國立交通大學

電子工程學系 電子研究所

碩士論文

實現在 40 奈米製程下可操縱在低電壓的四讀四寫多執行序暫存器叢集設計

Low VDD_{MIN} 4R4W Multi-Thread Register File Design and Implementation in 40nm CMOS Process

研 究 生 林弘璋

指導教授：黃 威 教授

莊景德 教授

中 華 民 國 一〇一年 九 月

實現在40奈米製程下可操縱在低電壓的四讀四寫多執行序暫存器叢集設計

Low VDD_{MIN} 4R4W Multi-Thread Register File Design and Implementation in 40nm CMOS Process

研 究 生：林弘璋　　Student：Hon-Jarn Lin

指導教授：黃　威 教授　　Advisor：Prof. Wei Hwang

莊景德 教授　　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 & 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 in

Electronics Engineering

June 2012

Hsinchu, Taiwan

中華民國一０一年九月

實現在40奈米製程下可操縱在低電壓的四讀四寫多執行序暫存器叢集設計

學生：林弘璋　　　　　　指導教授：黃　威　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

隨著攜帶式電子產品，像是手機、筆記型電腦、影像通訊和眾多電腦商品越來越廣泛的運用，一個低功率消耗且可提供 SoC 晶片平行處理的記憶體是非常重要的課題。在這篇論文中分別探討了兩個主題，第一個是具有二讀二寫 8Kb 靜態隨機存取記憶體，另一個則是具有四讀四寫 2Kb 多執行序暫存器叢集設計，兩者皆實現在 TSMC 40nm 製程上。為了達到高頻寬以及高效能，傳統單一讀寫隨機存取記憶體無法提供夠高的效率，因此我們提出了一顆具有二讀二寫多重埠的靜態隨機存取記憶體，此設計不僅可以解決同時列選取干擾並且可以使用於位元交錯結構，其他設計像是相鄰共用寫入模組、CLK 閘控和電壓偵測器阻隔皆可提供節省更多能量消耗。一個 8Kb 的測試晶片設計與實現在 TSMC 40nm 製程下，經由電路布局後的模擬顯示，在 0.9 伏特可操作在 475 百萬赫茲。另一個設計為具四讀四寫多執行序暫存器叢集設計，新的技術像是一周期兩次寫/讀、支援四個序列並行、資料空位轉移和共用讀取模組。此設計提供廣泛電壓使用，可從 0.4 伏特到 1.2 伏特讓使用可以更佳有彈性。考慮低能量消耗技術像是沒有仿造讀取動作、可降低一半讀取電路和字元線保持在高電位。藉由這些設計可大幅降低動態能量消耗以及靜態能量消耗分別為 50%和 25%。 經由電路布局後的模擬顯示在 0.9 伏特可操作在 238 百萬赫茲。

Low VDD_{MIN} 4R4W Multi-Thread Register File Design and Implementation in 40nm CMOS Process

Student: Hon-Jarn Lin Advisors: Prof. Wei Hwang

Prof. Ching-Te Chuang

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

ABSTRACT

There are wide-ranging usage of portable mobile device (PMD) such as cell phone, notebook and video product and many different types of computers in today markets. It is crucial important to emphasis energy efficiency, low power consumption and parallel memory design in system-on-chips (SoC) recently. In thesis, two topics will be presented. First topic is the low power 2R2W 8Kb multi-port SRAM design, second topic is the low power 4R4W 2Kb multi-thread register file design and implementation in TSMC 40nm CMOS technology. In order to gain high bandwidth and high performance, conventional single-port SRAM design is not efficiency. In this way, we proposed a new structure 2R2W multi-port bit-cell structure, this cell not only eliminate the half select distribute problem but also support bit-interleaving structure. Low power technology such as share WBL structure, CLK gating and SA power gating are included. An 8K test chip is designed and implemented in TSMC 40nm general purpose CMOS process. Post-layout simulation results demonstrate operating frequency of 475 MHz at 0.9V. Another work is 4R4W multi-thread register file design, with double pump, four threads, data slot switch control and share RBL structure technology are proposed. Wide range supply voltage operation form 0.4V to 1.2V, it supply designer has more flexibility. No dummy read operation, reducing RBL to 1/2 and RWL keep VVSS are design for low power consideration. In this work, active power reduction is more than 50% and standby power reduction is less than 25%. Post-layout simulation results demonstrate operating frequency of 238 MHz at 0.9V.

致　謝

可以順利完成這篇論文，有許多要感謝的人。首先，我的兩位指導老師黃威教授和莊景德教授，感謝能給我許多寶貴的意見和優渥的研究資源，讓我在研究時可以全力以赴不會有後顧之憂。兩位老師資深歷練，除了平時間指導我們有關研究方面的難題，也常常教導我們人生的哲理，讓我們每每受益良多。

接著是要感謝一起打拼的學長們王道平、張銘宏、黃柏蒼、楊皓義，在研究的路上崎嶇不平，對於一個轉組的學生更是困難重重，感謝他們不餘遺力教導和適時的給我許多想法，讓我得以度過難關解決層層關卡。感謝 LPMD 的同學們，因為有了你們，讓原本乏味的研究生活了增添了不少色彩，一起熬夜一起歡樂，此外 Digital VLSI Lab 實驗室的大家，有幸大家在這一路上互相扶持跟成長，也由衷的感謝。最後就是那些常常給我大力支持以及傾聽我苦水的工科學弟妹們，由衷的感謝大家。

最後，感謝家人對我的鼓勵與支持，也是我論文最大的推手，當我最堅實的後盾，讓我可以一心一意的完成我的研究，在此獻上無限的感謝。

Contents

Chapter 1 Introduction

1.1 Background 1

1.2 Challenges 2

1.3 Motivation...3

1.4 Thesis Organization 4

Chapter 2 Previous Low-Power SRAM Designs

2.1 Introduction 5

2.2 Power Dissipation 6

2.2.1 Dynamic Power 6

2.2.2 Leakage Power 7

2.2.3 Short Circuit Power 12

2.3 SRAM Bit-cell Stability and Write-ability..12

2.3.1 Static Noise Margin (SNM)..13

2.3.2 Write Margin (WM)..14

2.3.3 Impact of Variation on SRAM in Low Voltage Differential 6T SRAM...........15

2.4 Previous Read/Write Assist Peripheral Circuit 19

2.4.1 Keeper Tracking Circuit Assist for SRAM Design 19

2.4.2 Charge Pump Circuit Design...22

2.5 Previous Low Voltage SRAM Design..25

2.5.1 SRAM Bit-cell..25

2.6 Previous Low Power Register File Design..32

2.6.1 Register File Bit-cell 32

2.7 Summary 35

Chapter 3 Low Power 2R2W Multi-Port 8Kb SRAM Design

3.1 Introduction 37

3.2 Conventional Dual-Port 8T SRAM 38

3.2.1 Two Kinds of Access Mode in DP-SRAM 38

3.2.2 Write and Read Disturb Issue in 8T DP-SRAM 39

3.2.3 Read/Write Conflict of Dual-port 40

3.3 A New 2R2W Bit-cell 42

3.3.1 Bit-cell Schematic and Layout View 42

3.3.2 Share WBL Structure 44

3.3.3 Bit-interleaving (8 to 1) 45

3.4 Write Assist Technology 46

3.4.1 Negative VVSS 46

3.4.2 Inverter Feedback Loop Cut-off 48

3.5 2R2W Dual-port 8Kb SRAM Design 49

3.5.1 2R2W Multi-port SRAM Schematic 49

3.5.2 Data Transmission Path 50

3.5.3 New Technology Adaptive in 2R2W SRAM Design 51

3.5.4 Test Pattern and Simulation Waveform 52

3.6 Post-layout Simulation 54

3.6.1 Performance 54

3.6.2 Power Consumption 58

3.7 Summary 61

Chapter 4 Low-Power Register File Designs and New Bit-Cell Structure

4.1 Introduction 62

4.2 Previous of Low Power Register-File Design 63

4.2.1 Power Reduction 63

4.2.2 Banked Register File Architecture 64

4.2.3 Tri-state Register File Design 67

4.3 Multi-thread Register File Design 68

4.3.1 Multi-thread Application Design 69

4.3.2 The Parity Protected Multi-thread Register File 70

4.3.3 Thread Switching 72

4.4 Timing Sharing Technology 74

4.4.1 Pervious Work 74

4.4.2 Conflict Issues 76

4.5 This Work 76

4.5.1 Bank Structure 76

4.5.2. Read/Write Slot Controller 82

4.5.3 Switch Data Circuit 84

4.5.4 Thread Switch Control 85

4.5.5 Double Pump Operation 87

4.6 Summary 89

Chapter 5 Low VDD_{MIN} Multi-thread 4R4W Register File Design in TSMC 40nm CMOS Process

5.1 Introduction 90

5.2 4R4W Register File Structure 90
5.2.1 2R2W Register File Unit Cell & Layout View 90
5.2.2 Share WBL Structure 93
5.2.3 Share RBL Structure 93
5.3 Register File Assist Technology..95
5.3.1 Negative VVSS Design 95
5.3.2 Single-end Write Cut-off & Y_Cut for Floating Issues Free 96
5.4 Implementation of Multi-thread 4R4W RF 97
5.4.1 4R4W Register File Floor Plane 97
5.4.2 Design Implementation & Test-flow of Proposed 4R4W Register File 99
5.5 Post-layout Simulation Result 102
5.5.1 Performance..103
5.5.2 Power Consumption..107
5.5.3 Iso-Area SNM Simulation and Comparison..110
5.6 Summary 112

Chapter 6 Conclusion & Future Work

6.1 Conclusions 114
6.2 Future Work 115

List of Figures

Fig. 1.1 Voltage scaling and energy dissipation [1.4] 1
Fig. 1.2 Conventional dual-port 8T SRAM bit-cell 2
Fig. 2.1 Circuit diagram of inverter 6
Fig. 2.2 Leakage current of deep-submicron transistors 7
Fig. 2.3 Gate leakage current paths in a NMOS transistor 8
Fig. 2.4 Leakage current of deep-submicron transistors 9
Fig. 2.5 Conventional silicon dioxide gate dielectric structure compared to a potential high-k dielectric structure 10
Fig. 2.6 Fin-FET structure 11
Fig. 2.7 Standard setup for finding Hold SNM 13
Fig. 2.8 Standard setup for finding Read SNM 14
Fig. 2.9 Write margin of a SRAM bit-cell 14
Fig. 2.10 The read-disturb of 6T SRAM in different process [2.17] 15
Fig. 2.11 The β ratio of 6T SRAM bit-cell 16
Fig. 2.12 6T SRAM SNM loss at low voltages [2.18] 17
Fig. 2.13 6T SRAM write margin [2.18] 18
Fig. 2.14 Read-current distribution [2.18] 18
Fig. 2.15 I_{READ} is less than $I_{leakage}$ from un-accessed cells at low voltage [2.19] 19
Fig. 2.16 A conditional keeper with INV chain [2.20] 20
Fig. 2.17 A current mirror keeper [2.22] 20
Fig. 2.18 Cross couple keeper with INV chain (left) 21
Fig. 2.19 Rate sensing keeper with INV chain (right) 21

Fig. 2.20 Replica bias generator for RSK circuit........22

Fig. 2.21 The variation of the rates across different process corners........22

Fig. 2.22 8T SRAM cell with on-die RWL and WWL boosting [2.23]........23

Fig. 2.23 2SLS can Effective promotion boost ratio [2.23]........23

Fig. 2.24 Different boost frequency effect [2.23]........23

Fig. 2.25 2-step level-shifter reduce I_{LOAD} [2.23]........24

Fig.2.26 Dickson charge pump circuits........24

Fig. 2.27 Ker proposed CP circuit and waveform with four pumping stages........25

Fig. 2.28 (a) The standard 2-port 8T SRAM bit-cell with non-isolated read-port [2.28]
(b) An isolated read-port 7T SRAM bit-cell [2.28]........26

Fig. 2.29 Schematic diagram of the DCO 8T SRAM cell with dual VDD [2.29]........27

Fig. 2.30 DCO 8T cell shows 2x lower leakage at the same read current at 0.9V comparing to SCO 8T cell [2.29]........27

Fig. 2.31 Comparison of leakage components between DCO 8T cell and SCO 8T cell at 0.9V (Q = 1) [2.29]........27

Fig. 2.32 (a) Schematic diagram of the proposed 2-port 6T SRAM bit-cell with shared read and write assist transistors per word [2.30]........28

(b) The VTC and SNM obtained from butterfly curve for the standard ST, 7T and proposed 6T SRAM bit-cells [2.30]........28

Fig. 2.33 A 32-bit word organization of the proposed 2-port 6T SRAM bit-cell to eliminate simultaneous read/write disturbance problem [2.30]........29

Fig 2.34 Simultaneous R/W access issues in word-oriented array [2.30]........29

Fig. 2.35 Schematic of the proposed Z8T SRAM [2.31]........30

Fig. 2.36 Schematic of the proposed 10T SRAM [2.33]........31

Fig. 2.37 Schematic of the proposed 9T SRAM [2.34]........31

Fig. 2.38 Schematic of 8T SRAM Cell [2.35]........31

Fig. 2.39 Cell number on one bit-line is small [2.36] 33

Fig. 2.40 Two bit-cells share one bit-line [2.36] 33

Fig. 2.41 Cell number on one bit-line is small [2.37] 33

Fig. 2.42 IRF integrated memory cell circuit diagram [2.39] 34

Fig. 2.43 Standard 4R2W split to 2 copies of a 2R1W cell [2.41] 35

Fig. 3.1 Conventional 8T dual-port SRAM cell 38

Fig. 3.2 Different-row access mode [3.2] 38

Fig. 3.3 Access in the same row [3.2] 39

Fig. 3.4 Write operation disturbed by dummy read in the same row [3.3] 39

Fig. 3.5 Read operation disturbed by dummy read in the same row [3.3] 40

Fig. 3.6 2P-SRAM for image processing unit [3.6] 41

Fig. 3.7 Delay conflict waveform scheme [3.6] 41

Fig. 3.8 2R2W multi-port SRAM bit-cell 42

Fig. 3.9 2R2W metal layer organization 43

Fig. 3.10 2R2W multi-port SRAM bit-cell layout schematic 43

Fig. 3.11 Two 2R2W multi-port SRAM bit-cell layout view 44

Fig. 3.12 One port writes with no disturb issues 44

Fig. 3.13 Two write the same row with no disturb issues 45

Fig. 3.14 8 to 1 Bit-interleaved SRAM array 46

Fig. 3.15 Bit-interleaving select and Negative VVSS control 46

Fig. 3.16 Negative VVSS control circuit 47

Fig. 3.17 Negative level (a) Different supply voltage (b) Different corner 47

Fig. 3.18 Data write cut off scheme 48

Fig. 3.19 Conventional PMOS cut off structure 48
Fig. 3.20 Schematic of 8kb 2R2W SRAM Chip……………………………………..49
Fig. 3.21 Data transmission path in 2R2W SRAM chip 51
Fig. 3.22 2R2W 8kb SRAM array layout view 52
Fig. 3.23 Test pattern for 2R2W multi-port SRAM Chip 53
Fig. 3.24 Write test function for 2R2W multi-port SRAM chip 53
Fig. 3.25 Read test function for 2R2W multiport SRAM chip 54
Fig. 3.26 Read “0” speed with different voltage...54
Fig. 3.27 Write “0” speed with different voltage 55
Fig. 3.28 Write “1” speed with different voltage 55
Fig. 3.29 (a) Write conflict detect delay 56
Fig. 3.30 (b) Read “0” performance compare with write worst case..........................56
Fig. 3.31 Performance variation for different corner 57
Fig. 3.32 Write “1” speed with different voltage 57
Fig. 3.33 Active power consumption 58
Fig. 3.34 Power consumption with different voltage..58
Fig. 3.35 Power delay product with different voltage...59
Fig. 3.36 2R2W multi-port SRAM pin name 60
Fig. 3.37 2R2W multi-port 8K SRAM test chip 60
Fig. 4.1 Efficiency of power gating circuits..63
Fig. 4.2 Supply switching with ground collapse 64
Fig. 4.3 16-port SRAM architecture with 2-port banks and distributed crossbar........65
Fig. 4.4 Read-data path of the 2-port SRAM within each bank 66

Fig. 4.5 The bank-internal 1-to-8 read-port converter, ... 67
Fig. 4.6 Schematic design of novel register files 67
Fig. 4.7 CMT pipeline efficiency with 32 threads 69
Fig. 4.8 Representation of pipeline and enhancements for multithreading 70
Fig. 4.9 FRF double-pumped pulse clock generator circuit diagram 71
Fig. 4.10 SPICE simulations of both writes "0" and "1" 71
Fig. 4.11 Schematic of two threads switch cell 72
Fig. 4.12 The vector register file cell unit 73
Fig. 4.13 VRF cell capable of holding 2 bits and inter-bit thread decoder 74
Fig. 4.14 (a) Write replica circuit (b) Signal waveforms of this circuit 75
Fig. 4.15 Circuit cross-section of double-pumped write path 76
Fig. 4.16 4R4W Multi-Bank structure 77
Fig. 4.17 1st Data In/Out conflict detector 78
Fig. 4.18 1st Conflict WEN/REN detect waveform 79
Fig. 4.19 4R4W Multi-Bank structure 80
Fig. 4.20 Conflict finite state machine 81
Fig. 4.21 S0/S1 switch control circuit 84
Fig. 4.22 Power7 thread decoder control circuit 85
Fig. 4.23 Bit cell read out architecture 86
Fig. 4.24 A/B port threads switch control 86
Fig. 4.25 Slot detected circuit 87
Fig. 4.27 Register files replica circuit 88
Fig. 4.28 Register files slot control 89

Fig. 5.1 The 2R2W register file unit cell 91

Fig. 5.2 The metal layer implement of 2R2W register file unit cell 92

Fig. 5.3 Layout view of 2R2W register file 92

Fig. 5.4 Conventional dummy read operation in half select cell 93

Fig. 5.5 Share RBL structure in 4R4W register file..94

Fig. 5.6 Share RBL structure no dummy read power consumption.............................94

Fig. 5.7 Share RBL structure in 4R4W register file..95

Fig. 5.8 (a) Negative level in different voltage (b) Negative level in different corner 96

Fig. 5.9 4R4W 2kb multi-thread register file floor plan ……………………………97

Fig. 5.10 Test pattern and simulation waveform result...99

Fig. 5.11 Write post-layout simulation waveform result 100

Fig. 5.12 Read post-layout simulation waveform result..100

Fig. 5.13 Data transmission path in this 4R4W multi-threading register file 101

Fig. 5.14 4R4W multi-thread register file bank layout view 103

Fig. 5.15 Improve technologies of 4R4W register file design 103

Fig. 5.16 Read "0" performance for post-layout simulation.....................................104

Fig. 5.17 Write "0" performance for post-layout simulation 104

Fig. 5.18 Write "1" performance for post-layout simulation....................................105

Fig. 5.19 (a) Address conflict detect circuit delay under wide range 106

(b) Delay of write worst case compare with read "0" access time. 106

Fig. 5.20 Power consumption with different supply voltage 107

Fig. 5.21 Energy consumption with different supply voltage...................................108

Fig. 5.22 Read power consumption with different bit-interleaving structure............109

Fig. 5.23 Leakage power saving different voltage 109

Fig. 5.24 Hold static noise margin simulation 111

Fig. 5.25 Write trip point simulation 111

List of Tables

Table 3.1 Summary of the 8kb 2R2W multi-port spec 50

Table 3.2 Characteristic of 2R2W 8kb SRAM 52

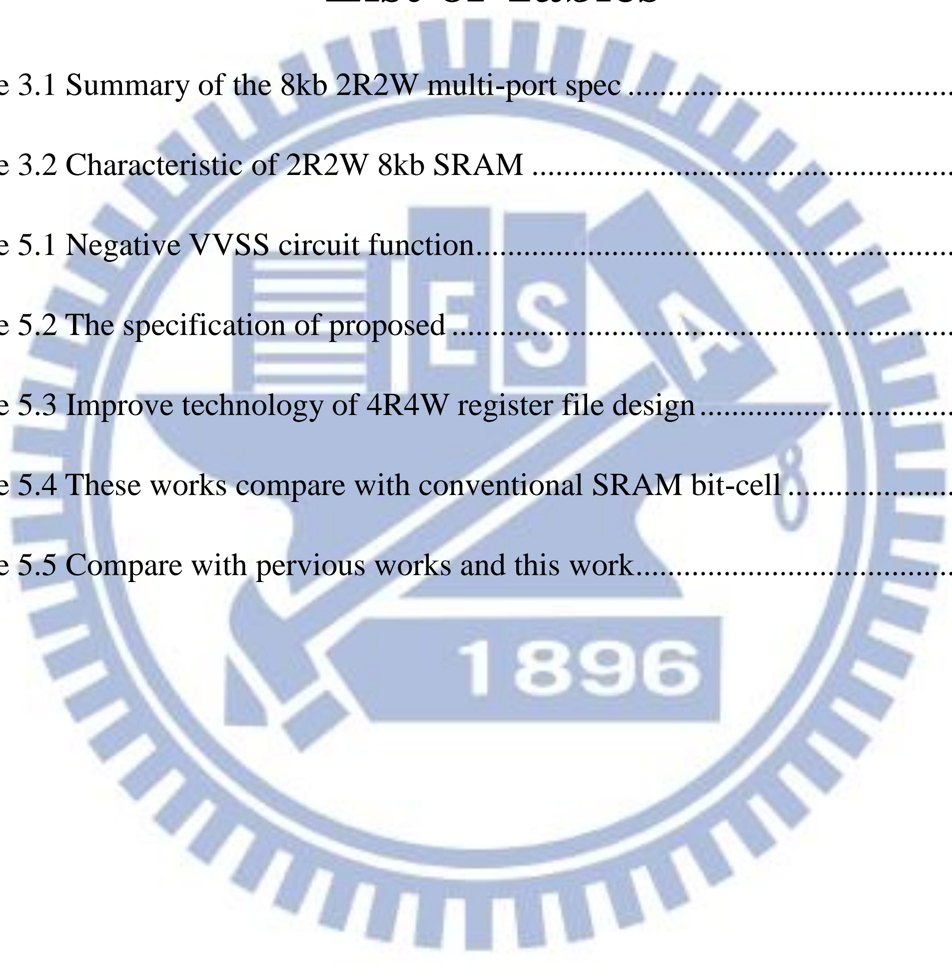
Table 5.1 Negative VVSS circuit function 96

Table 5.2 The specification of proposed 98

Table 5.3 Improve technology of 4R4W register file design 102

Table 5.4 These works compare with conventional SRAM bit-cell 110

Table 5.5 Compare with pervious works and this work 112



Chapter 1

Introduction

1.1 Background

Low power design for portable device such as cell phone, wireless device and notebook are rapidly growing in these years. A simple and effective way to reduce energy is to scale down supply voltage. Reduction of energy consumption is desirable in microprocessors to enable longer battery life and adequate heat dissipation. The active power saving is quadratic and leakage power reduction is linear [1.1]–[1.3]. The total energy consumption equation is showed in (1.1)

$$P_{total} = P_{dynamic} + P_{leakage} + P_{short-current} \quad (1.1)$$

Where $P_{dynamic} = \alpha \times f \times C \times VDD^2, P_{leakage} = VDD \times I_{leakage}, P_{sc} = VDD \times I_{mean}$

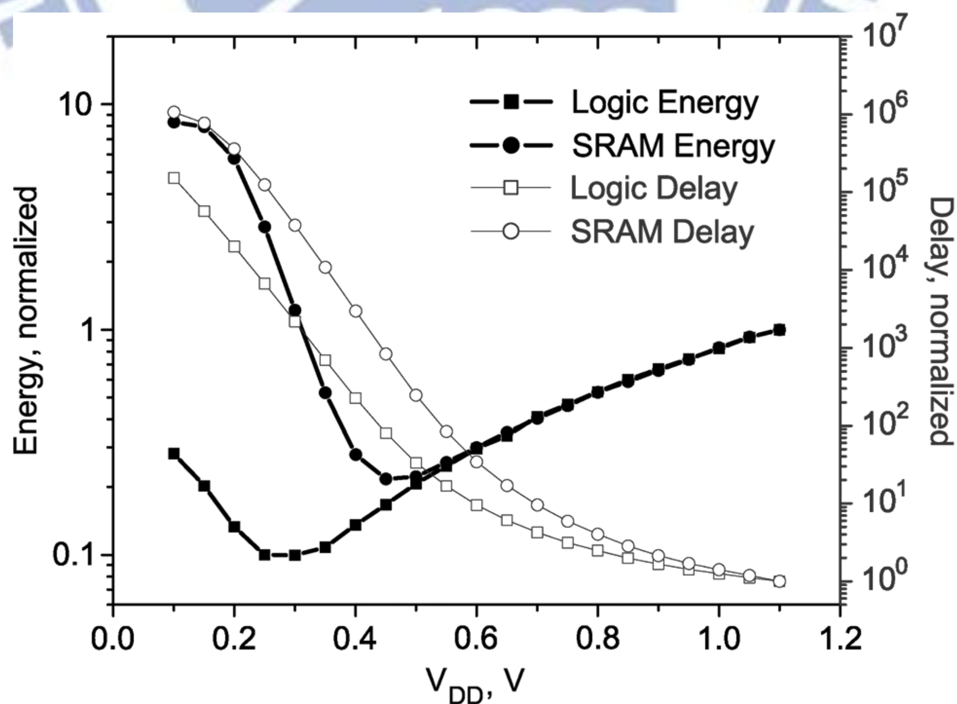


Fig. 1.1 Voltage scaling and energy dissipation [1.4]

Fig. 1.1 shows the SRAM Min. energy point not in the sub threshold but in near threshold region. Operation in sub threshold region, although power reduce only linear, delay rises significantly, and power delay product will not small anymore. This reason make the Min. energy point is shift to neat-threshold region.

