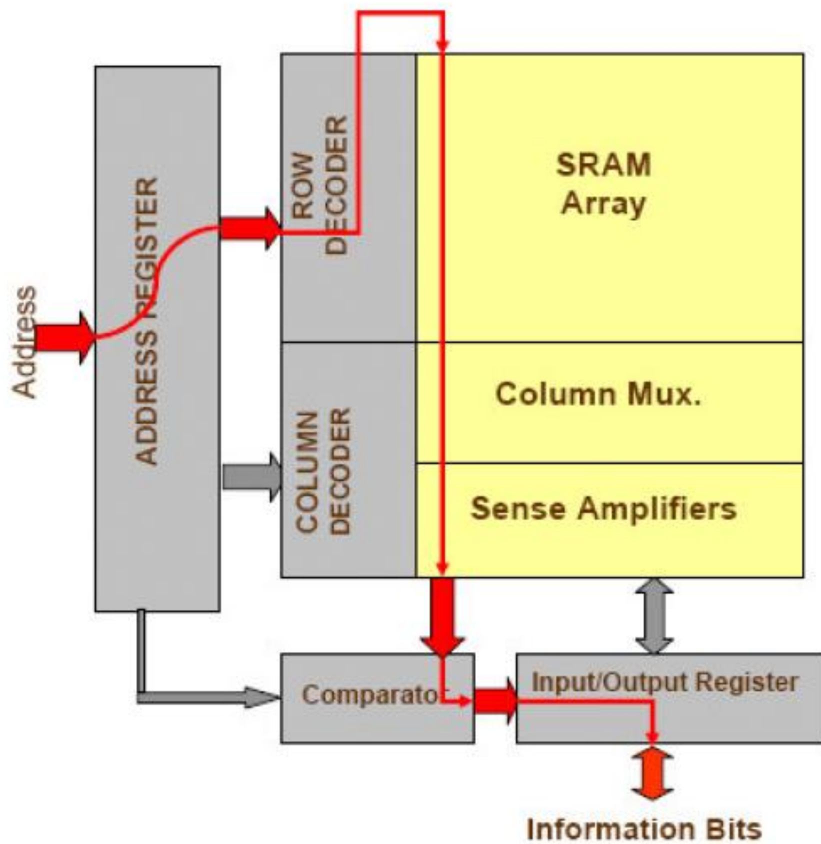


Fig. 2.26 SRAM critical path [2.9]

2.7 Variation Issue

2.7.1 Global and Local Variation

As we designed chip is manufactured from factories, anyway there must have some deviations in the real chip. For CMOS process, variations are able to be reflected on the threshold voltage (V_T) directly. Because of V_T is different from original we are designed, the current drive ability of transistor would be deviated and leakage probably become larger than initial device. The result is even serious enough to outcome failed or produce more power to run the chip. According to the advanced process is smaller in recently, the chip to make for lithography that must have diffraction be produced and comes to more critical.

In a deeper sight to face squarely the formation of variation, we are able to distribute the variation into Global and Local. Global variation has another saying that is called "inter-die variation". The variation is expressed from die between die each other. And then the Local variation is also called "intra-die variation". The definition is that the variations of transistors in one die. The two variations are shown in Fig. 2.27 [2. 10].

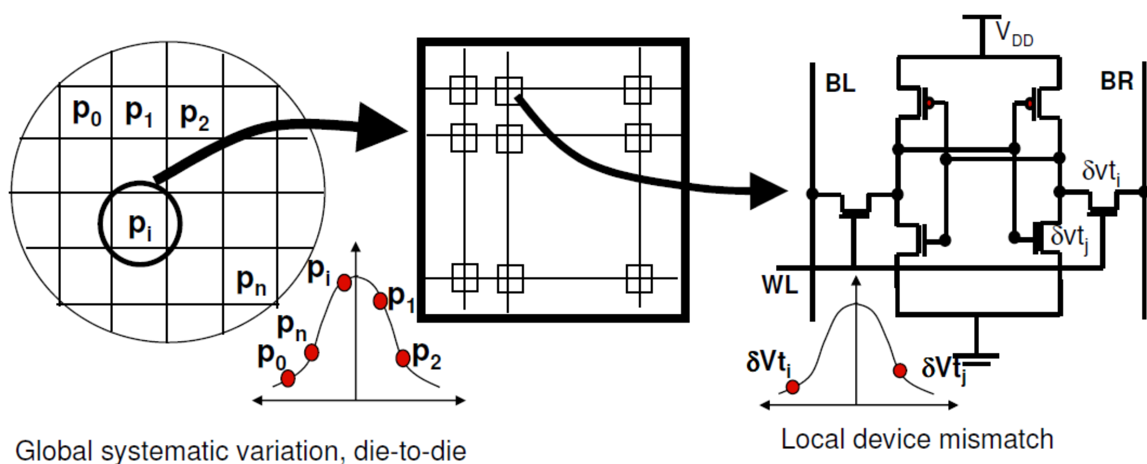


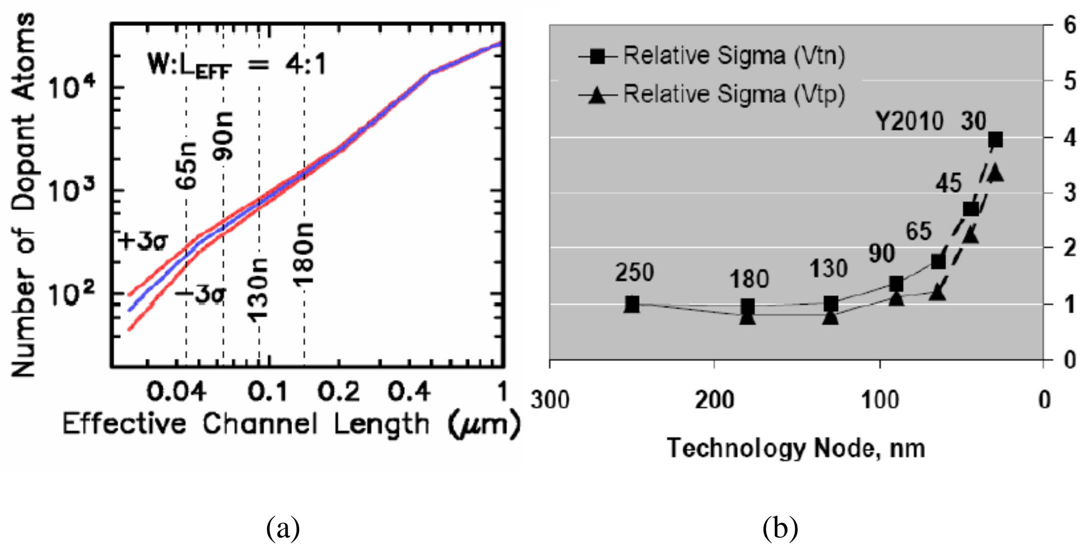
Fig. 2.27 Global variation and Local variation of V_T [2.10]

The figure of Fig.2.27 that δV_T can be equal to like this:

$$\delta V_T = \Delta V_{T_GLOBAL} - \Delta V_{T_LOCAL} \quad [2.11]$$

Besides, Global variation is the variation of environment like that temperature, lithography, machine settings, doping concentration, itself silicon film thickness and so on. We just only pick one selection of the variation to discuss that is doping concentration. At first, we generally realize the information for discrete dopant effect. For the 50nm L_{eff} (90/60nm node) that approximately has 200 dopant atoms in channel; for 12nm L_{eff} (32/28nm node) that approximately has 2.6 dopant atoms in channel; for 8nm L_{eff} (22/20nm node) that approximately has 1.7 dopant atoms in channel. This information tells us that if the process size is smaller, then the variation becomes larger and worse. We can look the Fig. 2.28 to understand that. In the 30nm process, the sigma (σ) of VT variation is raise by a factor of 4.

Fig. 2.29 shows an example of dopants positioned using this algorithm for a 25nm nMOSFET design. Using the technique, the threshold voltage fluctuations of many different MOSFET designs have been evaluated. The 0.225x0.2x0.1um simulated



(a) (b)

Fig. 2.28 (a) Number of dopant atoms in the channel as function of effective channel length (b) “sigma” of V_T variation of function of technology node

volume contained 224550104 Si atom positions, of which 62280 were converted to donors and 12703 were converted to acceptors. Darker dots are donors, lighter dots are acceptors. [2.12]

In addition to demonstrating the advantages of retrograde doping, we have also analyzed the separate effects of donor and acceptor descretiaztion in an aggressive 25nm channel length MOSFET, as shown in Fig. 2.30, which also illustrate that shallower source/drain doping yields smaller fluctuations. Results for a still smaller device design will also be shown, in addition to a study of the dependence of the V_T fluctuations on the device width and the nature of the boundary condition in the width direction. Besides, the source/drain doping profile is modeled as a Gaussian (10^{20} cm^{-3}, peaked at the surface) in both the vertical and lateral directions, with standard deviation d (depth parameter) in the vertical and 0.7d in the lateral direction [2.12].

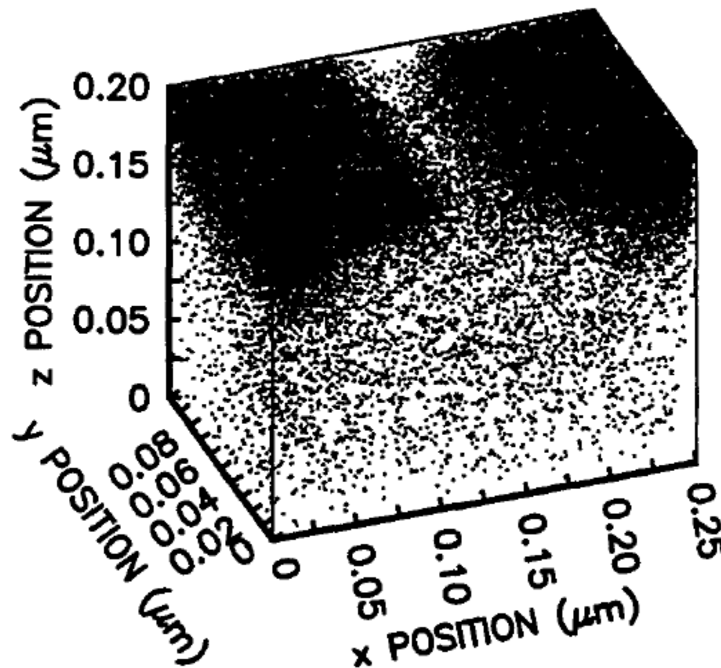


Fig. 2.29 Plot of randomly generated dopant positions in a 25nm MOSFET, viewed from the side. [2.12]

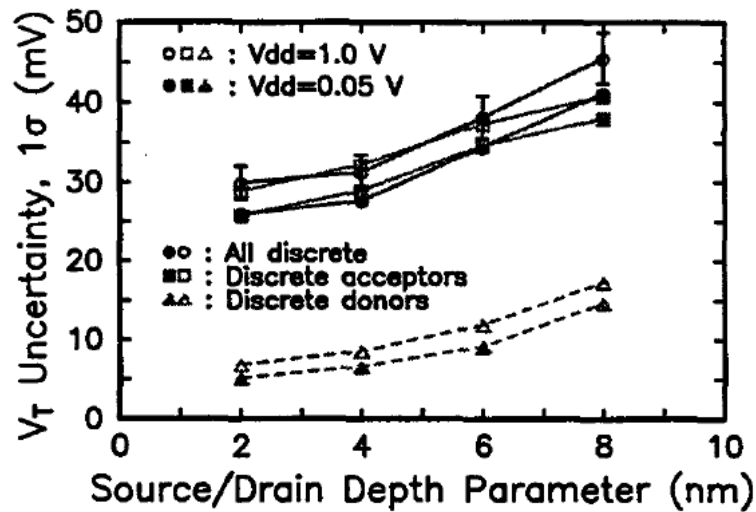


Fig. 2.30 Plot of threshold voltage uncertainty (1σ) versus the vertical depth parameter, d, of the source /drain doping for 25nm n-channel MOSFETs, comparing discrete donor effects to discrete acceptor effects. [2.12]

2.7.2 SRAM Cell Variation

Cause of low supply voltage become a trend in recent years, the V_T variation also comes to worse. If encounter large V_T variation, the design point has to tolerate V_T variation of 3% of VDD in 250nm process; the design point has to tolerate V_T variation of 20% of VDD in 90nm process; the design point has to tolerate V_T variation of 30% of VDD in 65nm process. At 45nm, design below 5σ that is impractical, and even useless because the redundancy needs for repair of projected array size.

V_T shift severely limits the scaling of SRAM cell size in that V_T mismatch in cell will exasperate the worse SNM even read or write two operations that one will fail. Fig. 2.31 [2.13] shows that owing to V_T variation, the cell switch-point and read-down level (read-disturb) begin to superimpose in 90nm process. Unfortunately, the technology node is going to lessen, and then the overlapping region comes to increase. In other words, SRAM stability gets worse in deep submicron technology.

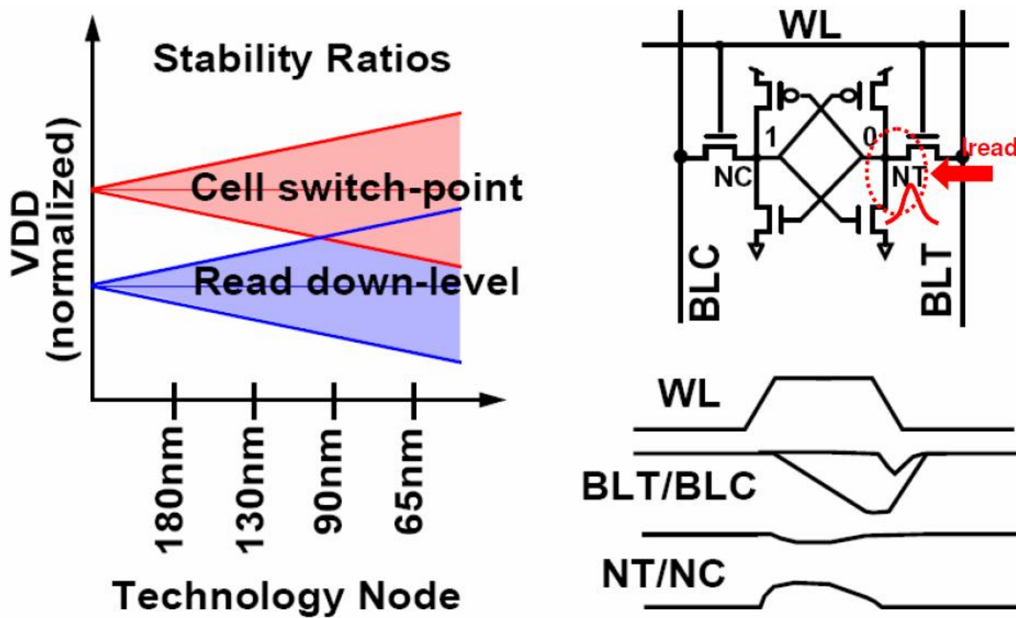


Fig. 2.31 Scaling and cell stability margin of 6T [2.13]

We start to discuss read operation at V_T variation and the SRAM uses 6T structure. The worse case in read mode corner is PSNF at Global variations that make the RSNM smallest in 6T cell. PSNF means the PMOS switching speed is slower than ordinary, and the NMOS switching speed is faster than general NMOS. The corner let the inverter trip voltage and read-disturb smaller than TT corner inverter. In addition, the Fig. 2.32 [2.11] shows the cross-couple inverter with V_T variation and the RSNM result. In the figure, V_T variation would move the voltage transfer curve vertically or horizontally along the side of the maximum nested square until the curves intersect at only one point [2.11]. While the curves overlap one point means the 6T cell cannot read regularly due to there is no RSNM.

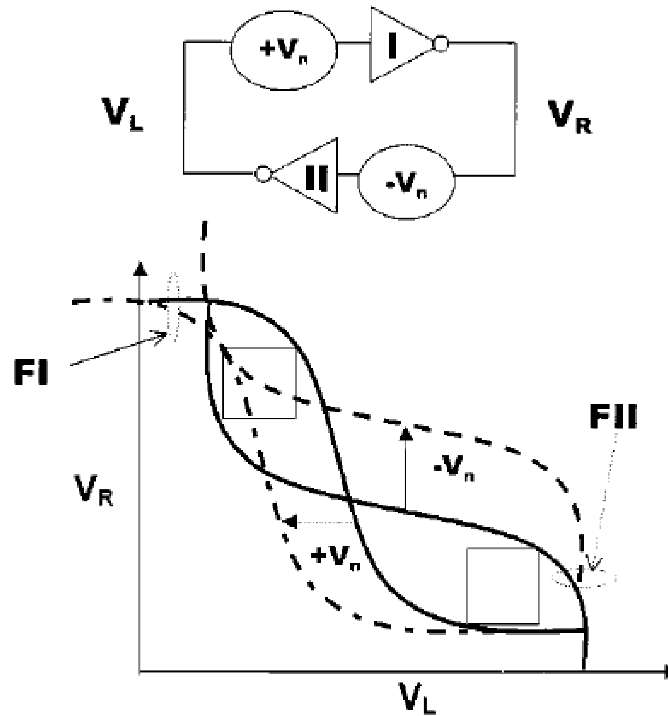


Fig. 2.32 6T RSNM with V_T variation [2.11]

Next on write mode, the worst corner at Global variation is PFNS that makes the cross-couple inverter of 6T cell trip voltage larger than normal. Another reason is pass-gate also NMOS composition that makes the external data transmits slower into cells. We consider the local variation in Fig. 2.33 which is the worst case in write mode. We assume the right-side becomes to write "0" and left-side write "1", which MN1 and M2 are used high V_T that writing trip voltage comes higher. In left side, the MP2 is used high V_T and MN2 is low V_T that condition let the trip voltage lower. We also show Fig. 2.34 that read operation in the worst V_T case.

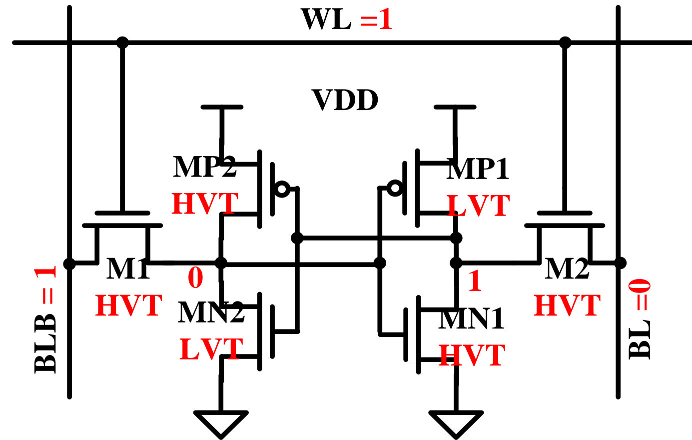


Fig. 2.33 The effect of local variation in write-mode with worse case

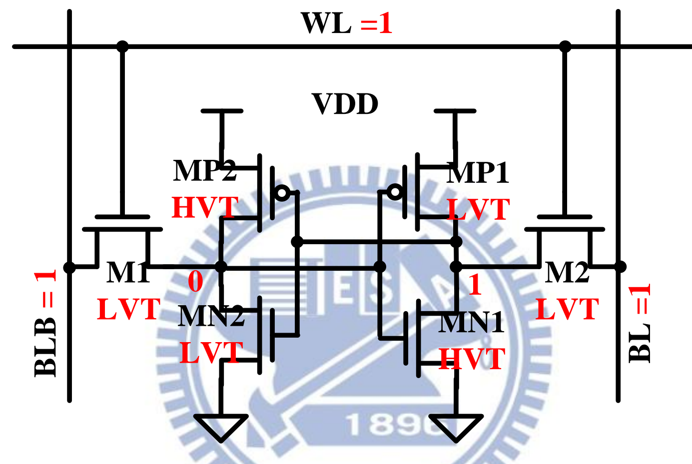


Fig. 2.34 The effect of local variation in read-mode with worse case

2.8 Modern SRAM Design Methodology

Because the process comes to smaller in recent years, the Global and Local variation, leakage, half selected, and SNM are very critical. To solve these problems and improve the system circuits, today use some design circuit methods or technological improvements on the process can clear up that.

2.8.1 Dual Supply Voltage

For improving the read/write ability, we can change the cell supply voltage at the appropriate time. If we wish to advance the read ability or RSNM, we can increase the cell supply voltage or come down the cell voltage from ground to negative. Else if we

want to improve the write ability or WSNM, we can reduce the supply of the cell. In Fig. 2.35 [2.14] which is the Dual-Vcc structure diagram.

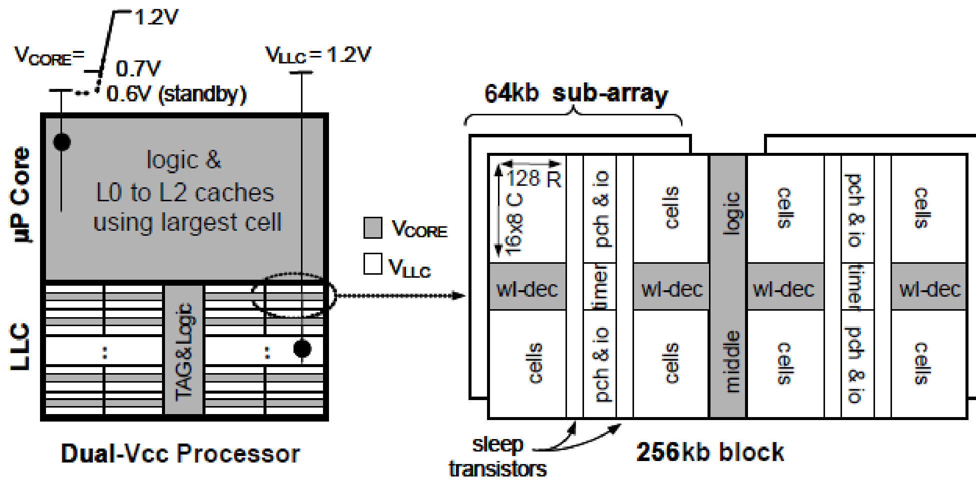
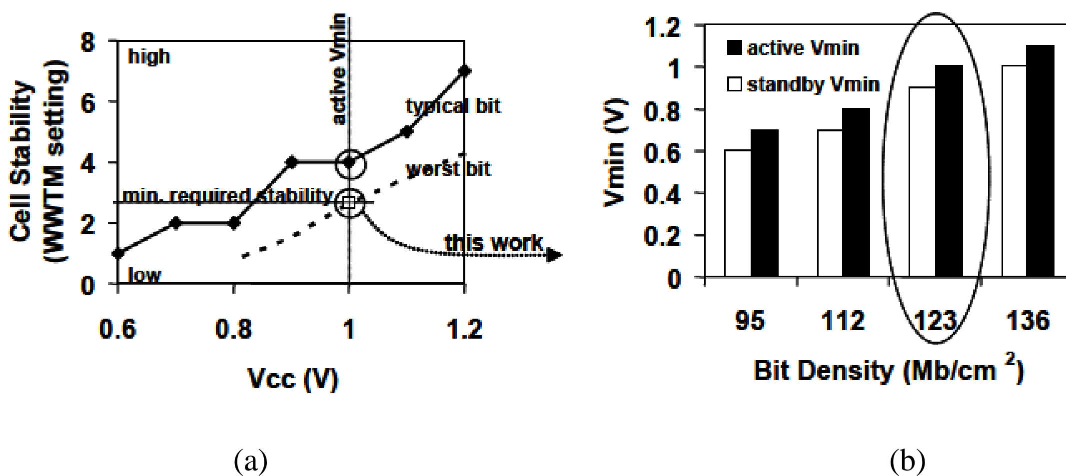


Fig. 2.35 Single-Vcc and Dual-Vcc processors [2-14]

Cells fixed at 1.2V in the last level cache (LLC) but cells voltage form 0.7 to 1.2V during operation in the core region that run 0.6V to achieve best energy efficiency in standby mode. The dual-Vcc design enables high-density SRAM than the conventional single-Vcc processor. We also see the Fig. 2.36 [2.14] to realize the cell stability compare at different Vcc and its operation Vmin in different bit densit (Mb/cm^2).



(a) (b)

Fig. 2.36 (a) Cell stability in the Vmin (b) Vnim in different bit desnsity [2.14]

In order to reach the goal to changes the cell supply voltage, we have many methods. However, we require second supply voltage that we have to design others power converter, power regulator, and so on. The cost is too expensive so we most of chips use just one supply voltage to generate second supply voltage by circuits.

The above description that the way changes the cell supply voltage to improve the read/write ability. In addition, in the Fig. 2.37 [2.15] the SRAM is used the cross-point 8T SRAM and the lowest supply voltage of the cell is “VSM”.

In read operation, the VSM, which connects to column’s VSS becomes negative bias by the read enable signal (RE). As a result, the enlarged SRAM cell bias improves the static noise margin (SNM). In general, a simple application of the negative VSS technique to the 6T SRAM accompanies with the power increase because all of the cell current in the half-selected columns increases simultaneously. [2.15]

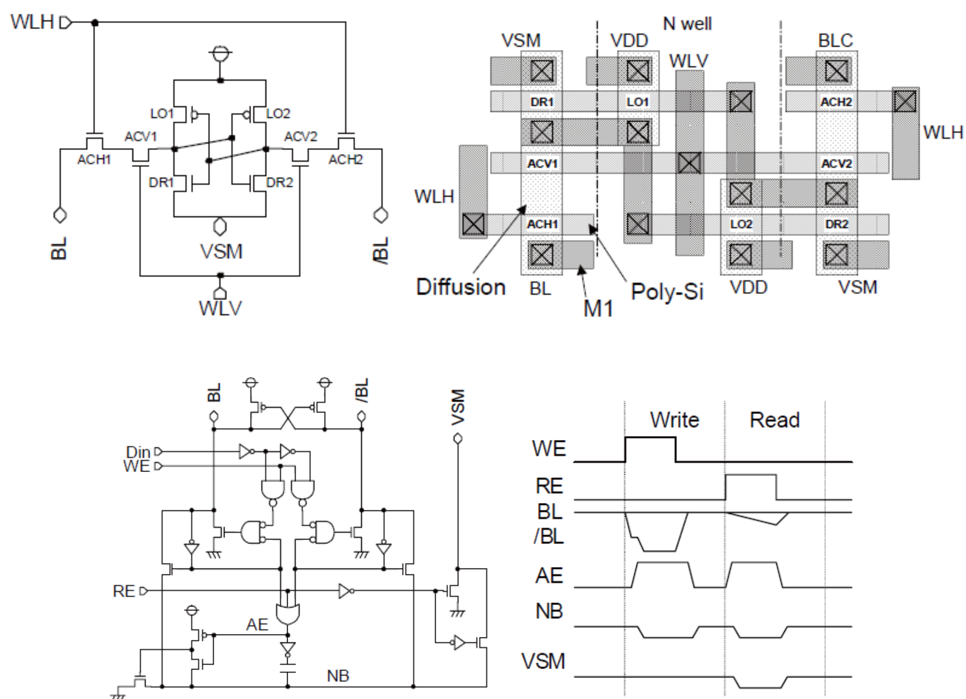


Fig. 2.37 The cross-point 8T SRAM schematic, layout, read/write negative-biased circuit, and waveform [2.15]

2.8.2 Negative Bit-Line

At the Fig. 2.37 [2.15], in write operation, the negative-based circuit will pull down the BL voltage to the negative. This method main improve the write ability. Lower bit-line voltage makes the cell written successfully in insufficient write margin. Fig. 2.38 [2.16] shows the circuit and waveform of Constant-Negative-Level Write Buffer. Its working use a negative-bootstrap circuit in a SRAM cell. The bootstrap capacitance and the signal boost-enable running depending on the memory cell count on a bit-line. It means that the charge in C_boost is proportional to each cell bit-line. Besides, the target bias range is -0.15V±0.05V. The design does not be much negative to hold the data in non-chosen cells and writing the data into the select cell shown in Fig. 2.39 [2.16].

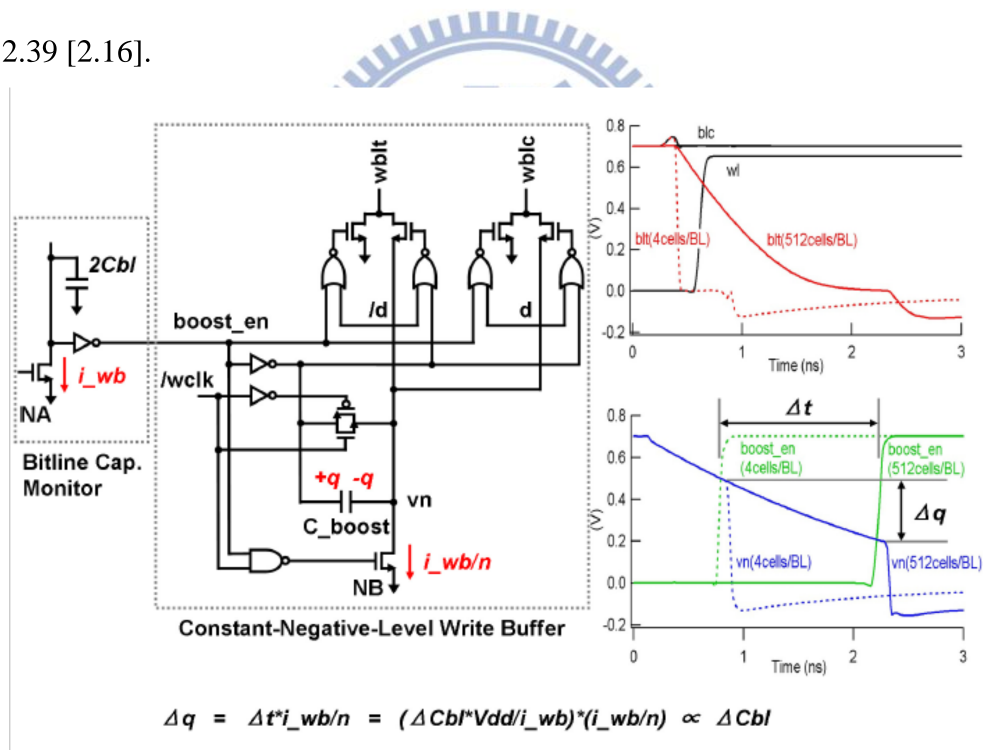


Fig. 2.38 Constant-negative-level write buffer. [2.16]

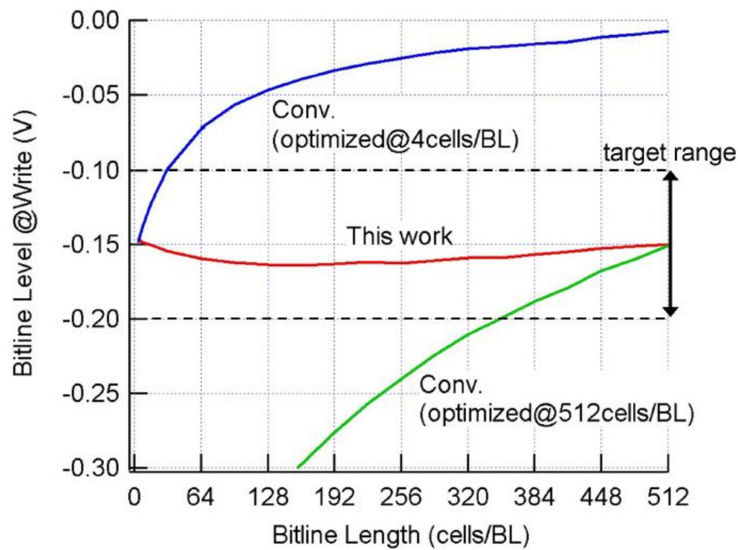


Fig. 2.39 Simulated negative bit-line level. [2.16]

Fig. 2.40 [2.17] shows that the write drive also use the capacitance "Cboost" to boost the negative voltage. The node "Nboost" connects to eight physical BL pairs, including the upper (Ntu/Ncu) and lower half (Ntl/Ncl) partitions. The Nboost node is per-charged to GND and the WS1n voltage is VDD before into write cycle. As gets into write cycle, WS1n node is pulled down to GND and the Nboost node voltage is turned into a negative voltage instantly in order to achieve bit-line voltage negative. The result (Fig. 2.41 [2.17]) of the designed circuit is worse in high voltage than in low voltage for the effect of negative bit-line.

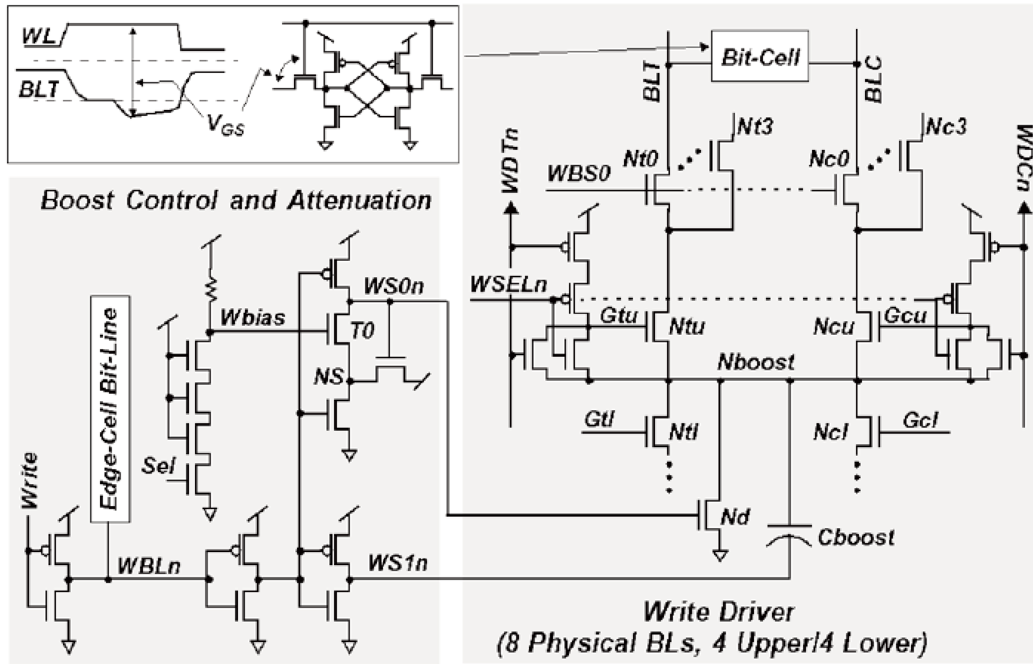


Fig. 2.40 Write driver with boost control [2.17]

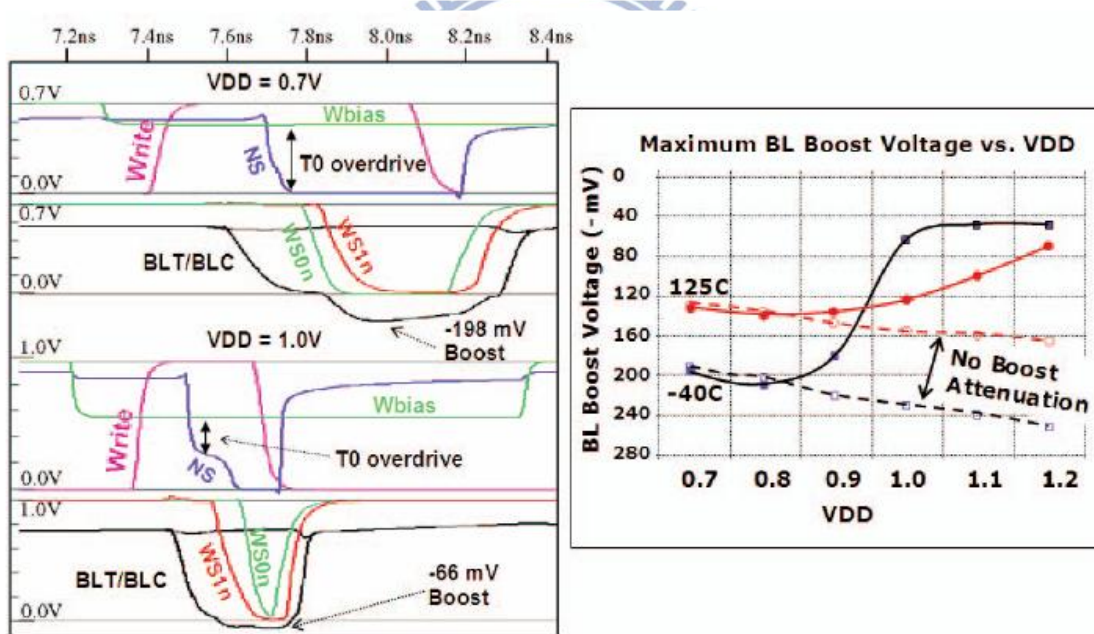


Fig. 2.41 Write cycle simulation waveforms and results [2.17]

2.8.3 Dynamic Word-Line Voltage

To control the WL voltage also is the way to improve the read/write ability. As the WL voltage increase, the pass-gate is able to be stronger due to the read/write speed better than original. Nevertheless, in 6T SRAM the read-disturb will raise and RSNM comes into lower. Fig. 2.42 [2.18] shows the RSNM with WL voltage with/without local V_T variation. Expressly, the Fig. 2.42(a) verities the argument, which WL voltage lower and the SNM is better in symmetry margin. However, we should consider the variation of V_T so the result (Fig. 2.42 (b)) cannot run in lower voltage. In this paper, the WL level should be lowered by more than 20% compared to the supply voltage of 1.0V.

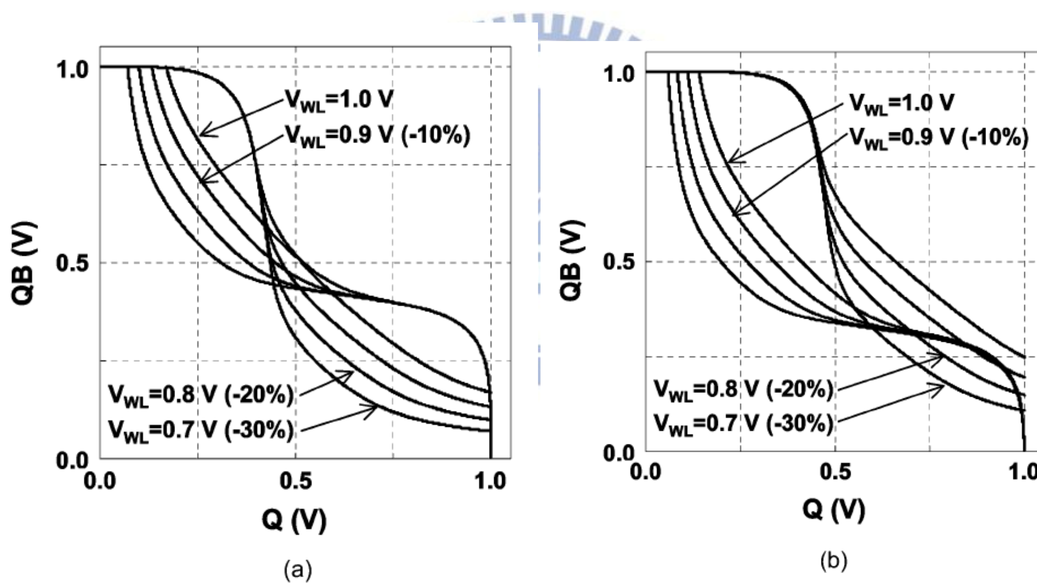


Fig. 2.42 SNM improvement by lowering WL voltage in 6T SRAM in case of (a) with and (b) without local V_T variations. [2.18]

Cause of the WL voltage lower is better in read mode on the 6T SRAM, Fig. 2.43 [2.18-19] shows the read assist circuit (RAC) to control the WL voltage and solve the problem. In the conventional circuitry, which use multiple pull-down NMOS transistors (called replica access transistors, RATs) to reduce the voltage of WL. WL voltage versus V_T simulated results for two temperatures we found two problems. The

first problem: in the FS condition, WL voltage drops too much at -40℃ that degrades the operation speed and write margin. The second problem: WL voltage at SS corner, which is the worst condition of the cell current. To solve the problem the paper design another circuit that is proposed circuitry. According to the simulated results the temperature dependence of a resistance is generally smaller than that of MOS transistors.

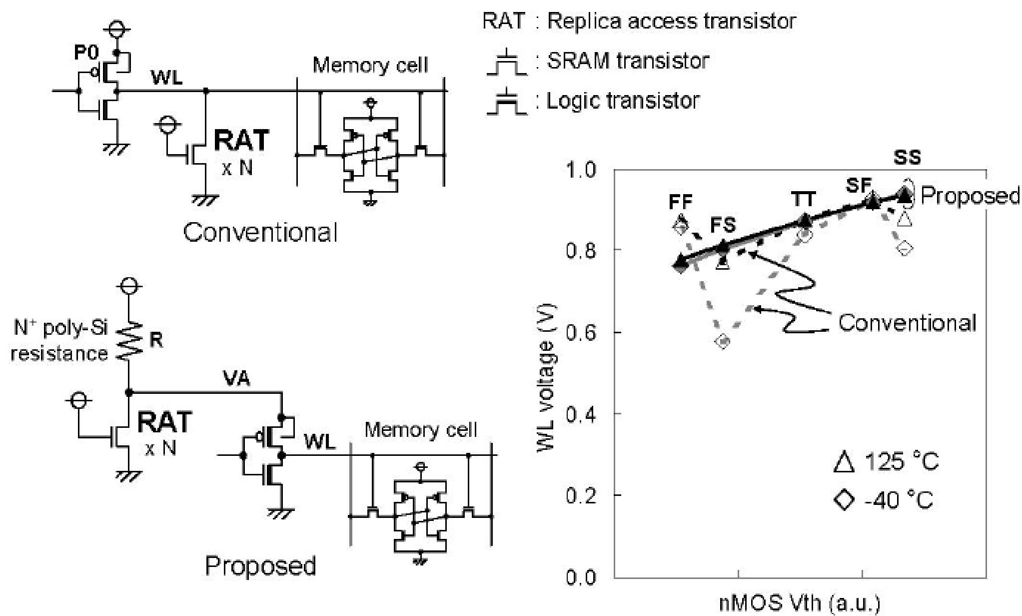


Fig. 2.43 Improving read stability depending on process and temperature variations. [2.18-19]

There is a write assist circuit (WAC) in Fig. 2.44 [2.18-19]. Lowering the voltage level of power line in the memory cell array (ary-VDM) is one of the effective ways of ensuring the SRAM write margin. To enhance the write margin against the increasing variation accompanied by the scaling it is necessary to lower the voltage of ary-VDM immediately. The ary-VDM lowering in several cases of the segment division number (#div=1, 2, 4, 8) and displays the waveform in Fig. 2.45(a) [2.19]. Also shows the DC simulation result of improvement of write ability defined by

write-trip-point in Fig. 2.45(b) [2.19]. We realize the division number larger and the write-trip-point becomes higher to write easily.

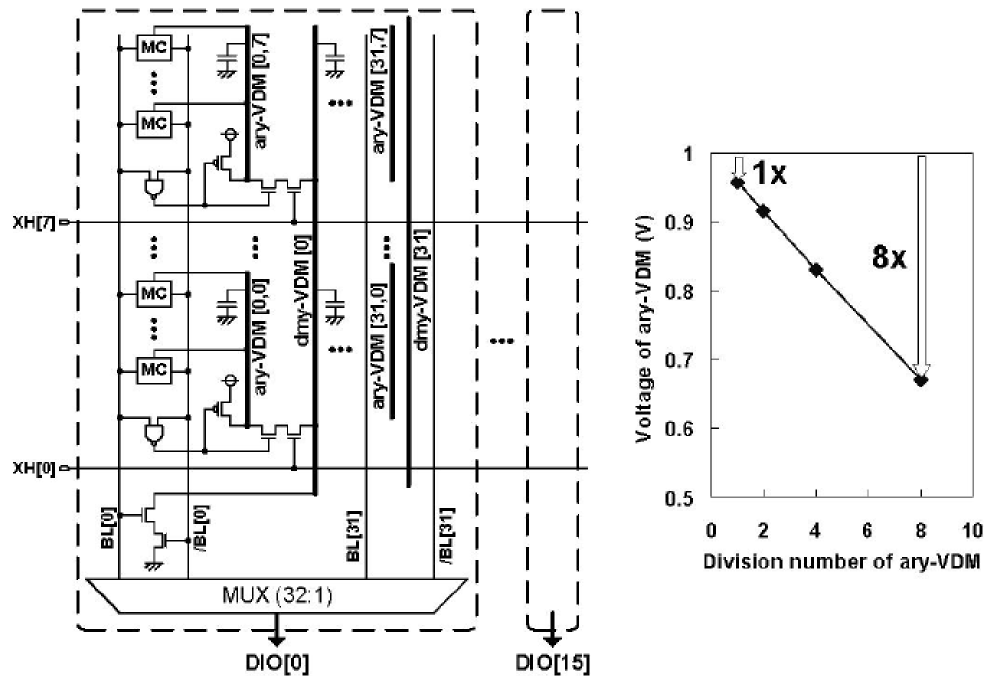
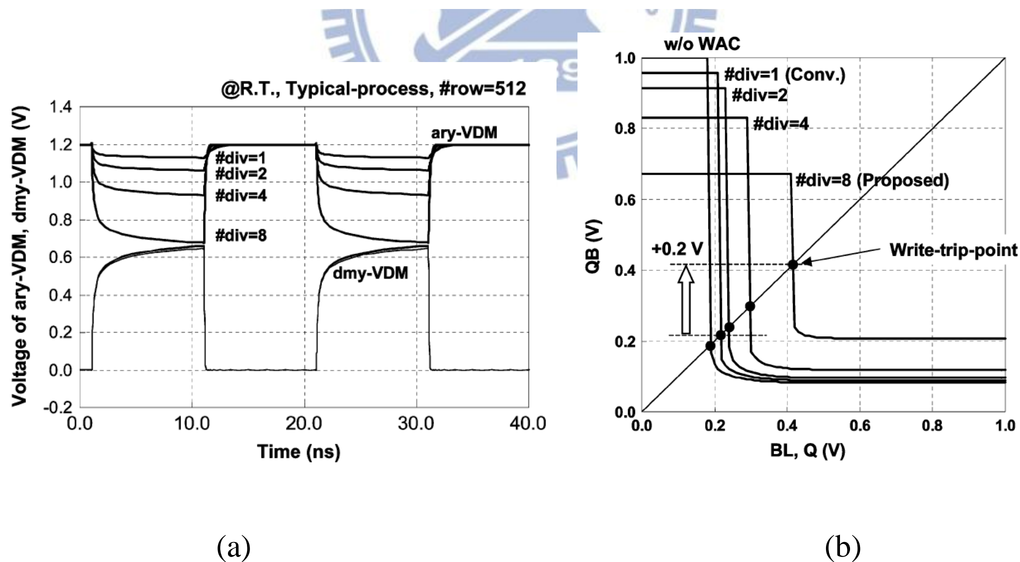


Fig. 2.44 Write assist circuit (WAC) improving write stability [2.18-19]



(a) (b)

Fig. 2.45 (a) Simulated waveform of the ary-VDM and dmy-VDM in the write status. (b) Comparison of the write ability by DC simulation result of the write-trip-point. [2.19]

2.9 Power Consumption

Due to the thesis designed chip main operates in low voltage, so the power also reduces naturally. The total power is composed of dynamic power, short-circuit power, and static power dissipation. We will in order to discuss these three power dissipations in below description.

2.9.1 Dynamic Power Dissipation

Fig 2.46 [2.20] is a CMOS schematic of inverter. The output currents I_{DP}, I_{DN}, and the load capacitance C_{int} are the major effect of dynamic power dissipation. Dynamic power dissipation usually defines average power of the period T as V_{IN} signal changes VDD from GND at moment and V_{IN} signal changes GND form VDD transiently.

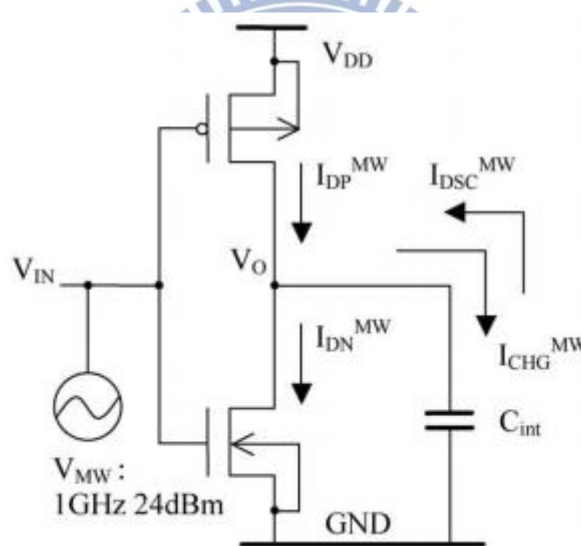


Fig. 2.46 The inverter schematic [2.20]

We drive the average dynamic power dissipation formula like that:

$$P_D = \frac{1}{T}\int_0^T [\, I_{DN} * Vo \; + I_{DP} * (VDD - Vo)]dt \qquad \text{, according to } I_{DN}=I_{DP}=I_C=C\frac{dV}{dt}$$

$$= \frac{1}{T}\int_0^T [\, C_{int}\frac{dVo}{dt} * Vo \; + C_{int}\frac{dVo}{dt} * (VDD - Vo)]dt$$

$$= \frac{Cint}{T}\int_0^{VDD} [\, Vo \; + (VDD - Vo)]\, dVo$$

$$= \frac{Cint}{T} * VDD \int_0^{VDD} dVo$$

$$= \frac{Cint}{T} * VDD^2 = f * C_{int} * VDD^2 \quad (2.1)$$

We get the equation in above derivation; f is represented to the frequency for the input signal. In addition, gates usually do not switch every cycle we have to think the probability of switching. For this reason, we superadd the factor α in the equation 2.1. The final Dynamic power is expressed as:

$$P_D = \alpha * f * C_{int} * VDD^2 \quad (2.2)$$

We can consider the (2.2) equation to realize that dynamic power is proportional to the switch factor, frequency, loading capacitance, and square of supply voltage.

2.9.2 Short-Circuit Power Dissipation

A CMOS inverter working in digital usually turns on only one transistor. If input signal is at high level (VDD), the NMOS transistor conducts, and else if input signal is at low level, which the PMOS will conduct. However, there will be a time period in which both the NMOS and PMOS turn on during a transient on the input. Because of a short-circuit (I) to flow from supply to ground as shown in Fig. 2.47 [2.21] for an inverter without loading.

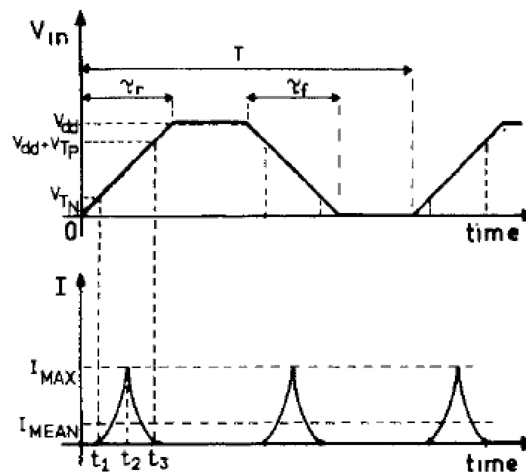


Fig.2.47 Current behavior of an inverter without load. [2.21]

Short-circuit power in general can be expressed as

$$P_{SC} = I_{mean} * VDD \qquad (2.3)\ [2.21]$$

Where I_{mean} is represented to the mean current during a time T, which equals to one period of the input signal. It can be written as

$$I_{mean} = \frac{1}{12} * \frac{\beta}{VDD} * (VDD - 2V_t)^3 * \frac{\tau}{T} \qquad (2.4)$$

Where τ is the time by input rise and fall time, and β is the gain factor (uA/V^2) of a transistor. In the equation, we realize to decrease supply voltage and rise/falling time of input signal that can mitigate the short-circuit power dissipation.

2.9.3 Static Power Dissipation

There is no DC current path for logic device in ideal condition stand-by mode. Nevertheless, there is current on MOS for standby in fact, which main is leakage current. Then, we will introduce major leakage currents of MOSEFT in below description [2.22-23]. Most of leakage can be divided into three parts: gate leakage, sub-threshold leakage, and reverse-bias PN-junction current as shown in Fig. 2.48 [2.22].