1.2 Challenges

Conventional dual-port (Fig. 1.2) or multi-port design is power hungry by accessing many ports parallel at the same time. High frequency and multi-port read out by register file often domain the whole chip power consumption. In order to gain high data transmission bandwidth, this design is needed and not to be lack. Applying for portable device, low power design is very important which can save obvious power dissipation, and enables longer battery life.

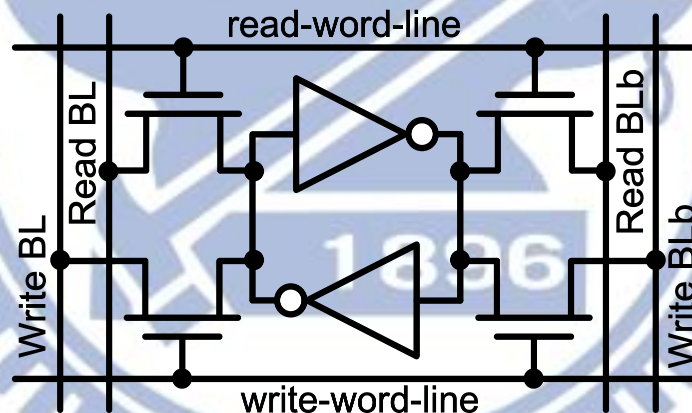


Fig. 1.2 Conventional dual-port 8T SRAM bit-cell

Voltage scale down is not easy in conventional dual-port 8T design because read disturb problem and read/write conflict issues will degrade cell stability. Not only cell stability problem, write–ability is another big challenge for low voltage operation. When supplying voltage scaling down, area of dual 8T must to enlarge cell size for more stability. Besides, as the supply voltage is reduced, the effect of Ion/Ioff ratio is smaller than operation in super-threshold region. Driver current normally degrades several orders of magnitude compare MOS active on strong inversion.

In addition to CMOS technology process scaling down, there is more physic solid state effect generated on device. Moreover, these minimum geometry transistors are vulnerable to inter-die as well as intra-die process variations. Intra-die process variation includes random dopant fluctuation (RDF), line edge roughness (LER). This may lead to threshold voltage mismatch between the adjacent transistors in a memory cell giving asymmetrical characteristics [1.5] [1.6].

1.3 Motivation

Try to do a low power and high retention SRAM cell [1.7], and no conflict distribute problem is this work goal. Conventional dual-port already not suit for novel technology process. A new structure must be proposed to solve distribute problem and conflict problem. Operation under near-threshold voltage to gain the Min. energy consumption, and this cell will apply to low power device such as wireless senor or mobile phone. In addition to low power and high reliability, high bandwidth support is another primary circuit concern. Multi-thread and multi-bank structure design may include in this work for performance improve.

Except scaling down the supply voltage, peripheral circuit will use low power design such as power gating, CLK gating, and DVS technology may added for active or standby power reduction [1.8]. Boost or Negative circuit can improve write-ability when operation under low voltage. Previous work like cut off feedback in single-end design also can gain more write ability. In this design, power reduction is more important than high speed operation.

For more robust consideration, bit-interleaving structure can eliminate soft error rate damage the SRAM bit-cell. Leakage problem in BL can't be ignored in low voltage operation region. If leakage current is too over reliability, read operation will fail by sensing logic.

1.4 Thesis Organization

Following is the main contents of this thesis. In Chap 2 we will discuss the recent work about low power SRAM design. A conventional 6T SRAM basic operation and stability are introduced at first. After that low power SRAM assists circuit design, multi-port SRAM and register file design are discuss step by step. Chapter 3 shows conventional dual-port operation conflict problem and read disturb issues. A new 2R2W multi-port SRAM structure is proposed. Share write bitline and X and Y cut control line can do bit-interleaving structure and no need any others periphery circuit. In Chap. 4, register file with multi–thread and double pump technology introduced at beginning. New technology “Data Slot Switch” and conflict detect circuit can help no disturb issue. In Chap 5, a new share read bitline is proposed and reduce dummy read of bit-interleaving structure. Active power saves by share RBL and leakage power reduces by keep RWL high in standby mode. In the end, Chapter 6 finally concludes this thesis.

Chapter 2
Previous Low-Power SRAM Designs
2.1 Introduction

In recent microprocessors, the capacity of on-chip memory is rapidly increasing to improve overall performance. According to ITRS roadmap in 2002 [2.1] [2.2], memory chip will occupy 90% of chip area in 2013. In such a memory rich chip, the leakage current of an SRAM, which comprises the vast majority of on-chip transistors, dominates the standby current because leakage power is proportional to the number of transistors. Thus, it becomes important to focus on SRAM standby leakage current reduction for ultra-low power application.

Low power, minimum transistor count and fast access SRAM is essential for embedded multimedia and communication applications realized using system on a chip technology. Hence, simultaneous or parallel read/write (R/W) access multi-port SRAM bit cells are widely employed in such embedded systems. Multi-port has many advantages like high performance and high bandwidth, but it also consumers more percentages of power and area.

This chapter begins with the analysis of power dissipation of SRAM circuit and technique for leakage reduction will be shown in section 2.2. In section 2.3, stability issues of SRAM cell, including hold stability, read stability, and write ability will be defined and the impact of variation on SRAM in low voltage will be presented. In section 2.4, 2.5 and 2.6, Conventional dual port SRAM and Multi-port SRAM cell are showed. Finally, In 2.7 the previous Multi-port register file cell design and peripheral circuit technology will be described.

2.2 Power Dissipation

This chapter begins with an analysis of power dissipation of CMOS circuit and circuit technique for power dissipation. Power dissipation combines with dynamic power ($P_{dynamic}$), leakage power ($P_{leakage}$), and short circuit power ($P_{short\text{-}circuit}$). Power could be expressed as, where $P_{dynamic} = \alpha C_L V_{DD}^2 f$, $P_{leakage} = V_{DD} I_{leakage}$, and $P_{short\text{-}circuit} = I_{mean} V_{DD}$

$$P_{total} = P_{dynamic} + P_{leakage} + P_{short-circuit} \quad (2.1)$$

2.2.1 Dynamic Power

Fig. 2.1 show a CMOS inverter, the average dynamic power dissipation can be obtained by summarizing the average dynamic power of N/P MOS. The cause of dynamic power is logic transition of CMOS circuits which charges or discharges its load capacitance and parasitic capacitance (C_L). As can be seen in (2.1), the dynamic power dissipation is direct proportion to switching activity factor (α), capacitance load (C_L), squire of supply voltage (V_{DD}^2), and operating frequency (f).

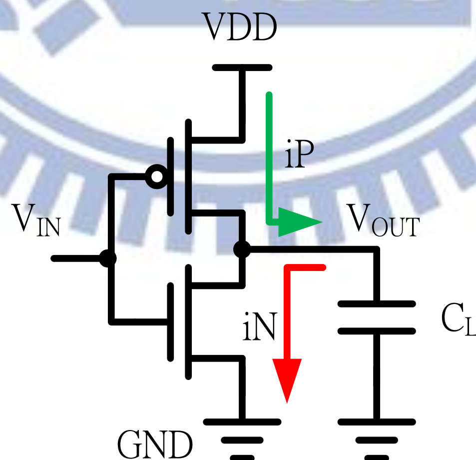


Fig. 2.1 Circuit diagram of inverter

2.2.2 Leakage Power

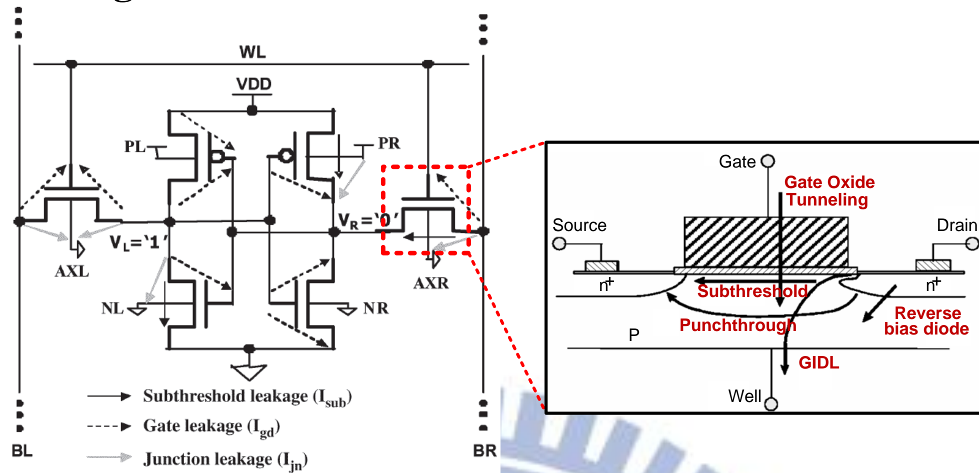


Fig. 2.2 Leakage current of deep-submicron transistors

In advanced CMOS technologies, embedded SRAM leakage current becomes dominant compared to the dynamic current. The majority of SRAM macro leakage current is from its bit cell array [2.3]. Leakage current is composed of reverse-biased junction leakage current (I_{REV}), gate induced drain leakage (I_{GIDL}), gate direct-tunneling leakage (I_G), and sub-threshold leakage (I_{SUB}) in a CMOS transistor [2.4] [2.5].

Fig. 2.2 shows reverse-biased junction leakage, sub-threshold leakage, gate direct-tunneling leakage, injection of hot carriers from substrate to gate oxide, gate induced drain leakage, and punch through leakage in the deep scaling transistor.

Junction Leakage

In Fig. 2.3, leakage in reverse biased transistors and diodes includes the effects of carrier generation, related to residual damage density and location relative to the junction boundary, as well as structure and bias dependent effects of gate oxide leakage, band-to-band tunneling at the drain junction and thermionic emission from metal contacts. All of these effects depend on process conditions, through dependence on dopant activation and profile shape, junction location and local electric fields.

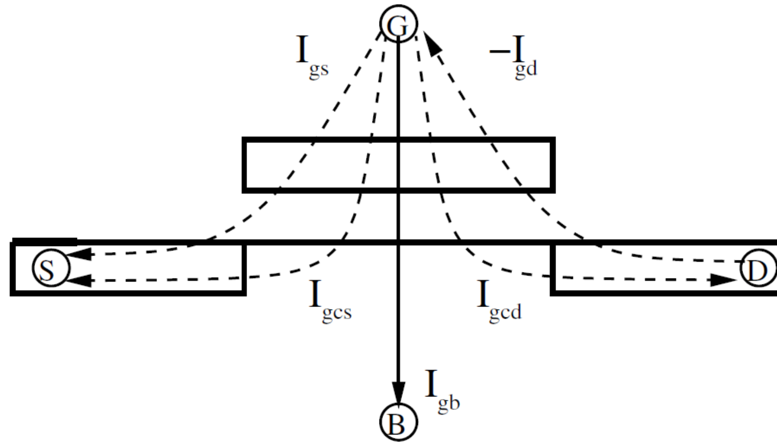


Fig. 2.3 Gate leakage current paths in a NMOS transistor

In the steady-state ON region both the gate and drain of the device are held at high with the source being grounded. In this state a well-formed channel exists and three separate components of the gate tunneling current Igs, Igcs and Igcd are active. The component from gate to drain overlap (Igd) is absent due to the almost zero electric field in that region of the oxide. The overall current flow is from gate to source and channel, opposite to the flow in the OFF state. In the steady-state OFF region both gate and source are at ground while the drain is at high (VDD) voltage. Since no channel is formed in this condition, the only active component is Igd [2.6].

Gate-induced drain leakage (GIDL):

As the electric field in and around the gated p-n junction is increased by the applied gate voltage, all the high-field effects, such as avalanche multiplication and band-to-band tunneling, can increase very dramatically (Fig. 2.4). Thus, the leakage current of a reverse-biased gated diode can increase dramatically when the gate voltage begins to cause field crowding in and around the junction region.

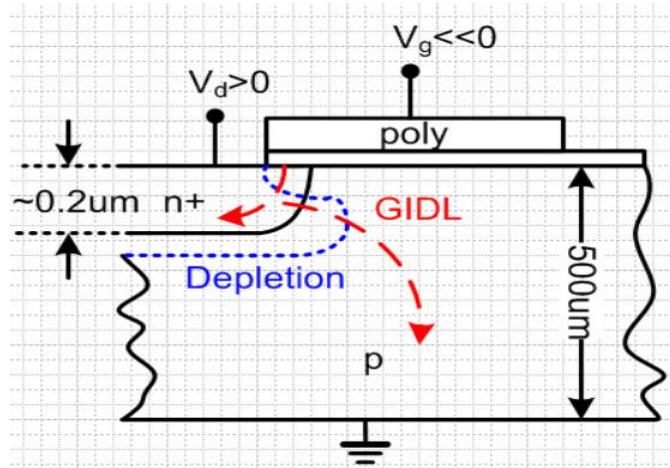


Fig. 2.4 Leakage current of deep-submicron transistors

Sub threshold Leakage

When gate voltage is below the threshold voltage, sub-threshold leakage or weak inversion current occurs between source and drain. For example, an off state inverter, although the V_{gs} of the NMOS is 0V, there is a light current (leakage) flowing from the drain to source due to the voltage V_{DD} across V_{ds} [2.7].

Sub-threshold behavior can be modeled physically as show in the following [2.8]

$$I_{ds} = \mu \frac{W}{L}\left(\frac{kT}{q}\right)^2 C_{sth} e^{\frac{V_g - V_T + \eta V_{ds}}{mkT/q}}\left(1 - e^{-\frac{V_{ds}}{kT/q}}\right), m = 1 + \frac{C_{sth}}{C_{ox}} \qquad (2.2)$$

Where W and L denote the transistor width and length, μ denotes the carrier mobility, $C_{sth} = C_{dep} = C_{it}$ denotes the summation of the depletion region capacitance and the interface trap capacitance both per unit area of the MOS gate, η is the drain induce barrier lowering (DIBL) coefficient, and C_{ox} denote the gate input capacitance per unit area of the MOS gate.

Sub-threshold leakage increases exponentially with the reduction of the threshold voltage and DIBL would lower threshold make leakage even worse. On the other hand, sub-threshold can be drop with increasing the threshold voltage. In low power technology we can use high V_{th} technology transistor to reduce sub-threshold leakage in off state.

High-K Metal Gate

In order to reduce gate leakage, a new material is used for replace the conventional SiO2. Silicon dioxide has been used as a gate oxide material for decades. As transistors have decreased in size, the thickness of the silicon dioxide gate dielectric has steadily decreased to increase the gate capacitance and thereby drive current, raising device performance. As the thickness scales below 2 nm, leakage currents due to tunneling increase drastically, leading to high power consumption and reduced device reliability (Fig. 2.5). Replacing the silicon dioxide gate dielectric with a high-κ material allows increased gate capacitance without the associated leakage effects. The 2.3 rule showed that we can add high k material and extended thickness to get the equal capacitive. By thickness oxide, leakage problem can reduce significantly [2.9].

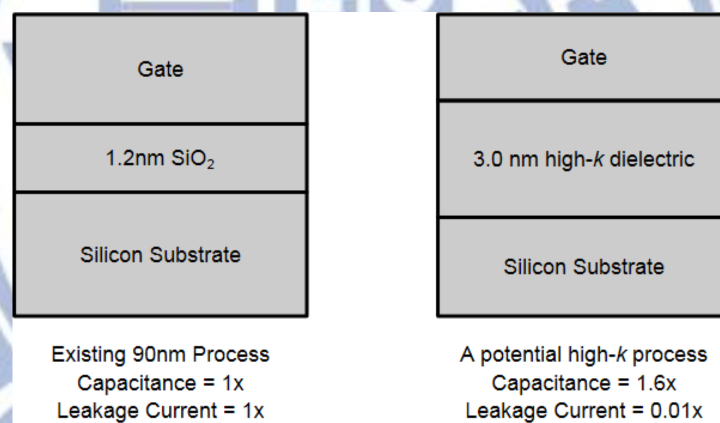


Fig. 2.5 Conventional silicon dioxide gate dielectric structure compared to a potential high-k dielectric structure

$$C = \frac{k\epsilon_0 A}{t} \qquad (2.3)$$

- A is the capacitor area
- κ is the relative dielectric constant of the material (3.9 for silicon dioxide)
- ε_0 is the permittivity of free space
- t is the thickness of the capacitor oxide insulator

Fin FET Structure

Fig.2.6 shows Fin FET device has especially faster switching times and higher current density. Not like conventional MOS structure, a new better gate control device is developed by IBM. Vertical gate has more area cover the channel, so better control ability is approach. Due to its superior gate control, electrostatic integrity, and variability, Fin FET has demonstrated satisfactory scalability and feasibility for mass production of post-22-nm technology node [2.10] [2.11].

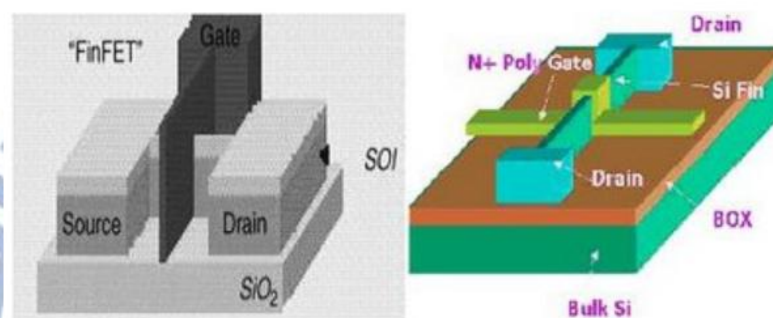


Fig. 2.6 Fin-FET structure

Punch-through Leakage

Finally, in short-channel devices, due to the proximity of the drain and the source, the depletion regions at the drain-substrate and source-substrate junctions extend into the channel. As the channel length is reduced, if the doping is kept constant, the separation between the depletion region boundaries decreases. An increase in the reverse bias across the junctions (with increase in V_{DS}) also pushes the junctions nearer to each other. As the combination of channel length and reverse bias leads to the merging of the depletion regions, punch through leakage occurs.

Punch through will bring a high current, and make the device short down. Hot and power dissipation by high current, so designer should very care about this effect.

2.2.3 Short Circuit Power

When CMOS switch frequently, a path from vdd to gnd will short together. This dc path makes external power consumption. Short circuit power can be expressed as rule (2.4). I_{mean} is the mean value of the short circuit current [2.12].

On the circuit-level, there have been a number of articles describing the short circuit power. From the short circuit power articles by Veendrick [2.13], and Hedenstierna and Jeppson [2.14], these power dissipation rules are showed below.

$$P_{short\text{-}circuit}=I_{mean} \times V_{DD} \qquad (2.4)$$

$$P_{short-circuit} = \frac{\beta}{12}(V_{DD} - 2V_t)^3 f\tau \qquad (2.5)$$

P: The device transistor conductance

τ: The ramp time

β: The gain factor of a transistor,

f: The operating frequency

2.3 SRAM Bit-cell Stability and Write-ability

When CMOS technology process is scaling down, process variation is become more and more important. PVT variation is the major effect on cell stability, such as global variation and local variation. Therefore, how to use the simulation information to accurate the true threshold is very important. The worst cast must be consider and usually use Monte Carlo simulation to detect it. The following of this section will state the most widely adopted SRAM cell stability definition.

2.3.1 Static Noise Margin (SNM)

The best common way to measure the stability of cross-coupled inverters is the static noise margin (SNM). Hold static noise margin is defined as the maximum value of static DC voltage noise which can be tolerated by the SRAM bit-cell without flipping the storage node when word-line turns off. Fig. 2.7 shows the normal test Hold SNM simulation in 6T SRAM cell. Give a two noise in the Q and Qb then find max voltage noise can maintain the storage data of the SRAM. In this case, WL is zero and two BL keep high [2.15].

Fig. 2.8 shows the standard setup for modeling Read SNM. Compare with HSNM mode, in this case WL is turn and simulation read operation. The node “0” will raise a little voltage because of the voltage dividing effect between the pass transistor and pull-down transistor. Once the disturb voltage rise near to the trip point of the inverter, data will be flipped. The curve is small than HSNM because read distribute issues and it reduce node stability significantly. Fig. 2.7 and Fig. 2.8 also show the example of butterfly curves during hold and read, revealing the degradation in SNM during read.

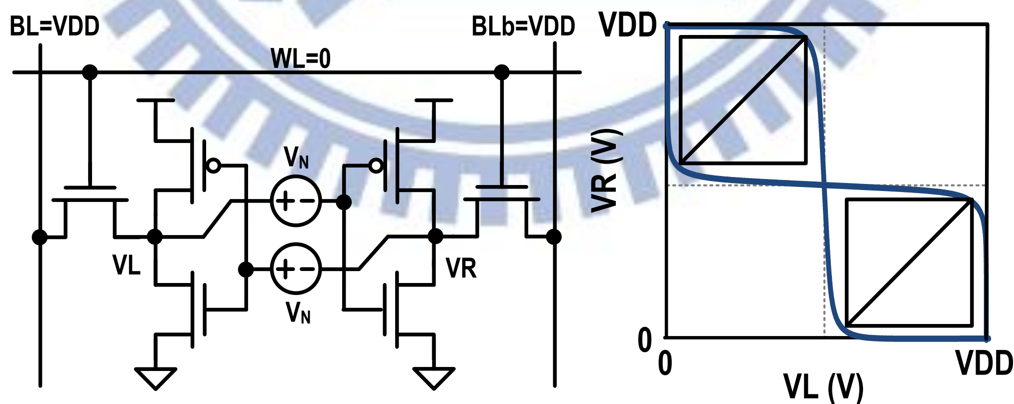


Fig. 2.7 Standard setup for finding Hold SNM

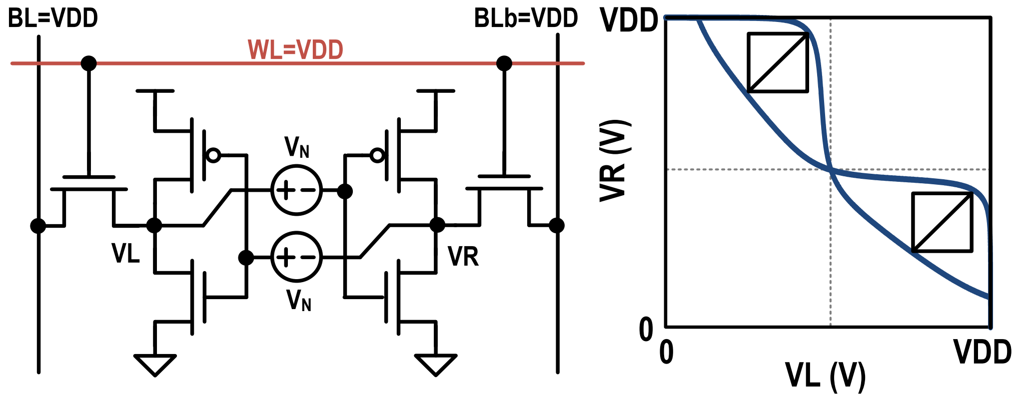


Fig. 2.8 Standard setup for finding Read SNM

2.3.2 Write Margin (WM)

There are many way to measure the write ability of SRAM bit-cell, the simple one is find the write trip point (WTP). Write margin is defined as$V_{DD} - MIN[V(WWL)]$. $MIN[V(WWL)]$ is the minimum write-word-line voltage required for flipping the bit-cell. In this write margin test mode, sweep WL voltage from VDD to Zero. The higher write margin, the easier the data is written into bit-cell. Fig. 2.9 shows a corresponding example of finding write margin. The write margin is defined as the V_{DD} - VWL value at the point when VR and VL flip. The write margin value and variation is a function of the cell design, SRAM array size and process variation. A cell is considered not writeable if the worst-case write margin becomes lower than the ground potential.

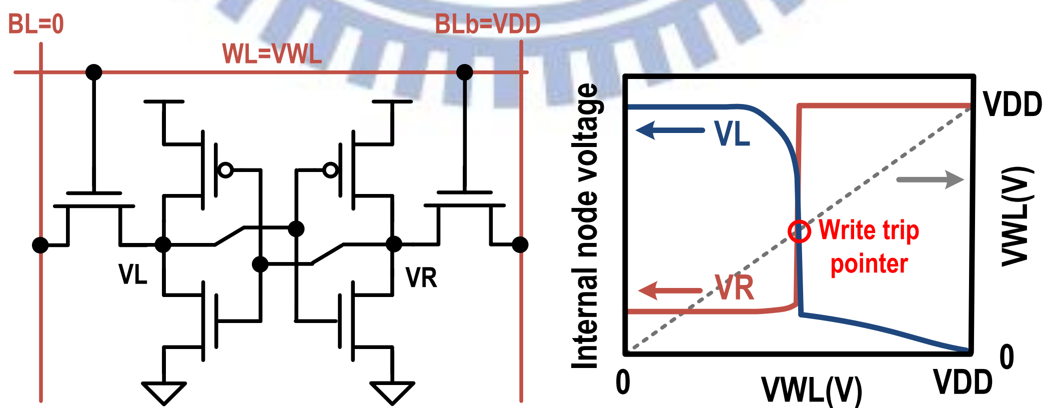


Fig. 2.9 Write margin of a SRAM bit-cell

2.3.3 Impact of Variation on SRAM in Low Voltage Differential 6T SRAM

6T bit-cell is not applied for process scaling down, and also not suitable in low-voltage operation. If 6T cell want to operate under novel technology, area of N/PMOS has to enlarge to gain more W/R ability. Process problem to 6T cell is very sensitive, such as random dopant fluctuation (RDF) and line edge roughness (LER). This may result in the threshold voltage mismatch between the adjacent transistors in memory cell [2.16] [2.17].