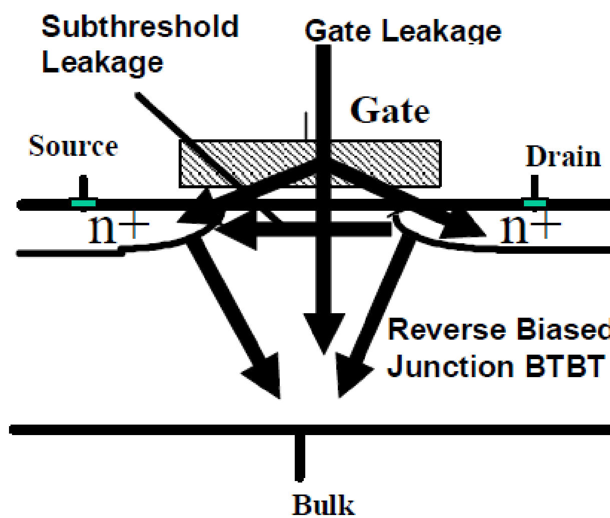


Fig.2.48 Leakage components in a scaled transistor [2.22]

GATE LEAKAGE

Gate leakage is composed of tunneling into and through gate oxide, injection of hot carriers form substrate to gate oxide, and gate-induced drain leakage (GIDL). At first, we discuss the oxide tunneling current. Today reduction of gate oxide thickness gets to increase the field across the oxide. The high electric field coupled with low oxide thickness makes electrons tunneling from substrate to gate and also from gate to substrate through the gate oxide. [2.23]

The second, we introduce the leakage of hot-carrier injection. Due to high electric field near the Si-SiO_2 interface in a short-channel transistor, electrons or holes can gain sufficient energy form the electric field to cross the interface potential barrier and enter in the oxide layer as shown in Fig. 2.49. The effect is known as hot-carrier injection. [2.23]

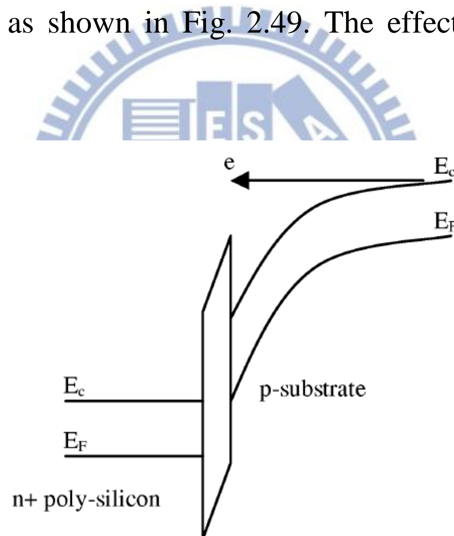


Fig.2.49 Injection of hot electrons from substrate to oxide [2.23]

Final gate leakage is Gate-Induced Drain Leakage (GIDL). As a result of GIDL is high filed effect in the drain junction of an MOS transistor. As the gate is biased to form an accumulation layer at the silicon surface, the silicon surface under the gate has almost same potential as the p-type substrate. Due to presence of accumulated holes at the surface, the surface behaves like a p region more heavily doped than the substrate. This causes the depletion layer at the surface to be much narrower than

elsewhere. While the negative gate bias is large shows in Fig. 2.50, the n+ drain region under the gate can be depleted. The condition causes more filed crowding and peak field increase, resulting in a dramatic increase of high field effects. [2.23]

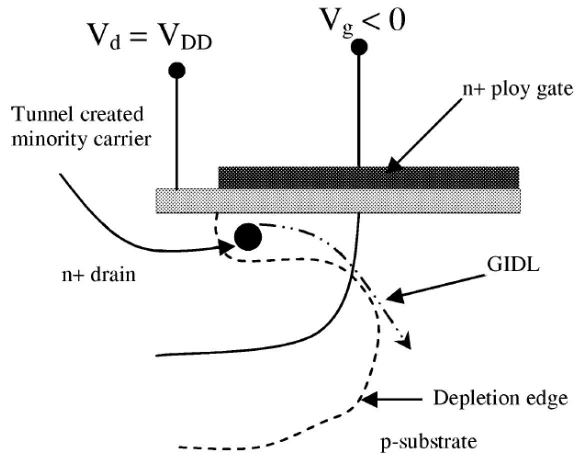


Fig.2.50 As n+ region is depleted or inverted with high negative gate bias, condition of the depletion region near the drain-gate overlap region of an MOS transistor [2.23]

SUBTHRESHOLD LEAKAGE

Subthreshold or weak inversion conduction current between source and drain in an MOS transistor occurs when gate voltage is below V_T. Fig. 2.51 shows the variation of minority carrier concentration along the length of the channel for an n-channel MOSFET biased in the weak inversion region. We let the source is grounded, $V_g<V_T$, and the drain to source voltage $|V_{ds}| \geq 0.1V$. For the condition, V_{ds} drops almost entirely across the reverse-biased substrate-drain pn junction. [2.23]

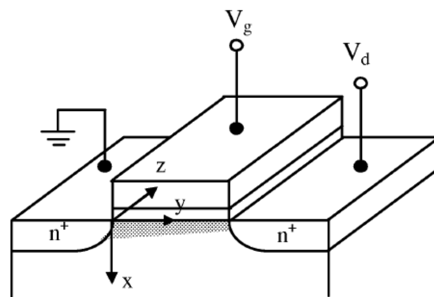


Fig.2.51 Variation of minority carrier concentration in the channel of a MOSFET biased in the weak inversion. [2.23]

PN JUNCTION REVERSE-BIAS CURRENT

A reverse-bias pn junction leakage has two main components: one is minority carrier diffusion/drift near the edge of the depletion region; the other is due to electron-hole pair generation in the depletion region of the reverse-biased junction. Pn junction verse-bias leakage is a function of junction area and doping concentration. If both n and p regions are heavily doped, band-to-band tunneling (BTBT) dominates the pn junction leakage. [2.23]

Chapter 3

40nm 512Kb Pipeline Low VDD 8T SRAM

3.1 Introduction

To make the chip performance better and cause to the thesis major to reduce the supply voltage, before beginning design the 512Kb-memory, we read some paper to realize and learn the skips to use that. [3.1-10]

As the process is getting smaller and smaller, we started to care about the power eliminate as much as possible to reduce. So, this chip used 8T SRAM is not a conventional single-ended 8T cell (Fig. 3.1(a)) [3.11]. We use the 8T cell with adaptive VVSS control (Fig. 3.1 (b)) [3.12] and add the differential data-aware power supplied above a set of 16-bits 8T cells. Both techniques are able to reduce SRAM active power and leakage power. Moreover, 8T cell with adaptive VVSS control has the same function of non-read-disturb with conventional single-ended 8T cell.

To improve working in higher frequency, we use the concept of pipeline insert to the whole chip. Then, wish for lower VDD_{MIN}, we put into the boosting word-line (WL) scheme in the 512k-bits memory. The function as well as improve write/read ability. However, boosting WL scheme is primarily used improve write ability; read ability just an additional effect. When simulated boosting WL in higher VDD (1.1V), we found the write/read ability to increase the trending less obvious. Inversely, it is more useful at low VDD (0.65V), so designed a voltage detector to control whether turn on WL boosting or not. These above schemes, we will do a careful description in the following article.

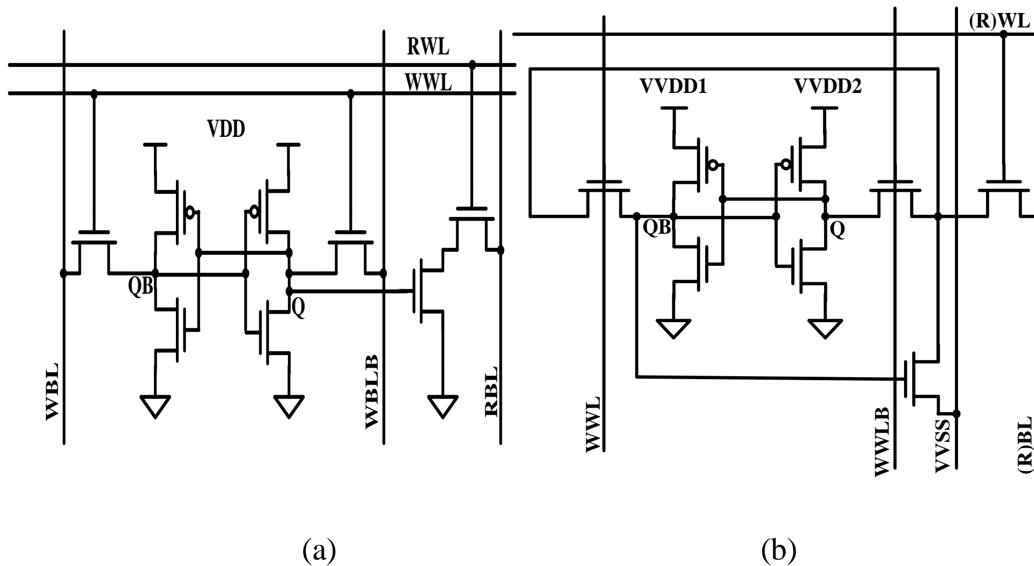


Fig. 3.1(a) A conventional single-ended 8T SRAM cell.[3.11] (b) A disturb-free DAWA single-ended 8T SRAM cell. [3.12]

3.2 8T SRAM Operation

3.2.1 Conventional Single-Ended 8T SRAM

We restart again introduce the conventional 8T SRAM. This 8T cell (Fig. 3.1 (a)) [3.11] adds two-transistors read stack in the conventional 6T cell. The below transistor gate of the stacking-transistor is connected to the node of Q. While write operation, we just turn on the WWL directly and RWL not turn on. Then the working way is the same by the conventional 6T SRAM cell. Afterwards, the read operation is turning on RWL and turning off WWL that the current only through two stack transistors if the Q value is logic “1”, then RBL is pulled down to logic “0”. If Q is logic “0” that RBL voltage is a floating “1”. In addition, the two cross-coupled inverters of 8T SRAM condition is alike standby state so that don’t have read-disturb problem.

3.2.2 Disturb-free DAWA Single-Ended 8T SRAM

This 8T cell is used to our chip (Fig.3.1(b)) [3.12]. Due to improve the write-ability, we adjust the PMOS of the cross-coupled inverter to be used High-Vt, another transistors are used regular-Vt. Two internal pass-gates are controlled WWL and WWLB by column-base respectively. The external pass-gate is handled row-base signal (R)WL and the source/drain end is connect to (R)BL, the other end is linked to internal pass-gate. The VVDD1 and VVDD2 are controlled by DAWA we next section will tell them.

Because we use the 8T SRAM cell design the chip, we will realize the 8T SRAM basic operation in detail. In stand-by mode, the WWL, WWLB, and (R)WL don't turn on. We set the VVSS is logic 0 and (R)BL is logic "1", as shown in Fig. 3.2(a). The stand-by mode major purpose is pre-charge the (R) BL.

In read mode, the WWL and WWLB also don't turn on that can improve the Read Static Noise Margin (RSNM). Moreover, the VVSS is set logic "0" and the (R)WL turn on let the (R)BL floating. If the QB is logic "0", then the NMOS connect by the VVSS don't turn on that the (R)BL is going to floating "1". On the other hand, the QB is logic "1", the same NMOS change to switch on that the (R)BL discharge to the floating "0", as shown in Fig. 3.2(b). We just pass through two transistors to get the read data.

In write mode, we use the basis of node Q to write 1 or write "0". While write "1", it means the Q node data will change from "0" to "1". The same write-mode operation is turning on the (R)WL and discharging the (R)BL to the logic "0". Supposing write "1", WWL switch on and WWLB switch off. VVSS become logic "1", and then the discharging path is shown in Fig. 3.2(c). If write "0", WWL turn on and WWLB turn off, VVSS become logic "0" that the path of discharging as shown in Fig. 3.2(d).

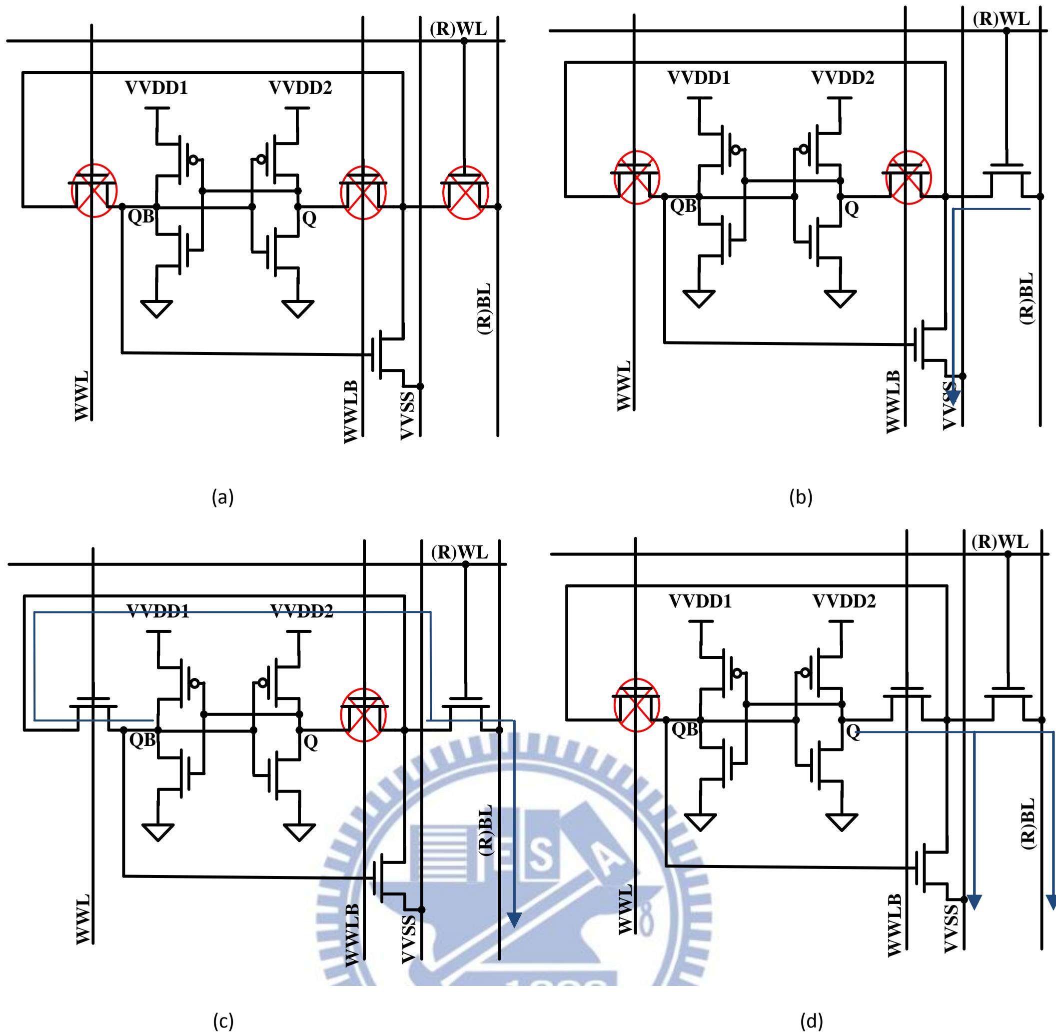


Fig. 3.2 (a) The operation for stand-by mode. (b) The read-mode operation. (c) The operation for wirte-1 mode. (d) The operation for write-0 mode.

Fig. 3.3 is the disturb-free new single-ended 8T cell layout. We place the WWL and WWLB between two gates in row-based from original 6T cell layout. There are three row gates at right-side and left-side. The middle layout don't change that stand for two cross-couple inverters. The 8T SRAM cell layout length is 1.44um and width is 0.59um. So the cell size is 1.44*0.59=0.85um^2 in UMC 40nm LP CMOS process.

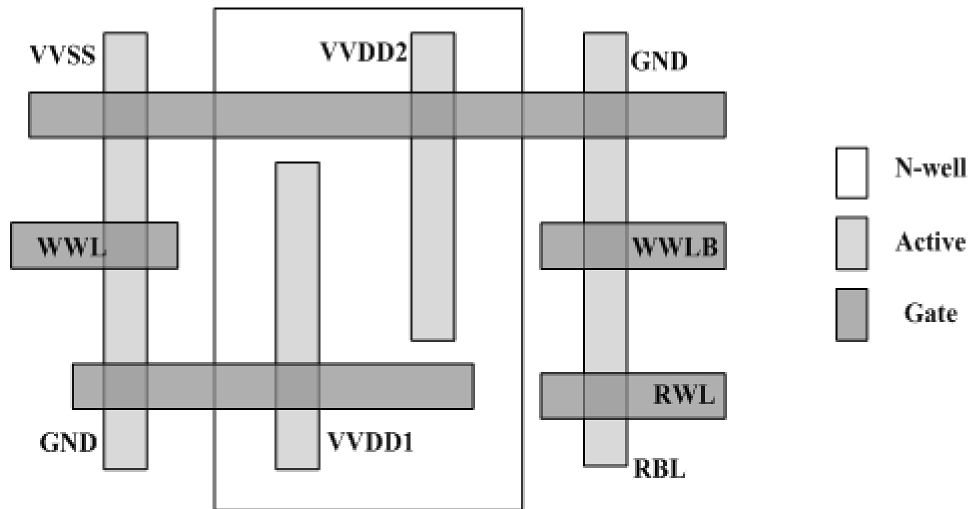


Fig.3.3 Cell layout for disturb-free DAWA single-ended 8T SRAM

Finally, we summarize the new 8T SRAM operation in the table 3.1. There are shown three modes at standby, write, and read conditions. The operation is worth noting that the (R)BL is floating in the read state.

mode	Standby	Write 0	Write 1	Read
(R)WL	0	1	1	1
WWL	0	0	1	0
WWLB	0	1	0	0
VVSS	0	0	1	0
RBL	1	0	0	floating

Table 3.1 Disturb-free new 8T SRAM operation

3.3 Data-Aware Write-Assist (DAWA) and Interleaving

Fig. 3.4 shows one data-aware write-assist (DAWA) set shares the column with 16 SRAM cells due to the chip local bank array is 16*32 bits. The reason is that we have a boosting function the circuit own a capacitor. The capacitor size don't use big thus local bank divide up size not too large, too. The DAWA switch control by WWL and WWLB. Fig. 3.4 tell us the operation of writing "1" and writing "0" that only one turn on between WWL and WWLB. Furthermore, the DAWA turn-on side the voltage is floating smaller than VDD that can improve write ability. In other words, cause of floating voltage of the PMOS of cross-couple inverter that be able to let the discharging current easily and faster in selected cell. The cell data "1" is not VDD just become to "floating 1".

In addition, we consider the signal VVSS in column. If we meet writing operation, that must have column half-selected problem. We use the critical write condition in the Fig. 3.4. In write "1" state, the VVSS is logic "0" in case that occurs disturb at column half-selected cells. So we choose VVSS is logic "1" that don't happen the above problem. Likewise in write "0" state, the VVSS signal is set logic "1" that come about disturb in column half-selected cells. We should change the signal to logic "0" that can solve the problem. Design the VVSS operation in write mode that also clear up the half-selected discharging problem for DAWA working. The signal DAWA_WE will be handle from DAWA tracking circuit that describe in following article. This chip uses the interleaving array configuration shown as Fig. 3.5 that enhance totally the speed of date processing.

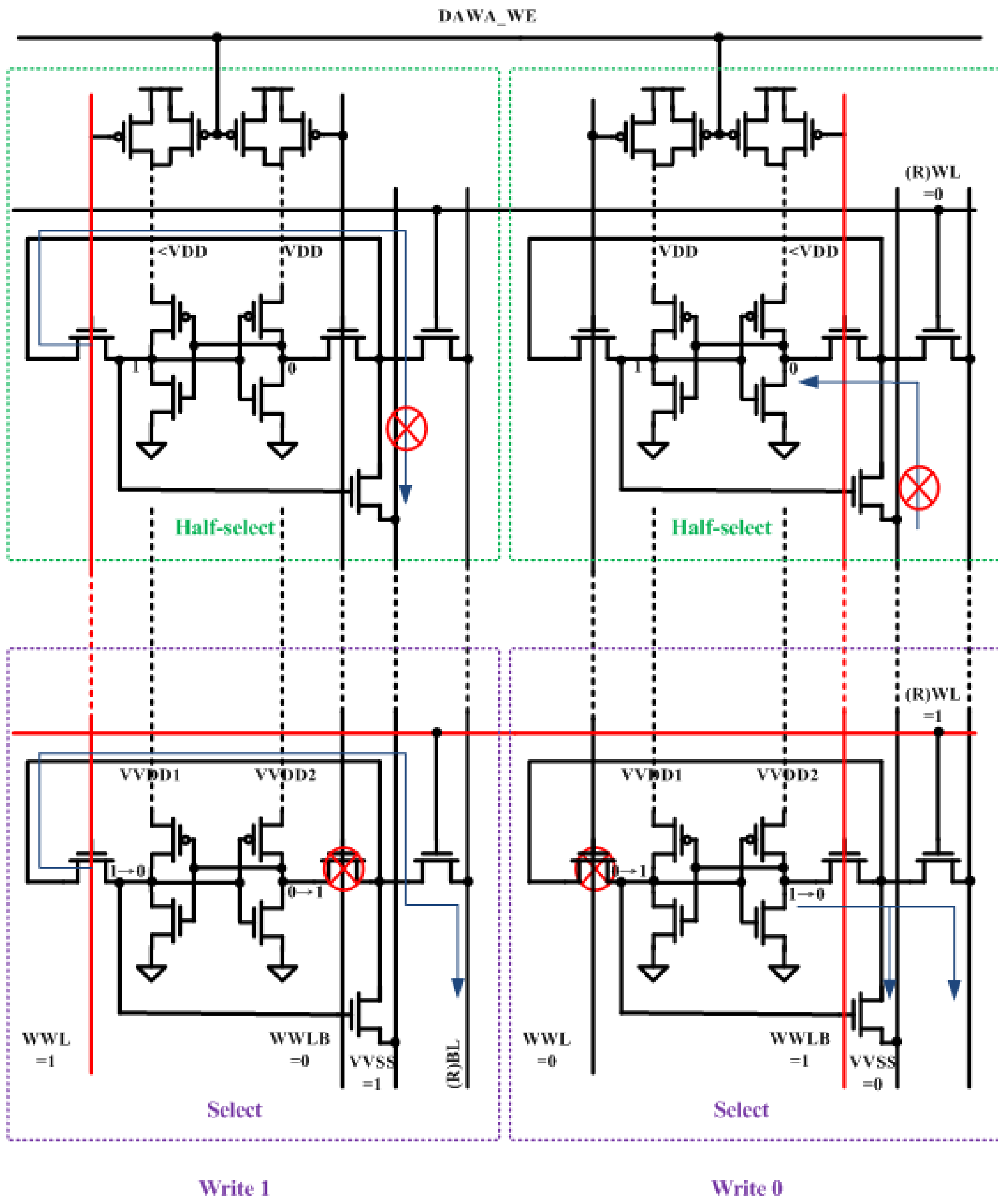


Fig.3.4 Write-1 mode and write-0 mode with half-select

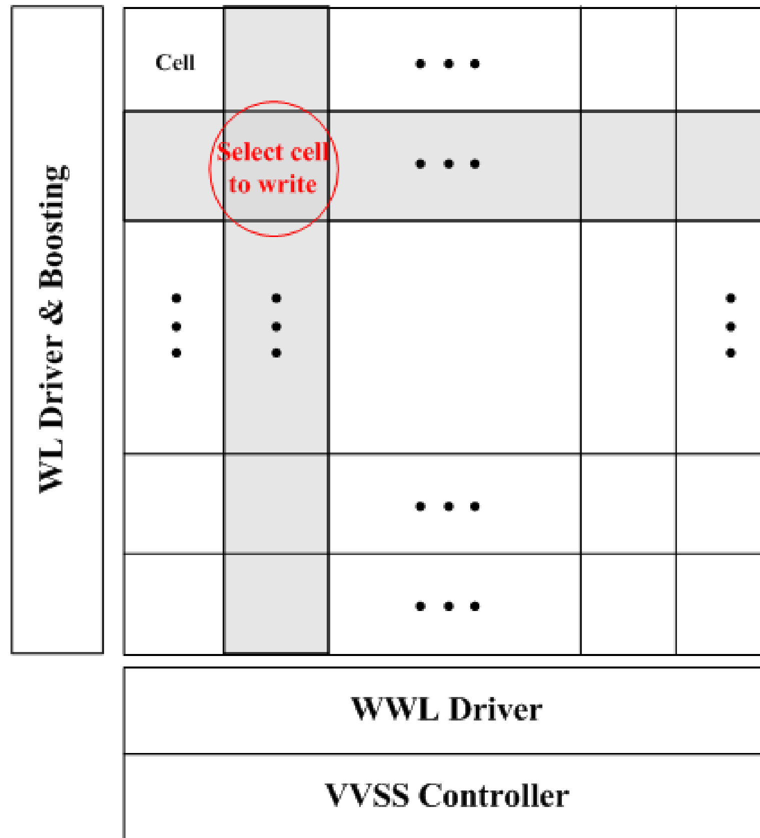


Fig.3.5 Interleaving array structure

3.4 Pipeline Structure and Clock Distribution

First of all, we will discuss how to design the distribution of clock (CLK). Due to the 512K-bits chip do not separate apart and insert into pipeline latch, we have to plan the CLK way carefully. A chip must have the nearest and the farthest cell by the signal transmission. The timing difference of signal larger would affect the redundancy seriously. To advice signal passed the nearest and the farthest cell timing difference is too large, we need to use the H-tree method to solve this problem (Fig. 3.6) [3.13-15]. Afterwards, we should design a clock distribution carefully, these reflection effects that corrects deviations of duty-cycle as well as reduces sensitivity to PVT variations [3.16].

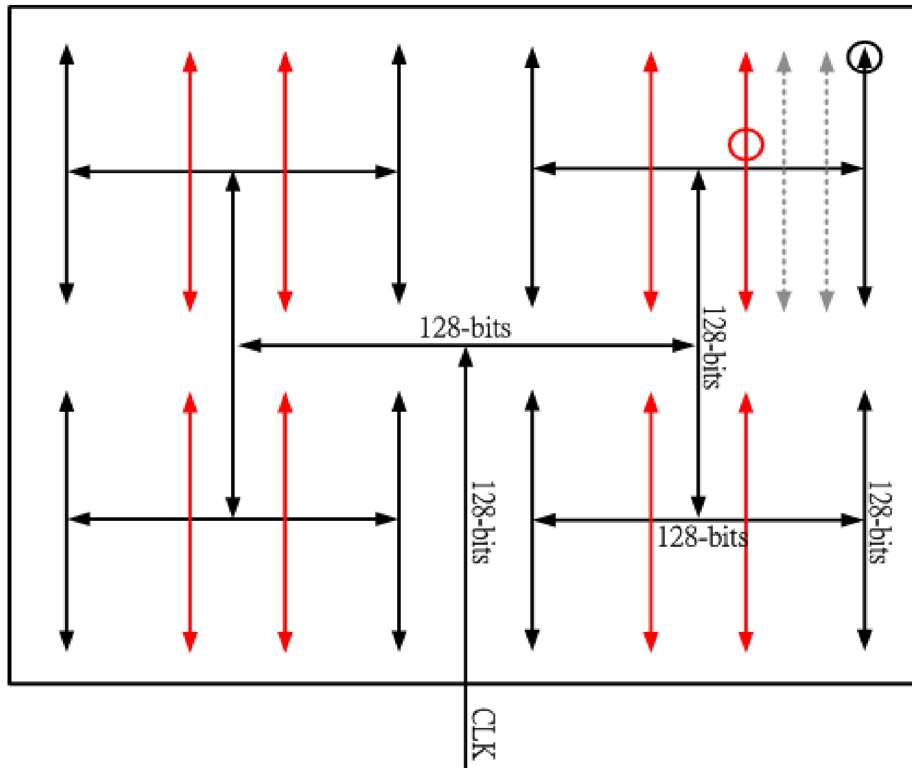


Fig.3.6 The CLK signal use H-tree way. The black line represents the farthest path of CLK and the black cycle stand for the farthest cell. And the red line represents the nearest path of CLK and the red cycle stand for the nearest cell.

Besides, we major put three latches into the chip. The first latch is placed at the beginning of input signals and called that input latch; the middle latch is putted in front of the WL driver; the final latch is droved out on the signal out and called that output latch. Both input and output latch are master-slave latches. We are able to watch the Fig. 3.7 to realize that easily.