Half select disturbs Failure

In Nano-device scaling down, threshold voltage variation is become larger. By process variation NMOS Vt is not a constant value anymore, if disturb voltage is larger than bit-cell trip voltage, the data will flip and error happened. Conventional 6T with bit-interleaving structure will have half select problem. Fig. 2.10 shows the half select disturbs failure and waveform, if pull-down NMOS Vt is too high, and access NMOS Vt is low. Current is stack on the Qb, a probability data will flip by this current path.

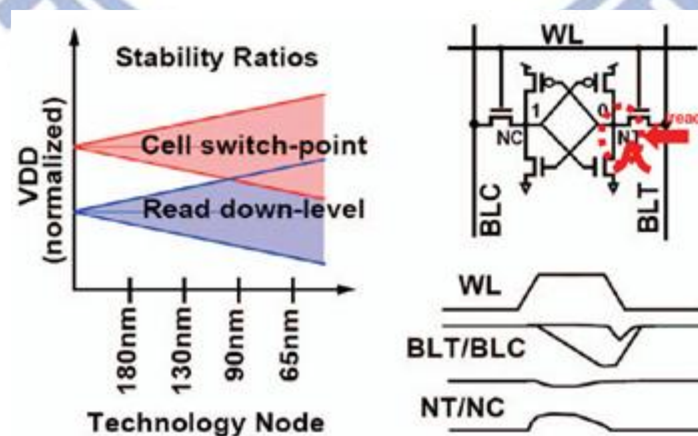


Fig. 2.10 The read-disturb of 6T SRAM in different process [2.17]

Read/Write conflict issues

In this configuration, both read and write accesses are opposite making it highly difficult to overcome the severe effect of variation and manufacturing defects. Fig. 2.11 shows the β ratio of 6T SRAM bit-cell and the β ratio conflict will be described afterward.

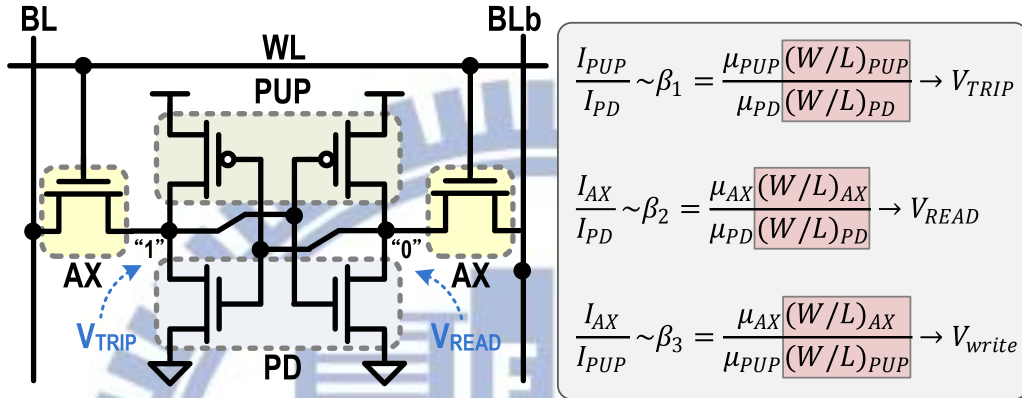


Fig. 2.11 The β ratio of 6T SRAM bit-cell

During read access the cell must remain bi-stable to ensure that both data logic value can be held and read without being upset by read disturb that occur at the internal nodes. In order to facilitate read and minimize read disturb, the β_2 ratio should be small enough by strong *PD* NMOS and weak *AX* NMOS. During write access the cell should be made mono-stable to write the desired data. For improving writability, the β_3 ratio must be large by strong AX NMOS and weak PUP PMOS.

For improving writability and minimizing read disturb simultaneously, the transistor can be sized as *PD* > *AX* > *PUP*. However, it would degrade the β_1 ratio hence the V_{TRIP} result in poor read SNM. Therefore, these three β ratios are conflict to each other, simply sizing could not solve 6T SRAM failures.

Hold and Read Failure

Hold failure happens if the destruction of the cell content in the standby mode at a low supply voltage. Therefore higher trip point of back-to-back makes the cell easier to flip, thereby increasing the hold failure probability. As shown in Fig. , it is preserved to

very low voltages and will form the basis for several of the ultra-low voltage bit-cell design described in section 2.5 and 2.6.

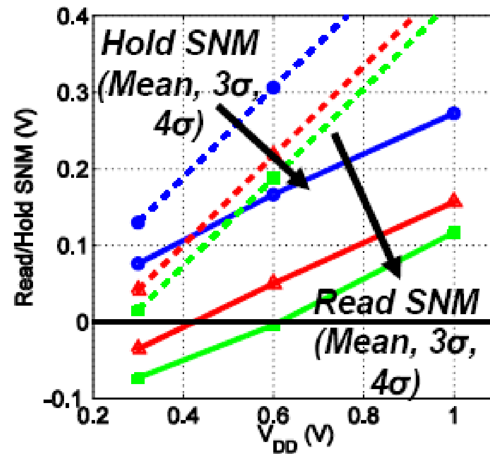


Fig. 2.12 6T SRAM SNM loss at low voltages [2.18]

If the data stored in an SRAM cell flips during reading, there is a read failure. If the voltage rise at the node storing "0" and higher than the trip point of the back-to-back inverter, then the data stored in the cell would flip over. Fig. 2.12 shows that the 6T SRAM bit-cell fails to operate at low voltages because of reduced signal levels and increased variation. At low voltages, the read SNM is negative, indicating loss of stability.

Write Failure

If the data stored in an SRAM cell can't be flip during writing, there is a write failure. While writing "0" to node storing "1," the voltage at the node need to be discharged below the trip point of the back-to-back inverter. As shown in Fig 2.13, it is also lost at low voltage, where a positive value, in this case, indicates write failures.

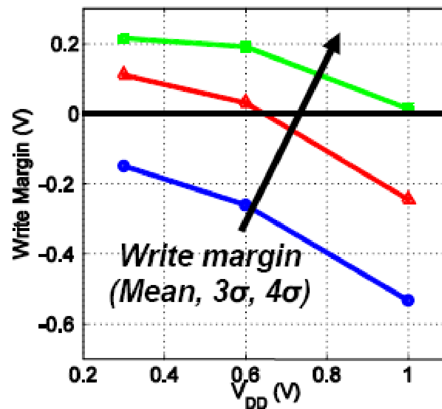


Fig. 2.13 6T SRAM write margin [2.18]

Access Failure

If the voltage difference between the two bit-lines (dual-end) or the voltage drop of the single bit-line (single-end) can't be sensed by the sense amplifier during the access time, there is an access failure. The cause of access failure can be ascribed to read-current degradation and data-dependent bit-line leakage.

The cell read-current, I_{READ}, is the current sunk from the pre-charged bit-lines during a read access when the access devices are enabled. At ultra-low voltages, we expect a significantly reduced read-current because of the lower gate-drive voltage. However, the increased effect of threshold voltage variation severely degrades the weak cell read-current even further. Fig. 2.14 normalizes the read-current distribution by the mean read-current to highlight just the further degradation due to variation.

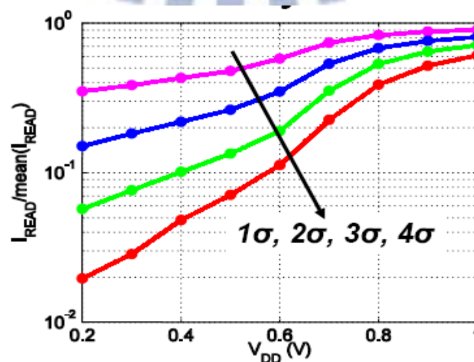


Fig. 2.14 Read-current distribution [2.18]

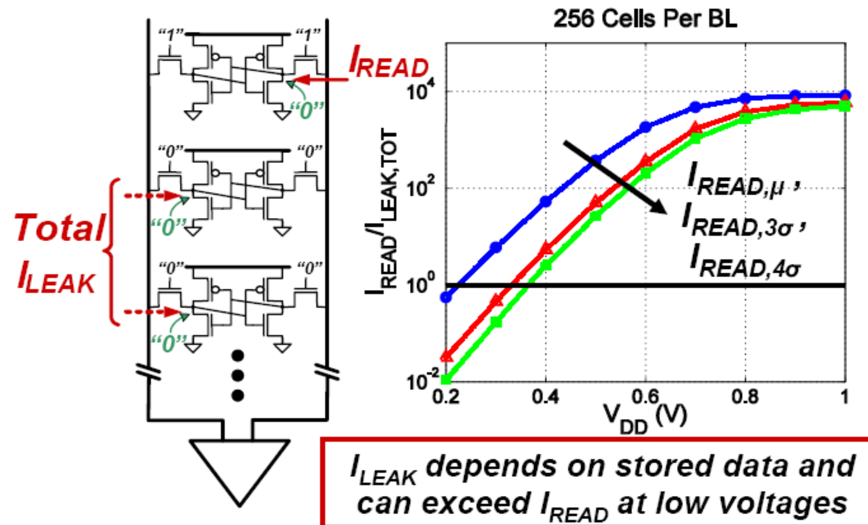


Fig. 2.15 I_{READ} is less than $I_{leakage}$ from un-accessed cells at low voltage [2.19]

An implied consequence of the reduced read-current is that the aggregate leakage currents from the un-accessed cells on the same bit-lines can make conventional data sensing impossible. Because of the reduced I_{ON}-to-I_{OFF} ratio and severe degradation from read-current variation, these can exceed the actual read-current of the accessed cell. Fig. 2.15 shows $I_{READ}/I_{LEAK,TOT}$ of 256-row SRAM array loss of functionality at low voltages. At ultra-low voltage the bit-line leakage exceeds the read signal, making the accessed data indecipherable.

2.4 Previous Read/Write Assist Peripheral Circuit

2.4.1 Keeper Tracking Circuit Assist for SRAM Design

Wide or structures are typically used in the read path of register files, L1 caches, match lines of TCAMs, flash memories and PLAs. In most of the applications the worst case requirement would be to sense the difference between the leakage state where all the pull-down legs are leaky and the ON state where only one of the legs is ON. The increase in the variability and magnitude of the leakage current has become a major bottleneck in realizing such wide OR gates [2.21] [2.22].

In the conventional design, the keeper being PMOS and it does not track the

leakage currents in the pull-down NMOS logic for the FNSP and SNFP corners. These results in performance degradation, higher short-circuit power dissipation and limit the number of pull down legs.

An ideal keeper is expected to have minimum contention, good noise robustness, good process tracking, less power and area overhead and should support wide fan-in gates.

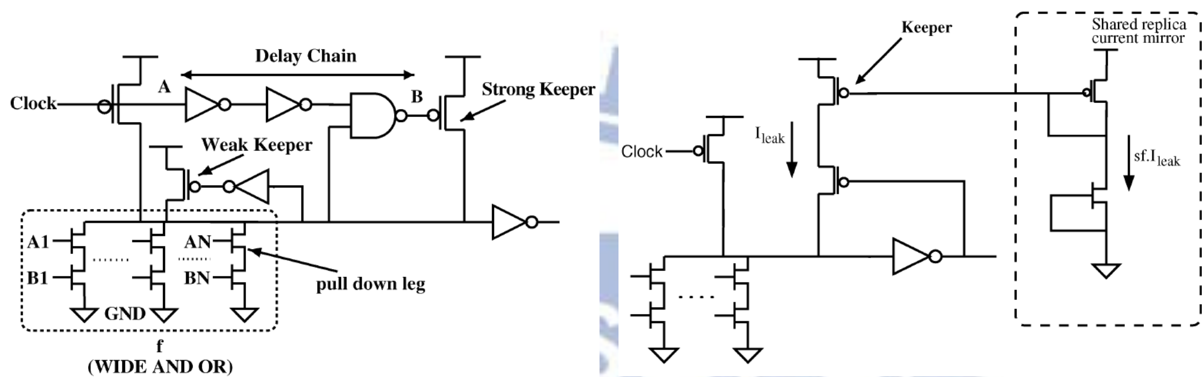


Fig. 2.16 A conditional keeper with INV chain [2.20]

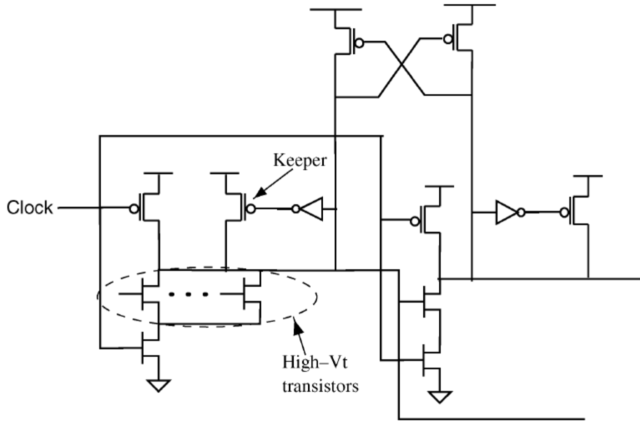
Fig. 2.17 A current mirror keeper [2.22]

Conditional keeper (CKP)

A weak keeper holds the state of the dynamic node during the transition window and a strong keeper is conditionally activated based on the state of the dynamic node after a certain delay Fig. 2.16 This reduces contention during the evaluation period, thereby enabling high speed and reducing the short circuit power dissipation.

Current mirror keeper (LCR)

Current mirror-based keeper technique Fig. 2.17 was proposed for better process tracking. This technique provides excellent tracking of the delay, and the contention is still high because the keeper is strongly ON during the beginning of the evaluation phase. Further the replica transistor does not track the leakage due to noise (as Vgs=0) and DIBL (as the drain voltage of the replica NMOS varies across process corners) in the pull-down NMOS logic.



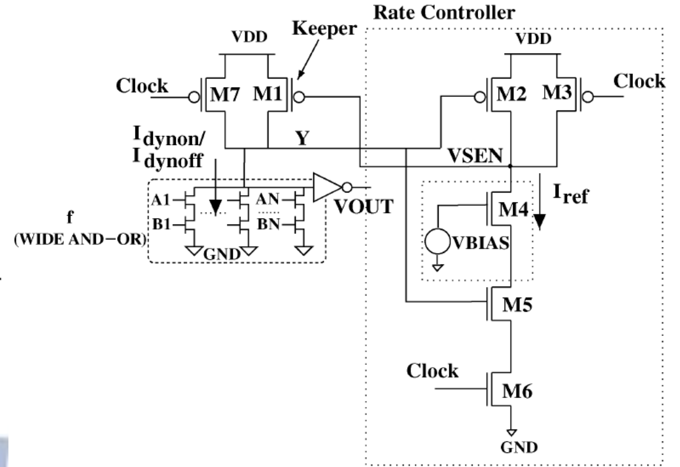


Fig. 2.18 Cross couple keeper with INV chain (left)

Fig. 2.19 Rate sensing keeper with INV chain (right)

Cross couple keeper (CSK)

Fig. 2.18 is based on cross coupled structures, and has two switch steps. Using a cross coupled structure based on SCL and feedback PMOS transistors to provide additional noise immunity to the dynamic node without much performance degradation.

Rate sensing keeper (RSK)

Fig. 2.19 is the rate sensing keeper technique works based on the difference in the rate of change of voltage at the dynamic node of the gate during the ON (R_{dynon}) and the leakage (R_{dyoff}) condition. A reference rate (R_{ref}), which is the average of the two rates, is used to control the state of the keeper. The fact that the keeper is OFF during the start of the evaluation phase and the adaptive control of the keeper strength based on the process corner helps RSK to achieve higher speed and better tracking, respectively.

$$\begin{aligned} \text{Keeper ON} : \frac{I_{\text{dyn}}}{C_{\text{dyn}}} < \frac{I_{\text{ref}}}{C_{\text{ref}}} \\ \text{Keeper OFF} : \frac{I_{\text{dyn}}}{C_{\text{dyn}}} > \frac{I_{\text{ref}}}{C_{\text{ref}}} \end{aligned} \tag{2.6}$$

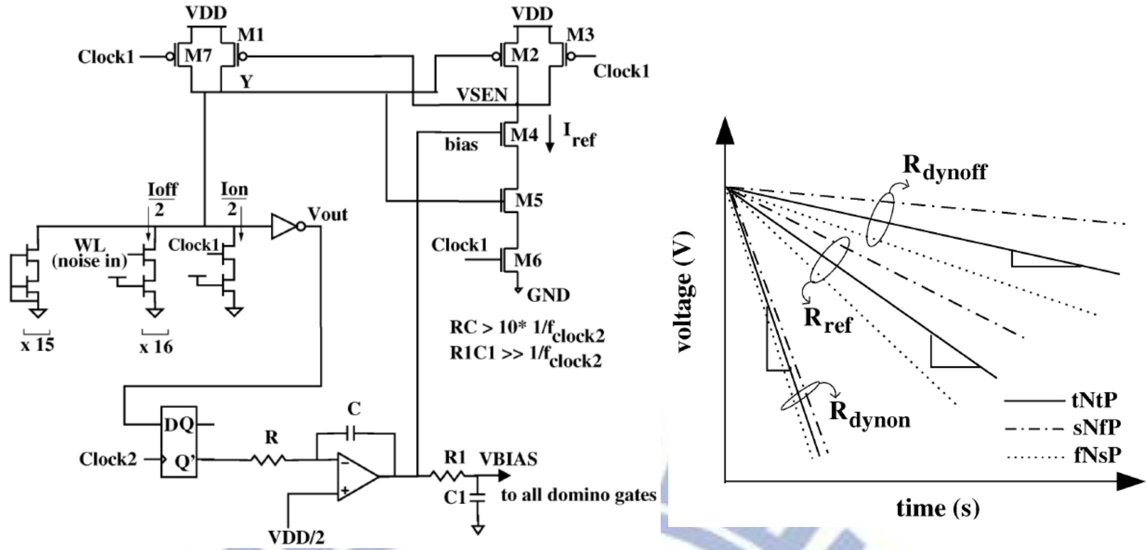


Fig. 2.20 Replica bias generator for RSK circuit

Fig. 2.21 The variation of the rates across different process corners

$$R_{\text{ref}} = \frac{\frac{I_{\text{on}}}{2}}{C_{\text{dyn}}} + \frac{\frac{I_{\text{off}}}{2}}{C_{\text{dyn}}} = \frac{R_{\text{on}} + R_{\text{off}}}{2} \quad (2.7)$$

2.4.2 Charge Pump Circuit Design

In this section, I will to introduce about the boosting method for SRAM assist. How to do can gain the more efficiency is very important and save extra power dissipation. While the demand for aggressive low-power is ever increasing thus demanding a lower V_{MIN} of the SRAM cell. Before, people may do a SRAM cell sizing in order to scale down the VMIN, but only 10% total V_{MIN} reduction is attained and upsizing at a cost of ~25% increase in array area. Therefore, RD and WR circuit assists that can achieve VMIN reduction at a minimal area impact are necessary.

Boosting RWL enables larger read "ON" current without forcing a larger PMOS keeper. Boosting WWL helps WR V_{MIN} for 2 reasons – improving contention without upsizing Nx (or lowering its V_{TH}), and improving completion by writing a "1" from

the other side. At iso-array area, increase on-die boosting achieves twice as much V_{MIN} reduction as simple cell upsizing Fig. 2.21.

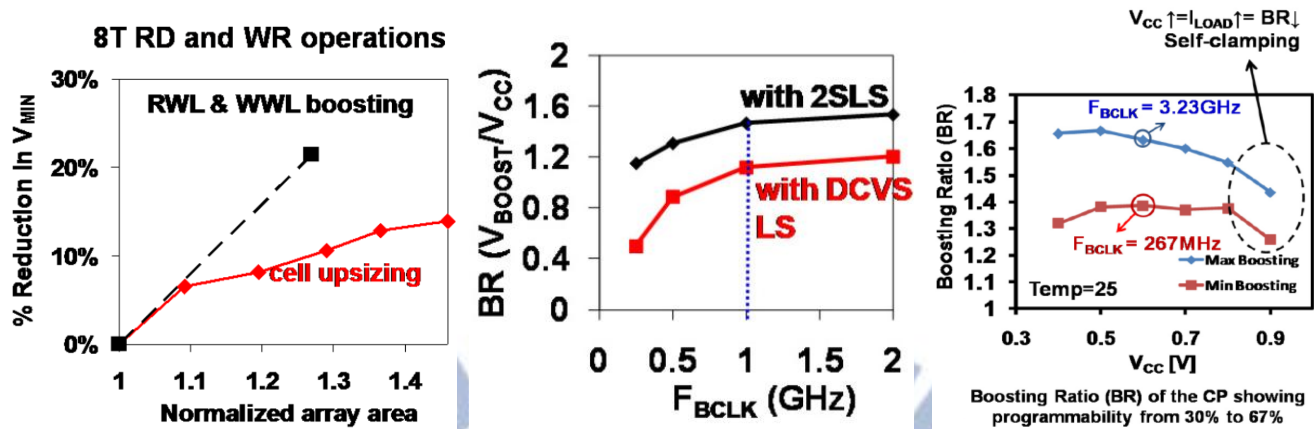


Fig. 2.22 8T SRAM cell with on-die RWL and WWL boosting [2.23]

Fig. 2.23 2SLS can Effective promotion boost ratio [2.23]

Fig. 2.24 Different boost frequency effect [2.23]

Boost Ratio Optimum

Ideal boosting ratio (BR = V_{BOOST}/Vcc) under no load current (I_{LOAD}) is 2VCC. Actual BR is lower, however, as determined by I_{LOAD} from all active & inactive level-shifters, boosting clock frequency (F_{BCLK}) Fig. 2.22, 2.23, and boosting capacitance (C_{CP}). At a given phase of BCLK, one of the two CP paths alternately supply charge to the V_{BOOST} rail. In order to maintain gate oxide & junction reliability of devices connected to V_{BOOST}, CP is enabled (i.e. BCLK is toggling) if [BR x VCC < V_{MAX}] is met. The CP is turned off otherwise, with transistor MX turned on to short V_{BOOST} to VCC [2.23].

The 2SLS minimizes dynamic I_{LOAD} current that needs to be supplied by the CP. Fig. 2.24 unlike conventional (DCVS) LS where a “0”-to-V_{BOOST} transition is all supplied by the V_{BOOST} rail, the 2SLS performs this transition in 2 steps Fig. 2.24. In the first step, “0”-to- VCC is supplied by MP1 at which point MP2 kicks in to supply the remaining VCC to- V_{BOOST}. The circuit is proposed in [2.24].

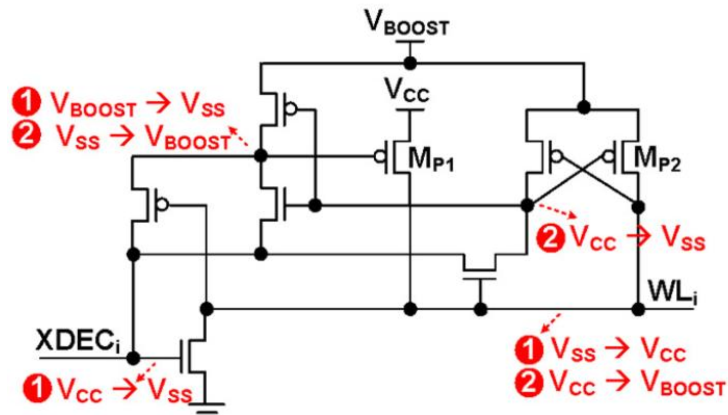


Fig. 2.25 2-step level-shifter reduce I_{LOAD} [2.23]

Charge Pump Circuit

Fig. 2.26 shows the four-stage Dickson charge pump circuit, where the diode-connected MOSFETs are used to transfer the charges from the present stage to the next stage [2.25] [2.26]. The voltage difference between the drain terminal and the source terminal of the diode connected MOSFET is the threshold voltage when the diode-connected MOSFET is turned on. Therefore, the output voltage of the four-stage Dickson charge pump circuit has been derived as

$$V_{\text{out}} = \sum_{i=1}^{5} \left(\text{VDD} - V_{t(M_i)}\right) \tag{2.7}$$

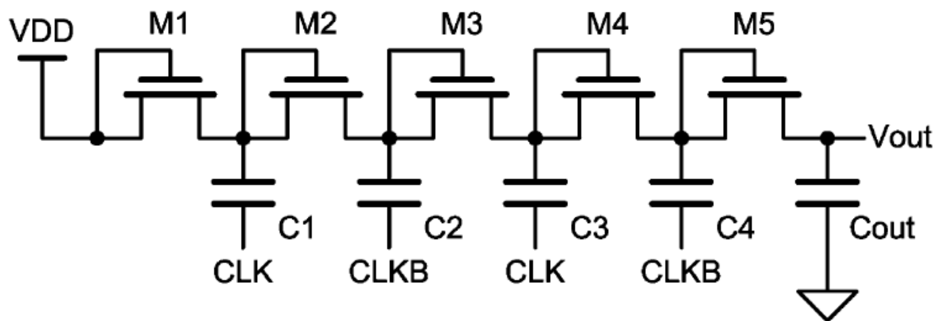


Fig.2.26 Dickson charge pump circuits

The threshold voltage (Vt) of the diode-connected MOSFET becomes larger due to the body effect when the voltage on each pumping node is pumped higher. Therefore, the pumping efficiency of the Dickson charge pump circuit is degraded by the body effect when the number of pumping stages is increased.