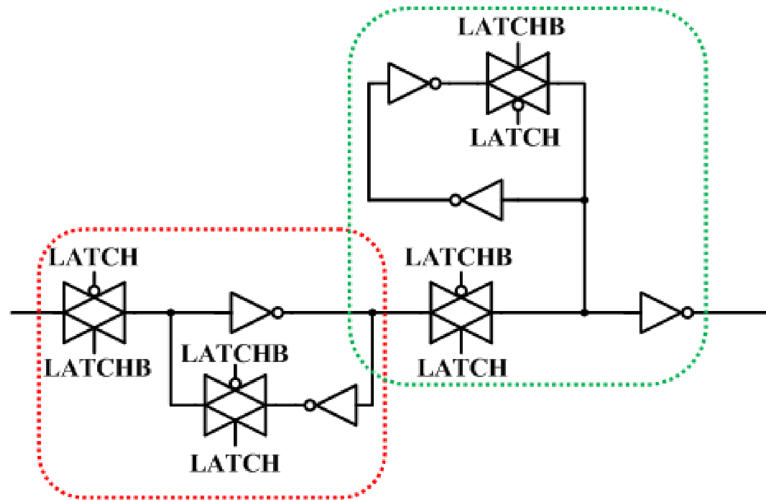


Fig.3.7 The master-slave latches. The master latch part enclosed by the red wireframe; then slave latch part enclosed by the green wireframe. The master latch capture data in the positive clock, we called the latch of L1. Also, the slave latch to do is launching the data, we called the latch of L2.

The Fig. 3.8 show that the signal how to operate by external CLK in detail. A cycle can be divided into L1 and L2 latch, while encounter the clock positive-edge that means capture data and run into the clock negative-edge that signifies launch data. The first cycle at the clock of positive is done decoding address, but if decoding time isn't enough we can delay to the next cycle of negative clock and don't lead to function failed. In other words, the decoding time just is handled to not overtake the next positive clock that is not effect at totally work function. The middle latch has a local clock buffer (LCB) and a L1 latch makes gating to turn on WL we choose [3.17]. Afterwards, the second cycle is enabled WL to run write/read operation and the final result will appear in the next cycle. As input the address of write/read we want, the overcome would produce after two cycles. The concept of pipeline is while the one data finished the first step which work is decoding and started to operate write/read action, and the next one can input the address followed the front address after one cycle.

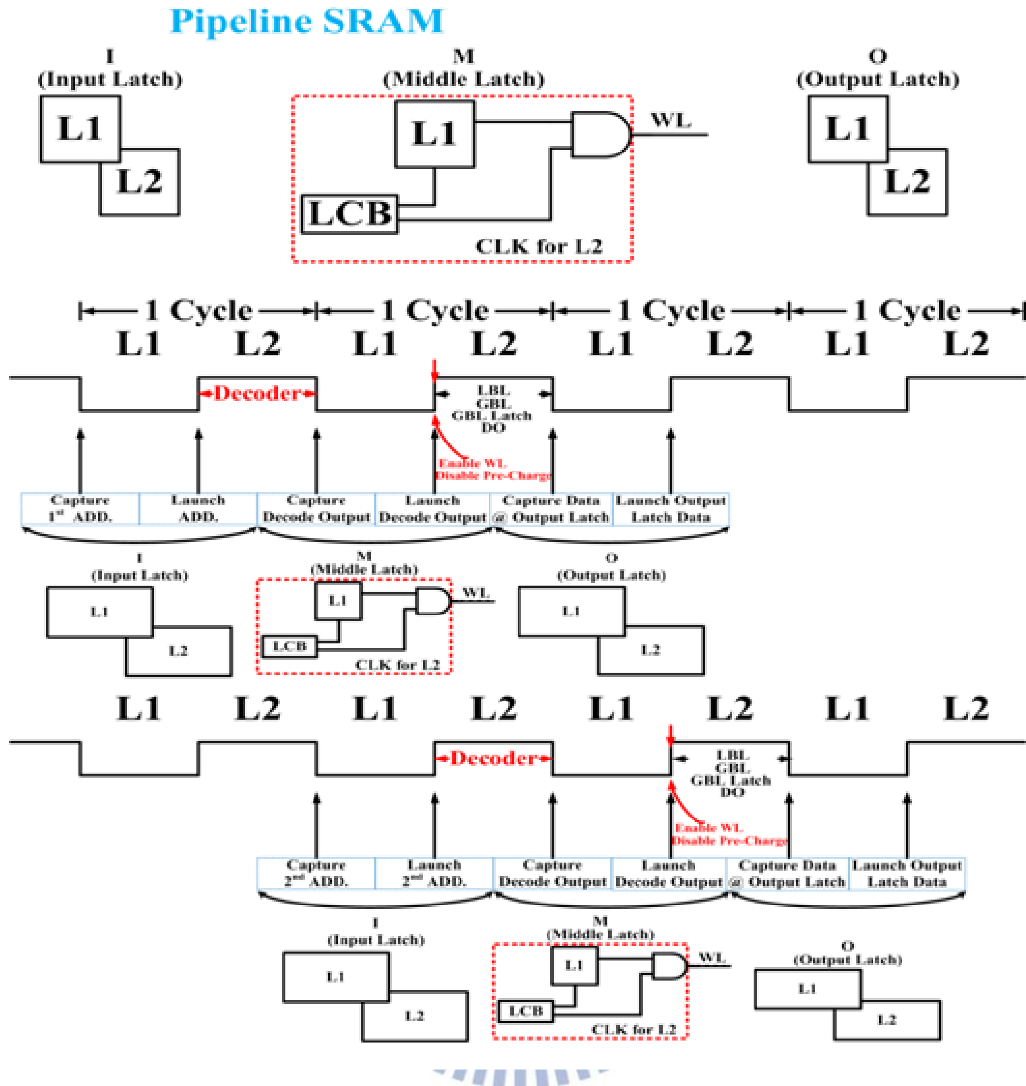


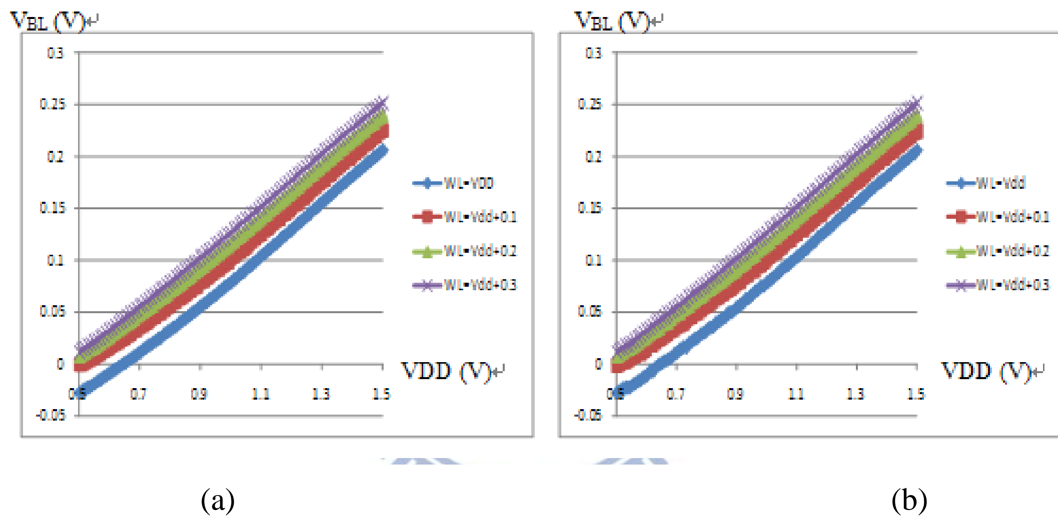
Fig.3.8 The pipeline SRAM operation in whole chip

3.5 Word-Line Booster

3.5.1 Word-Line Booster Circuit

We design the WL booster that improve the write-margin (WM) and write cell speed [3.18]. We use the way to set the BLs at high level, and sweep down one BL form VDD to ground. As the BL voltage let the cell storage values will be suddenly flipped that the voltage of BL is defined as WM. See the Fig. 3.9 that WM obvious increase with boosting WL at low VDD. Basically, wish boosting the voltage of ΔV to WL must have a capacitor [3.19-21]. We choose a PMOS as a capacitor, because its

leakage less and easier to boost WL than the NMOS in simulation. As a result of the capacitor size can't too large, so we design one WL booster share with four WLs. The booster scheme shown in Fig. 3.10, which operation is controlled two signals: BST_EN and BST_TRIGGER. As the booster standby, the BST_EN signal goes low, whether BST_TRIGGER is high or low which will not affect the results. Then T2 become going to low and VVDD is VDD. On the other hand, if WL has to boost voltage ΔV, the signal BST_EN goes high and T2 is changed to high that lead to VVDD voltage become VDD+ΔV.



(a) (b)

Fig. 3.9(a) Write margin of writing 0 with boosting. (b) Write margin of writing 1 with boosting.

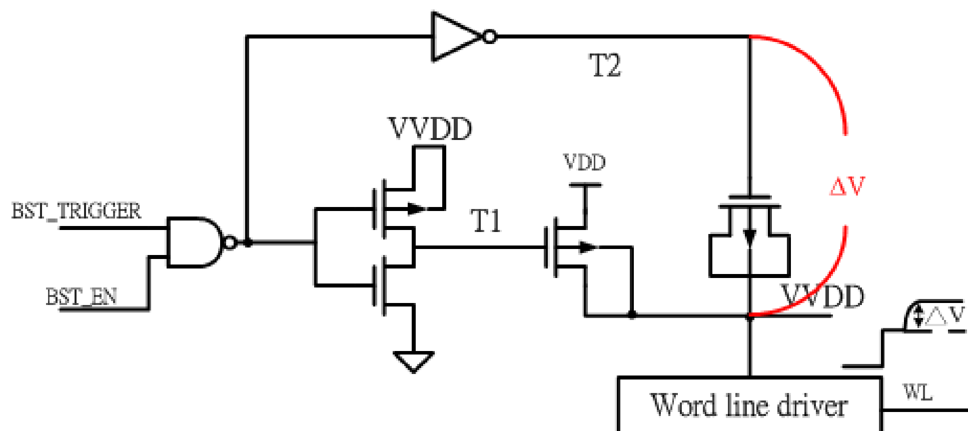


Fig. 3.10 The wordline booster which BST_EN come from voltage detector. If we want to boost WL, then the BST_EN signal goes high, vice versa.

3.5.2 Booster Capacitance design

By the way, how to compute the MOS capacitor size? Just use the below equation and get it.

$$\frac{\Delta V}{V} = \frac{C_{boost}}{C_{VVDD} + C_{boost}} \quad (1)$$

The C_{VVDD} is represented the capacitor which see into the VVDD. We can see Fig. 3.11-12 to realize the boost WL principle. Fig. 3.11 shows total the capacitance of C_{VVDD} estimation in detail. Cause to booster has to use a capacitance to increase the WL voltage; we also consider the capacitance on the WL. In fact, composed the WL capacitance one is NMOS of pass gate for 8T SRAM cells by pass through, the other is C in Fig. 3.11, which is related with the WL length, so we don't make the WL length too long due to the C_{VVDD} too large and not easily to boost. That is the reason why we design the sub-bank cell array is smaller with 32 columns.

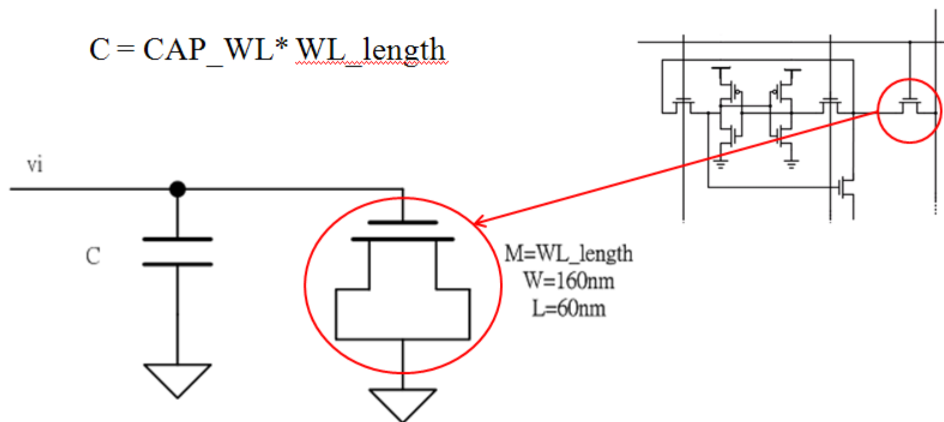


Fig. 3.11 The WL capacitance (C_{VVDD}) estimation

Fig. 3.12 shows how to charge Δv in the C_{VVDD} and is sized the C_{boost} from the equation (1). At first, the C_{boost} one end connects to GND in standby and begins charging the voltage Δv to capacitances for C_{boost} and C_{VVDD}. As the C_{boost} GND end change to VDD voltage, the VVDD node in the Fig. 3.10 is charge to Δv successfully.

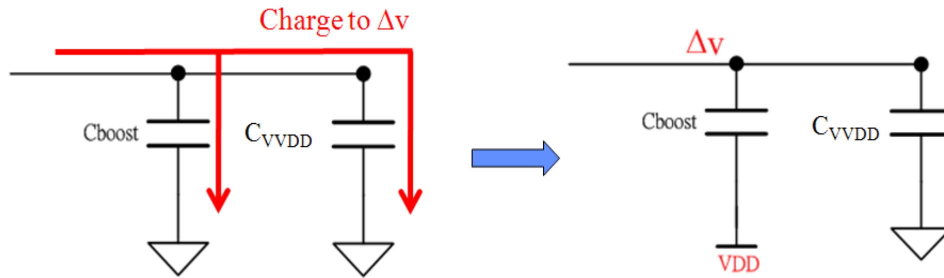


Fig. 3.12 Boosting Δv steps.

According to equation (1), we can derive the C_{boost} size as the Δv is fixed.

$$C_{boost} = \frac{\Delta V}{V-\Delta V} C_{WL} \qquad (2)$$

In the booster circuit, we wish the Δv is 0.2V in supply low voltage. The suitable boosting capacitance size it can boost each side (right/left) 16 cells and once boosts for four row WLs. The capacitance can be used NMOS and PMOS, and because the NMOS leakage larger than PMOS and boosting ability worse than PMOS by simulation. Finally we decide use the PMOS in the boosting circuit.

3.5.3 Simulation Result

The Fig. 3.13 is shown the simulation of WL voltage smaller than 1V. The cycle time is 20ns and the VDD is 0.9V, cause of smaller than 1V through the voltage detector to determine the WL have a ΔV produce finally and let the WL voltage to rise. The simlation opretaion step is writing-1, writing-0, read-1, and read-0. The data result displays after two cycle as the benchmark is data inputs the cycle.

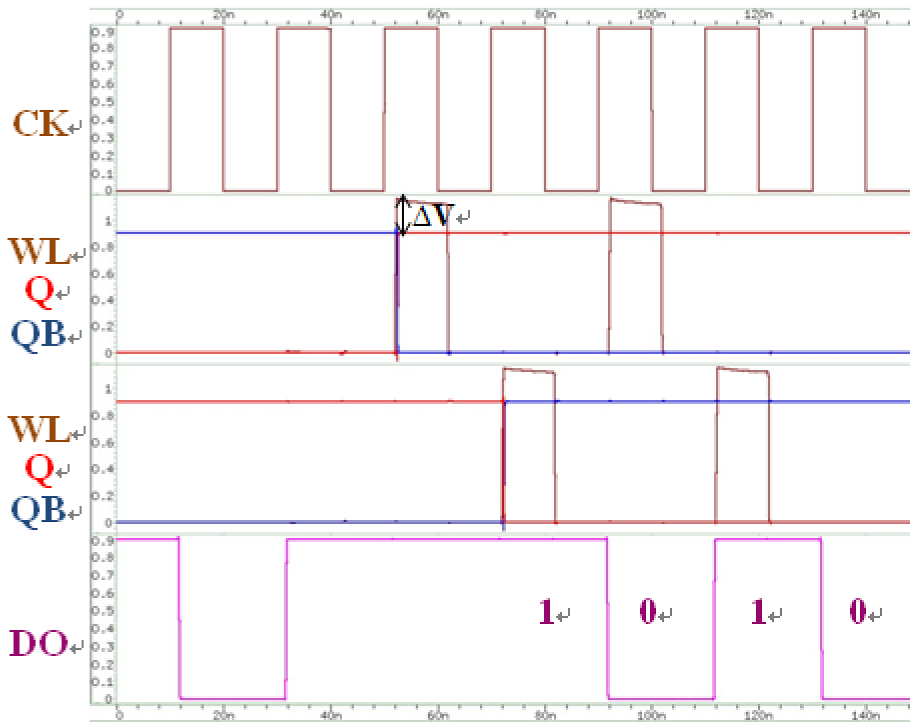


Fig. 3.13 The simulation of WL, farthest cell array write/read, and data-out (DO)

3.6 Voltage Detector

3.6.1 Voltage Detector Design

Owing to above description that voltage goes high not affect evidently as the WL boosting. Inversely, boosting WL at low VDD is useful relatively. Base on this reason, we design the new scheme: voltage detector (Fig. 3.10) and the boundary voltage is 1V. When voltage >1V the result SET_EN will go low and communicate the signal to the WL booster, vice versa.

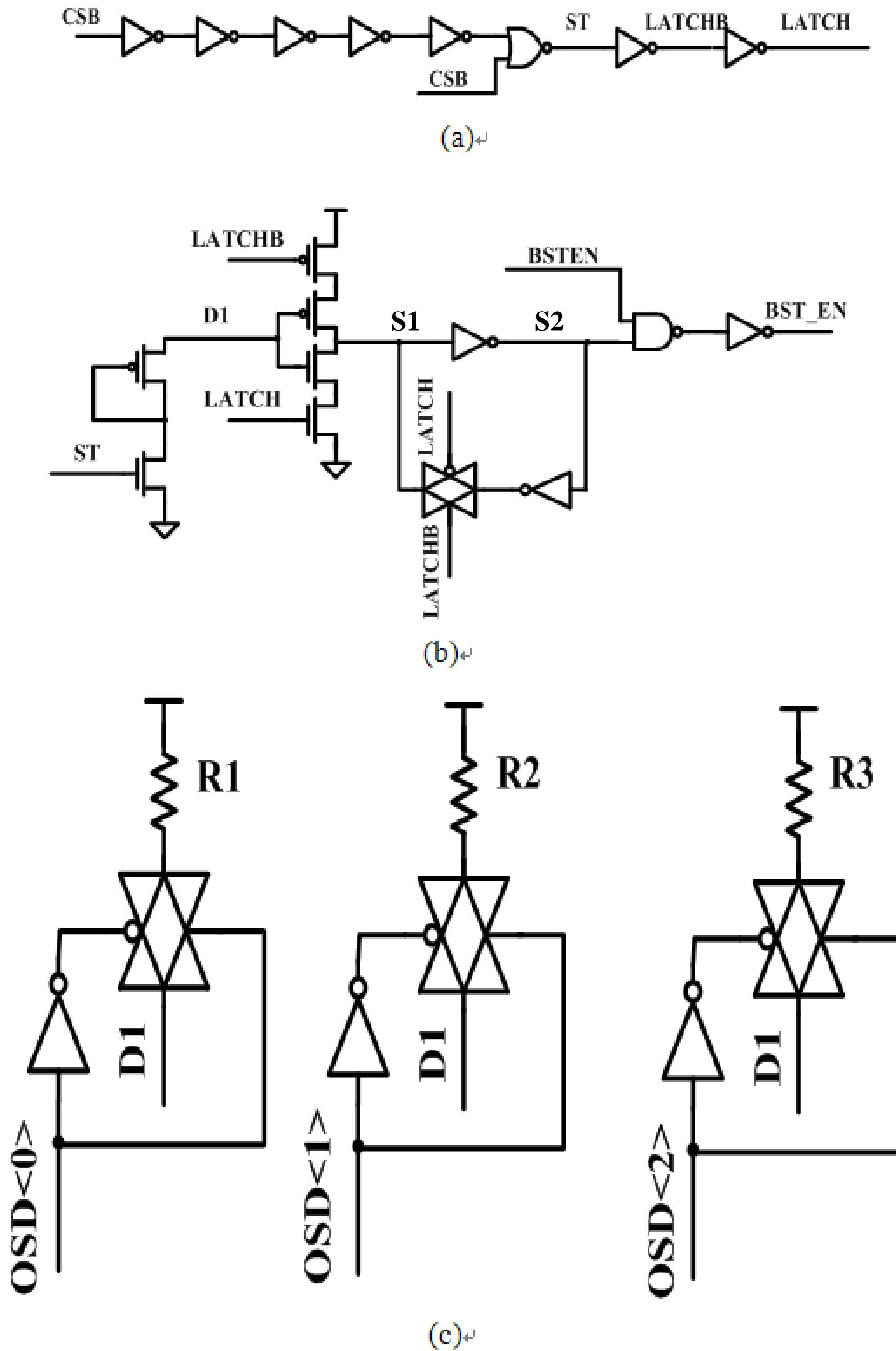


Fig. 3.14 (a) The delay chain by the CSB controlled. (b) The section by an external voltage to decide whether or not as boosting WL. (c) Three different corner (SS, TT, FF) be controlled by OSD<0>~<2>. voltage at VDD is 0.9

We separate three parts to explain the voltage detector. The first part (Fig. 3.14 (a)), which is a simple delay chain to produce a pulse by the chip select bar (CSB). As the chip start to work, we set the option CSB always logic “0” until using different frequency the CSB signal will reset again. Besides ST, LATCHB, and LATCH also bring a pulse to next part (Fig. 3.14 (b)). The D1 net would connect the external options by three pins to control major three different corners (SS, TT, FF) at 25℃ (Fig. 3.14 (c)). We choose three different resistors (R1, R2, and R3) to handle three trip point of inverter at distinct corners. OSD<0~3> turn on individually represent the SS, TT, and FF corner. To improve measurement speed whether boosting or not, the part of Fig. 3.14(b) and (c) use the LVT totally. In addition, the BSTEN external signal is controlled to the voltage detector circuit whether turn on by a strong hand.

3.6.2 Simulation Result

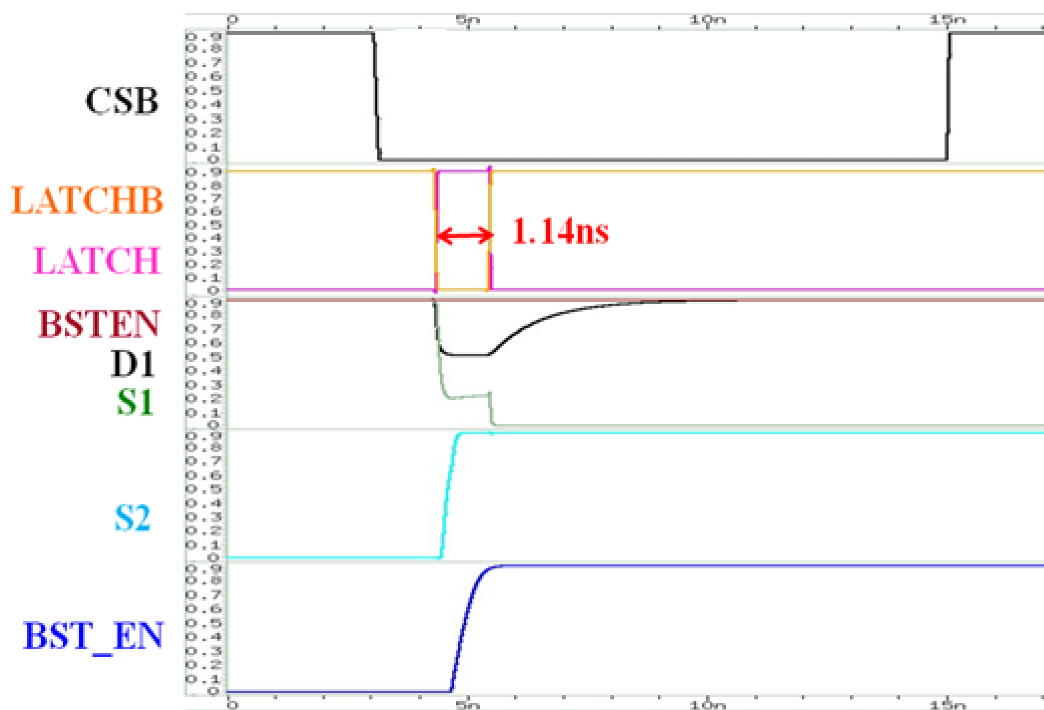


Fig. 3.15 shows the voltage detector simulation waveform. As the above

Fig. 3.15 The simulation waveform of voltage detector.

description, the SCB comes to a constant signal for “0” that reflects the LATCH and LATCHB signal changed. S1 voltage through pass the inverter and the S2 response to the result until followed the signal to BST_EN.

3.7 Data-Aware Write-Assist Tracking Circuit

3.7.1 DAWA Tracking Circuit Design

The chip operation is controlled external CLK. To advice the cycle at the lower frequency that data-aware write-assist (DAWA) is turned on for long time lead to the cell data flipped and leakage current. We designed the circuit to hold in check the time of DAWA turn on in a constant and show the circuit at Fig. 3.16. As a result of the chip local bank array is 16x32 (bits), we use a column of 16-bits as the dummy loading, and the 1-bit dummy cell loading shown at Fig. 3.17. The 1-bit dummy cell loading we let the status at the stand-by and Q is saved 1, QB is saved 0. And 1-bit dummy cell some nodes connect to tied-0 and -1 schema (Fig. 3.17).

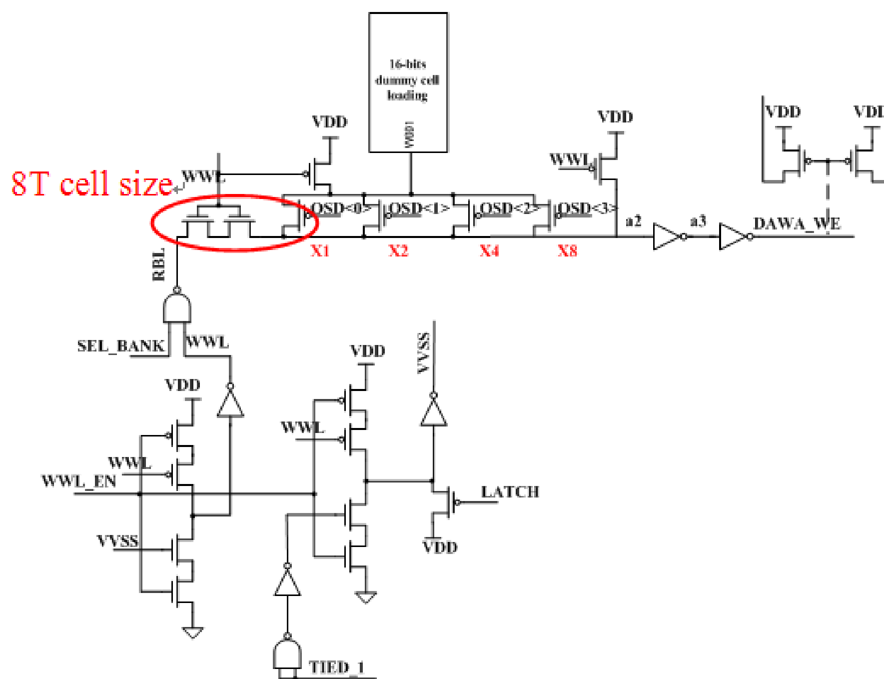


Fig. 3.16 DAWA tracking circuit

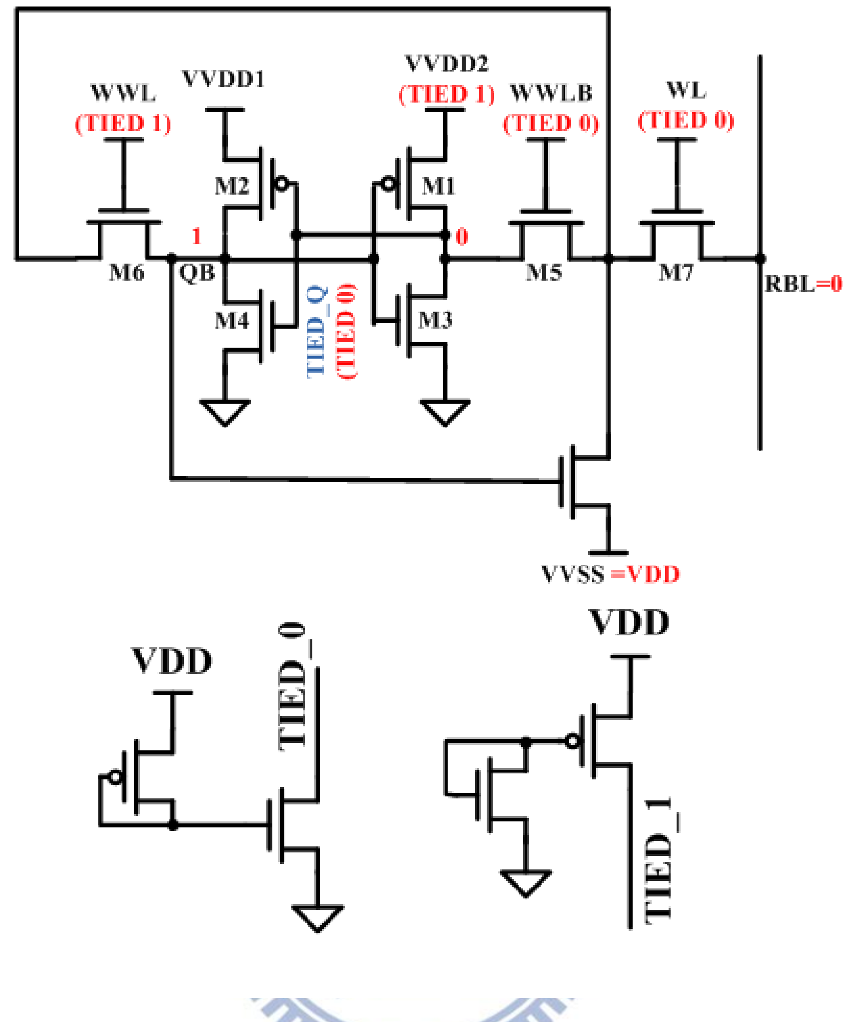


Fig. 3.17 1-bit dummy cell and tied 0/1 schematic.

The DAWA tracking circuit have two NMOS is placed in left and they gates end connect WWL (Write Word Line) that use the size of 8T cell pass gate. Besides, the OSD<0>, OSD<1>, OSD<2>, OSD<3> sizes is one, two, four, and eight times of 8T put-up size individually that option can control the DAWA turn-off time. Then the signals will transmission to the LEV and handle 8T cell DAWA time. The longest time turning off the DAWA time option is all OSD are logic “1”. Because of the capacitance is the biggest looking from the charging node that the charging time must spend enough time to charge to VDD. By the way, the LEV dummy condition uses the write “1” in the Fig. 3.16 lower left corner, because the path is the longest and slowest.

3.7.2 Simulation Result

We display the simulation waveform for DAWA tracking Circuit at TT corner, 25 ℃, VDD is 1.1V and only turns on OSD<0> for Fig.3.18. Only switch on OSD<0> is the worst case for DAWA that means the shortest time to turn off the DAWA signal. If we did not design carefully, we have the problem for this: while the data writes into cell for the period, it is happened the DAWA mechanism turns off early. In fact, we want to DAWA always turns on until the writing operation writes completely.

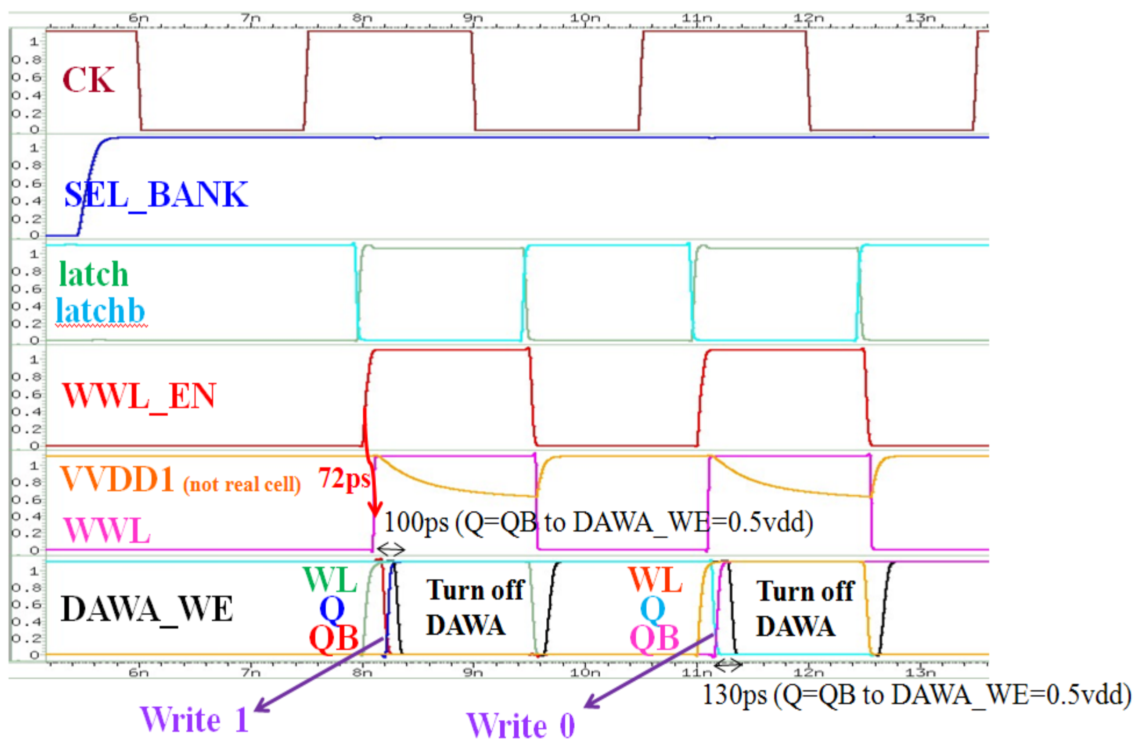


Fig. 3.18 DAWA tracking circuit waveform @ TT, 25℃, VDD=1.1V, OSD<0>: on

We can see distinctly the simulation for the nethermost waveform no matter write "1" or "0", the data writes perfectly before turned off the DAWA mechanism. The signals **CK**, **latch**, and **latchb** are clock signals. **CK** is the external clock and **latch/lathcb** is the delay **CK/CKB** signal in internal circuit.

3.8 Ripple Bit-Line (BL) and Multiplexer

3.8.1 Ripple BL and Multiplexer Designs

The schematic of ripple buffer proposed read path is shown as Fig. 3.19. The path pass four local-evaluation (LEV) read parts at most to spread the data to Multiplexer. The LEV read part has two types in the structure, which are shown in Fig. 3.19 separately. The main composition uses a PMOS that pre-charge the local bit-line (LBL), then it cascade two NMOS for write operation at LEV_TOP and the result will pass through the ripple buffer to the next LEV read part. The LEV_DN different from LEV_TOP is that adds the inverter the input connect previous LBL_OUT, then the output link to the ripple buffer.

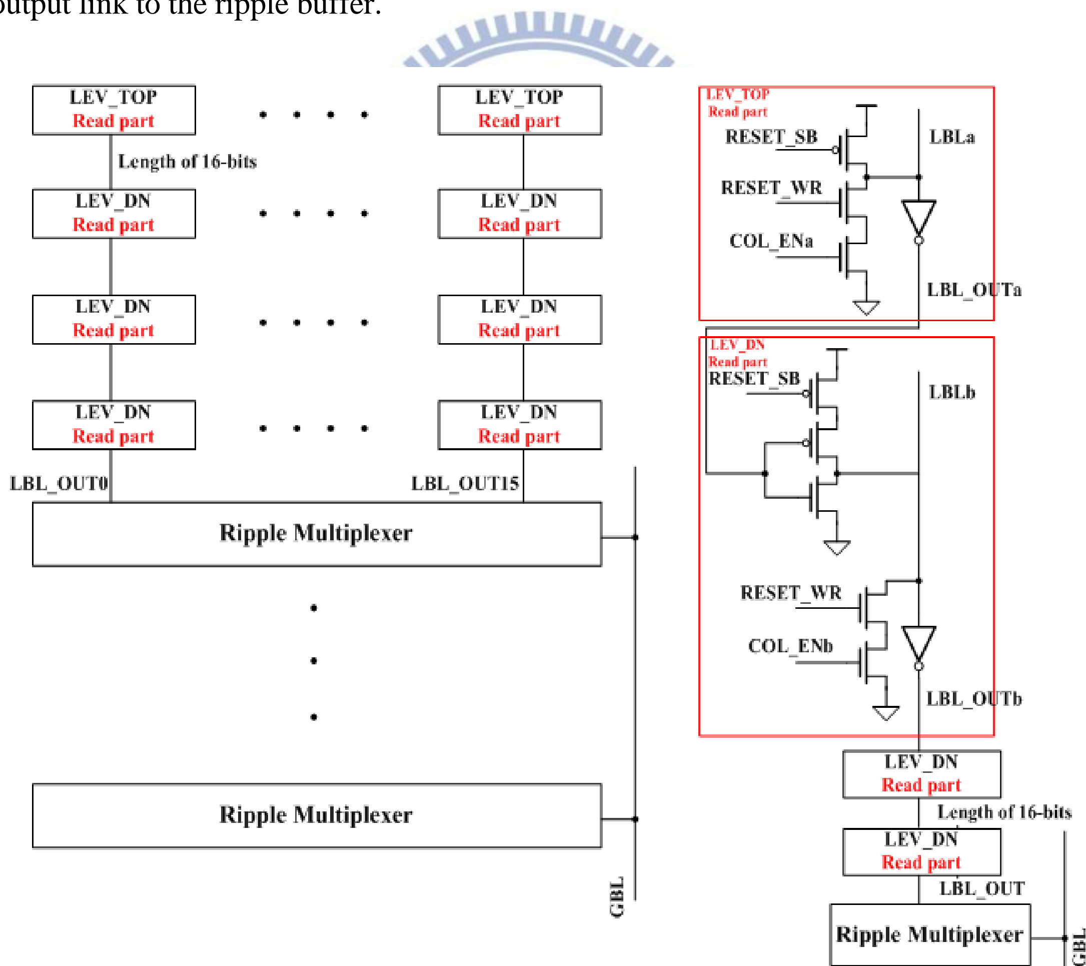


Fig. 3.19 The ripple path structure and the LEV_TOP and LEV_DN read part circuit

In stand-by mode, the RESET_SB, RESET_WR, and COL_EN signals all are logic "0" and the segment LBL voltage pre-charge to VDD. When we want to read the cell data in selected local bank, the signals of RESET_SB, RESET_WR, and RESET_WR change to logic "1". This LBL voltage is floating at the read time and rippling the data stage to stage. If the cell store data is logic "1", the LBL will become VDD. Conversely, the cell storage is logic "0"; the LBL voltage will discharge to GND for long time if the CLK frequency is very slow.

Because of LBL should get across the ripple buffer, the cell data and final LBL_OUT logic are opposite. In other words, while the date is "1" and the LBL_OUT is logic "0", vice versa. Next the signal transforms the multiplexer that circuit show in Fig. 3.20. Each LBL_OUT would cascade NMOS which gate-end link to each corresponding COL_EN. The MP3 PMOS is controlled over signal LATCH, which signal phase is the same with external CLK. The two cascade PMOS, MP1 and MP2 are the GBL keeper design. A combination of the NOR gate and NM1 is a sense amplifier structure.

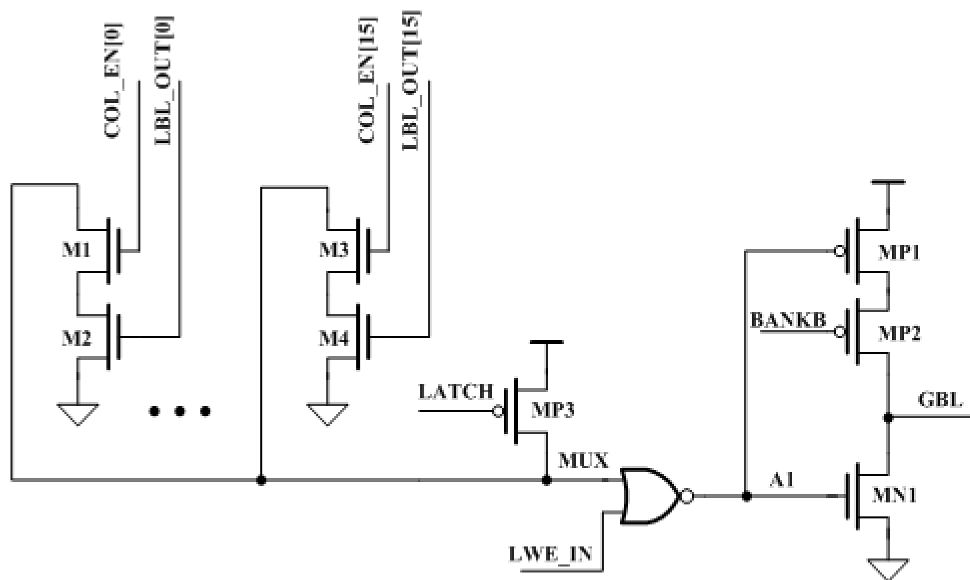


Fig. 3.20 The ripple multiplexer circuit

In stand-by mode, the signal LATCH is logic “0” then the node MUX voltage per-charge to VDD. The MP2 is handled by signal BANKB, which logic is opposite with select bank signal. Namely, most banks don’t choose so the bank signals are logic “0”, then the BNAKB is logic “1” and the MP2 turn off. The keeper circuit isn’t working that the GBL voltage is floating. Otherwise, only one select bank due to the corresponding BNAKB signal is logic 0 gets the MP2 turn on. Because of MP2 turns on that means the keeper by GBL switch on in selected bank.

In read mode, the local write enable (LWE) signal is logic “0”, so go through a latch the signal LWE_IN also become to logic “0”. The signal LATCH is logic “1” turns off MP3 don’t charge at MUX node. The MUX floating voltage desponds on the one COL_EN of 16 column lines. The one column the address chooses turns on and passes the data to MUX. If the data original is logic “1”, cause of through the ripple buffer the signal comes to logic “0”. Then the working switch off M1 if we choose the COL_EN[0] that makes the MUX voltage is floating “1”. Next, the A1 node is going to become logic “0” that turns on the keeper circuit by MP1 and MP2. The keeper working make the GBL keep VDD in reading time. Similarly, while the original data is logic “0” passes the ripple buffer becomes logic “1”. The discharging path produce by M1 and M2 that we assume the selection of one column of 16 lines is COL_EN[0] again. The MUX node voltage gets to GND that makes the A1node voltage change to VDD and turn on the MN1. As a result of that GBL also discharge to GND and the keeper turns off.

3.8.2 Simulation Result

The simulation (Fig. 3.21) used the period for 3ns at TT corner, 25℃, and VDD=1.1V. The movement is doing write option by writing-1 and write-0 at first. Then read the cell data individually which has written for a moment ago. In read mode, the LWE_IN signal comes to logic “0”, so the result follows the MUX node voltage. While the MUX is logic “1”, the GBL result is logic “1”. On the other hand, for the MUX is logic “0”, then the voltage is going low for GBL node.

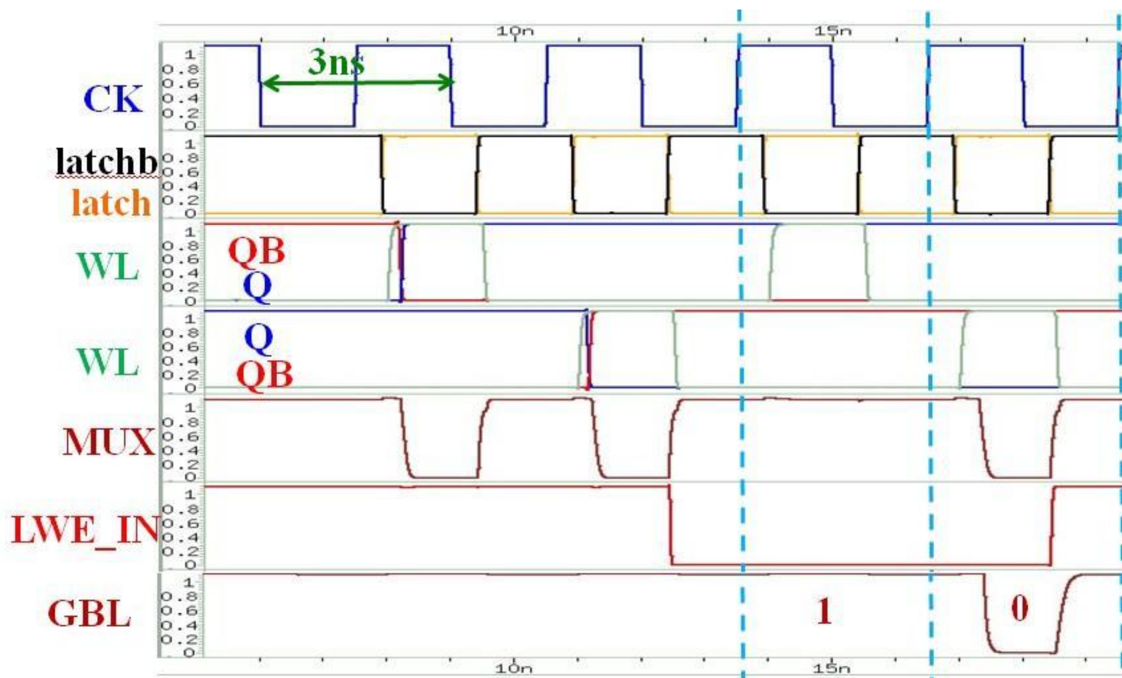


Fig. 3.21 The waveform for the ripple multiplexer

3.9 The Keeper Design of LBL

3.9.1 The Keeper Design

There are two places will occur leakage and discuss them. First place, when in the read time the MUX in Fig. 3.20 voltage is floating; we have a problem that is the leakage current if the MUX node is floating 1. There are 16 leakage paths of two cascade NMOS, for example: M1 and M2 (Fig. 3.20). The structure of cascade also can reduce the NMOS leakage better than use only one NMOS. But we wish the chip working speed faster, so all transistors are using LVT process in read path that increase the leakage current unfortunately.

Second place, while one cell array select the LBL voltage in read mode is floating. The LBL length is 16-bits, in other words, 1 bit cell choose but another 15 bit cells don't do that. There are 15 leakage paths in this local bank. Cause to signal VVSS is logic 0 whether in read mode for selecting cell or in stand-by mode for none-selecting cells, we can see the Fig. 3.1 (b) to realize the NMOS is controlled over (R)WL producing the leakage path through pass this NMOS to VVSS. We know where major positions of leakage are, we just place the keeper in these locations that the problem will clear up as shown in Fig. 3.22 (a), (b), and (c).

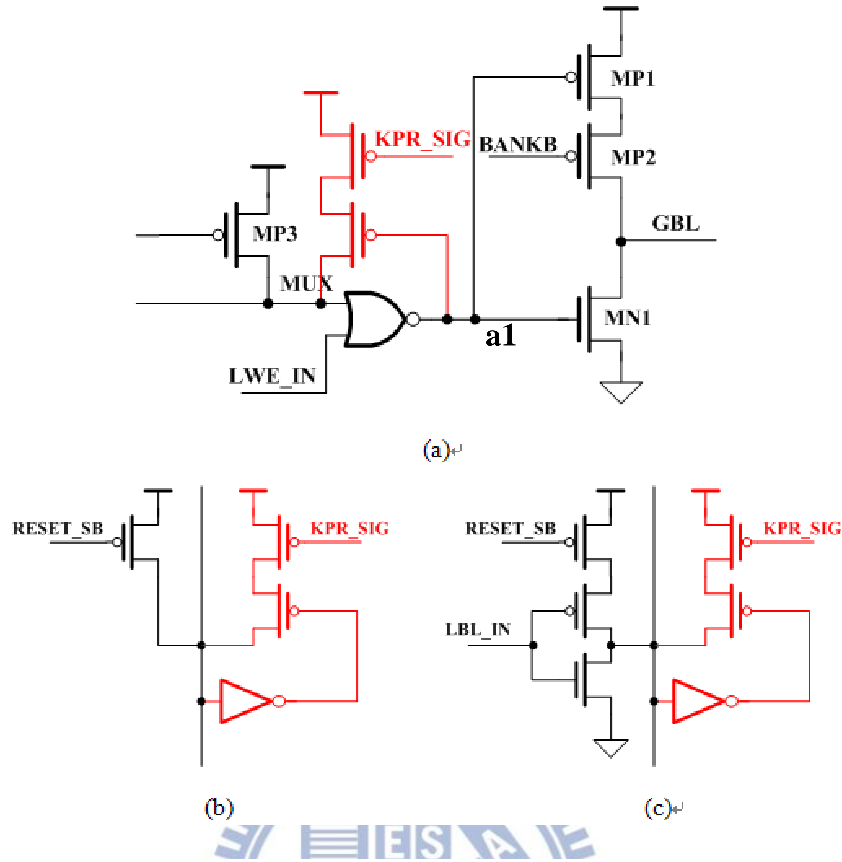


Fig. 3.22 (a) The keeper of ripple multiplexer, (b) the keeper of LEV_TOP, and (c) the keeper of LEV_DN in red part

In the beginning for Fig. 3.22(a) circuit KPR_SIG node connect the replica keeper [3.22]. The replica keeper circuit is shown as Fig. 3.23 that principle is used a current mirror by resupply current for leakage of the designed circuit. And we also use the replica keeper design to connect the Fig. 3.22(a) KPR_SIG point. Order to separate differences, we change node name from KPR_SIG to GBL_KP which is used replica keeper. We simulate the GBL_KP voltage the any corner as shown in Table 3.2.

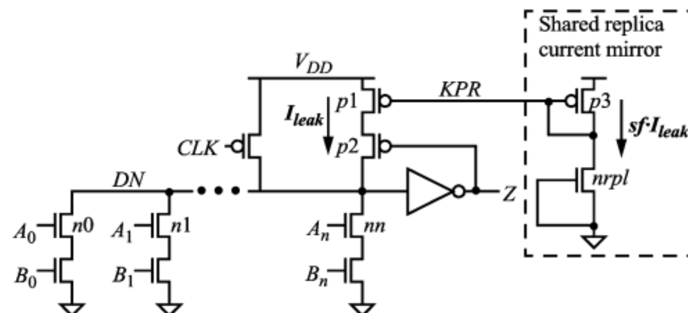


Fig. 3.23 LCR keeper dynamic gate topology [3.22]

VDD 125℃	GBL_KP @ psns	GBL_KP @ ptnt	GBL_KP @ pfnf	GBL_KP @ pfns	GBL_KP @ psnf
1.2V	0.954V	0.97V	0.974V	1.1V	0.84V
1.1V	0.86V	0.88V	0.89V	0.992V	0.75V
1.0V	0.77V	0.78V	0.789V	0.896V	0.652V
0.9V	0.669V	0.686V	0.697V	0.8V	0.56V
0.8V	0.574V	0.592V	0.604V	0.705V	0.467V
0.7V	0.479V	0.498V	0.512V	0.609V	0.376V
0.6V	0.384V	0.404V	0.42V	0.52V	0.248V

Table 3.2 GBL_KP voltage on different corner

The table shows VDD is the supply voltage and use 125℃ that temperature is the worst condition for none-leakage. We can see clear the voltage hold high best is the pfns corner and psnf is the lowest. To look back Fig. 3.22 (a), GBL_KP (KPR_SIG) voltage is reflected by switching the PMOS. That has a problem for using replica keeper. Due to the GBL_KP voltage is not the logic “1”, in the other words; total corners are floating “1”. The PMOS cannot turn on completely, so supplied current speed is not faster that probably slower than leakage speed. The phenomenon will be worse obviously in low supply voltage. Fig. 3.24 shows the waveform in pfnf, 125℃, and VDD=1.1V the GBL voltage result. The GBL_KP is a fixed voltage and as

Fig. 3.24 The waveform for GBL by replica keeper @ ff, 125℃, VDD=1.1V.

reading data “1” for low frequency, the leakage is larger than keeper added current due to the MUX voltage cannot hold logic “1” for cycle time is 100ns.

Hence, we use other method to solve this problem. The way is simple and do not use complex design [3.23]. The method is used a multiplexer with delay chain that shown as Fig. 3.25 [3.23].