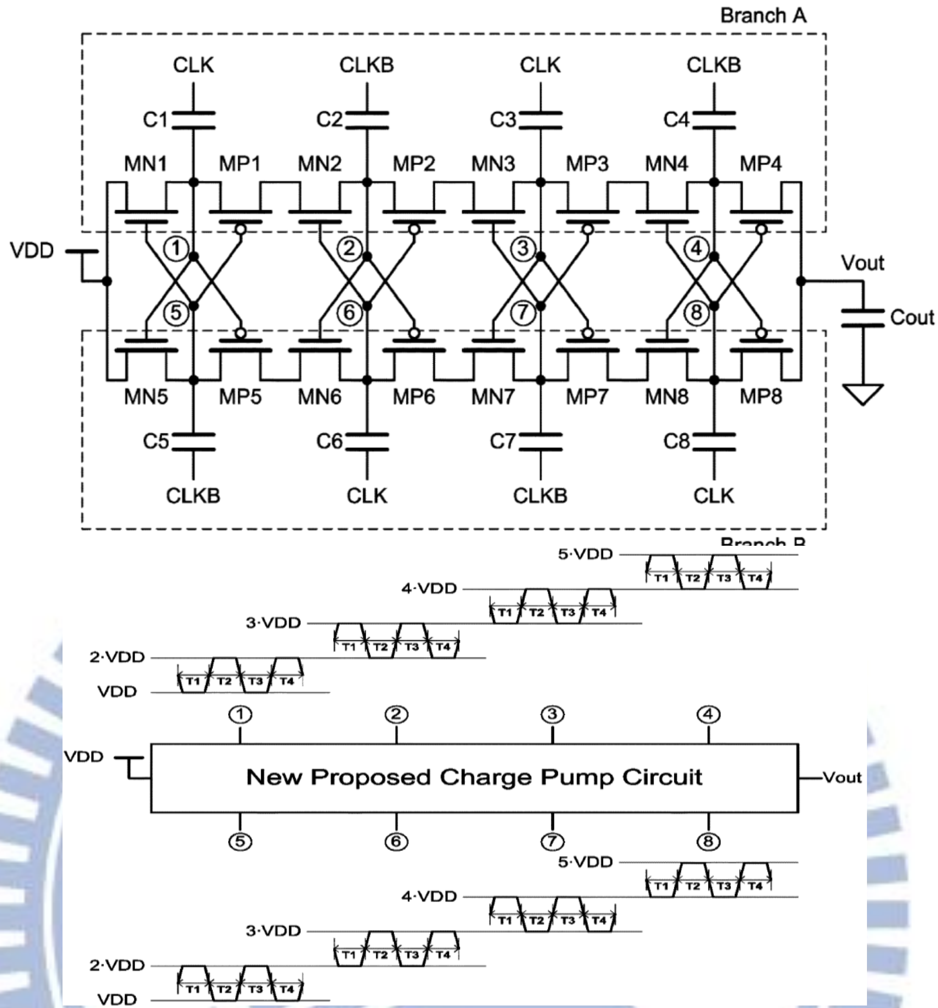


Fig. 2.27 Ker proposed CP circuit and waveform with four pumping stages

The circuit and waveform of the new proposed charge pump circuit with four stages are shown in Fig. 2.27 [2.27]. To avoid the body effect, the bulks of the devices in the proposed charge pump circuit are recommended to be connected to their sources respectively if the given process provides the deep n-well layer. Clock signals CLK and CLKB are out-of-phase but with the amplitudes of VDD.

2.5 Previous Low Voltage SRAM Design

2.5.1 SRAM Bit-cell

Differential VSSM 7T SRAM Bit-cell

The standard non-isolated read and writes 2-port 8T SRAM bit-cell is shown in Figure 2.28(a) & (b) shows a 2-port (1R/1W) single-ended 7T bit-cell, with an

isolated read-port comprising of two transistors M_{1R}, M_{2R}, and a single read bit-line to directly sense the data from node Q. By separating write port consisting of a single ended write bitline and write word-line , this design offers a static-noise-margin-free read operation, since it isolates the read current path (shown in dotted) from the data storage nodes (Q or QB).

In the reason of separating R/W ports, the isolation of read-ports provides more than 2 times better read SNM that cannot be achieved in standard 6T bitcell like the HSNM.

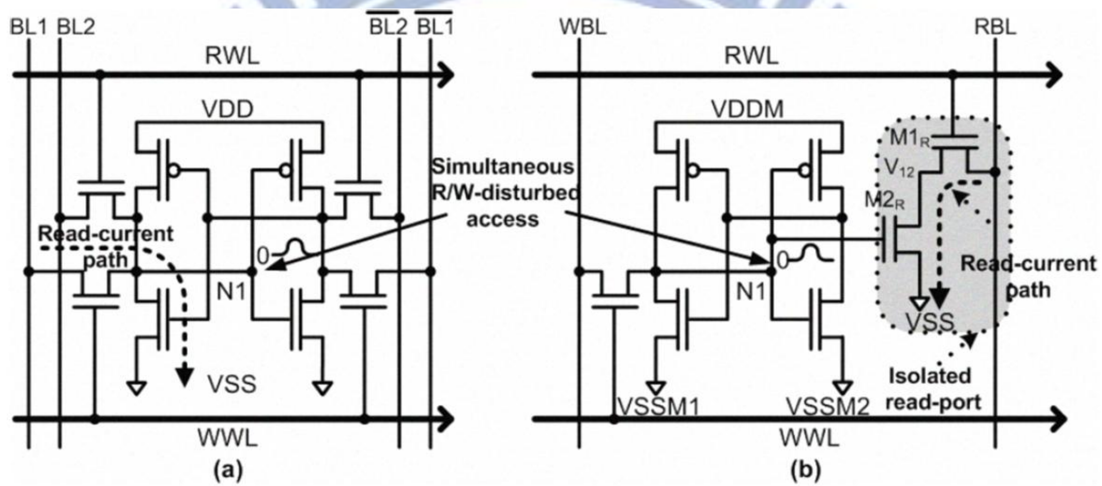


Fig. 2.28 (a) The standard 2-port 8T SRAM bit-cell with non-isolated read-port [2.28]

(b) An isolated read-port 7T SRAM bit-cell [2.28]

DCO 8T SRAM Cell Design

In this paper [2.29], the authors try to use two kind core oxide structures for power reduction and low VCC_{MIN}. A new structure 8T cell with dual core oxide (DCO) in 45LPG triple gate oxide CMOS process is proposed for high performance low leakage mobile applications. The DCO 8T SRAM operates under dual voltage supplies with write assist. Compared to traditional single-end 8T cell, DCO 8T SRAM showed the same performance with only half the standby leakage, and lower VCC_{MIN}.

The DCO 8T cell is designed in 45nm LPG CMOS process which shows in Fig. 2.29. Different WWL operation voltage for power saving. VddM and VddM1 are

normally at 0.9V and 1.1V, respectively. For example, during low voltage operation, only VddM will be lowered to 0.6V while VddM1 remains unchanged.

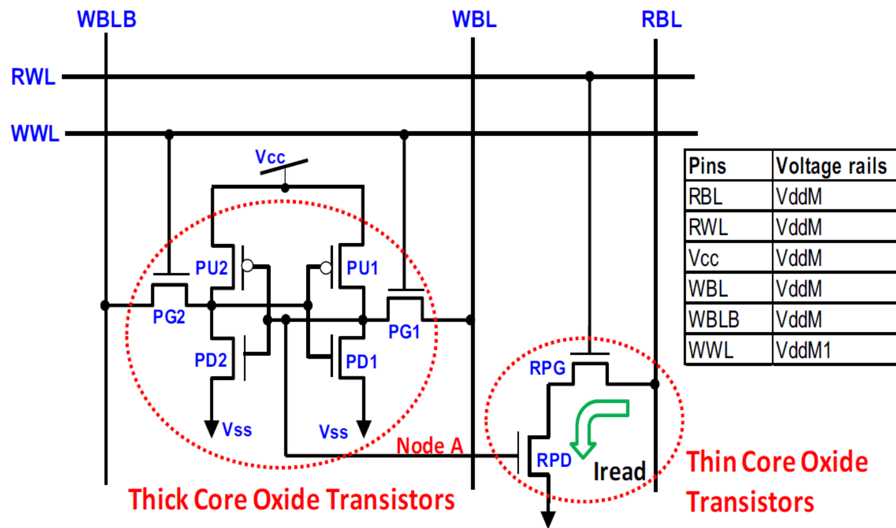


Fig. 2.29 Schematic diagram of the DCO 8T SRAM cell with dual VDD [2.29]

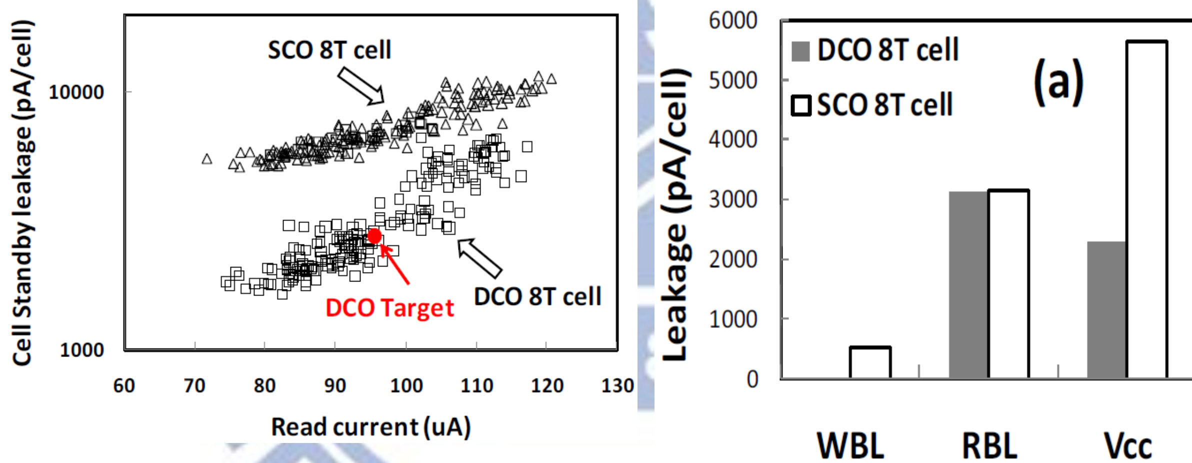


Fig. 2.30 DCO 8T cell shows 2x lower leakage at the same read current at 0.9V comparing to SCO 8T cell [2.29]

Fig. 2.31 Comparison of leakage components between DCO 8T cell and SCO 8T cell at 0.9V (Q = 1) [2.29]

Fig. 2.30 shows read current vs. standby leakage comparison between these two cells across process corners. DCO SRAM standby leakage at 0.9V is only 3nA, which is half of the SCO cell at the same 98uA I_{read} performance. This is mainly due to the gate leakage and sub-threshold leakage reductions by using LP transistor in 6T write

port. Silicon data showed the DCO cell read BL leakage (sub threshold leakage) and read pull-down gate leakage are dominating leakage source, as shown in Fig.2.31.

A new 2-port SRAM it-cell

In this paper [2.30], a new structure dual-port 6T cell is proposed. Fig. 2.32 combine with assist MOS and more one global WL, there are three merits show in the list compare with convention dual port design.

1. A new 2-port 6T memory bit-cell and its word-oriented array organization is proposed to eliminate simultaneous read and write access disturbances due to column select functionality in neighboring bit-cells or words.
2. The poor read-noise margin and conflicting read-write problems are handled by isolating the read and write-ports to achieve higher stability margins.
3. The process variation sensitivity analysis shows that the proposed design has significantly low process variation sensitivity as compared to existing ones, hence a better parametric yield.

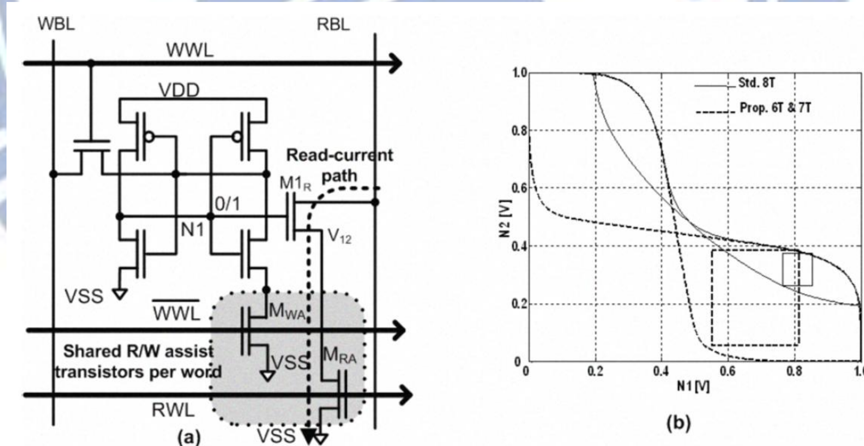


Fig. 2.32 (a) Schematic diagram of the proposed 2-port 6T SRAM bit-cell with shared read and write assist transistors per word [2.30]

(b) The VTC and SNM obtained from butterfly curve for the standard ST, 7T and proposed 6T SRAM bit-cells [2.30]

Fig. 2.33 shows a 32-bit word-oriented SRAM array organization of the proposed 2-port 6T bit-cell. Because this 6T cell couldn’t do bit-interleaving, a bank

is divided into many parts block for do this feature. Each word also has a sub-wordline driver to activate the local wordlines, and a set of read and write-assist transistors. In a word-oriented SRAM array organization, all the bit-cells of a word are kept together, which facilitates the sharing of read and write-assist transistors.

Not only reasons said before, multi-divide word and bitline techniques are commonly used to reduce the charging and discharging capacitance of wordlines and bitlines, or in other words to minimize the read/write delay for improving the array performance.

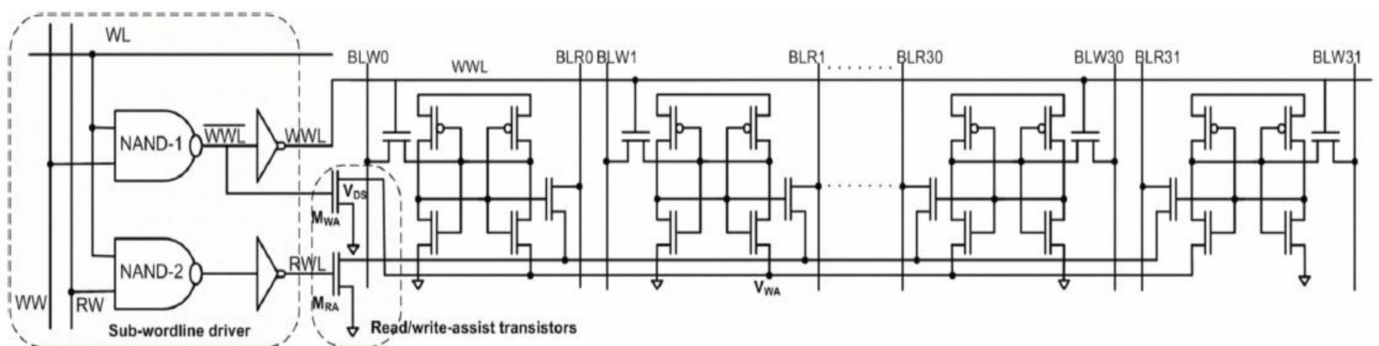


Fig. 2.33 A 32-bit word organization of the proposed 2-port 6T SRAM bit-cell to eliminate simultaneous read/write disturbance problem [2.30]

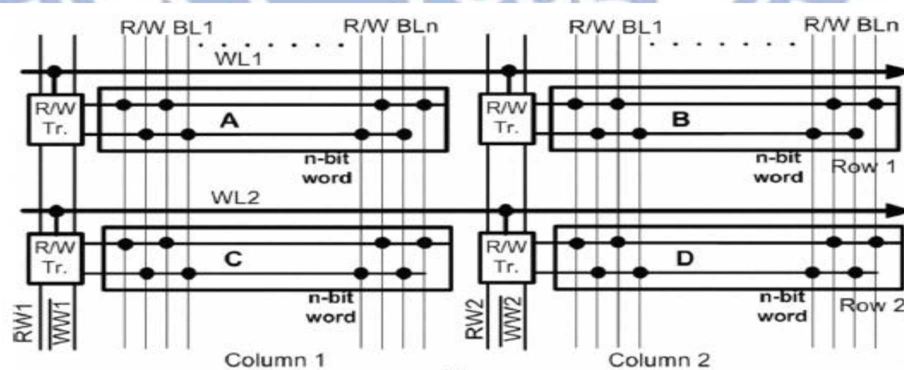


Fig 2.34 Simultaneous R/W access issues in word-oriented array [2.30]

Fig. 2.34 shows the schematic diagram of a 2-port 6T SRAM bit-cell memory module, with word-oriented array organization having four n-bit words (A, B, C and D) arranged in 2-rows and 2-columns. By this way, the cell array can do simultaneous read and write accesses influence the states of the neighboring bits or words.

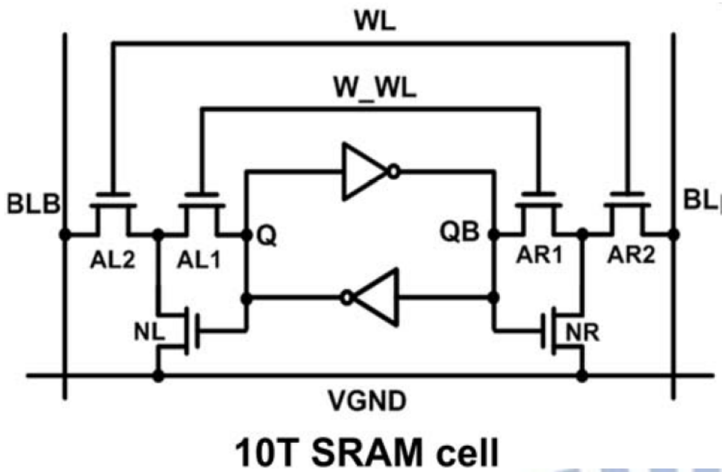
Zigzag 8T-SRAM

Previously 8T/10T has obvious drawbacks of slower write back or wasteful layout in implementing schemes, even if they are much better than the 6T cell. A decoupled single-ended 8T (DS8T) [2.32] suffers slower read first and WB due to its single-ended sensing. The CP10T [2.33] cell has larger area penalty because it uses a 5-poly pitch layout, and suffers degraded write ability due to its serial access-gates. Decoupled differential 9T (D9T) and 10T cells improve read speed but require large area. Poorer area-cost effective cells lead to an increasing σV_{TH} due to limited resorting to transistor upsizing (Fig. 2.37 & Fig. 2.36).

This paper demonstrates for the first time quantitative performance advantages of a zigzag 8T-SRAM (Z8T) [2.31]. Fig. 2.35 shows cell over the decoupled single-ended sensing 8T-SRAM (DS8T) with write-back schemes, which was previously recognized as the most area-efficient cell under large σVTH/VDD conditions. Since Z8T uses only 1T for each decoupled read-port, faster 2T differential sensing (D2S) can be implemented within the same area as the single-ended DS8T. Thanks to D2S, Z8T cell enables much faster R/W speed at VDDmin than DS8T. For the same VDDmin/speed, Z8T save the cell area by 15%. Compare with conventional DS8T area is 14% smaller and 53% faster read. In this work, a low VDDmin can down to 250mV.

		This Work	9T Ref [6]
RWL level	Select.	GND	GND
	Un-Select.	VDD	Floating
Impacts on RWL design		Enables RWL sharing in the column w/o column data collision	Need split of RWL & one more 1T to prevent column data collision

Fig. 2.35 Schematic of the proposed Z8T SRAM [2.31]



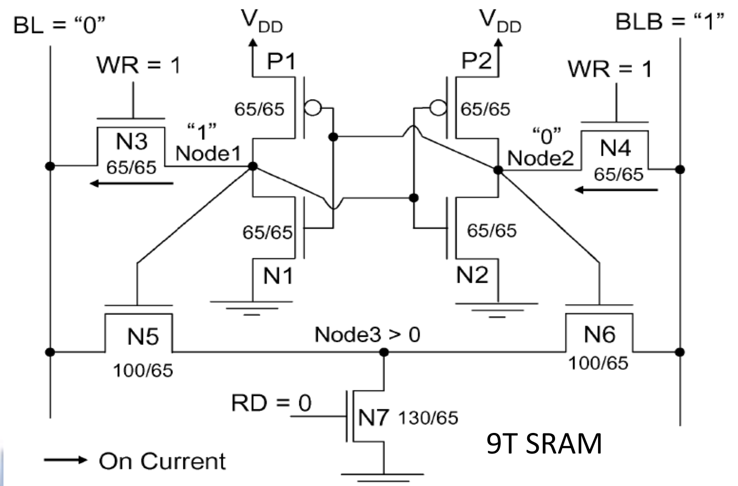


Fig. 2.36 Schematic of the proposed 10T SRAM [2.33]

Fig. 2.37 Schematic of the proposed 9T SRAM [2.34]

A New Low Leakage 8T-SRAM [2.35]

Figure 2.38 shows the architecture of new 8T SRAM cell. It consists of two extra transistors MNLL and MNWL as compared to conventional 6T SRAM cell. Transistor MNLL is used to reduce gate leakage while transistor MNWL is used to make cell SNM free in the zero state. There are three characteristics in this cell design. First is a novel read ''0'' static noise margin free eight transistors SRAM cell is proposed that reduces gate leakage power in the zero state. Second, this new high VT 8T SRAM cell reduces total leakage by 60% in zero state at highest temperature. Finally, new cell improves SNM by 2.2 times as compared to conventional 6T SRAM cell in read operation and standby mode for the case when cell stores logic '1'.

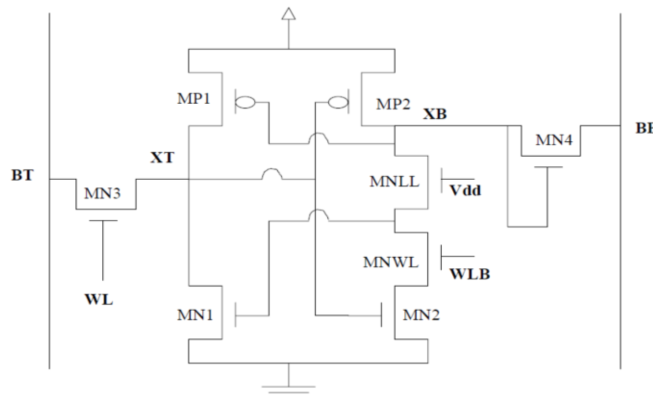


Fig. 2.38 Schematic of 8T SRAM Cell [2.35]

2.6 Previous Low Power Register File Design

2.6.1 Register File Bit-cell

The Fig. 2.39 shows a RF bit-cell which can work in sub-threshold region [2.36], the disadvantage is that the read port limits the cell number on bit-line due to a little fan-in/out. The author replaced the conventional cell with the bottom right cell Fig. 2.40. It provides a solution to reduce the capacitance. However, the speed will degrade and cause large area in array. In this mux cell design, select one cell of two will spend an extra time.

A likely design in Fig. 2.41 also uses the same combinational circuit to reduce the loading on RBL [2.37]. In this paper, the Double-DICE storage element, which reduces charge sharing and collecting between the sensitive nodes of sensitive pairs in a Dual Interlocked Cell (DICE) storage cell. If a radiation particle strikes a sensitive node (drain of a NMOS or a PMOS in off mode), and it loses its charge, the redundant nodes restore the state of this affected node and prevent an upset in the storage cell logic. The DICE design provides excellent protection against SEU for sub-micron technologies, where a single radiation strike results in charge collection at only one node. So the author combines this cell and reduce fan in technology by two NAND-OR gate to get a low capacitor design.

In [2.37], he proposed a Dual-DICE design, which interleaves two DICE storage cells to make them more resistant to upsets caused by charge sharing and creation of lateral parasitic bipolar transistors in multiple PMOS devices in deep submicron technologies. The design provides an area savings as compared to the alternative approach and results in a very small Clock to Q delay overhead.

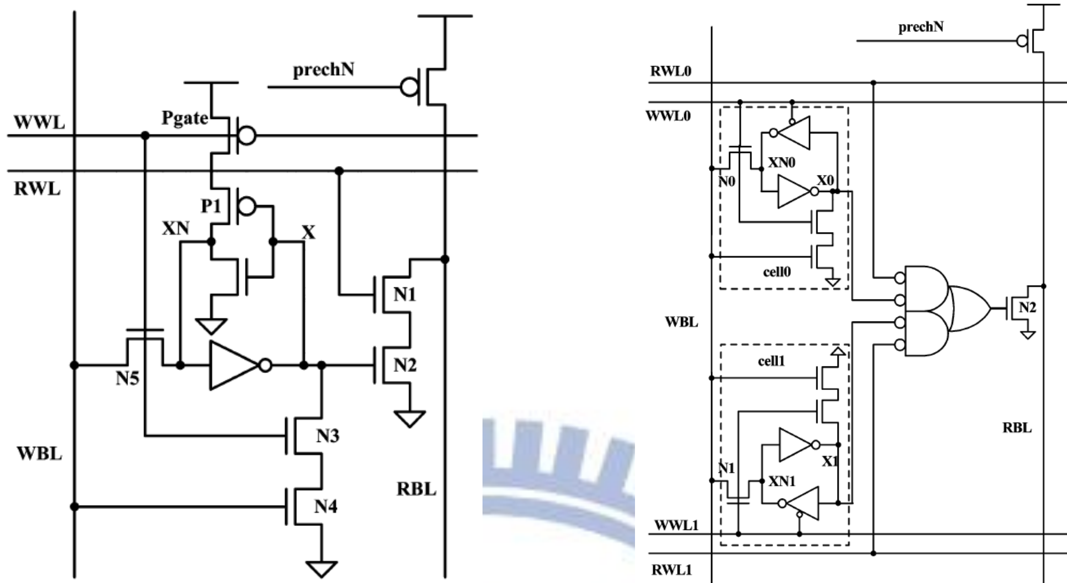


Fig. 2.39 Cell number on one bit-line is small [2.36]

Fig. 2.40 Two bit-cells share one bit-line [2.36]

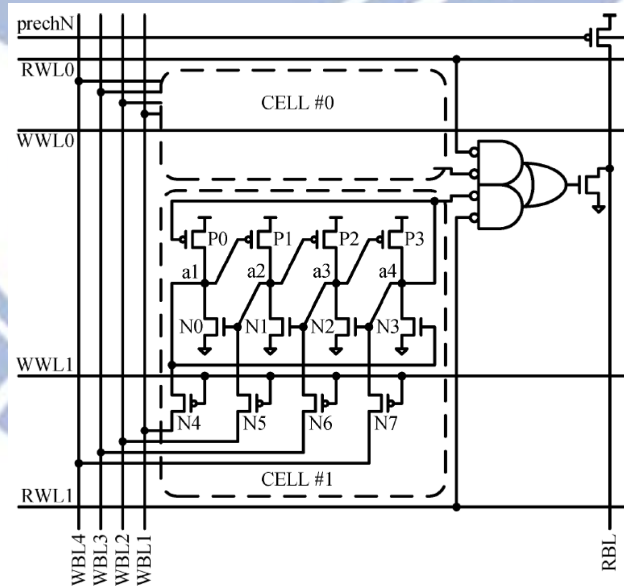


Fig. 2.41 Cell number on one bit-line is small [2.37]

The IRF design presented several challenges with the large number of multi-ported registers required to support the four threads in the core [2.38] [2.39]. The design goal was to satisfy the performance needs with competitive area and power consumption. Performance-wise, the pipeline requires a read access immediately after a restore operation to be completed in half a cycle.

The IRF in this design supplies a maximum of three operands per instruction for the single active thread. Therefore, the read ports for all four threads are merged into a compact structure with shared read bitlines (Fig. 2.42) to reduce area and power. The 32 entries are folded into two columns with only 16 read cell pull downs on the bitline for optimal performance and array aspect ratio [2.40].

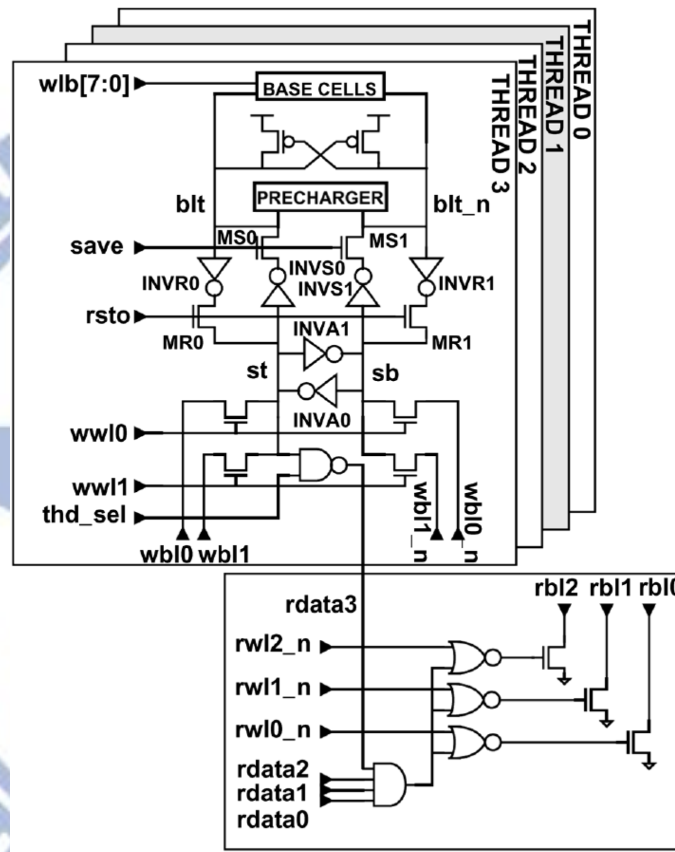


Fig. 2.42 IRF integrated memory cell circuit diagram [2.39]

Multi-port separation

Due to more efficient wiring and contact sharing, a 2R1W register file cell is ~3 to 4× smaller than a 4R2W cell (Fig. 2.43), which reduces cell dimensions and thus both wordline and bitline lengths by nearly a factor of two [2.41] [2.42].

Instead, the 2R1W cell subarray is replicated (with common write operations occurring on two duplicate copies of the data) so that four read ports are functionally

achieved while still maintaining low word and bitline capacitances. Even with subarray duplication, the 3 to 4× smaller cell size achieves a near 2× macro-level area reduction over a traditional 4R2W design. This area reduction also results in a corresponding decrease in leakage power. Due to reduced read bitline capacitance and smaller drivers, read power and read bitline latency can both be improved by ~2×. Write power is not dramatically affected as reduced write bitline capacitance balances subarray duplication [2.43].

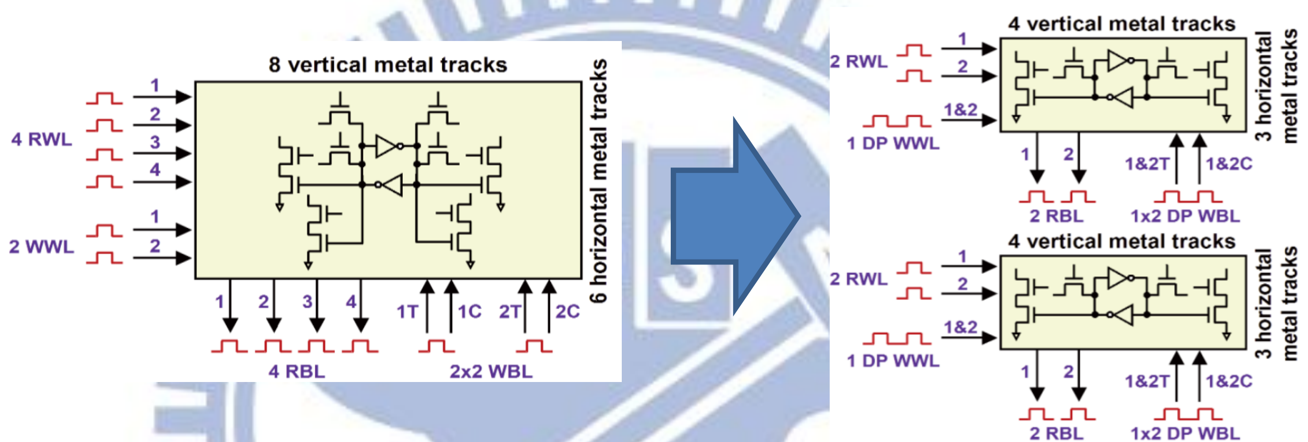
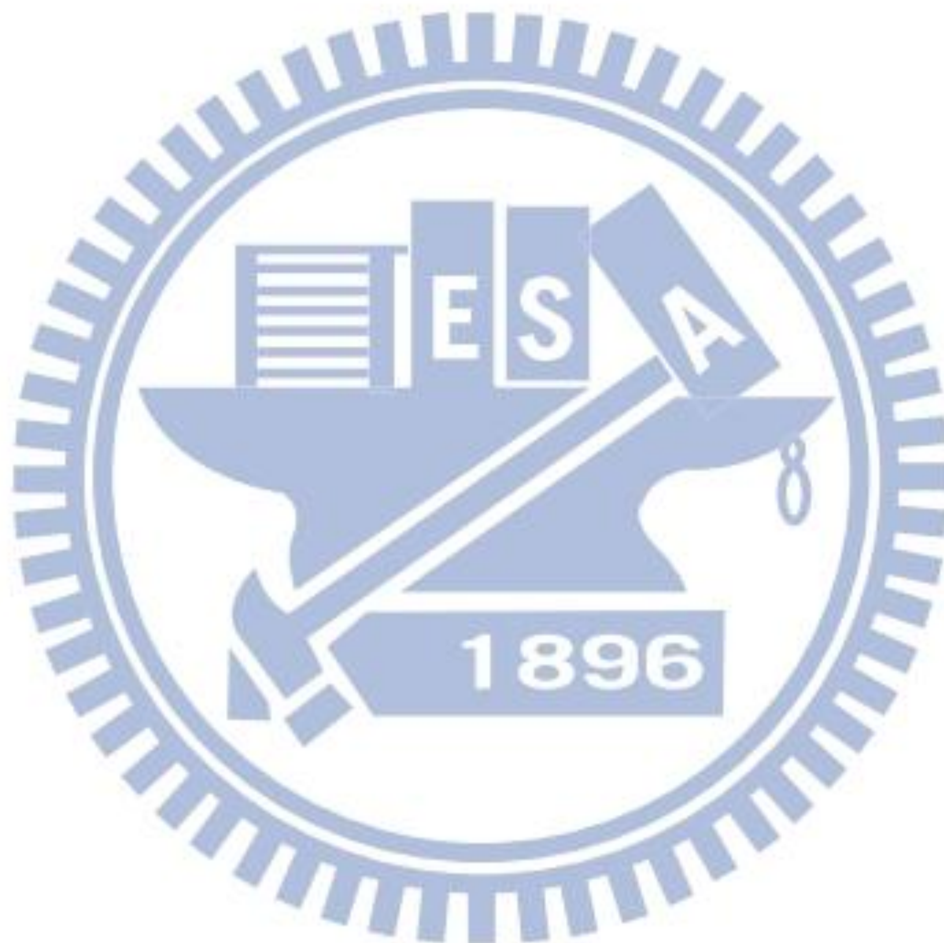


Fig. 2.43 Standard 4R2W split to 2 copies of a 2R1W cell [2.41]

2.7 Summary

In the beginning of this chapter I introduce the power consumption model and device geometric effect. Nowadays, leakage power is domain the whole chip power consumption and how to reduce power dissipation is a very important issue. Standby power and leakage current are discussed in the section 2.1, then CMOS device design and new technology such as FinFET, High-K metal gate are also introduced. After that the basic operation of conventional 6T SRAM and introduce the basic concept and measurement of stability and write ability in SRAM bit-cell. By technology process scaling down, the process variation is already damage the SRAM cell stability significantly. Global variation and local variation are discussed in section 2.3. Then,

we introduce some new assist technologies for SRAM design or improve SNM, such as boosting circuit, keeper design and negative BL …etc. Finally, new cell or share WWL structures for low power purpose are discussed. Besides, new register design bit-cell and concept also listed in the 2.6.


ESA 1896

Chapter 3
Low Power 2R2W Multi-Port 8Kb SRAM Design

3.1 Introduction

In this chapter, a new low power 13T 2 Read 2Write (2R2W) multi-port SRAM bit cell is proposed. Combine with wide range operation and multi-port and multi-port goodness, it very suit for portable device or mobile phone. A new sharing WBL structure and cross Y_Cut & X_Cut can help cell more robustness and improve write ability and WBL driver power Reduction. Negative VVSS technology is embedded for low voltage write success. Using this technology, a shorter write 1 time is approached.

In order to gain higher bandwidth, multi-port design becomes more important in media application. No like conventional single port, multi-port SRAM design can do synchronous or asynchronous operation, because it with two independent ports. Parallel operation is got more bandwidth at same time, but a new conflict issues must be take care.

At first, I discuss conventional problem in Chap.3.2. In this section conflict problem will be specific introduced. In Chap3.4, in order to improve write "1"ability, there are two technology used in this design. Single-end write is low power reduction but write ability is drop compare with convention differential write. By use negative VVSS and cut off feedback loop can improve write strength Chap. 3.5 shows post layout simulation, performance and power analysis. The TSMC 40nm general purpose 2R2W multi-port 8K chip is tape out by CIC on Aug. 22.

3.2 Conventional Dual-Port 8T SRAM

3.2.1 Two Kinds of Access Mode in DP-SRAM

Fig. 3.1 shows conventional dual-port SRAM bit-cell, it has two port can read / write at the same time. Compare with conventional single port design, dual port structure give designer more control flexible. Dual-port SRAM provides high bandwidth and asynchronous CLK timing control property. Conflict problem is a very important in dual-port, there are many technologies to improve the Vmin of DP-SRAM against a disturb condition [3.1] [3.2].

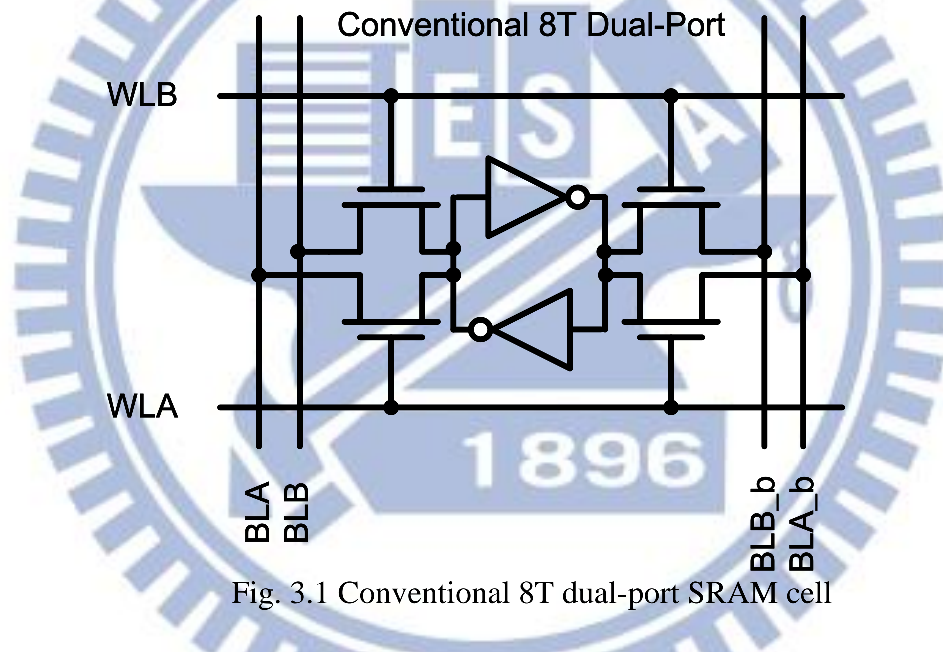


Fig. 3.1 Conventional 8T dual-port SRAM cell

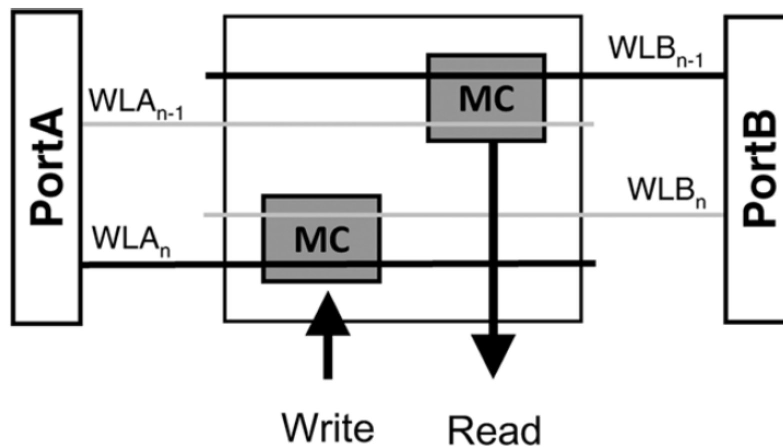


Fig. 3.2 Different-row access mode [3.2]

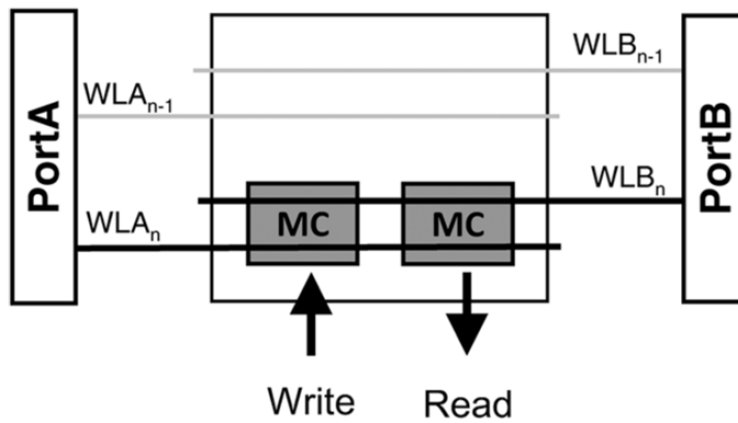


Fig. 3.3 Access in the same row [3.2]

3.2.2 Write and Read Disturb Issue in 8T DP-SRAM

There are two access modes in two port operation, first is different row which will no disturbed problem [3.3]. Second is two ports access in the same row simultaneously (Fig. 3.2 & Fig. 3.3). The case 1 (Fig. 3.4): If write for the left cell in the same row, and read for the right cell. Dummy read is happened for the left side, which is referred to as "write disturbed" The dummy read operation prevents the internal "1" node from begin flipped by BLA, so the write-ability for the left memory.

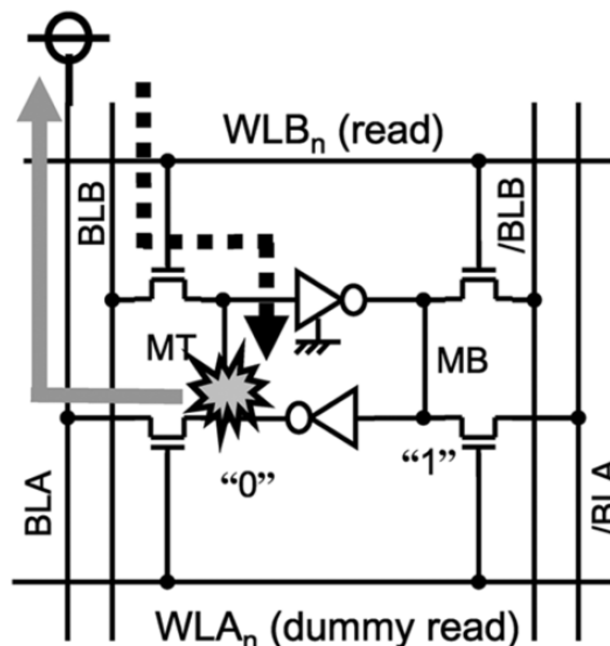


Fig. 3.4 Write operation disturbed by dummy read in the same row [3.3]

The case 2 (Fig. 3.5): If read for the left cell on the same row, and another read port is pointed to the right cell. Dummy read operation occurs for the left cell. The internal "0" node is ramped up though BLA, causing a reduction of the cell current. Consequently reduction in the cell in the cell current leads to a read failure due to lack of BL swing (read disturb).

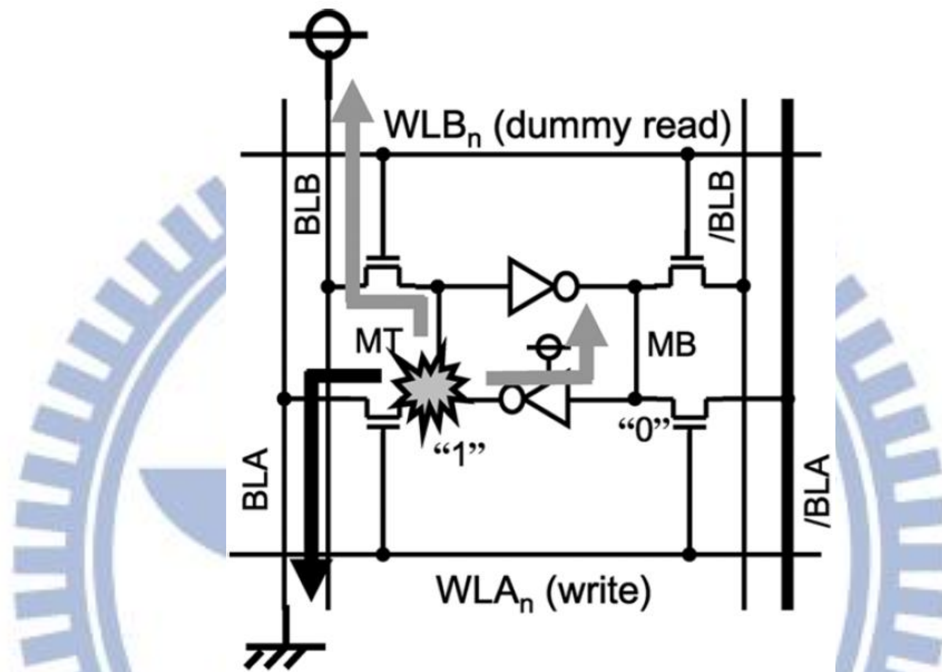


Fig. 3.5 Read operation disturbed by dummy read in the same row [3.3]

Timing control with CLK skew disturbed on dual-port also discussed in [3.4]. Timing variation is relative wire line in whole chip, if positive skew or negative skew is happened, write / read disturbed maybe caused function failed.

3.2.3 Read/Write Conflict of Dual-port

Fig. 3.6 shows new technologies that can solve the conflict problem by using timing sharing technology [3.5] [3.6]. Normally, Read and write operation is forbidden at the same time. In conventional design, if read/write is point to the same bit, large conflict power consumption and it needs more wide WL pulse have to finish the operation.

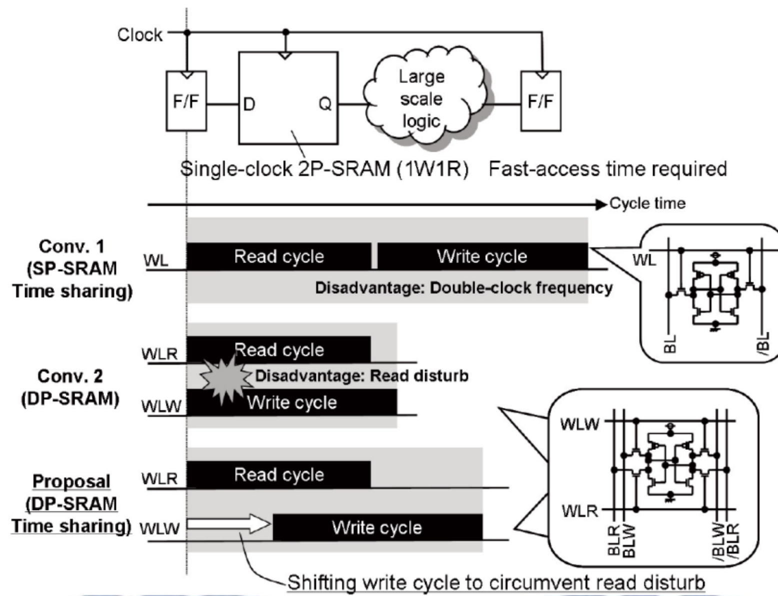


Fig. 3.6 2P-SRAM for image processing unit [3.6]

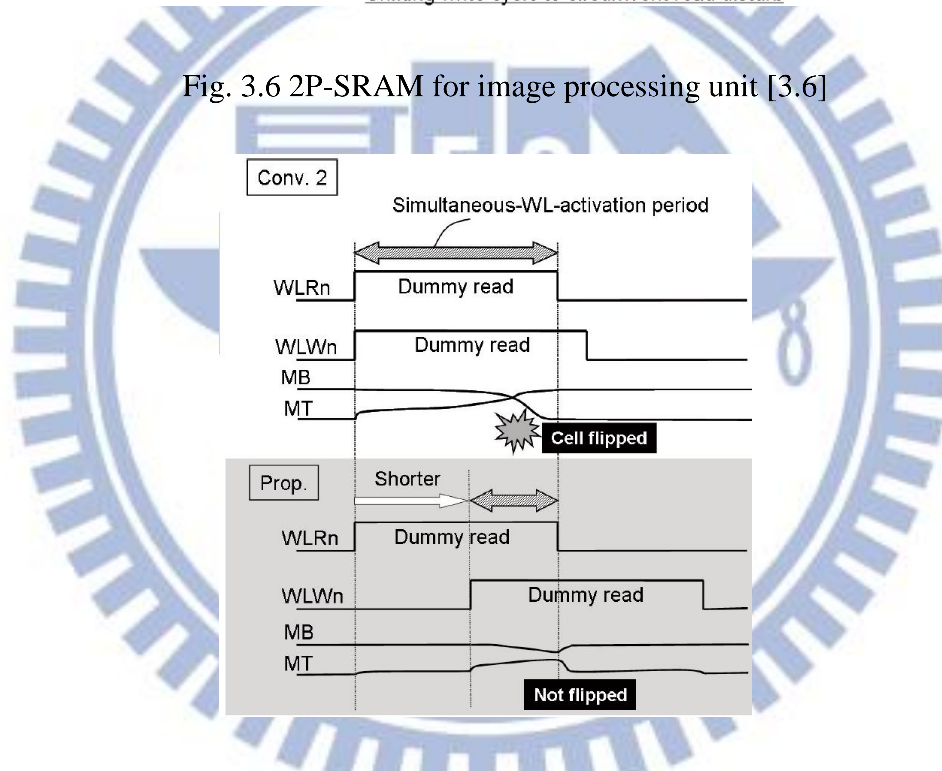


Fig. 3.7 Delay conflict waveform scheme [3.6]

In this timing sharing design, if read/write happened at the same time, one of two WL pulse will delay turn on and reduce conflict problem. Consequently, power and time delay smaller than conventional design. Not only power consumption, but also cell stability improved by short disturbed time. Fig. 3.7 shows operation waveform compare with conventional and this work [3.6].