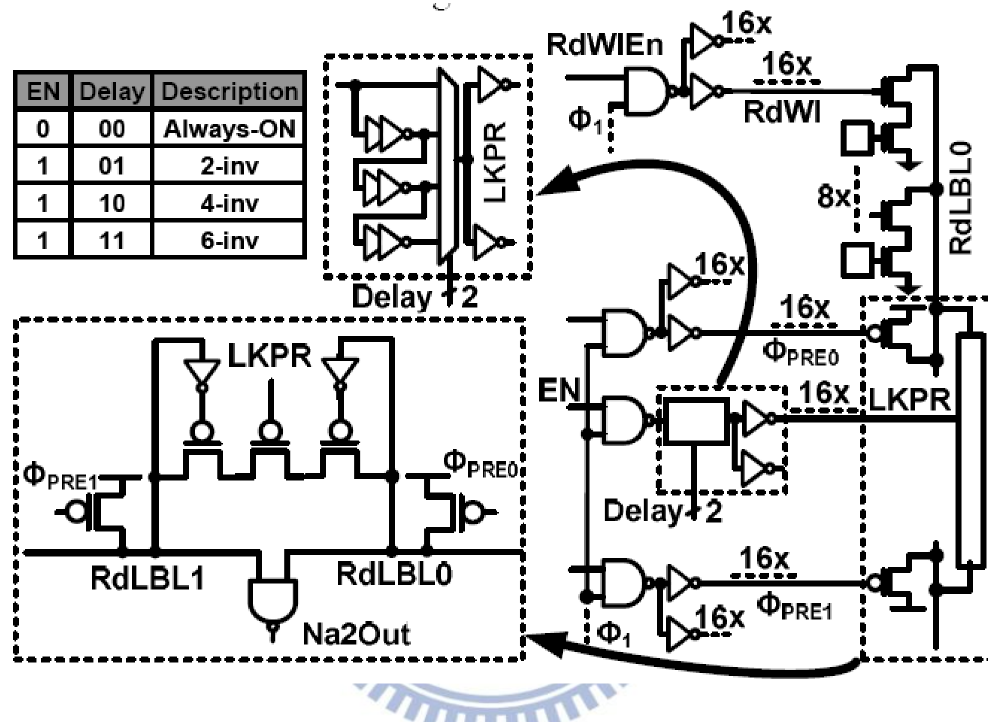


Fig. 3.25 Read local bit-line with contention free shared (CFS) keeper and global keeper delay element. [3.23]

The two major leakage place we use the design shown as Fig. 3.26. The structure is followed the Fig. 3.25 and the GBL_KP for Fig. 3.24 (or KPR_SIG for Fig. 3.22(a)) would keep MUX node stay in floating “1” and read “1” does not fail. We show then waveform in Fig. 3.26 at PFNF, 125℃, and VDD=1.1V with the Fig. 3.25 conception. We can compare the before and after designs result in DO (Data Out).

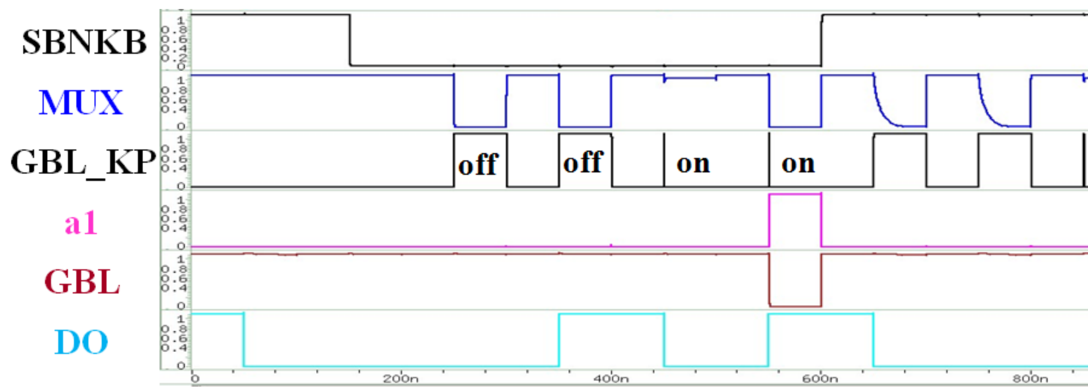


Fig. 3.26 The waveform for GBL by CFS keeper @ ff, 125℃, VDD=1.1V.

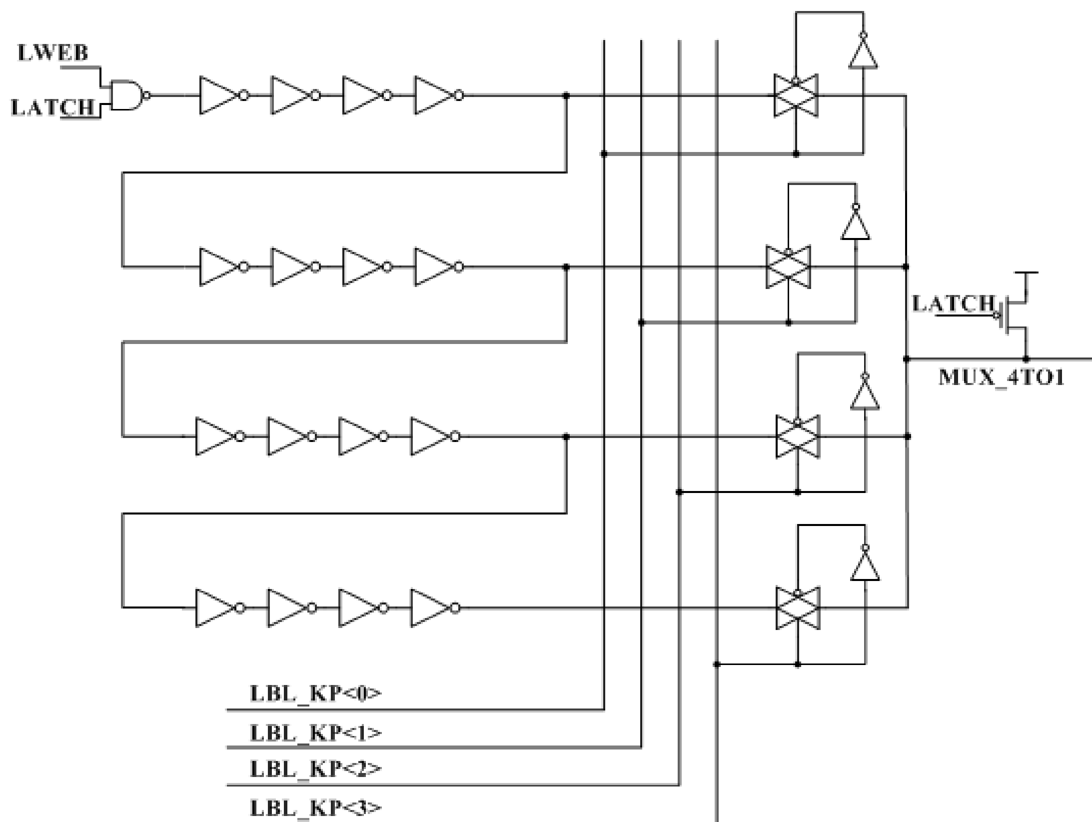


Fig. 3.27 The local bit-line keeper design in detail

The signal KPR_SIG in Fig. 3.22 is controlled over the circuit [3.23] that handle the keeper what time turn on or turn off as the Fig.3.27. The circuit work function only at read mode. Because of LWEB and LATCH two signals make a NAND gating in the circuit beginning. The LBL_KP<0>~<4> signals handle the 4-to-1 multiplexer that are controlled of the external option through pass a decoder which only choose one to logic “1”, others are logic “0”. There are using four sets to compose of a

CMOS switch and an inverter to accomplish the multiplexer like as Fig. 3.12(c) in voltage detector that handle corner. The signal LATCH phase is the same with external CLK. As LATCH in low level, the circuit is not running, so the MUX_4TO1 node voltage charge to VDD and transmit the result the keeper that turn off. As the LATCH in high level, the MUX_4TO1 node is floating and the voltage will be decided by the LWEB signal. In write mode, the MUX_4TO1 node is logic “1”, then the result make the keeper turn off; in read mode, the MUX_4TO1 node voltage is logic “0” that handle the keeper working. Fig. 3.28-29 shows that the selection of LBL_KP which is controlled over a decoder and the delay path will be different by turning on keeper period.

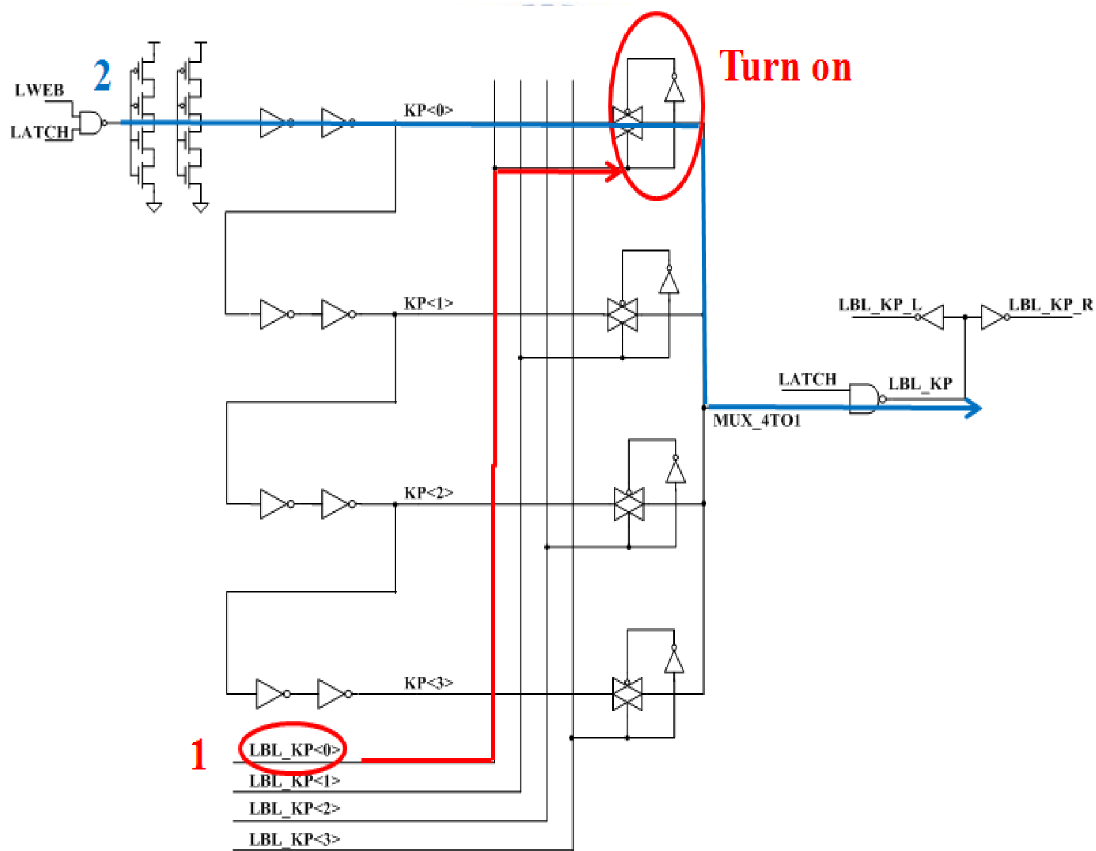


Fig. 3.28 The local bit-line keeper design is chose from the LBL_KP<0>.

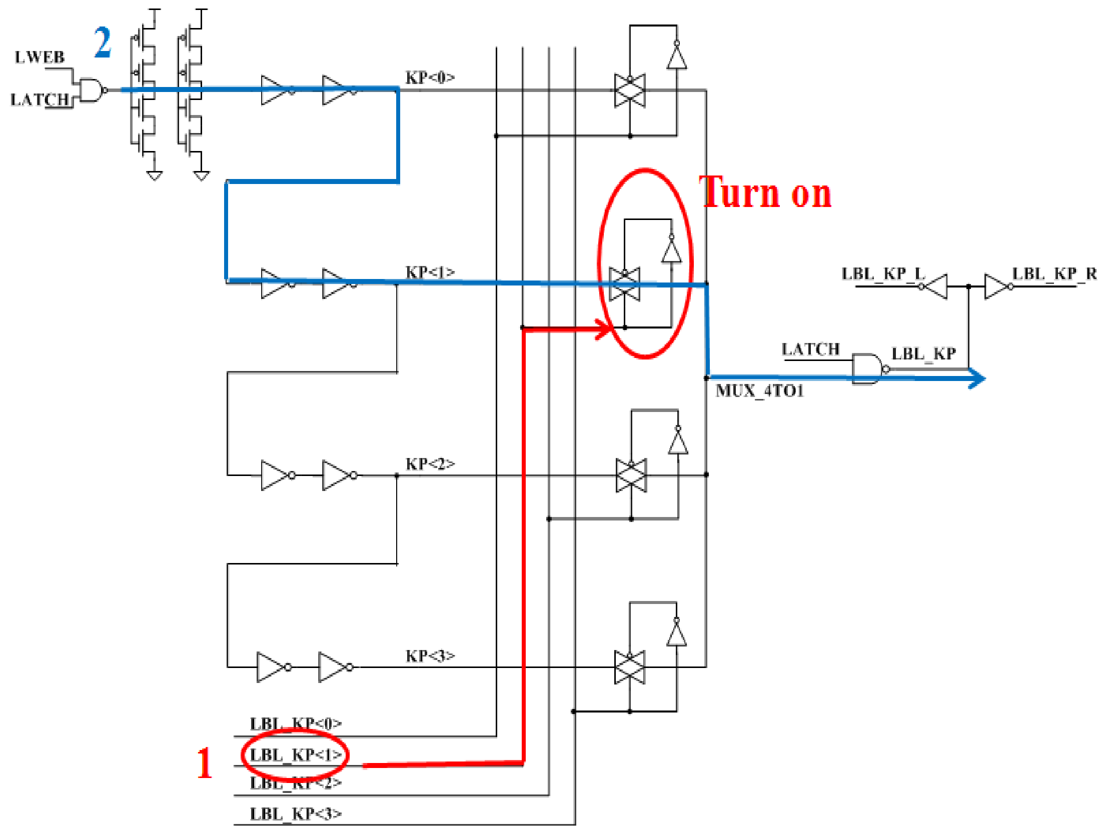


Fig. 3.29 The local bit-line keeper design is chose from the LBL_KP<1>.

3.9.2 Simulation Result

Fig. 3.29 displays the pre-sim waveform for keeper design at SS corner, 125℃, VDD=1.1*0.9=0.99V. In read operation, the LBL_KP will turn on (logic "0") and then we design the switching time is according to the RBL voltage around 5% (50mV) by VDD and transmits the signal to LBL_KP must have 50ps for the shortest delay path. Besides, as two buffer delay we simulate the value around the 150ps for pre-sim.

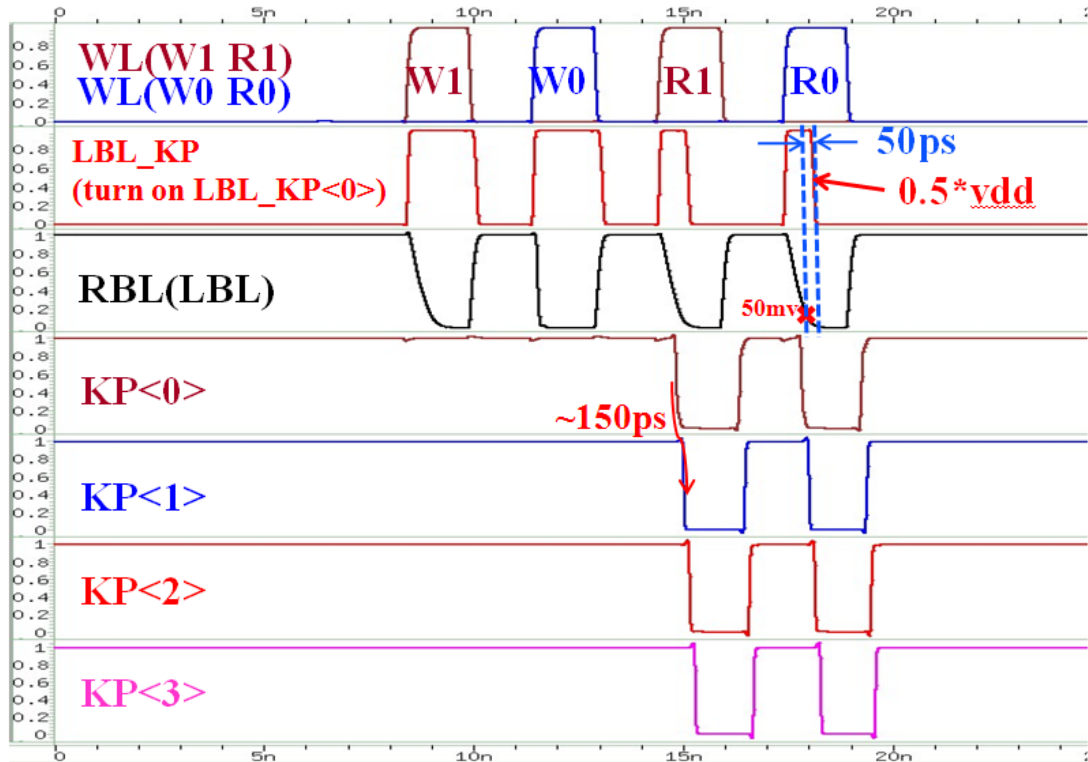


Fig. 3.30 The waveform for local bit-line keeper design @SS, 125℃, VDD=0.99V

3.10 Data-In Data-Out (DIDO) Design

3.10.1 GBL Latch

The DIDO compose main two structures in this chip: GBL latch and Write through schematics. At first, we start discuss the GBL latch design and the circuit shown as in Fig. 3.31. GBL_IN is represented the data by transmitted in read mode and depends on the latch responded whether sense GBL data or latch GBL data. If into sense-GBL -data mode, the PMOS above the GBL_IN node will go high and be not working that make the GBL_IN signal pass through the latch directly to DIDO flip-flop. Else if above GBL_IN PMOS signal is logic "0" then charges the GBL_IN voltage to VDD that inters the latch-GBL-data mode. In addition, Fig. 3.32 is the simulation of GBL latch waveform, which is doing read-0 operation. In GBL_IN row for Fig. 3.32, the waveform begins falling down to GND that reflects the result of DO_INB and also feedbacks A node to latch data.

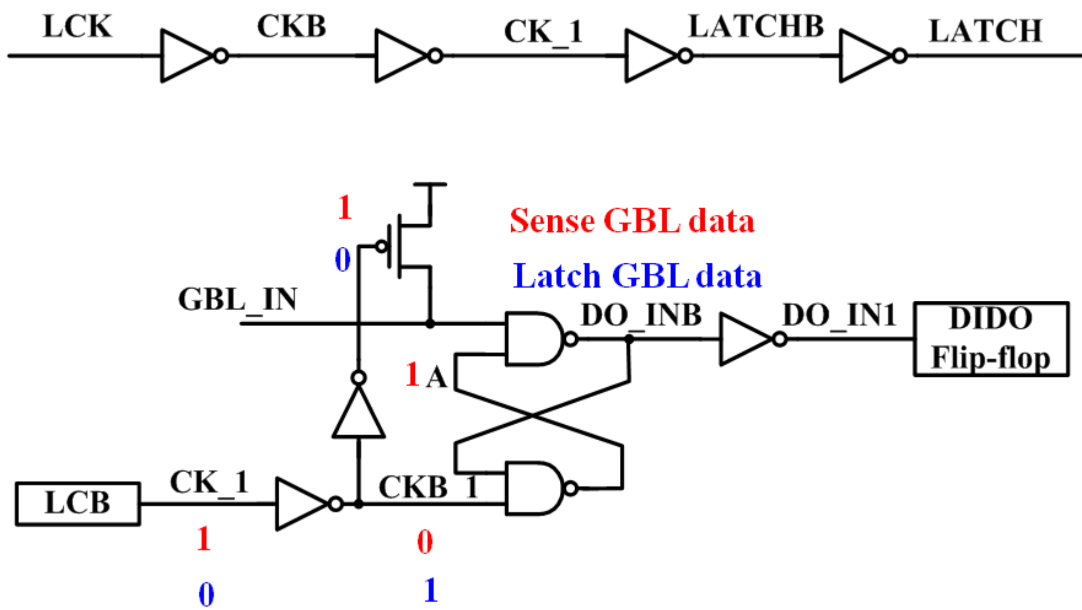


Fig. 3.31 The GBL latch circuit

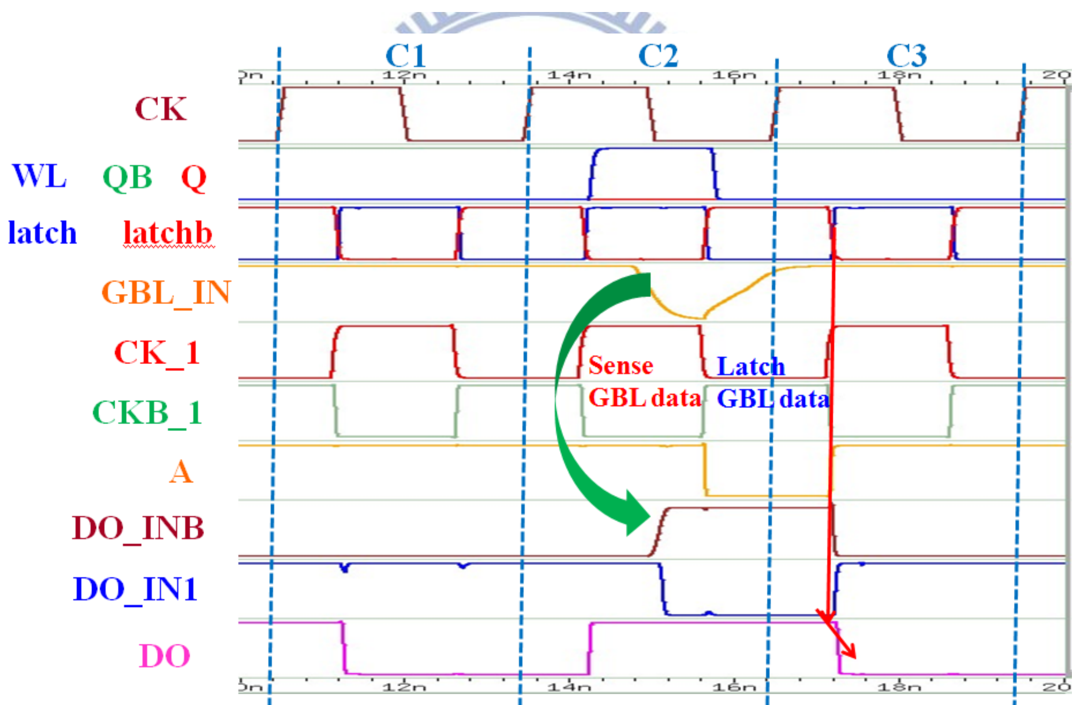


Fig. 3.32 The waveform of GBL latch circuit

3.10.2 Write through Circuit

DIDO the other main circuit is write-through design that shown as Fig. 3.33. The circuit includes the GBL latch back-end. By the way, GBL latch works in read mode and write through circuit operates in write mode. The circuit running starts to DI and WEB two signals. After operating in C2 (Fig. 3.34) with L1 latch, we produce the DIB and WE signal results and pass through a latch to switch the NMOS. Pay attention to the **B** node has two places in Fig. 3.33. One is located at CMOS switch right-side; another is located at the inverter input-end that controls the DO_IN1 whether turning on.

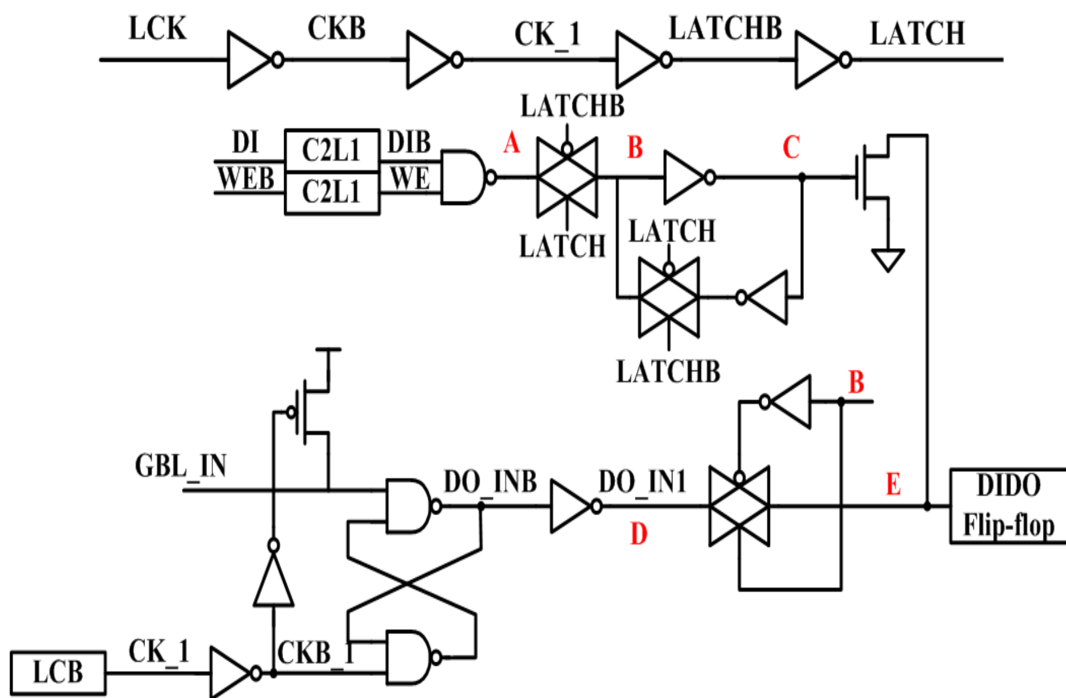


Fig. 3.33 Writer through design

Fig. 3.34 shows the simulation waveform by write through design. Cause to the write-0 operation can display clearer the variation of waveform, so we choose the write operation. In Fig. 3.34, **E** node waveform that depend on the **B** node change. Due to **B** node goes low and turns off CMOS switch, so the **E** result is decided by the

C node voltage. For the simulation, the **C** is logic “1” that turns on the NMOS making **E** node voltage to go low.