3.3 A New 2R2W Bit-cell

3.3.1 Bit-cell Schematic and Layout View

A new cell with share WBL structure and half select disturb free characteristic is proposed. Fig. 3.8 shows a new 2R2W multi-port bit-cell, the NMOS with green color is share with neighbor column. For area reduction, this cell is used single-end write and conventional 8T read buffer. Not only low area consumption, but also power is lower than conventional dual port SRAM. In write mode, Xsel and Ysel is cross couple, this structure is suit for bit-interleaving structure. Using this design, we will not need others technology such as read and write back in bit-interleaving architecture. This cell has 13 control lines, 5 is row base control and the others are column base.

The bit cell all takes regular Vt N/PMOS in this design. In order to improve read/write current, short channel effect is used. Not use the minima size 40nm, larger channel length can gain more Ion current. So MN1, MN3, MN2, MN4, MN5, MN5, MN7, MN8 all use 70nm to replace minima channel length. For cell robust issues, the cross couple is not used minima length either.

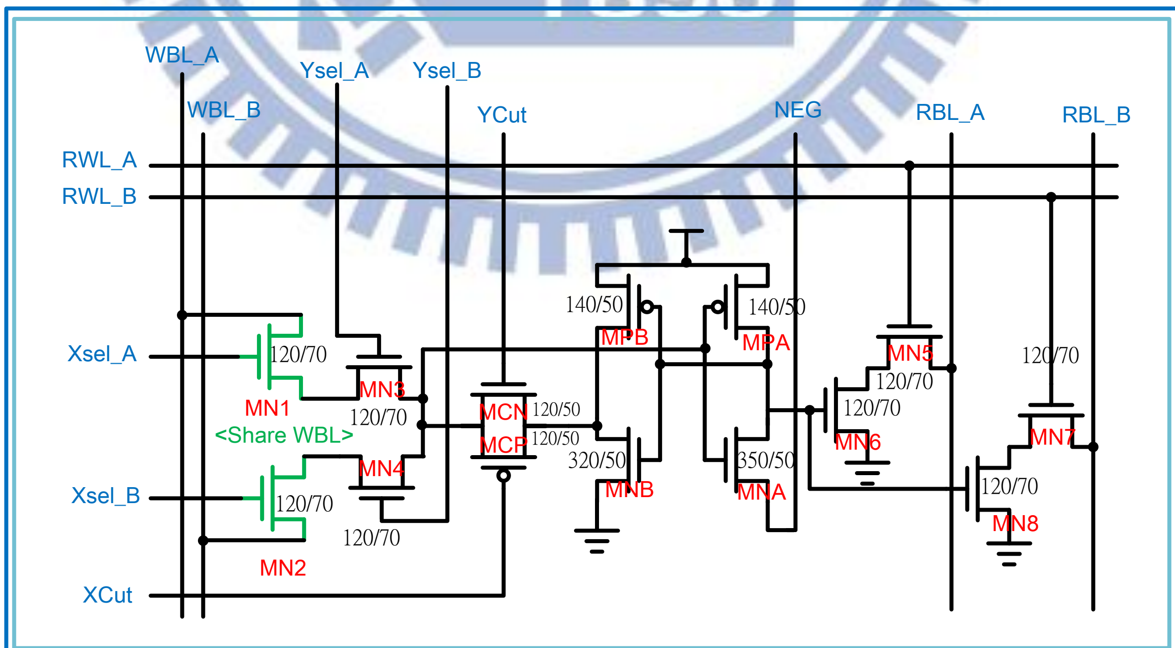


Fig. 3.8 2R2W multi-port SRAM bit-cell

The 2R2W SRAM bit-cell layout is take M1~M3. M1 & M3 is row base and M2 is column based. Metal layer and bit-cell schematic are showed in Fig. 3.9, Fig.3.10 and metal layer width is 90nm each one. In TSMC 40G technology process, dummy ploy is needed between poly and poly. In this reason, the cell layout must be larger than regular 65nm or others technology process. For area reduction, dummy poly in this design is share neighbor. Fig. 3.11 shows 2R2W multi-port SRAM layout view, the area size is 3.095um x 0.9um = 2.785um^2.

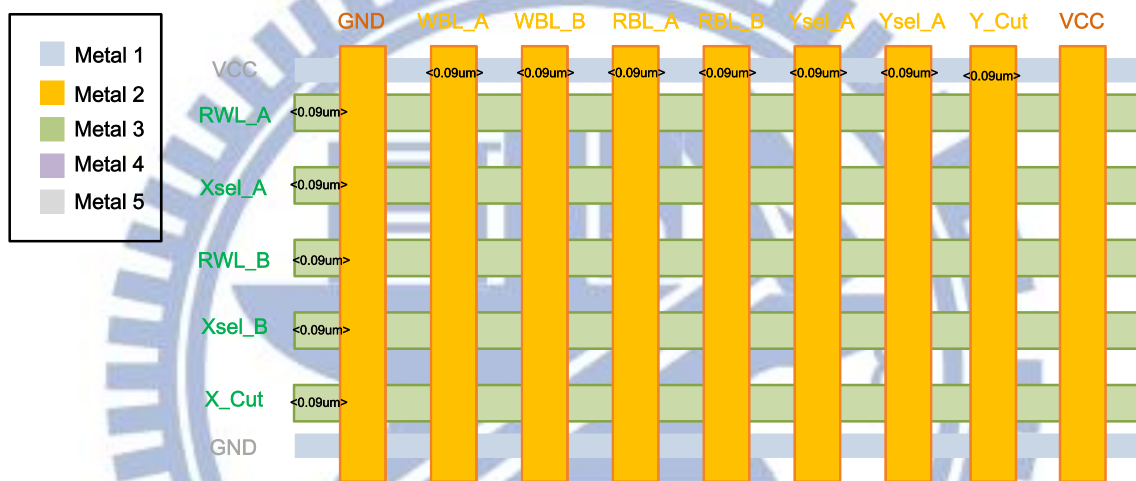


Fig. 3.9 2R2W metal layer organization

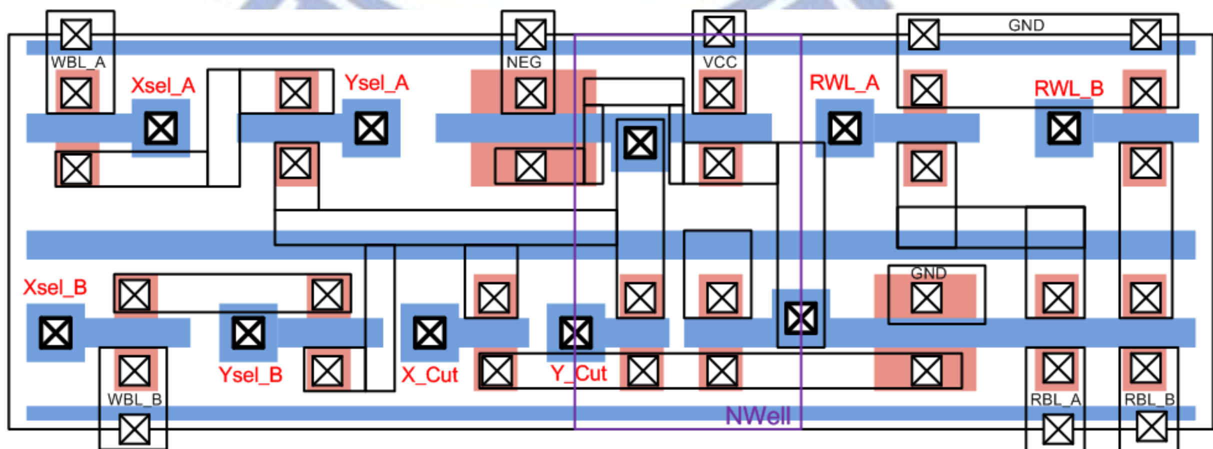


Fig. 3.10 2R2W multi-port SRAM bit-cell layout schematic

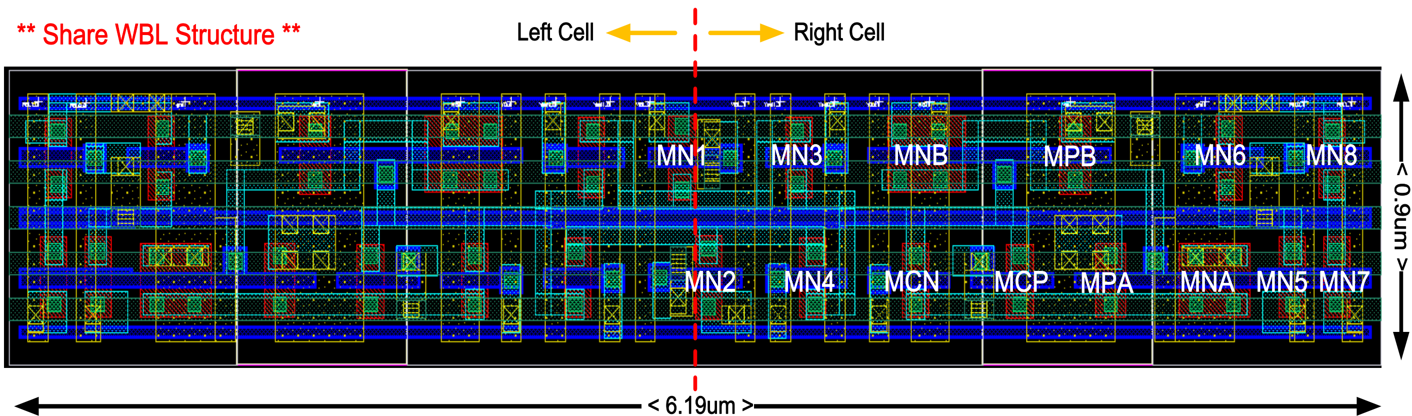


Fig. 3.11 Two 2R2W multi-port SRAM bit-cell layout view

The cell layout adaptive thin cell layout and left/right share the same WBL. Thin cell layout is popular in these years and has minima area consumption.

3.3.2 Share WBL Structure

For low power design, a new share WBL structure is proposed. Single write operation can see in Fig. 3.12, data from share WBL pass through two NMOS and write into Q. In write mode, X & Y are cut off for write ability improve. Fig. 3.13 shows two ports are access in the same row at the same time. Because in each side has one NMOS can isolated another WBL signal, there is no disturbed problem.

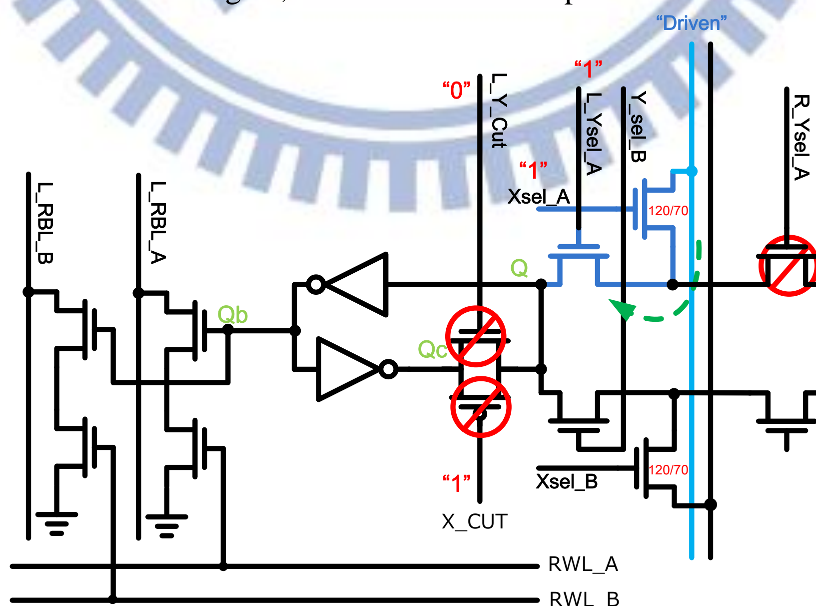


Fig. 3.12 One port writes with no disturb issues

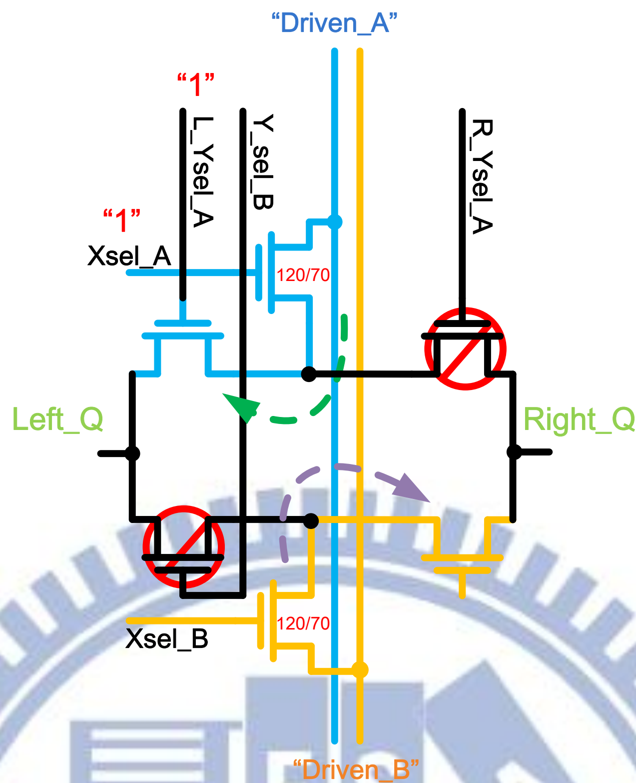


Fig. 3.13 Two write the same row with no disturb issues

3.3.3 Bit-interleaving (8 to 1)

SRAM array is more un-robust when technology process scaling down. Short channel effect modulation and soft error is easier can see everywhere in nowadays. Soft errors are caused by radiation of energetic particles, thermal neutrons, random noise, or signal integrity. A soft error is a signal or data which is wrong, but is not assumed to imply such a mistake or breakage. Since contiguous bit-cells could be corrupted at one radiation injection, the interleaving scheme takes a benefit that the effect of soft error will associated with different logical words.

A common 8-1 bit-interleaved SRAM array is illustrated in Fig. 3.14. In each row, bit-cells of words A, B, C, D…and H are interleaved and share one word-line. During a read/write operation, the column-multiplexers select the bit-lines of accessed columns among words A, B, C, D…and H.

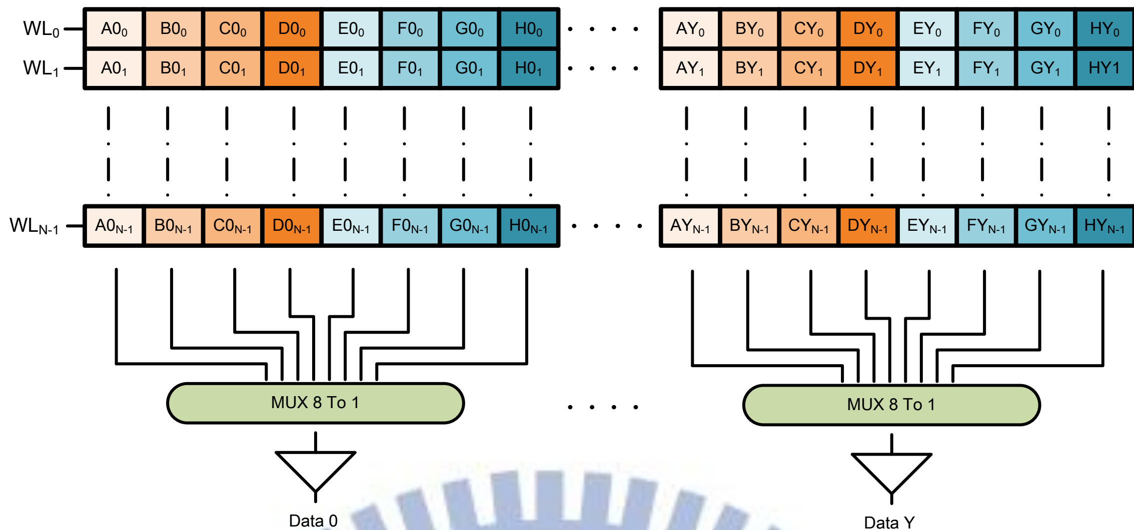


Fig. 3.14 8 to 1 Bit-interleaved SRAM array

3.4 Write Assist Technology

3.4.1 Negative VVSS

Write “1” is worst case in this single-end write bit-cell. Two stack NMOS poor the write “1” ability, voltage is drop form full VDD to VDD –Vtn. By this reason, if no any assist circuit to help write “1” operation, write function will failed in 0.7V.

Negative VVSS circuit can solve this problem, and use this technology the cell supply voltage can down to 0.5V. It is important for sizing capacitance and set a suitable negative level when write “1” operation. It will flip the data in half select on the same column when negative level is over triple point. Fig. 3.15 shows 8 to 1 bit-interleaving scheme with negative VVSS circuit output.

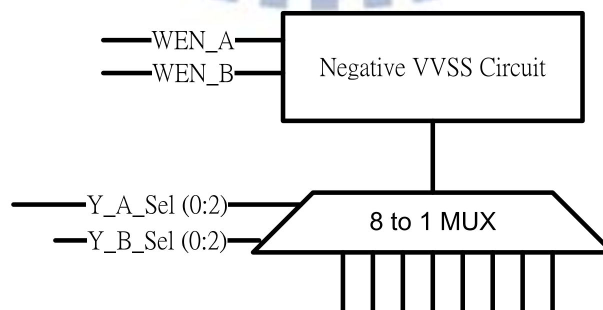


Fig. 3.15 Bit-interleaving select and Negative VVSS control

Fig. 3.16 shows negative VVSS control circuit, which use two PMOS capacitance to generative a negative level pulse. This circuit works only when write data “1” into bit-cell, others case VVSS is connect to ground. Fig. 3.17 shows negative level generative in different voltage and different corner.

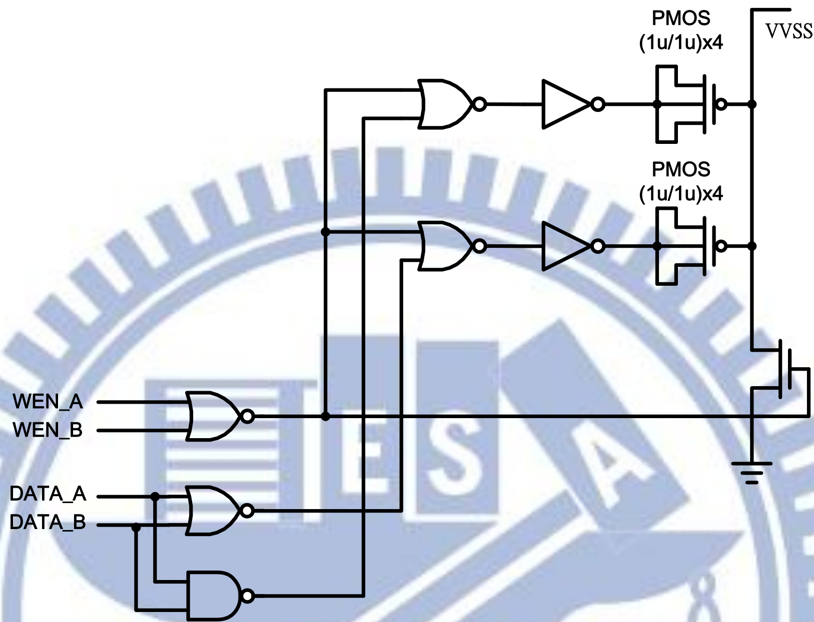


Fig. 3.16 Negative VVSS control circuit

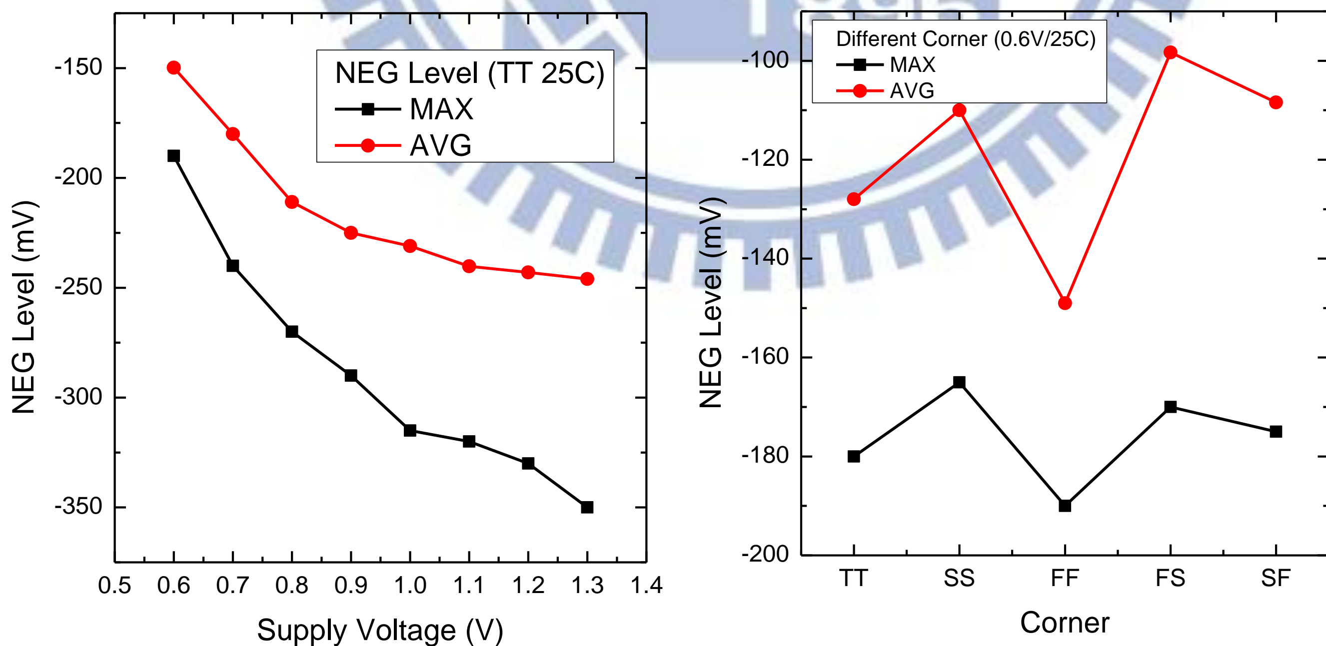


Fig. 3.17 Negative level (a) Different supply voltage (b) Different corner

3.4.2 Inverter Feedback Loop Cut-off

Not only use negative write assist technology, cut off feedback path NMOS also used in this 2R2W multi-port SRAM design. N/P MOS will cut off when write operation, they are separate control by Y_Cut and X_Cut signal. Fig. 3.18 shows the equivalent circuit when data write form share WBL to Qc. Conventional PMOS cut off is not suit for bit-interleaving structure, because it will floating the neighbor cell node when write operation (Figure 3.19). In my design, another NMOS control by column base specific cut off which one is ready to write. By this method, a more robust structure and no floating issues is achieved.