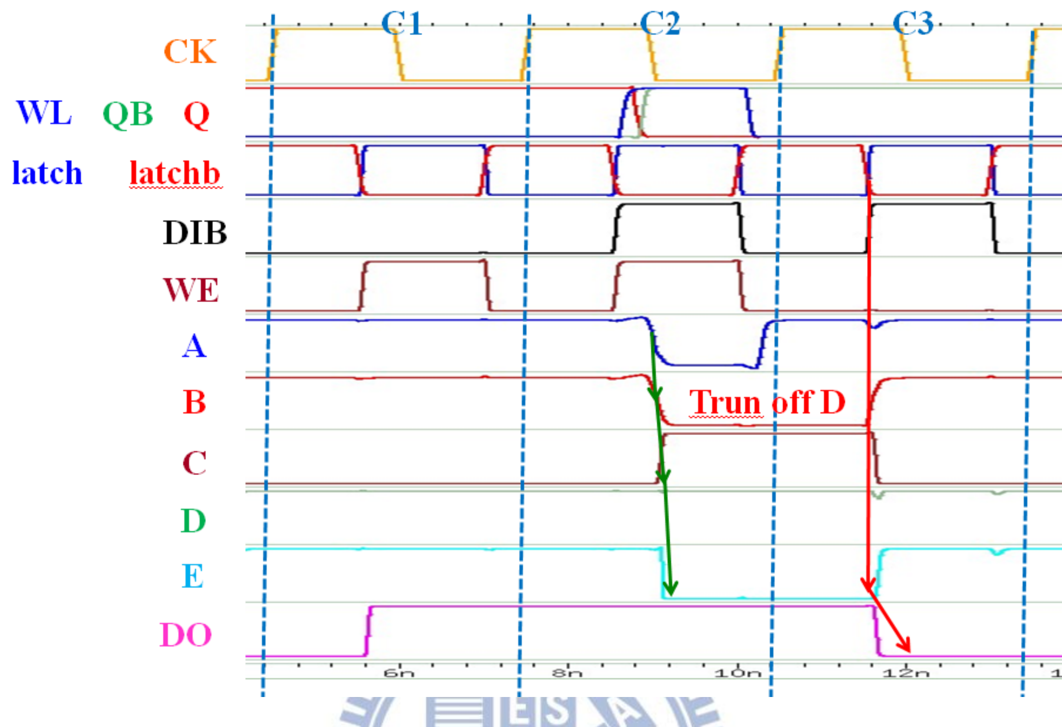


Fig. 3.34 Writer through design waveform

Chapter 4
The Chip Structure Result

4.1 Chip Spec.

We first introduce the chip pin definitions for below and Fig.4.1 shows the chip simply:

- Power pins : VDD, GND
- RESET pin: RESET
- Address pins: A<12:0>
- Data input pins: DI<63:0>
- Data output pins: DO<63:0>
- Chip select pin: CSB
- Write enable pin: WEB
- Data aware Write assist (DAWA) option pins: OSD<3:0>,
- LBL Keeper control pins : OSK<1:0>
- Voltage detector option pins: OSP<2:0> & BSTEN

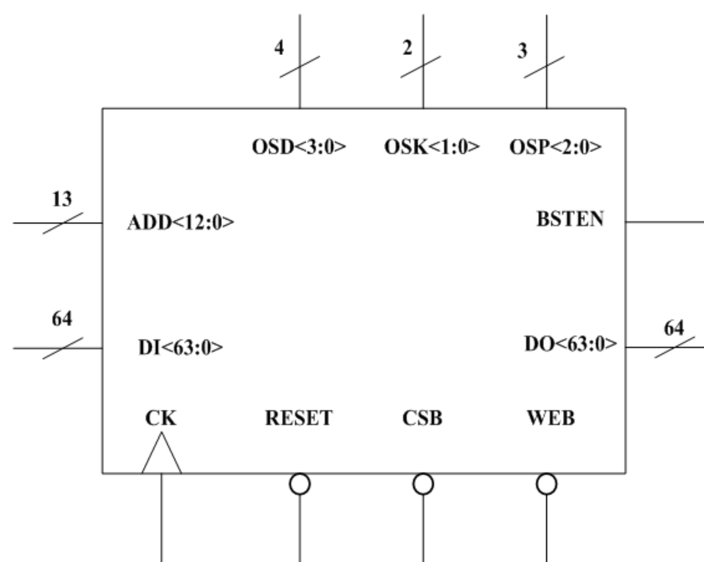


Fig. 4.1 The simple diagram of 512Kb 8T SRAM

Thesis pin descriptions we use a table to explain that (Table 4.1):

Pin Name	Description
CK	Clock signal of ADD, DI, WEB, CSB, positive edge
RESET	Negative phase RESET
A<12:0>	Address signals of width 13 bits
DI<63:0>	Input data of width 64 bits
DO<63:0>	Output data of width 64 bits
CSB	Chip select, active low
WEB	Write enable, active low
OSD<3:0>	DAWA control option pin
OSK<1:0>	LBL keeper control option pin
OSP<2:0> BSTEN	Voltage detector control option pin

Table 4.1 Pin descriptions for the chip.

Fig. 4.2-3 are stood for the cycle timing diagram for read/write cycle and signal line distribution in the chip. Whether in read or write mode, the data inputs the address and transmits to data-out has to wait for 2 cycle times. And the **t** in Fig.4.2 is represented a delay time for the Data-Out (DO) result.

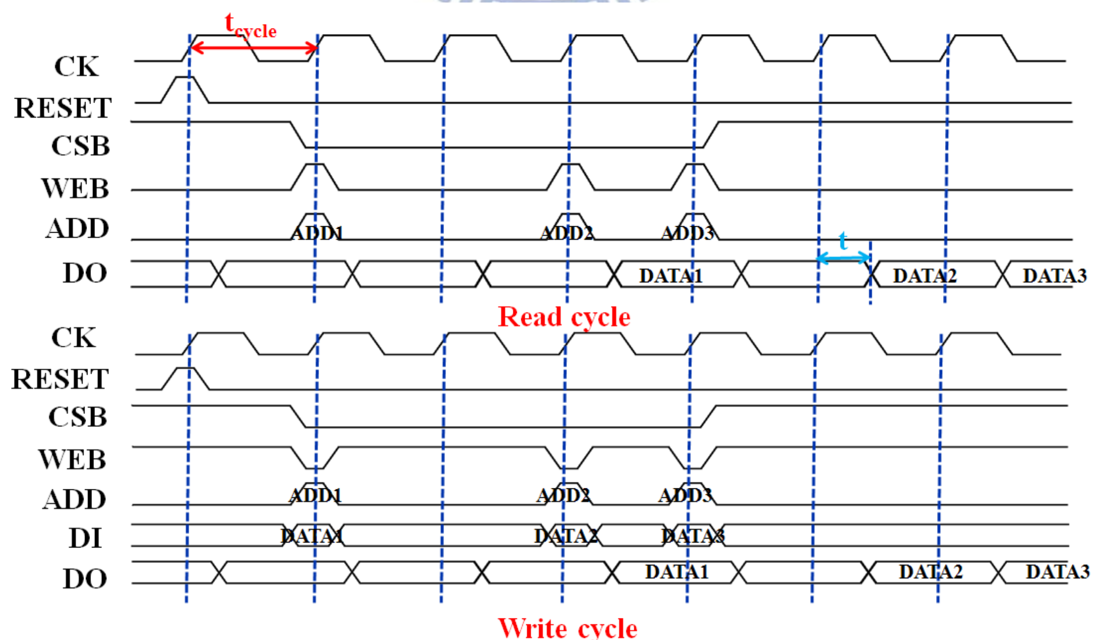


Fig 4.2 Cycle time diagram for read/write

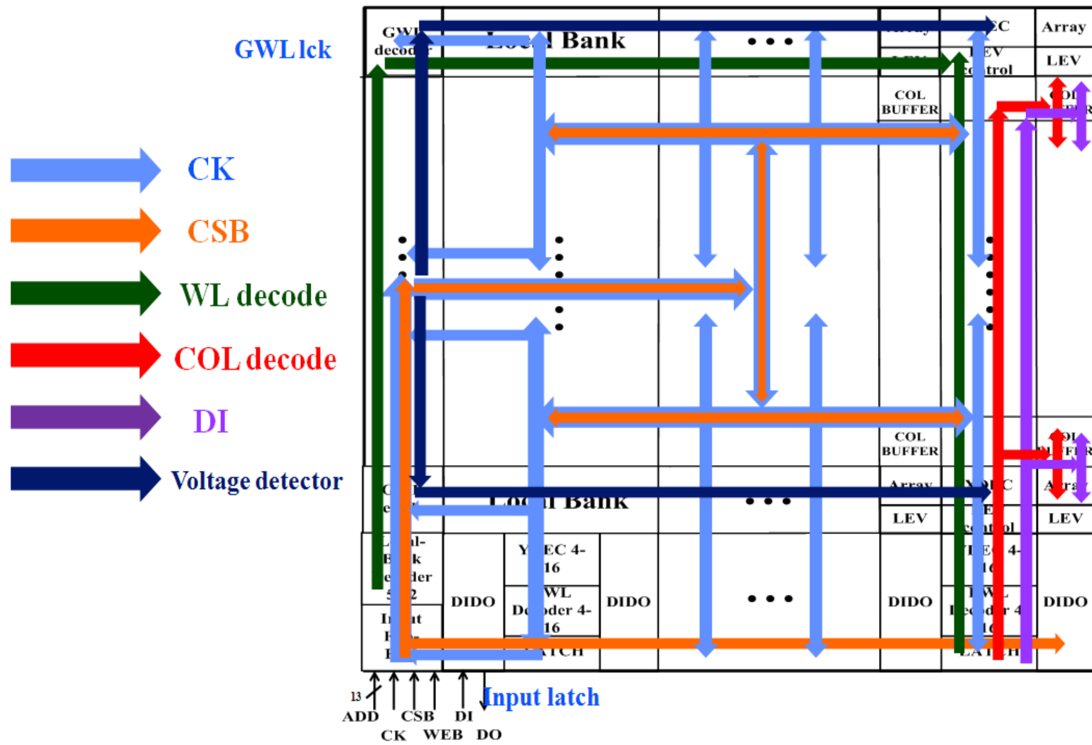


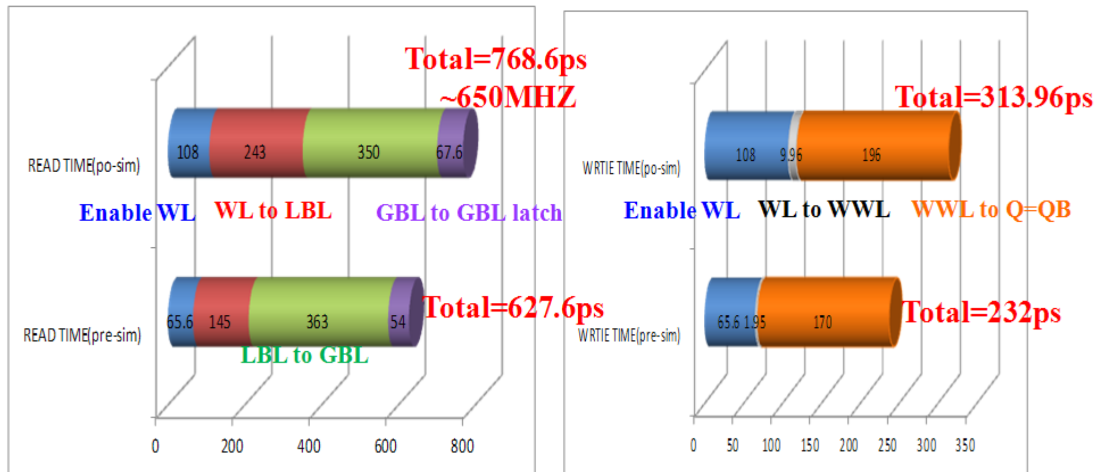
Fig 4.3 Signal line distribution in chip

4.2 Chip Per-Simulation vs. Post-Simulation

We will shows main simulations for different corners and temperatures separately in Fig. 4.4-6. Besides, we make the read/write time the chip needed by pre-sim and po-sim. The reading time, we define after the address passes through decoder decodes complete, then the signal enables one row-base WL the address selected which time is called **Enable WL**; the WL enabled and reads the selected cell data successful spreading to LBL, which time is called **WL to LBL**; and the data transmits from LBL to GBL which time is called **LBL to GBL**; finally, the data spread GBL to GBL latch placed in DIDO, which time is called **GBL to GBL latch**.

For the writing time, **Enable WL** definition is the same for reading time; the signal spread WL to WWL which time is celled **WL to WWL**; as the WWL turning on, the cell start to write until the Q value is equal QB, which time is called **WWL to Q=QB**. For above description, we have already realized the definition for Fig. 4.4-6. For

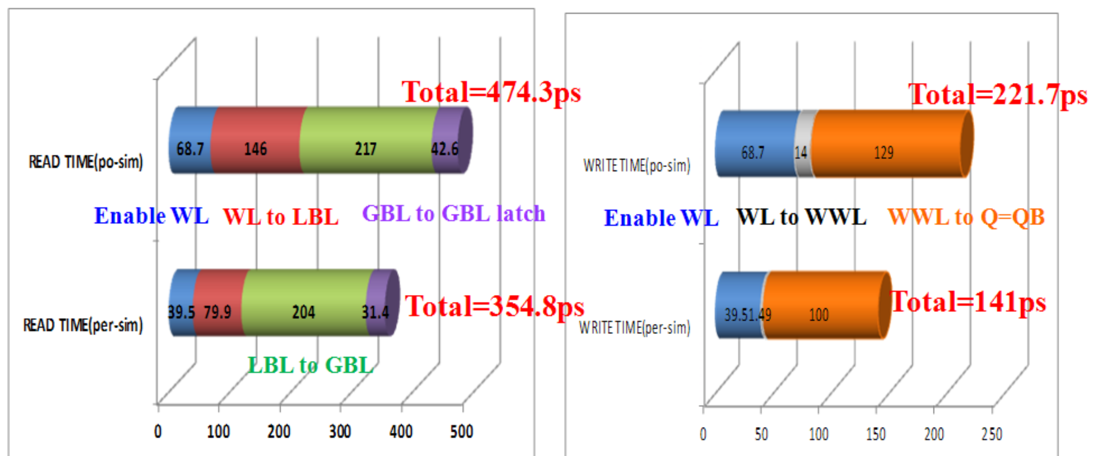
supply voltage is 1.1V, the general condition is TT, 25℃, VDD=1.1V (Fig. 4.4); the beast case is FF corner, -40℃, VDD=1.1*1.1=1.2V (Fig. 4.5); the worst case is SS corner, 125℃, VDD=1.1*0.9=0.99V (Fig. 4.6). Because of the chip is pipeline concept, we just choose the longest time to make the cycle time. In other words, we only care about the read time, although the write time degrade large than read time.



Performance degrade 22.5%

Performance degrade 32.2 %

Fig 4.4 Po-sim vs. pre-sim @ TT, 25℃, VDD=1.1V



Performance degrade 33.7%

Performance degrade 50%

Fig 4.5 Po-sim vs. pre-sim @ FF, -40℃, VDD=1.21V

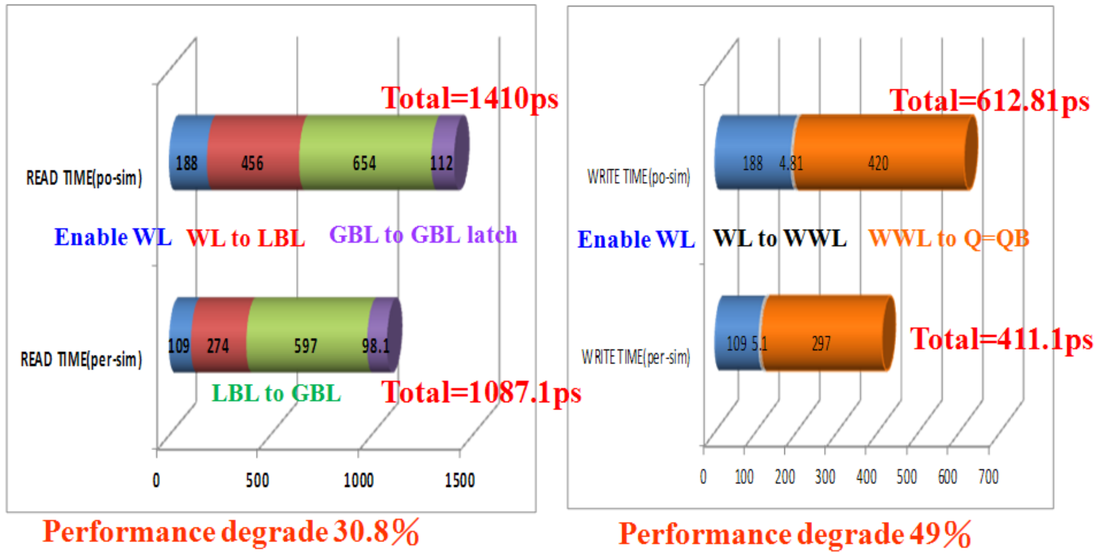


Fig 4.6 Po-sim vs. pre-sim @ SS, 125℃, VDD=0.99V

4.3 Test Flow

Fig. 4.7 shows the special circuits of half chip where they place. Only one voltage detector put the total chip center in LCB_1. Next, the DAWA enable design is placed at each **LEV** block. The keeper for local/global bit-line and DAWA tracing circuit are placed at each **LEV control** block.

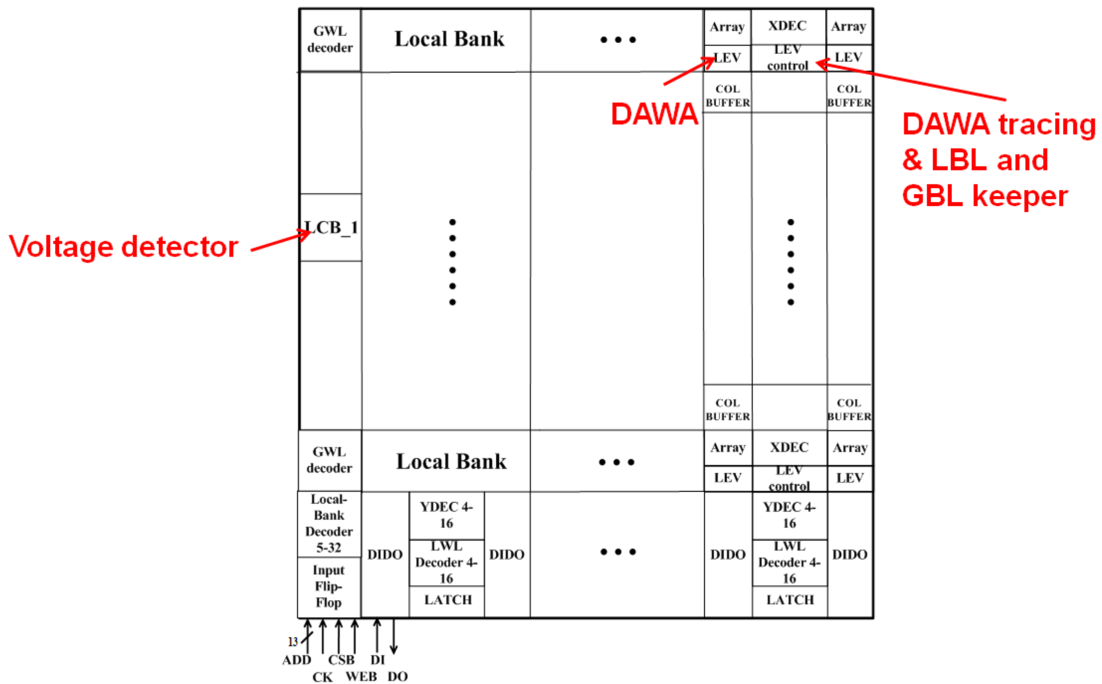


Fig 4.7 The design places on half chip.

Then, we introduce the pin information in detail, although the above article have realized generally. We look down the description of pins information:

- CK: CLK
- RESET: When CLK negative phase & RESET positive phase => Start RESET.
- CSB: Chip enable bar (CSB=0, Chip will be enabled)
- WEB: Write enable bar (WEB=0 => Write operation ; WEB=1 => Read operation)
- DI0~DI63: 64-bit Data input
- DO0~DO63: 64-bit Data output
- A12~ A0: Address (A12-A8=> Select Bank, A7-A4=> Select WL, A3-A0=> Select Column)
- BSTEN: Boosting enable (BSTEN=1 => turn on; BSTEN=0 => turn off)
- OSD3 ~ OSD0: To enable Data-Aware Write Assist(4-bits) => <OSD3,OSD2,OSD1,OSD0>
    - OSD3 =0, OSD2 =0, OSD1 =0, OSD0 =0: The Data-Aware Write Assist is "ON", startup time is **maximum**
    - OSD3 =0, OSD2 =0, OSD1 =0, OSD0 =1: The Data-Aware Write Assist is "ON", startup time is second maximum

      ↕

    - OSD3 =1, OSD2 =1, OSD1 =0, OSD0 =1 : The Data-Aware Write Assist is "ON", startup time is second minimum
    - OSD3 =1, OSD2 =1, OSD1 =1, OSD0 =0 : The Data-Aware Write Assist is "ON", startup time is **minimum**
- OSK1~ OSK0 : To enable LBL & GBL Keeper (2-bits)=> <OSK1,OSK0>
    - OSK1 =0, OSK0 =0 : The **fastest** speed of generating delay signal of keeper

- OSK1 =0, OSK0 =1 : The second fast speed of generating delay signal of keeper
- OSK1 =1, OSK0 =0 : The second slow speed of generating delay signal of keeper
- OSK1 =1, OSK0 =1 : The **slowest** speed of generating delay signal of keeper
- OSP2~ OSP0 : To enable Voltage Detector (3-bits)=> <OSP2,OSP1,OSP0>
- OSP2 =0, OSP1 =0, OSP0 =1 : The Voltage Detector is "ON", VDD<1.0V For **SS** Corner
- OSP2 =0, OSP1 =1, OSP0 =0 : The Voltage Detector is "ON", VDD<1.0V For **TT** Corner
- OSP2 =1, OSP1 =0, OSP0 =0 : The Voltage Detector is "ON", VDD<1.0V For **FF** Corner

Fig. 4.8 shows the test flow of the chip. We first test the option for chip that turns off the boosting mechanism and uses the fastest time to turn off DAWA operation. Then, both turn on the boosting and DAWA (for enough period). For different conditions, we measure the write/read minimum voltage. If unfortunately testing data failed, we should determine which operation failed. In case of writing failed, we extend time of DAWA to use. Else if reading failed, we have to adjust the LBL and GBL keeper. Finally, we must measure the power in these conditions.

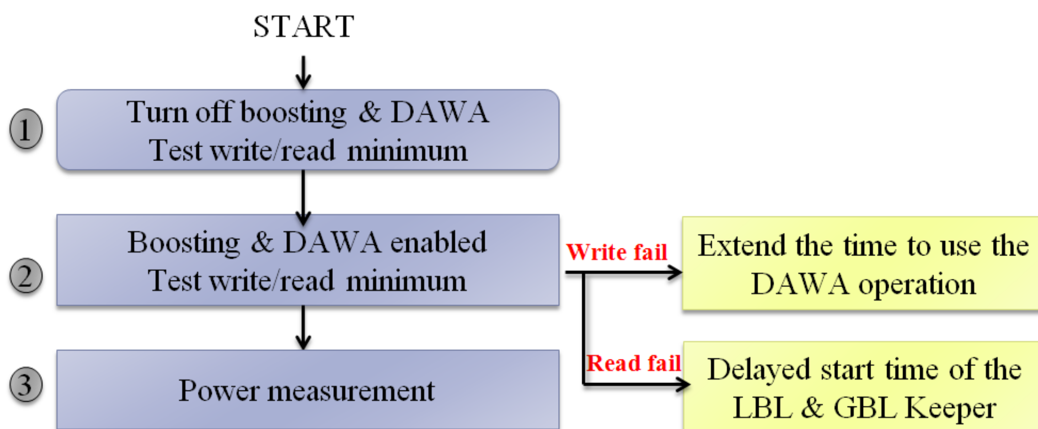


Fig 4.8 The chip testing flow.

4.4 Testing Result

We can see the chip layout at Fig. 4.9. The total width is 2810.76um and the length is 947.26um. We also watch the Fig. 4.10 that is the po-sim waveform for the chip. The function running with a sequence for writing-1, writing-0, reading-1, and reading-0 by clock started for first blue dotted line. And we can find the result to DO<63> and DO<39> that is represent with the farthest cell data and nearest cell data by CK transmission.

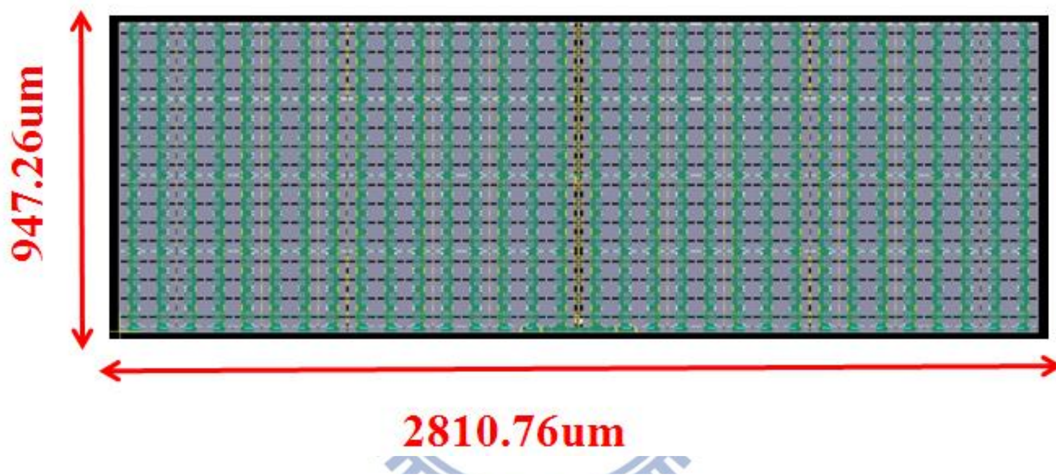


Fig 4.9 The layout of chip.

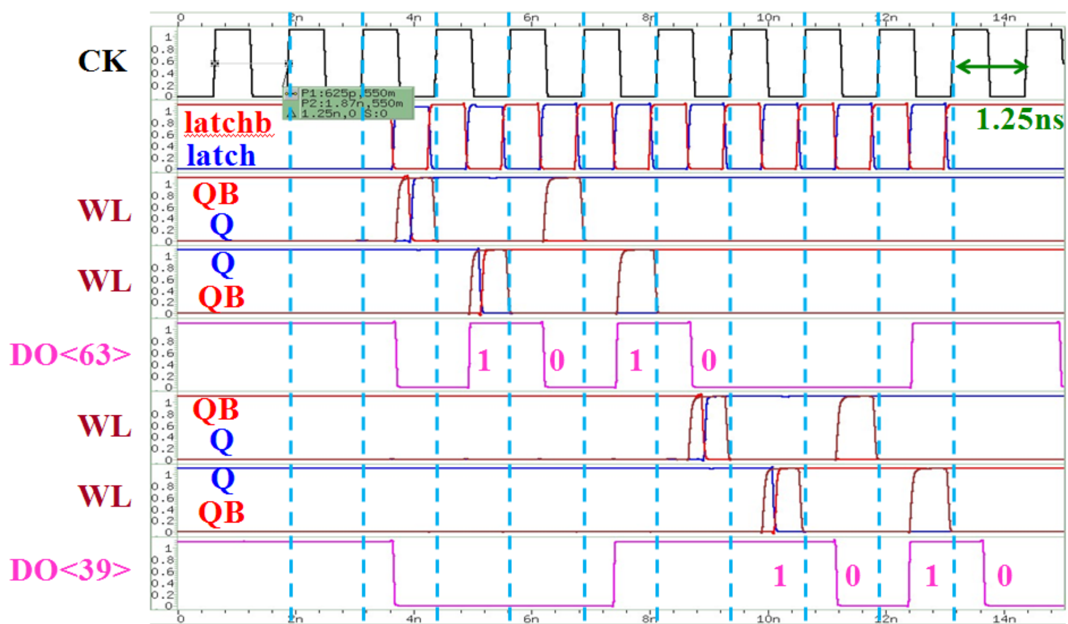


Fig 4.10 The function check for chip.

We use three options to test the chip that shows as Table 4.2. And the way of operation in detail has been described by above section. Afterwards, we start to see the chip result of pass rate at Table. 4.3-5. The chip testing results in three different corner there are include TT, FF, and SS corners at 25℃. There are 65 samples at TT corner; 58 samples at FF corner; 53 samples at SS corner.

	BSTEN	OSP[2]	OSP[1]	OSP[0]	OSK[1]	OSK[0]	OSD[3]	OSD[2]	OSD[1]	OSD[0]
op1	0	0	1	0	1	0	1	1	1	0
op2	1	0	1	0	1	0	1	1	0	0
op3	1	0	1	0	0	1	1	1	0	0
		Voltage detector			**Keeper**		**DAWA control**			

Table 4.2 Three options for the chip.

Boldface(粗體字) is default option	1.5V	1.475V	1.45V	1.425V	1.4V	1.375V	1.35V	1.325V	1.3V	1.275V	1.25V	1.225V	1.2V	1.175V	1.15V
SHLA40_8192X64X1CM16A_PIPE8T_op1.rst_Yield(%):	98.5	98.5	98.5	98.5	98.5	98.5	98.5	98.5	98.5	98.5	98.5	98.5	98.5	98.5	98.5
SHLA40_8192X64X1CM16A_PIPE8T_op2.rst_Yield(%):	98.5	98.5	98.5	98.5	98.5	98.5	98.5	98.5	98.5	98.5	98.5	98.5	98.5	98.5	98.5
SHLA40_8192X64X1CM16A_PIPE8T_op3.rst_Yield(%):	98.5	98.5	98.5	98.5	98.5	98.5	98.5	98.5	98.5	98.5	98.5	98.5	98.5	98.5	98.5

1.125V	1.1V	1.075V	1.05V	1.025V	1V	0.975V	0.95V	0.925V	0.9V	0.875V	0.85V	0.825V	0.8V	0.775V	0.75V	0.725V	0.7V	0.675V	0.65V
98.5	98.5	98.5	98.5	98.5	98.5	98.5	98.5	98.5	98.5	98.5	98.5	98.5	98.5	98.5	87.7	70.8	47.7	12.3	0
98.5	98.5	98.5	98.5	98.5	98.5	98.5	98.5	98.5	98.5	98.5	98.5	98.5	98.5	98.5	87.7	70.8	50.8	16.9	3.1
98.5	98.5	98.5	98.5	98.5	98.5	98.5	98.5	98.5	98.5	98.5	98.5	98.5	98.5	98.5	89.2	70.8	53.8	18.5	3.1

Table 4.3 The pass rate for TT, 25℃

Boldface(粗體字) is default option	1.5V	1.475V	1.45V	1.425V	1.4V	1.375V	1.35V	1.325V	1.3V	1.275V	1.25V	1.225V	1.2V	1.175V	1.15V
SHLA40_8192X64X1CM16A_PIPE8T_op1.rst_Yield(%):	94.8	93.1	94.8	94.8	94.8	94.8	94.8	94.8	94.8	94.8	94.8	94.8	94.8	94.8	94.8
SHLA40_8192X64X1CM16A_PIPE8T_op2.rst_Yield(%):	94.8	94.8	93.1	94.8	94.8	94.8	94.8	94.8	94.8	94.8	94.8	94.8	94.8	94.8	94.8
SHLA40_8192X64X1CM16A_PIPE8T_op3.rst_Yield(%):	93.1	94.8	93.1	93.1	93.1	94.8	94.8	94.8	94.8	94.8	94.8	94.8	94.8	94.8	94.8

1.125V	1.1V	1.075V	1.05V	1.025V	1V	0.975V	0.95V	0.925V	0.9V	0.875V	0.85V	0.825V	0.8V	0.775V	0.75V	0.725V	0.7V	0.675V	0.65V	0.625V	0.6V
94.8	94.8	94.8	94.8	94.8	94.8	94.8	94.8	94.8	94.8	94.8	94.8	94.8	94.8	94.8	93.1	93.1	91.4	77.6	56.9	48.3	24.1
94.8	94.8	94.8	94.8	94.8	94.8	94.8	94.8	94.8	94.8	94.8	94.8	94.8	94.8	94.8	93.1	93.1	91.4	81	58.6	48.3	25.9
94.8	94.8	94.8	94.8	94.8	94.8	94.8	94.8	94.8	94.8	94.8	94.8	94.8	94.8	94.8	93.1	93.1	89.7	81	56.9	48.3	29.3

Table 4.4 The pass rate for FF, 25℃

Boldface(粗體字) is default option	1.5V	1.475V	1.45V	1.425V	1.4V	1.375V	1.35V	1.325V	1.3V	1.275V	1.25V	1.225V	1.2V	1.175V	1.15V
SHLA40_8192X64X1CM16A_PIPE8T_op1.rst_Yield(%):	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
SHLA40_8192X64X1CM16A_PIPE8T_op2.rst_Yield(%):	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
SHLA40_8192X64X1CM16A_PIPE8T_op3.rst_Yield(%):	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100

1.125V	1.1V	1.075V	1.05V	1.025V	1V	0.975V	0.95V	0.925V	0.9V	0.875V	0.85V	0.825V	0.8V	0.775V	0.75V	0.725V	0.7V
100	100	100	100	100	100	100	100	100	100	98.1	92.5	88.7	69.8	50.9	28.3	11.3	5.7
100	100	100	100	100	100	100	100	100	100	98.1	94.3	92.5	77.4	62.3	43.4	26.4	9.4
100	100	100	100	100	100	100	100	100	100	98.1	94.3	94.3	81.1	66	43.4	28.3	11.3

Table 4.5 The pass rate for SS, 25℃

Table 4.3-5 pass rates mean that how many samples can read/write operation for all SRAM cells successful and data-out complete, then the number compare to the total sample number. This ratio is the yield in Tables. For see the result easily, we anew the above table data to transform Fig. 4.11-13.

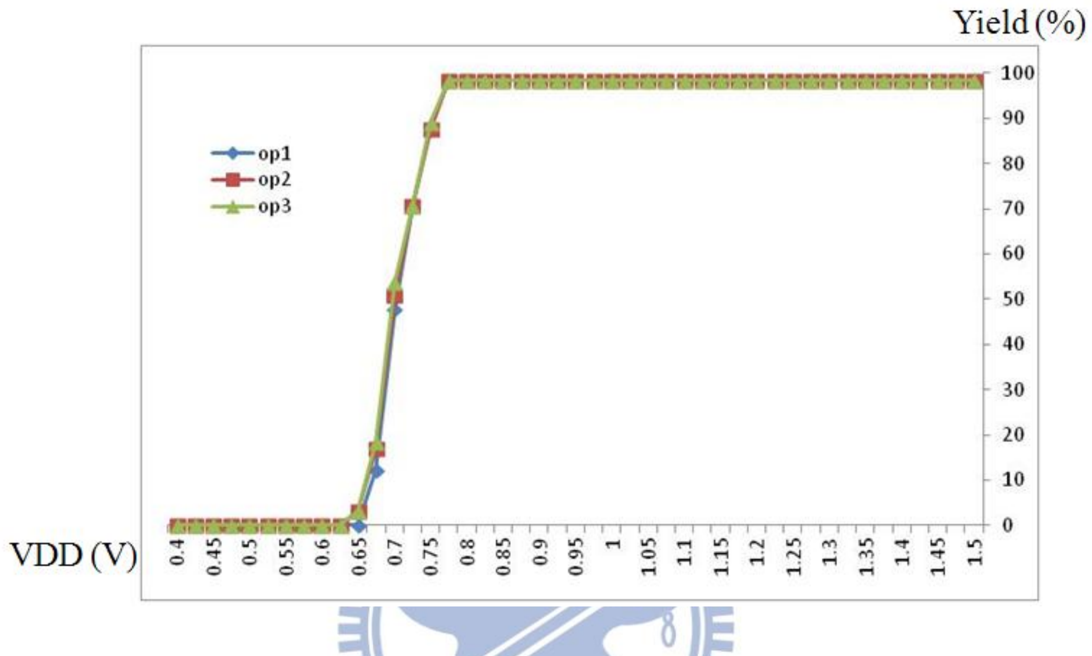


Fig. 4.11 The pass rate (Yield) for TT, 25℃

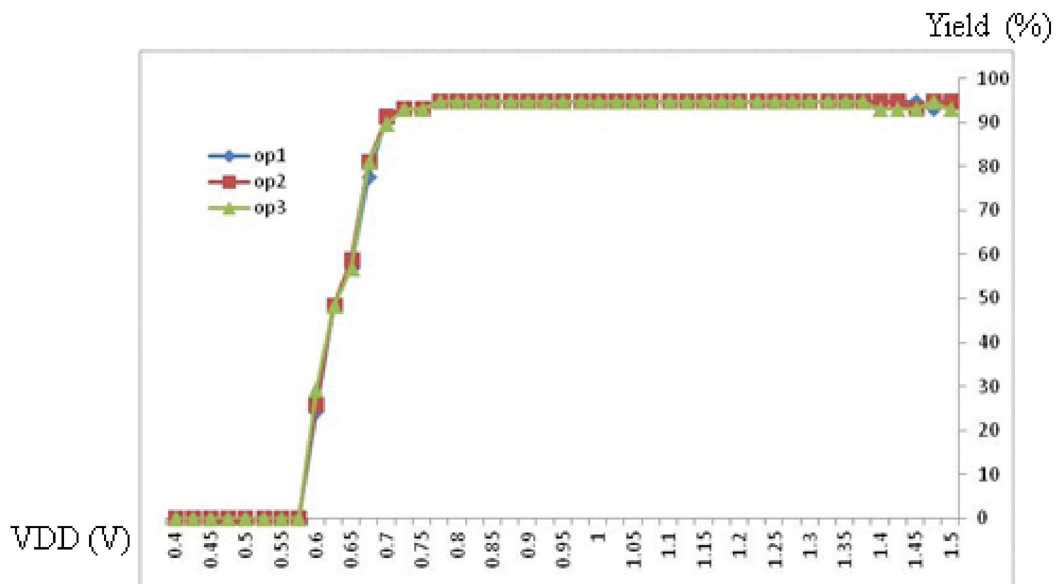


Fig. 4.12 The pass rate (Yield) for FF, 25℃

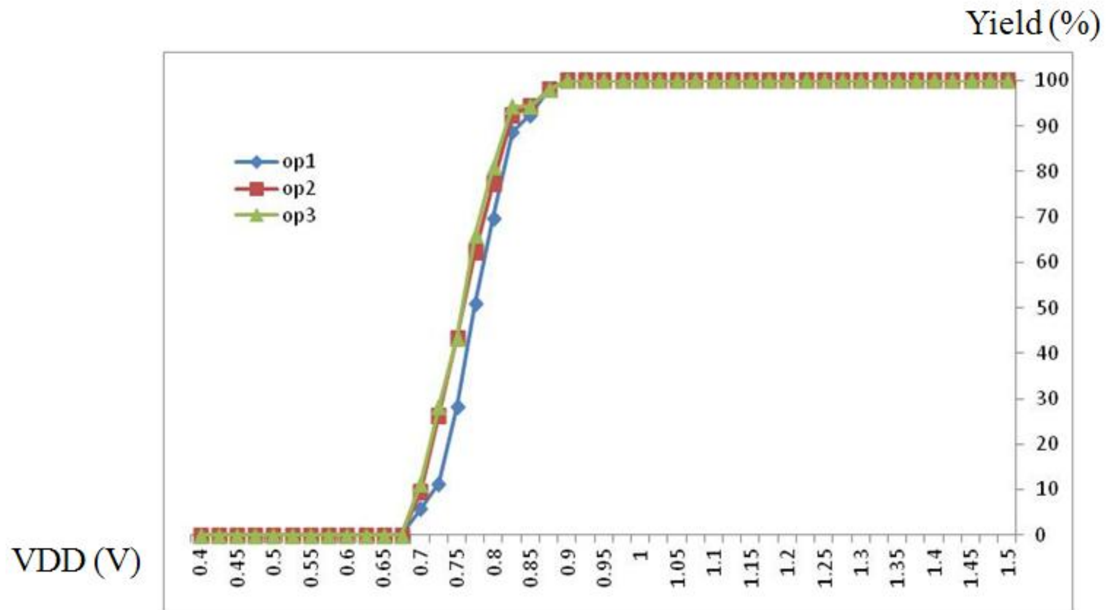


Fig. 4.13 The pass rate (Yield) for SS, 25℃

The boosting enabled (op2/3) and not enabled (op1) has a significant difference at SS corner (Fig. 4.13). The figure can prove the boosting circuit and voltage detector are working successfully. The boosting function increases the yield of 10% in low supply VDD, even up to 17% in SS corner.

Chapter 5
Conclusion
5.1 Conclusion

For the recent IC design, how to decrease the design chip power consumption is the trend. Low power benefit not only can increase number of times for battery-supplied products, but also save more energy for plugged 3C products. By the equation: Power = Current * VDD, we understand that just reduce the total current or supply voltage, even better the two things both drop off to get the goal. In modern IC design, SRAM use the area is the largest of whole SoC design. Because of reduced power, .we should careful design SRAM circuit, of course the performance should not too bad.

For come down the power, we choose the way to descend VDD. The 8T SRAM can working correct better than conventional 6T SRAM, because the 8T SRAM structure design and its operation make the RSNM better than 6T SRAM, although the area become bigger. The SRAM read supply voltage has already reduced; next deals with the write operate at low voltage. We use the data-aware write assist (DAWA) and boosting WL mechanism to help write data in cell accomplish. For more power saving, we also design the power gating to get the target. This design make only the working cell array provide the voltage, others cell arrays which do not work turn off the VDD to supply.

In addition, in order to make the performance better, put into the ripple bit-line-read-path composition and put pipeline structure in whole chip. And the preliminary testing result in the chip also verify the above design is correct and practicably.

Reference

Chapter 1

[1.1] L. Chang, R.K. Montoye, Y. Nakamura, K.A. Batson, R.J. Eickemeyer, R.H. Dennard, W. Haensch, D. Jamsek, "An 8T-SRAM for Variability Tolerance and Low-Voltage Operation in High-Performance Caches," *IEEE Journal of Solid-State Circuits,* vol.43, no.4, pp.956-963, April 2008.

Chapter 2

[2.1] Roberto Bez, Emilio Camerlenghi, Alberto Modell, and Angelo Visconti, "Introduction to Flash Memory," *Proceedings of the IEEE,* vol.91, no.4, pp.489-502, April 2003.

[2.2] Harry Pon, "Technology Scaling Impact on NOR and NAND Flash Memories and Their Applications", *Solid-State and Integrated Circuit Technology, 2006. ICSICT '06. 8th International Conference on,* pp.697-700, October 2006.

[2.3] Hari Ananthan, Aditya Bansal and Kaushik Roy, "FinFET SRAM – Device and Circuit Design Considerations", *Quality Electronic, 2004. Proceedings. 5th International symposium on,* pp.511-516, 2004

[2.4] Sherif A. Tawifk, Zhiyu Liu, Volkan Kursun, "Independent-Gate and Tied-Gate FinFET SRAM Circuits: Design Guidelines For Reduced Area and Enhanced Stability", *Microelectronics 2007. ICM 2007. International Conference on,,* pp.171-174, December 2007

[2.5] L. Chang, R.K. Montoye, Y. Nakamura, K.A. Batson, R.J. Eickemeyer, R.H. Dennard, W. Haensch, D. Jamsek, "An 8T-SRAM for Variability Tolerance and Low-Voltage Operation in High-Performance Caches," *IEEE Journal of Solid-State Circuits,* vol.43, no.4, pp.956-963, April 2008.

[2.6] .Adel S. Sdera, Kenneth C. Smith, "Microelectronic Circuits" 6rd ed.

[2.7] EVERT SEEVINCK, FRANS J. LIST, and JAN LOHSTROH, "Static-Noise Margin Analysis of MOS SRAM Cells", *IEEE J. Solid-State Circuit*, VOL. SC-22, NO. 5, OCTOBER 1987.

[2.8] "Introduction to papers presented", *ISSC 2006,* VOL. 54, March 2006.

[2.9] 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.

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

[2.11] 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.

[2.12] David J. Frank and Hon-Sun P. Wong, "Simulation of Stochastic Doping Effects in Si MOSFETs", *Computational Electronics, 2000. Book of Abstracts. IWCE Glasgow 2000. 7th International Workshop on,* 2000, pp.2-3.

[2.13] S. Natarajan, A. Shubat, "SE5 - SRAM design in the nanoscale era," *IEEE Iternational Solid-State Circuits Conference, Digest of Technical Papers,* pp.366-367, 10-10 Feb. 2005.

[2.14] M. Kheelah at al., "A 4.2GHz 0.3mm^2 256kb Daul-Vcc SRAM Building Block in 65nm CMOS", *IEEE International Solid-State Circuits Conference,* 2006, pp624-625.

[2.15] 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.

[2.16] Fujimura, Y.et at. "A Configurable SRAM with Constant-Negative-Level Write Buffer for Low-Voltage Operation with 0.149um2 Cell in 32nm High-k Metal-Gate CMOS", *IEEE Internal Solid-State Circuits Conference Digest of Technical Papers,* 2010, pp.348-349.

[2.17] H. Pilo, I. Arsovski, K. Batson, G. Braceras, J. Gabric, R. Houle, S. Lamphier, F. Pavlik, A. Seferagic, L.-Y. Chen, S.-B. Ko, C. Radens, "A 64Mb SRAM in 32nm High-k metal-gate SOI technology with 0.7V operation enabled by stability, write-ability and read-ability enhancements", *IEEE International SOC Conference,* pp.211-214, 26-29, Sept. 2007.

[2.18] K.Nii et al. "A 45-nm Bulk CMOS Embedded SRAM with Improved Immunity Aginst Process and Temperature Variations", *IEEE Journal of Solid-State Circuits,* vol. 43, issue 1, pp. 180-191.

[2.19] M. Yabuuchi et al, "A 45nm Low-Standby-Power Embedded SRAM with Improved Immunity Against Process and Temperature Variations", *IEEE International Digest of Technical Papers.* 2007, pp. 326-606.

[2.20] Kyechong Kim, Iliadis, A. A., "Impact of Microwave Interference on Dynamic Operation and Power Dissipation of CMOS Inverters", *IEEE Transactions on Electromagnetic Compatibility,* vol. 49, issue:2, pp. 329-338.

[2.21] H.J.M Veendrick, "Short-circuits dissipation of static CMOS circuitry and its impact on the design of buffer circuits", *IEEE Journal of Solid-State Circuits,* vol. 19, no. 4, pp. 468-473, Aug. 1984.

[2.22] Q. Chen, S. Mukhopadhyay, A. Bansal, K. Roy, "Circuit-aware Device Design Methodology for Nanometer Technologies: A Case Study for Low Power SRAM Design", *Proceedings, Design, Automation and Test in Europe,* vol.1, pp.1-6, 6-10, March 2006.

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

Chapter 3

[3.1] C. T. Chuang, S. Mukhopadhyay, J. J. Kim, R. Rao, "High-Performance SRAM in Nanoscale CMOS: Design Challenges and Techniques", *IEEE International Workshop on Memory Technology, Design and Testign,* pp. 4-12, Dec. 2007.

[3.2] H. I. Yang, S. Y. Lai, W. Hwang, "Low-Power Floating Bitline 8-T SRAM Design with Write Assistant Circuits", *IEEE SOC Conference,* pp. 239-242, Sep. 2008.

[3.3] B. D. Yang, "A Low-Power SRAM Using Bit-Line Charge-Recycling for Read and Write Operations", *IEEE International Solid-State Circuits Journal,* pp. 2173-2183, Jan. 2007.

[3.4] B. D. Yang, L. S. Kim, "A Low-Power SRAM Using Hierarchical Bit and Local Sense Amplifiers", *IEEE International Solid-State Circuits Journal,* pp. 1366-1376, June 2006.

[3.5] S. P. Cheng, S. Y. Huang, "A Low-Power SRAM Design Using Quiet-Bitline Architecture", *IEEE Internation Workshop on Memory Technology, Design, and Testing,* pp. 135-139, Aug. 2005.

[3.6] H. Qin, A. Kumar, K. Ramchandran, J. Rabaey, "Error-Tolerant SRAM Design for Ultra-Low Power Standby Operation", *ISQED Conference Quality Electronic Design,* pp. 30-34, March 2008.

[3.7] H. Qin, Y. Cao, D. Markovic, A. Vladimirescu, J. Rabaey, "SRAM Leakage Suppression by Minimizing Standby Supply Voltage", *ISQED Conference Quality Electronic Design,* pp. 55-60, 2004.

[3.8] S. Hattori, T. Sakurai, "90% Write Power Saving SRAM Using Sense-Amplifying Memory Cell", *IEEE Solid-State Circuits,* vol. 39, Issue 6, pp. 927-933, 2004.

[3.9] T. B. Hook, M. Bretwisch, J. Brown, P. Cotterll, D. Hoyniak, C. Lam, R. Mann, "Noise Margin and Leakage in Ultra-Low Leakage SRAM Cell Design", *IEEE Electron Device ,* pp. 1499- 1501,Aug. 2002.

[3.10] M. Margala, "Low-Power SRAM Circuit Design", *IEEE International Workshop on Memory Technology, Design, and Testing,* pp. 115-122, Aug. 1999.

[3.11] L. Chang, R.K. Montoye, Y. Nakamura, K.A. Batson, R.J. Eickemeyer, R.H. Dennard, W. Haensch, D. Jamsek, "An 8T-SRAM for Variability Tolerance and Low-Voltage Operation in High-Performance Caches," *IEEE Journal of Solid-State Circuits,* vol.43,no.4, pp.956-963, April 2008.

[3.12] Hao-I Yang, Shih-Chi Yang, Mao-Chih Hsia, Yung-Wei Lin, Yi-Wei Lin, Chien-Hen Chen, Shi-Shin Chang, Geng-Cing Lin, Yin-Nien Chen, Ching-Te Chuang, Wei Hwang, Shyh-Jye Jou, Nan-Chun :ien, Hung-Yu Li, Kuen-Di Lee, Wei-Chiang Shih, Ya-Ping Wu, Wen-Ta Lee, Chih-Chiang Hse, " A high-performance low VMIN 55nm 512Kb disturb-free 8T SRAM with adaptive VVSS control ", *SOC Conference (SOCC), 2011 IEEE International*, pp. 197-200.

[3.13] Restle, P.J.; McNamara, T.G.; Webber, D.A.; Camporese, P.J.; Eng, K.F.; Jenkins, K.A.; Allen, D.H.; Rohn, M.J.; Quaranta, M.P.; Boerstler, D.W.; Alpert, C.J.; Carter, C.A.; Bailey, R.N.; Petrovic, J.G.; Krauter, B.L.; McCredie, B.D., " A clock distribution network for microprocessors ", *VLSI Circuits, 2000. Digest of Technical Papers. 2000 Symposium on*, pp. 184-187

[3.14] Restle, P.J.; Deutsch, A., “Designing the best clock distribution network”,

VLSI Circuits, 1998. Digest of Technical Papers. 1998 Symposium on , pp. 2-5

[3.15] Restle, P.J.; Carter, C.A.; Eckhardt, J.P.; Krauter, B.L.; McCredie, B.D.; Jenkins, K.A.; Weger, A.J.; Mule, A.V., "The Clock Distriburion of the POWER4 Microprocessor", *Solid-State Circuits Conference, 2002. Digest of Technical Papers. ISSCC. 2002 IEEE International*, pp. 144-145 vol. 1

[3.16] Thomson, M.G.R.; Restle, P.J.; James, N.K., "A 5GHz Duty-Cycle Correcting Clock Distribution Network for the POWER6 Microprocessor", *Solid-State Circuits Conference, 2006. ISSCC 2006. Digest of Technical Papers. IEEE International* , pp. 1522-1529

[3.17] J. Pille, C. Adams, T. Christensen, S. Cottier, S. Ehrenreich, F. Kono, D. Nelson, O. Takahashi, S. Tokito, O. Torreiter, O. Wagner, D. Wendel, "Implementation of the CELL Broadband Engline in a 65nm SOI Technology Featuring Dual-Supply SRAM Arrays Supporting 6GHz at 1.3V", *Solid-State Circuits Conference, 2007. ISSCC 2007. Digest of Technical Papers. IEEE International*, pp. 322-606

[3.18] Raychowdhury, A.; Geuskens, B.; Kulkarni, J.; Tschanz, J.; Bowman, K.; Karnik, T.; Shih-Lien Lu; De, V.; Khellah, M.M., "PVT-and-Aging Adaptive Wordline Boosting for 8T SRAM Power Reduction", *Solid-State Circuits Conference Digest of Technical Papers (ISSCC), 2010 IEEE International* , pp. 352-353

[3.19] Tanakamaru, S.; Takeuchi, K., "a 60pJ, 3-Clock Rising Time, VTH Loss Compensated Word-Line Booster Circuit for 0.5V Power Supply Embedded/Discrete DRAMs", *Memory Workshop, 2009. IMW '09. IEEE International* , pp. 1-2

[3.20] Tanzawa, T.; Atsumi, S., "optimization of word-line booster circuits for low-voltage flash memories", *IEEE Journal of Solid-State Circuits,* vol. 34,

issue 8, pp. 1091-1098, Aug. 1999

[3.21] Masaaki Iijima, Kayoko Seto, Masahiro Numa, Akira Tada, Takashi Ipposhi, "Low Power SRAM with Boost Driver Generating Pulsed Word Line Voltage for Sub-1V Operation", *journal of computers*, vol. 3, no. 5, May 2008

[3.22] Y. Lih, N. Tzartzanis, W. Walker, "A Leakage Current Replica Keeper for Dynamic Circuits", *IEEE Solid-State Circuits Jurnal, Jan. 2007,* pp.48-55.

[3.23] A. Agarwal, S. Hsu, S. Mathew, M. Anders, H. Kaul, F. Sheikh, R. Krishnamurthy, "A 32nm 8.3GHz 64-entry × 32b variation tolerant near-threshold voltage register file", *IEEE Symposium on VLSI Circuits,* pp.105-106,16-18 June 2010

[3.24] Yi-Wei Lin, Hao-I Yang, Geng-Cing Lin, Chi-Shin Chang, Ching-Te Chuang, Wei Hwang, Chia-Cheng Chen, Willis Shih, Huan-Shun Huang, "A 55nm 0.55V 6T SRAM with Variation-Tolerant Dual-Tracking Word-Line Under-Drive and Data-Aware Write-Assist," *Proc. 2012 IEEE International Symposium on Low Power Electronics and Design (ISLPED), Redondo Beach, CA, USA,* pp. 79-84, July 30 - August 1, 2012.

Vita

Li-Wei Chu 朱俐瑋

PERSONAL INFORMATION

Birth Date: March 9. 1987

Birth place: Taipei, Taiwan, R.O.C.

Address: Department of Electronics Engineering

National Chiao Tung University

1001 Ta-Hsueh Road

Hsin-Chu, Taiwan 30010, R.O.C.

E-Mail Address: b9427106@gmail.com

EDUCATION

08/2010-09/2012 M.S. in Electronics Engineering, National Chiao Tung University

09/2005-09/2009 B.S. in Electronics Engineering, Chang Gung University

PATENT

"奈米靜態隨機存取記憶體之變異容忍升壓控制使用電壓(準位)偵測器" ,12(專)A050, 中華民國及美國專利由世界專利商標事務所準備申請中。(國立交通大學)。