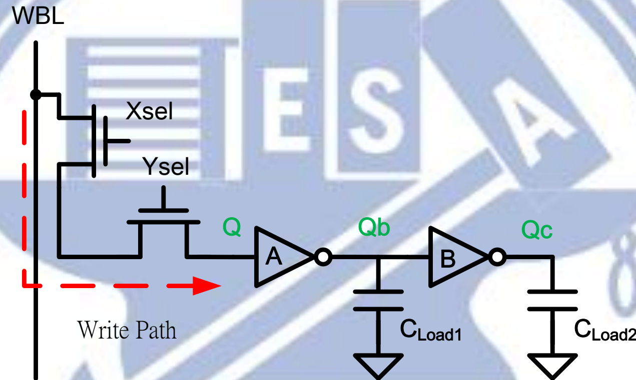


Fig. 3.18 Data write cut off scheme

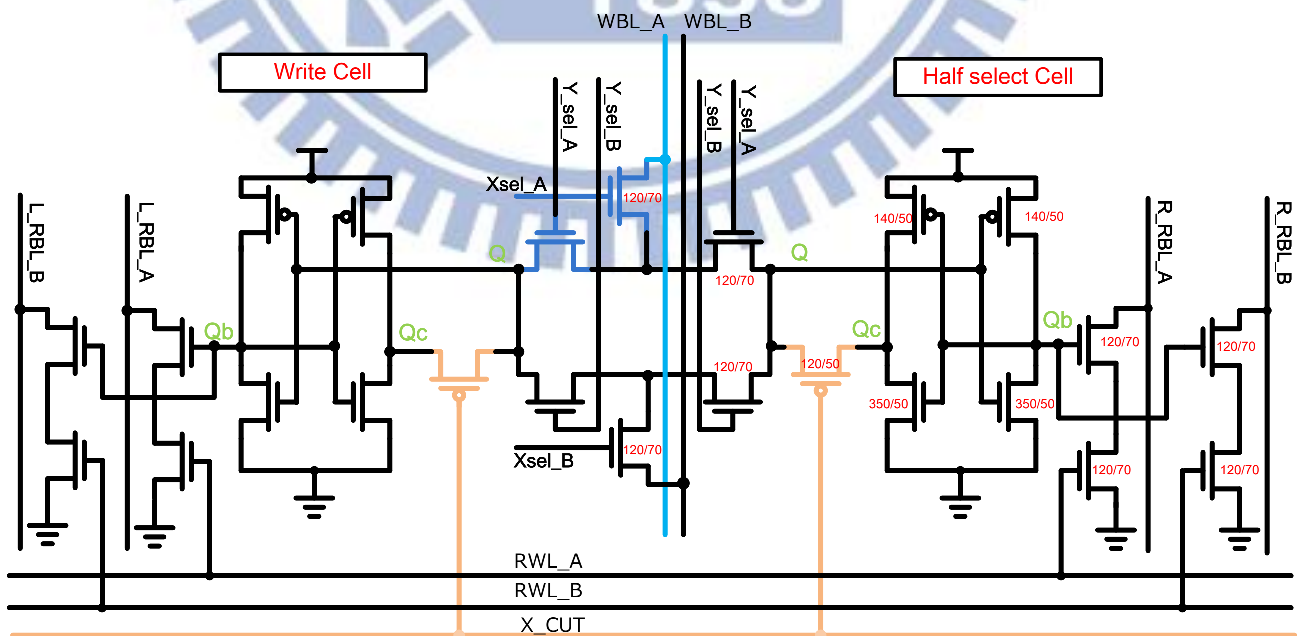


Fig. 3.19 Conventional PMOS cut off structure

3.5 2R2W Dual-port 8Kb SRAM Design

3.5.1 2R2W Multi-port SRAM Schematic

Fig. 3.20 shows 2R2W 8Kb multi-ports SRAM design schematic. There are two banks in this chip and each one size is 4Kb 64 bits x 64 bits. Write IO is place on the top of bank, and Read IO is place on the bottom of bank. Each bank has itself replica circuit and control RWL/WWL pulse width. The specific spec is showed on table 3.1, for low

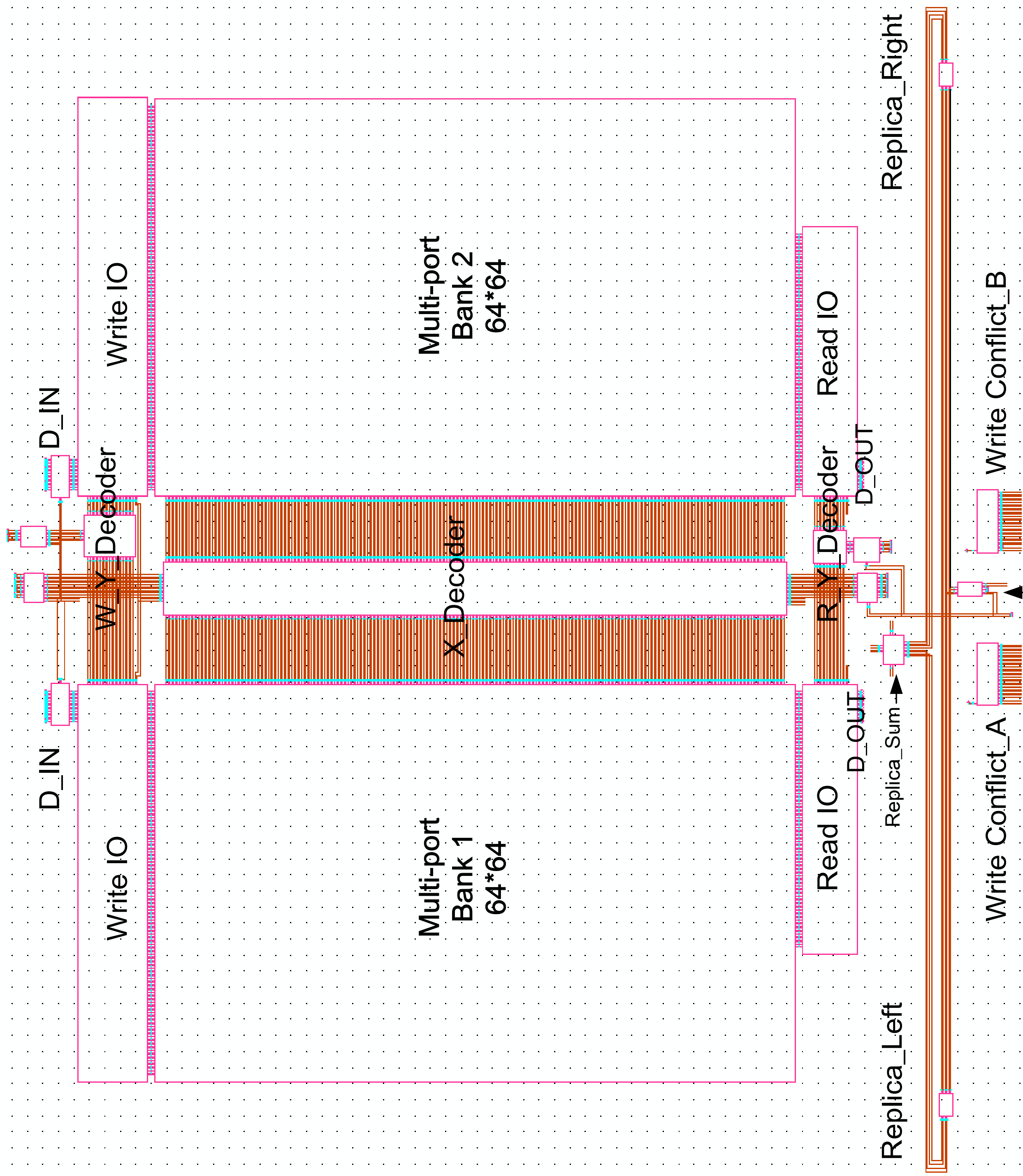


Fig. 3.20 Schematic of 8Kb 2R2W SRAM Chip

power consumption the chip is operation at 0.6V VDD. Power and performance analysis will show in the next section.

2R2W Multi-port SRAM TSMC 40nm General Purpose	
Memory Size	8Kbit (4K*2)
Data-width	16 bit
Read Address (each port)	8 bit
Write Address (each port)	8 bit
Bit-Interleaving	8 bit
Read / Write Port	2Read / 2Write
Metal layer	M1~M5 (1P9M)
Voltage range	0.5V~1.2V
Cell size (Bank)	3.095um x 0.9um
Access time @ 0.6V TT 25	5.93 ns
Cycle time @0.6V TT 25	6.1 ns (160 MHz)
Read power	0.0692uW/t (per bit-cell)
Write power	0.115uW/t (per bit-cell)

Table 3.1 Summary of the 8kb 2R2W multi-port spec

3.5.2 Data Transmission Path

In order to eliminate the conflict problem in conventional multi-port operation, a new conflict detect circuit is included in this design. Read operation is non-broken the store data in the cell and write will flip the data. I set read priority always higher than write, if read and write address are the same. Only read operation is turned on, and write operation is in stall. According to this conflict detect design, WEN signal need a more conflict detect time. After detecting is finish, intra-WEN signal is output to the replica and trigger the next stage. Fig 3.21 shows data transmission path in this 2R2W multi-port chip.

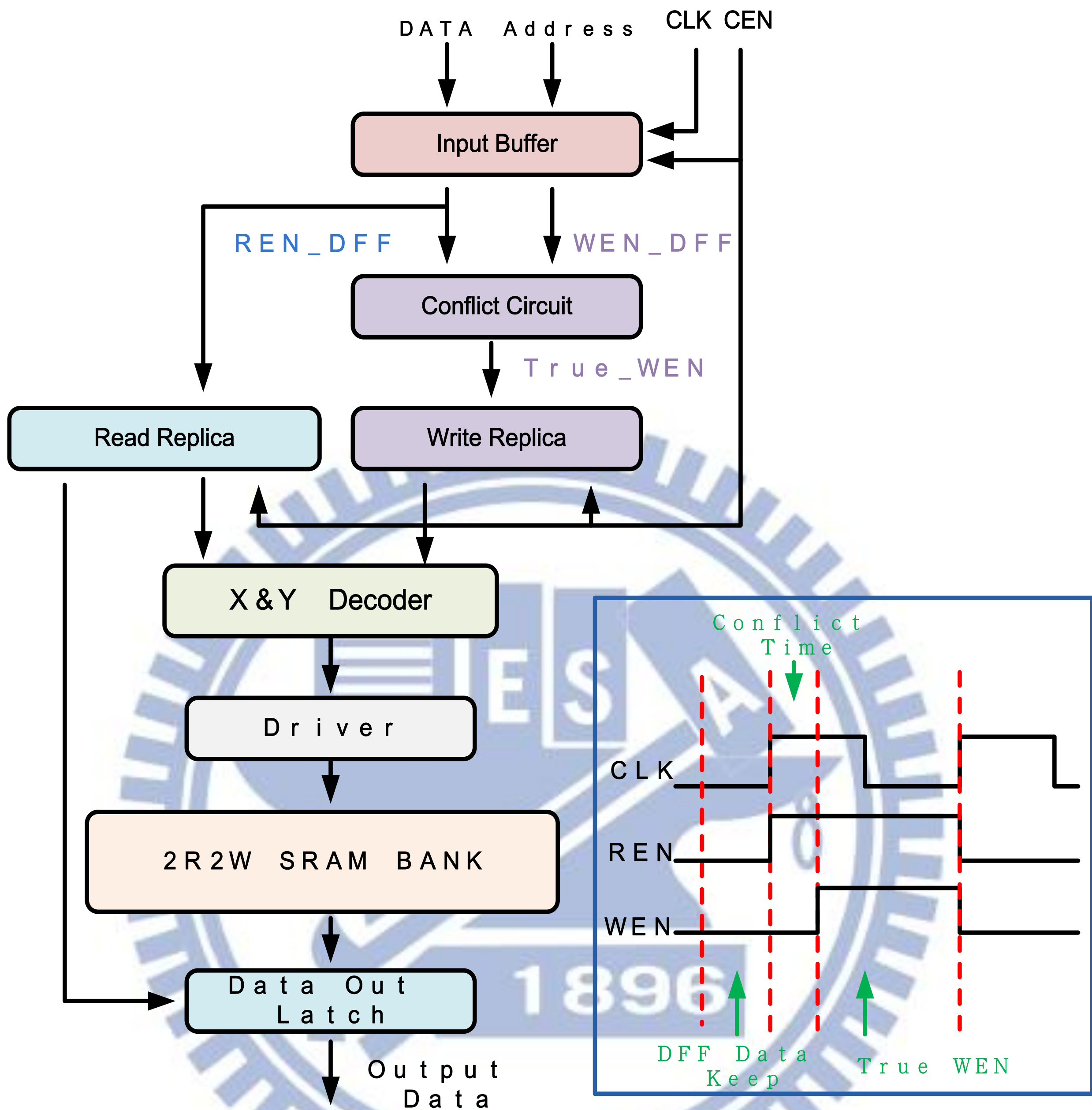


Fig. 3.21 Data transmission path in 2R2W SRAM chip

3.5.3 New Technology Adaptive in 2R2W SRAM Design

Fig. 3.22 shows whole chip layout view of the proposed 2R2W multi-port 8K SRAM. The proposed 2R2W multi-port SRAM is fabricated using TSMC 40nm general purpose process. The area of bit-cell is 3.095um x 1.8um = 5.571um^2 and the Whole chip size is 621.55um x 152.67um = 94.892 mm^2. Below is all of improved technology of this design (Table 3.2).

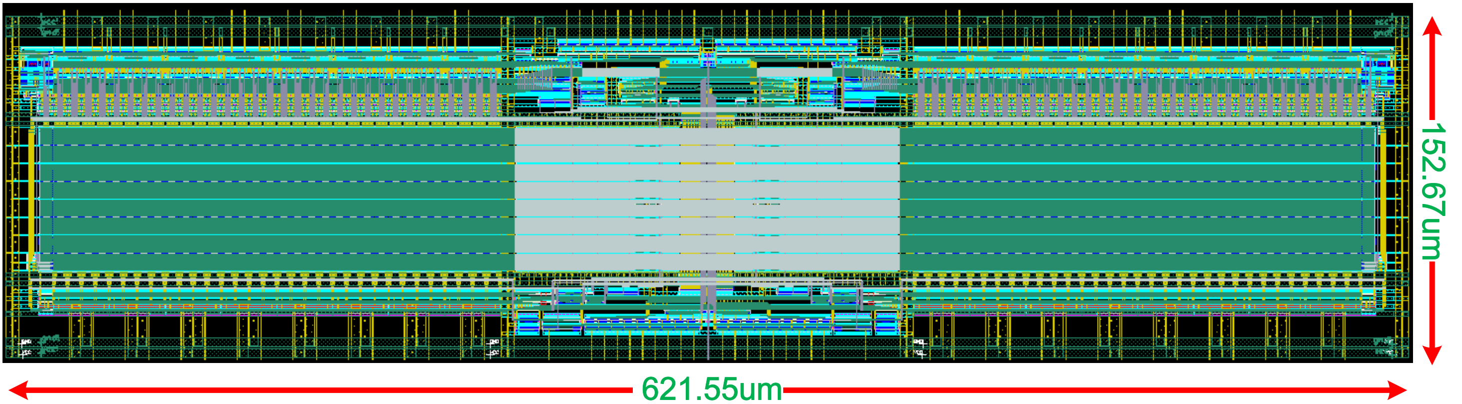


Fig. 3.22 2R2W 8kb SRAM array layout view

NO.	Technology
1.	**Wide range operation 0.4~1.2V**
2.	**Share WBL structure**
3.	**Power gating sensing circuit**
4.	**Negative VVSS for write assist**
5.	**Transmission gate cut off write assist**
6.	**Y_Cut NMOS for floating issue free**
7.	**Bit-interleaving structure**

Table 3.2 Characteristic of 2R2W 8kb SRAM

3.5.4 Test Pattern and Simulation Waveform

In order to test all function and worst case in this design, Fig.3.23 is my test function. There are 7 CLK cycle and every cycle test on different pattern. Such as 1W, 1R, 1W1R …etc., and the next cycle is test conflict detect circuit. Write data pattern is first nearest and next is furthest bit-cell, and try to find the critical path in this chip. Fig. 3.24 is post layout simulation write waveform and Fig. 3.25 shows read waveform by post simulation.

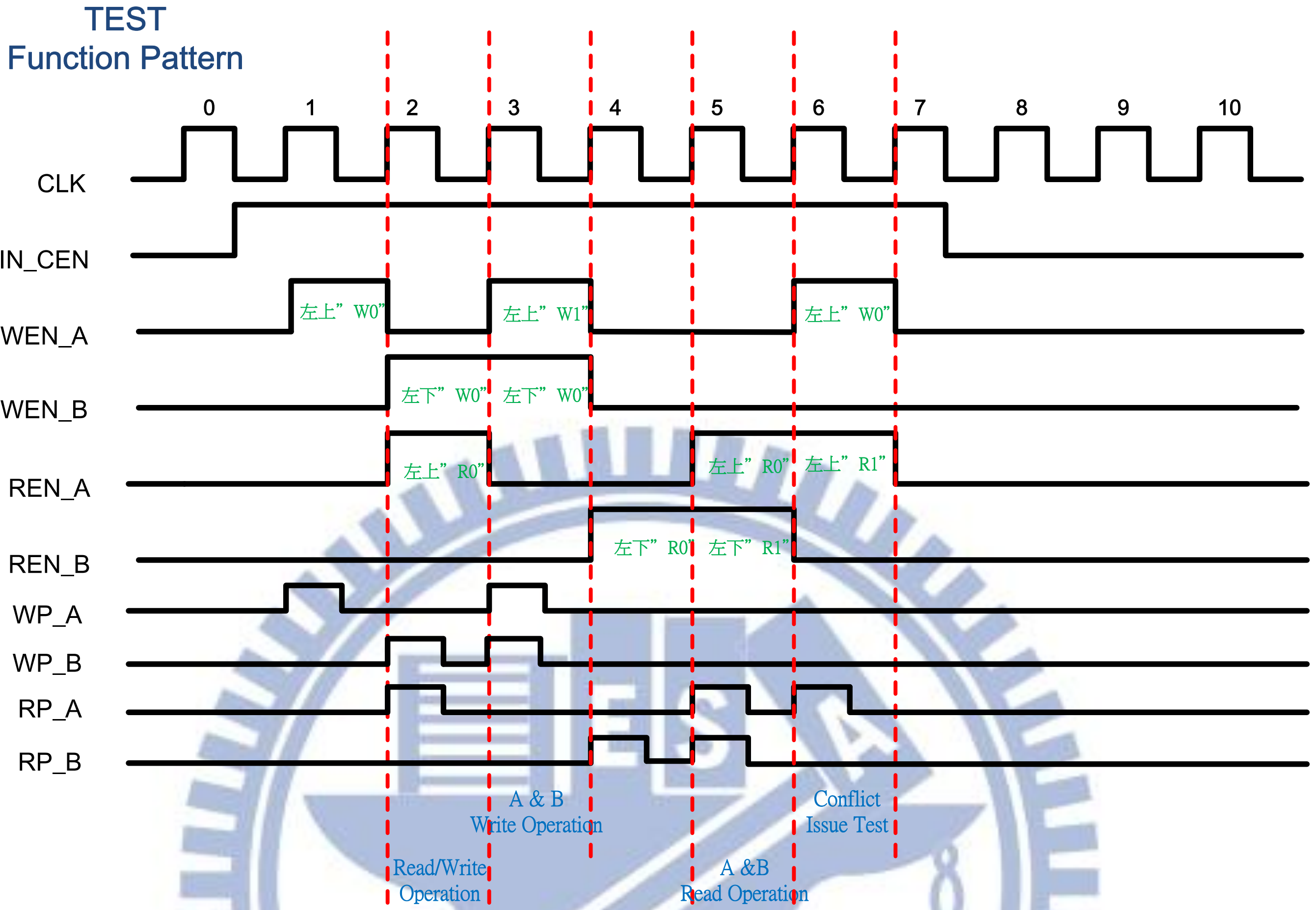


Fig. 3.23 Test pattern for 2R2W multi-port SRAM Chip

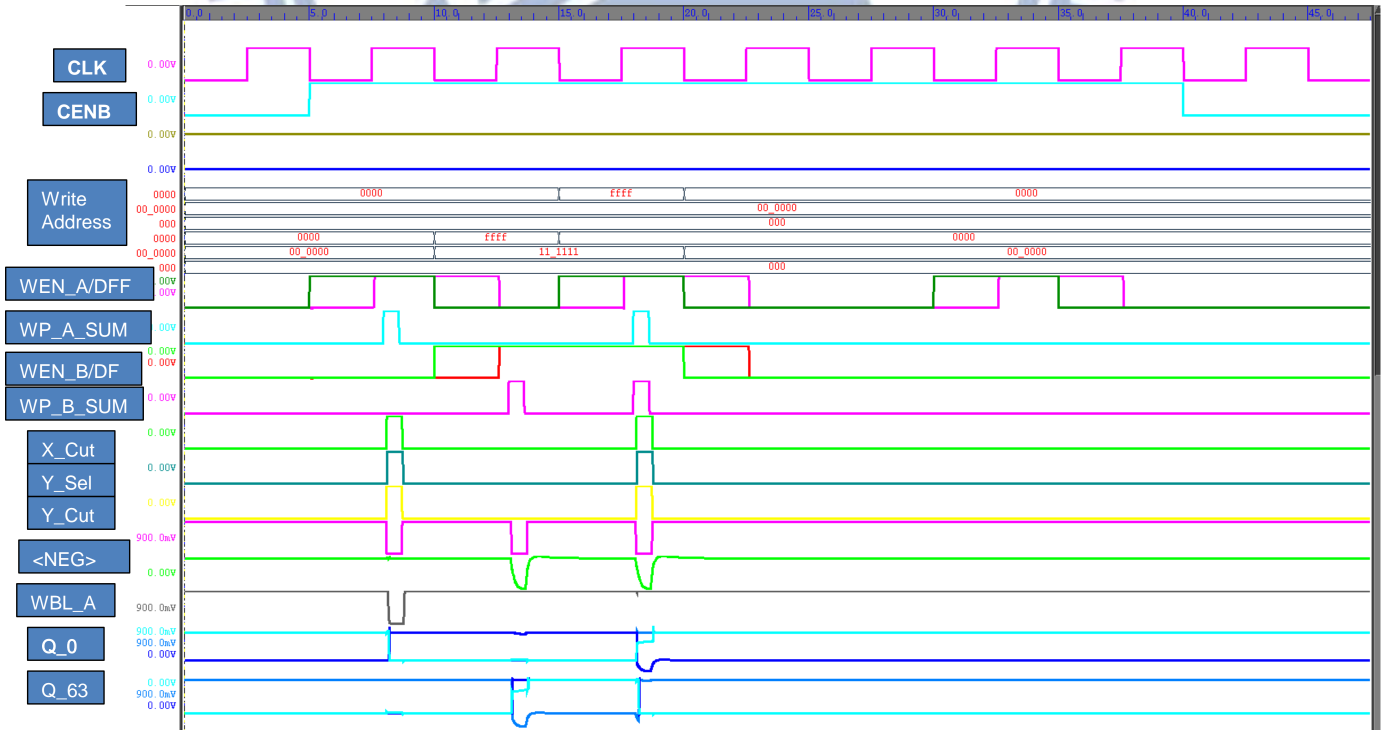


Fig. 3.24 Write test function for 2R2W multi-port SRAM chip

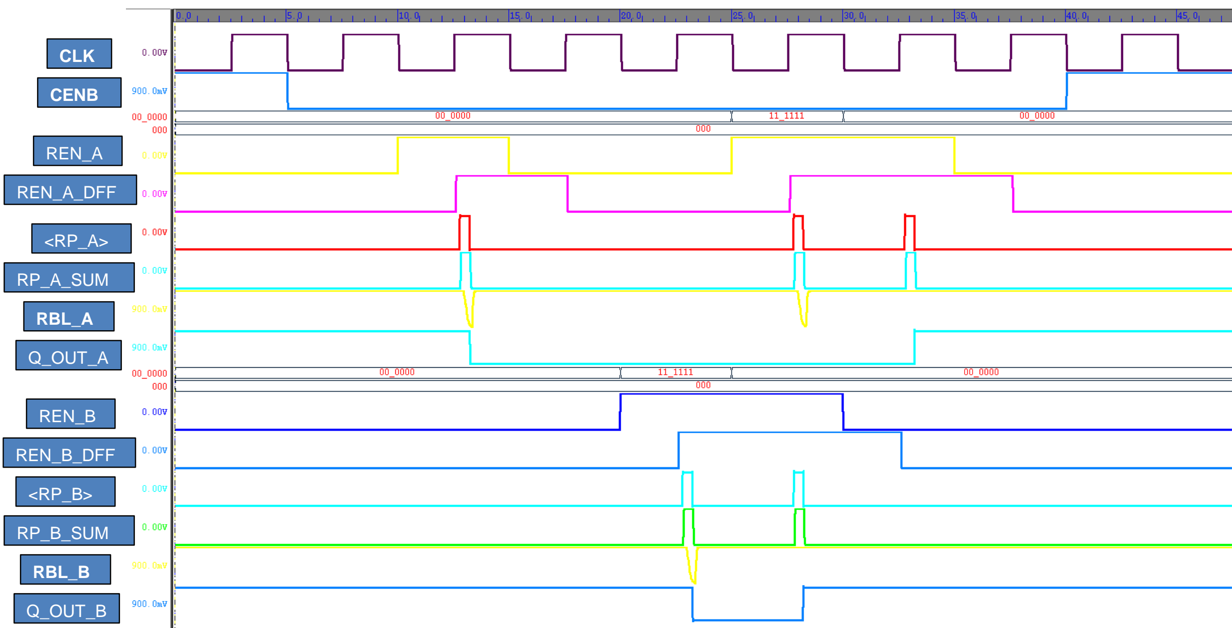


Fig. 3.25 Read test function for 2R2W multiport SRAM chip

3.6 Post-layout Simulation

3.6.1 Performance

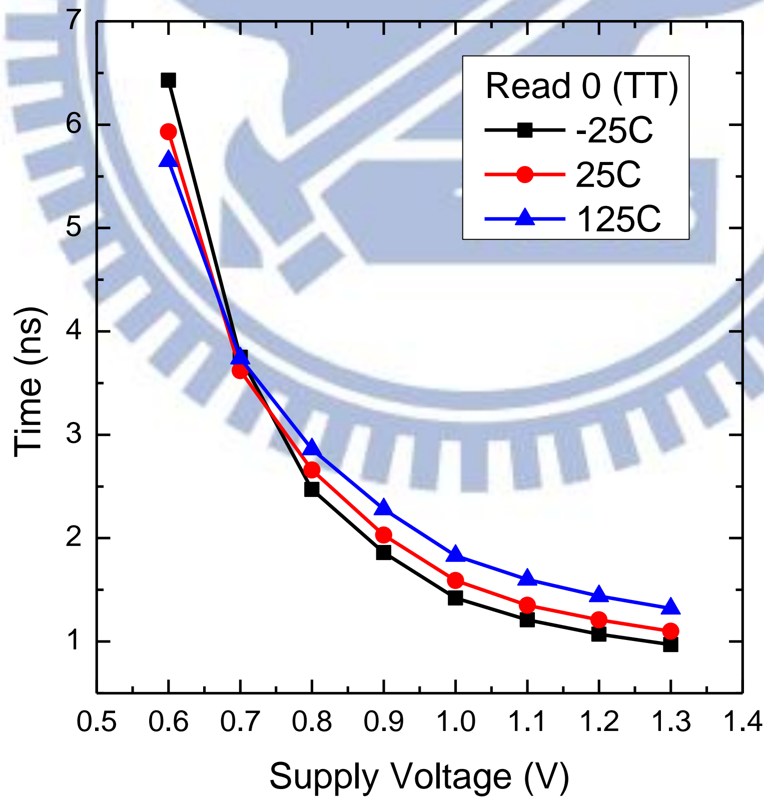


Fig. 3.26 Read "0" speed with different voltage

Fig. 3.26 shows read "0" performance in this chip, read time is domain the cycle time

in this 2R2W multi-port SRAM. Read buffer is like conventional 8T SRAM read buffer, two stack NMOS reduce the read current. Not like write in this design have two assist which is discussed before, read NMOS only uses short channel effect to gain more Ion current. Fig 3.27, 3.28 show write “0” and write “1” performance in different supply voltage.

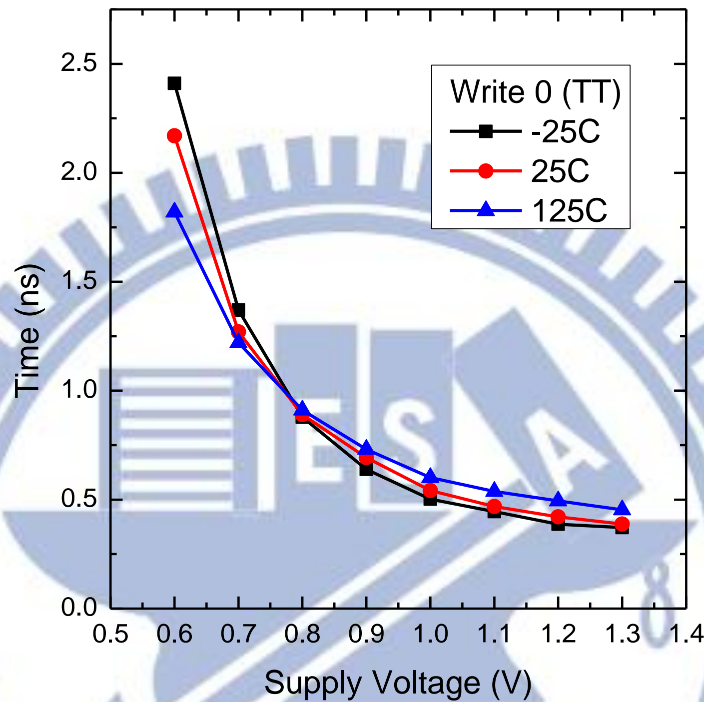


Fig. 3.27 Write “0” speed with different voltage

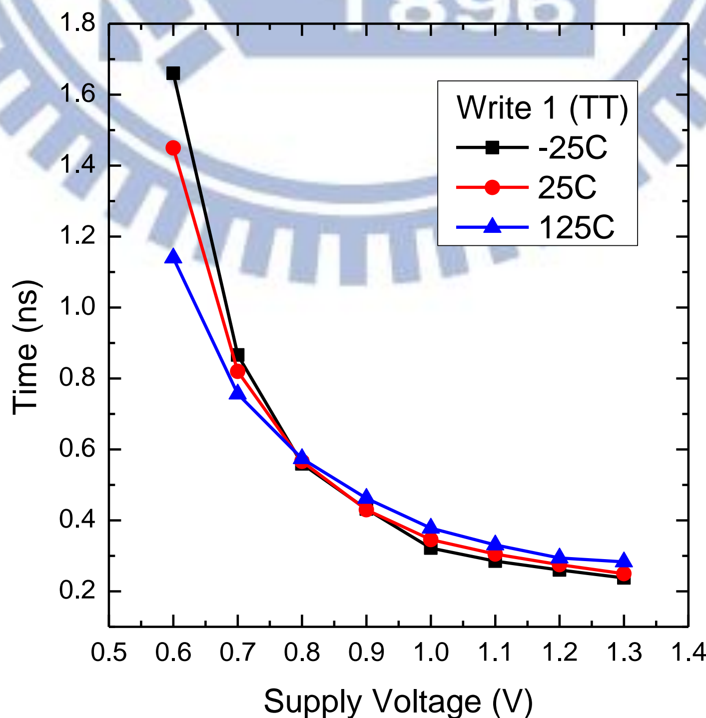


Fig. 3.28 Write “1” speed with different voltage

Fig. 3.29 (a) shows write address conflicts detect circuit delay, delay time raise significant in low voltage supply. Right figure 3.29 (b) is read time compare with worst write time (write operation time and wen conflict delay time). Time different between read "0" and write worst case is near double timing. So when moderation the W/R performance, it must be very take care of read "0" operation.

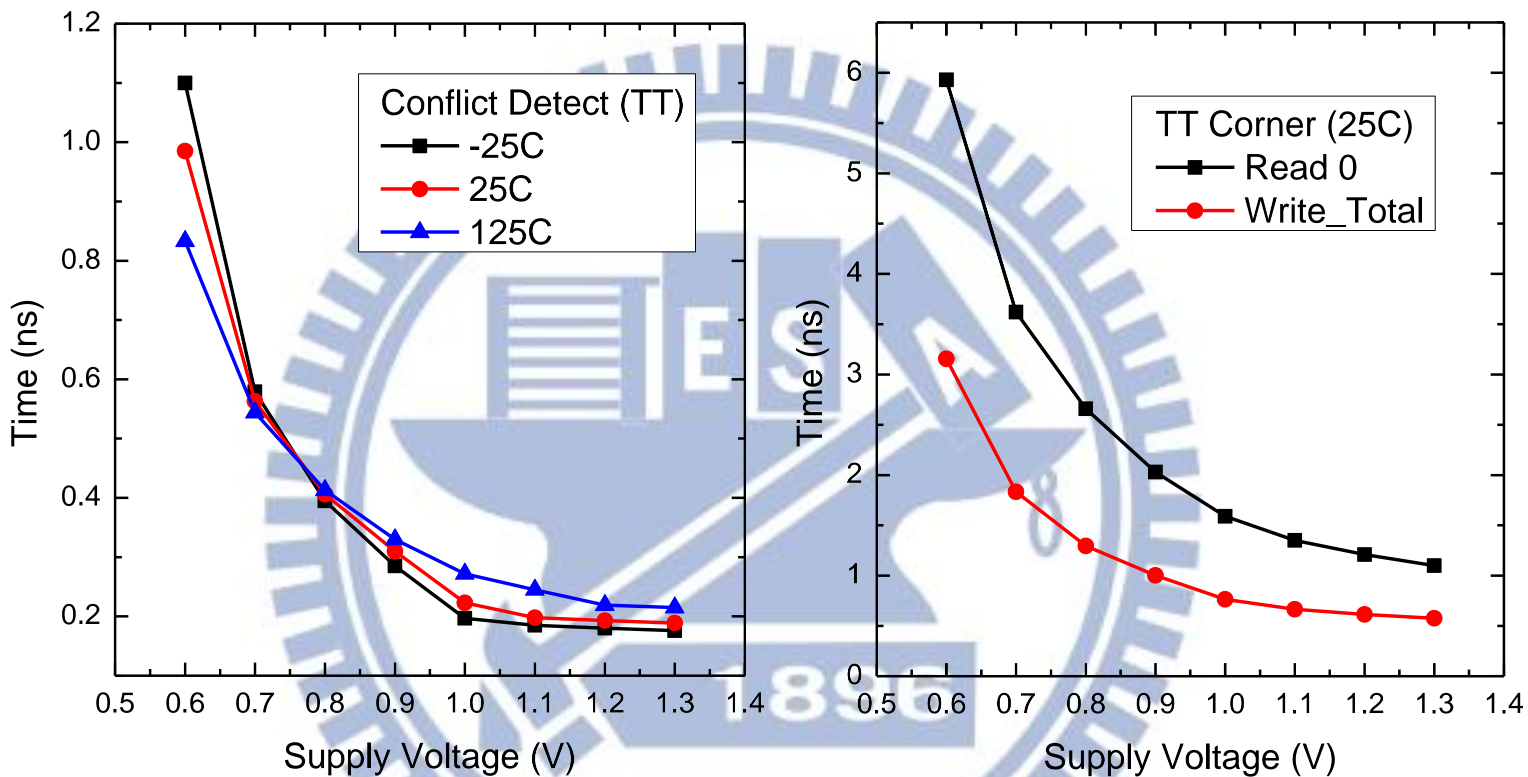


Fig. 3.29 (a) Write conflict detect delay

Fig. 3.30 (b) Read "0" performance compare with write worst case

Fig. 3.31 shows different corner verse Read/Write performance for supply voltage = 0.6V. FF speed is the faster one, and SS is the slowest on this case. SS change clear in different corner and SF and FS performance is near equivalent between SS &FF. Fig. 3.32 shows temperature effect on read/write operation, post simulation shows that temperature effect for 0.6V is smaller than corner effect.

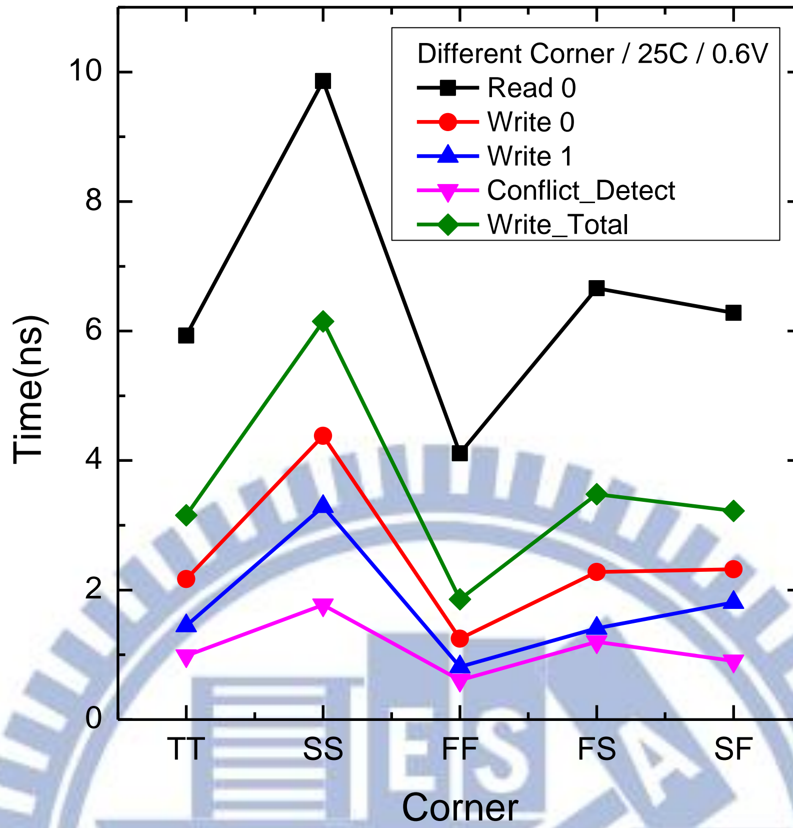


Fig. 3.31 Performance variation for different corner

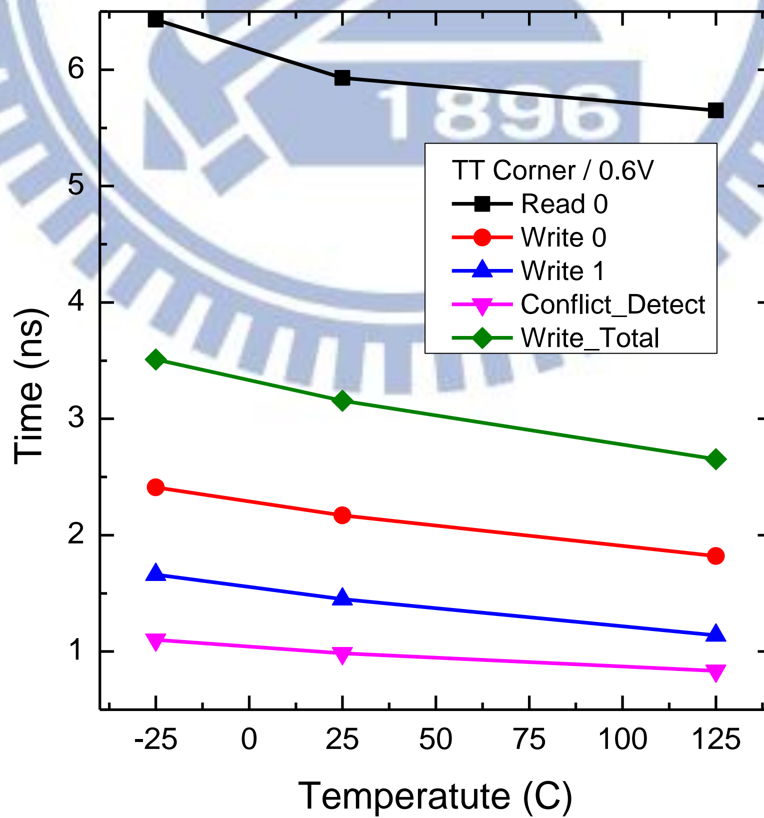


Fig. 3.32 Write “1” speed with different voltage

3.6.2 Power Consumption

Fig. 3.34 shows A/B port write operation and read operation power dissipation. Not like conventional, share WBL structure make the write driver reduce to 1/2. Not only share WBL structure, but also single-end bit-cell structure. A new bit-cell structure can supply bit-interleaving, no half select issues when W/R on the same row. A shorter access time reduces active power consumption and this design with faster operation frequency. Power reduction compare with conventional 8T cell operation reduce more than 30% in active mode (Fig3.33).

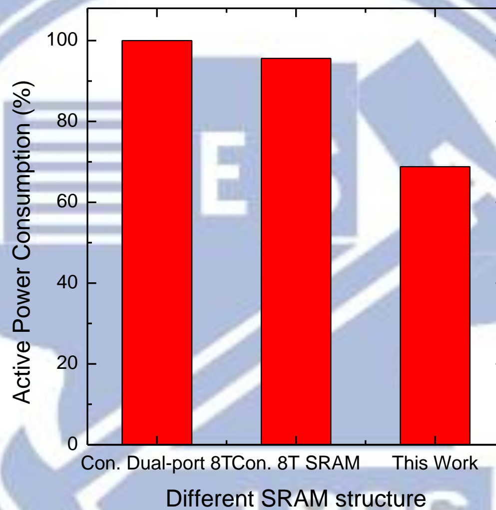


Fig. 3.33 Active power consumption

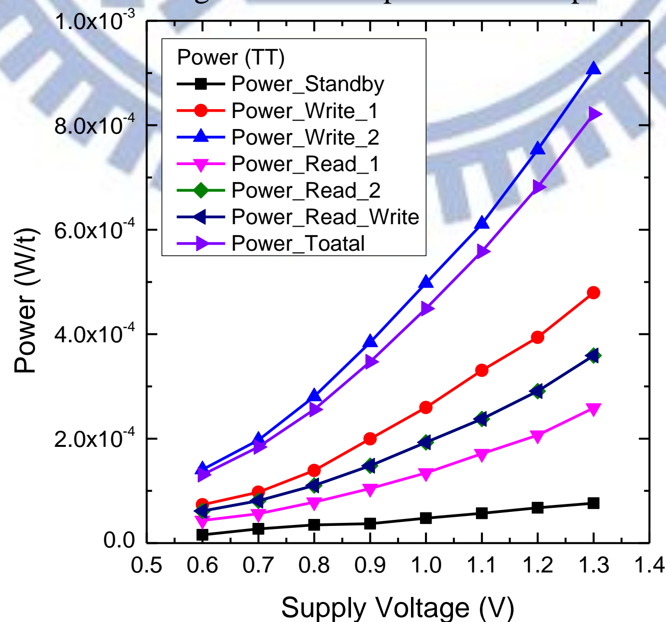


Fig. 3.34 Power consumption with different voltage

In order to find out the Min. energy point, P x T is showed in Fig.3.35. Although this cell can lower voltage down to 0.5V, energy is not the lowest one. Timing delay is raised significantly in low supply voltage, it make energy product more than 0.7 V. Read power is more than write because too longer delay make power delay product bigger. Operating in 0.7V voltage scale is the best choose for low power dissipation.

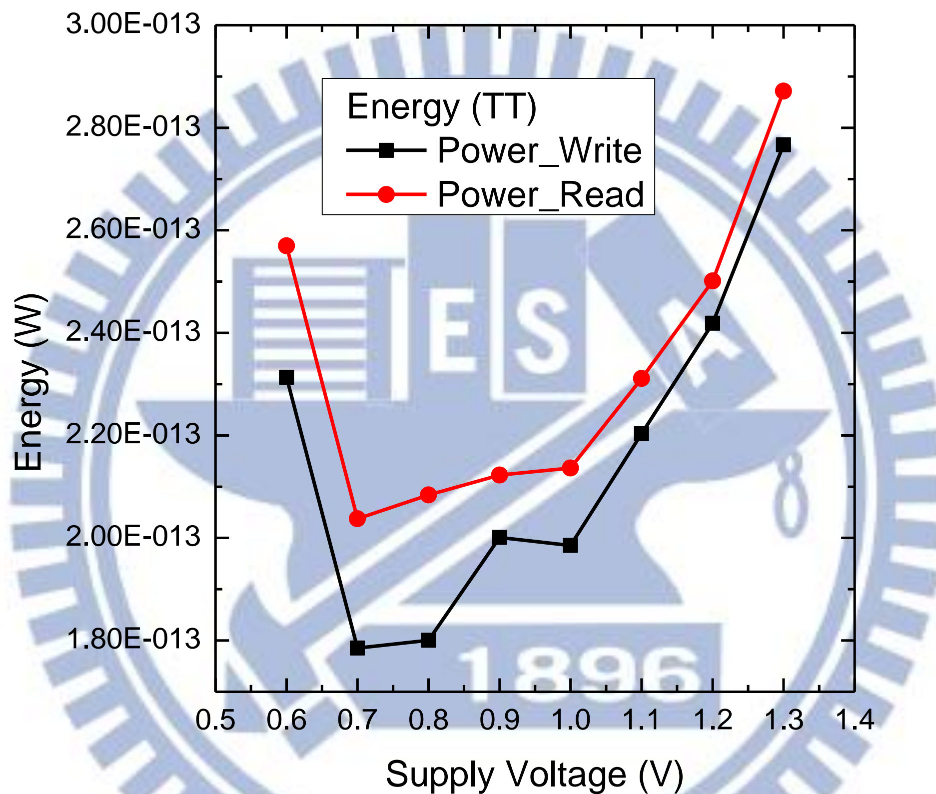


Fig. 3.35 Power delay product with different voltage

Next page is the tape out chip view, and pin name. Pin number in this test chip is 68 pins and total area is 890um x 1090um. This chip is tape out by CIC and use TSMC 40nm general purpose design. Pins are stacked for reducing area consumption, in this way pin and pin each other is very tightly.

		51	50	49	48	47	46	45	44	43	42	41	40	39	38	37	36	35		
		Q_A_SE	Q_A_R	Q_A_CKSEL	WADD_B_SN	WADD_B_R	WADD_B_D	VDDPST	VSSPST	AVDD	VDD	VSS	WADD_B_CKSEL	WADD_A_SN	WADD_A_R	WADD_A_D	WADD_A_CKSEL	RADD_B_SN		
52	Q_A_SI																		RADD_B_R	34
53	Q_A_SN																		RADD_B_D	33
54	Q_B_CKSEL																		RADD_B_CKSEL	32
55	Q_OUT_A<0>																		RADD_A_SN	31
56	Q_OUT_A_<15>																		RADD_A_SI	30
57	QOUT_B_<0>																		RADD_A_R	29
58	VDDPST																		VDDPST	28
59	VSSPST																		VSSPST	27
60	AVDD																		AVDD	26
61	VDD																		VDD	25
62	VSS																		VSS	24
63	QOUT_B_<15>																		RADD_A_D	23
64	VSS																		RADD_A_CKSEL	22
65	VDD																		DIN_B_SN	21
66	AVDD																		DIN_B_R	20
67	AVDD																		DIN_B_D	19
68	AVDD																		DIN_B_CKSEL	18
		ADD_SE	CEN	CLK	CLKSEL	CM	REN_A	VDDPST	VSSPST	AVDD	REN_B	TDELAY	WEN_A	WEN_B	DIN_A_CKSEL	DIN_A_D	DIN_A_R	DIN_A_SN		

Fig. 3.36 2R2W multi-port SRAM pin name

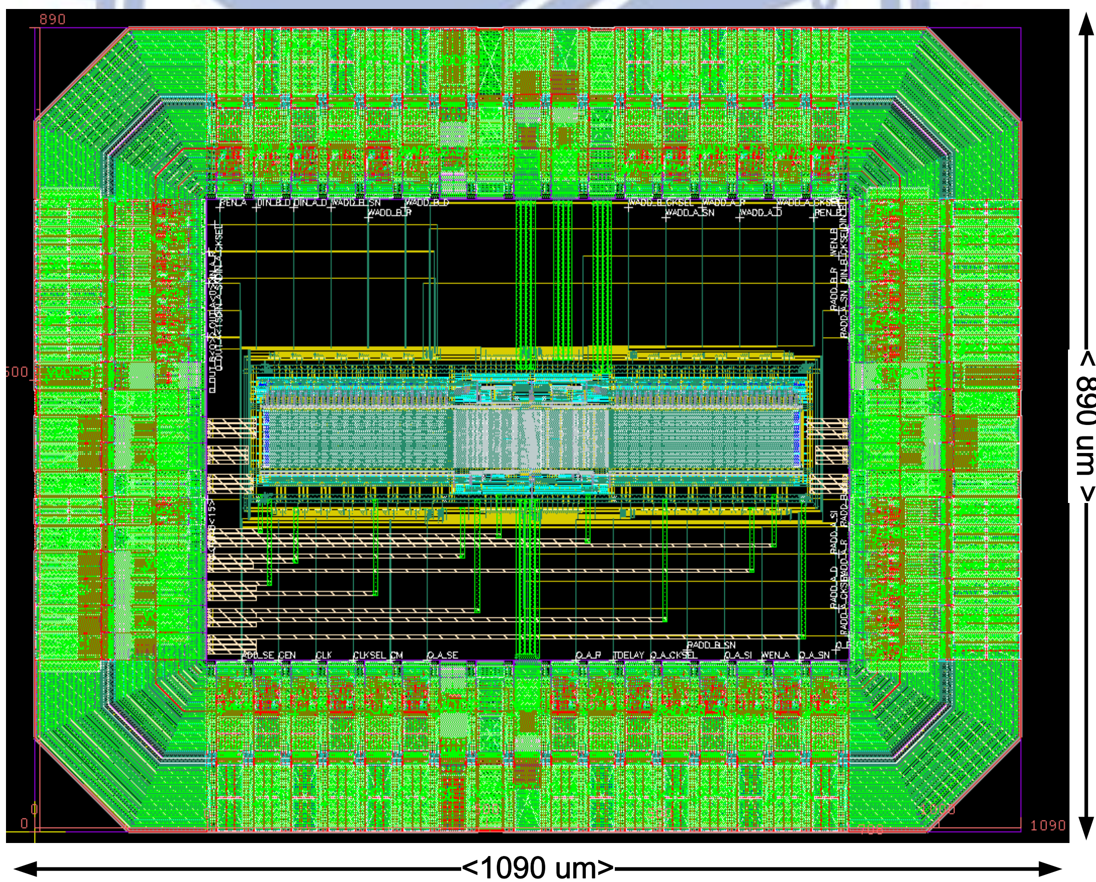


Fig. 3.37 2R2W multi-port 8K SRAM test chip

3.7 Summary

An 8K 2R2W 13T SRAM array design is presented in this chapter. The new cell structure can supply bit-interleaving structure, and no disturb issues compare with convention dual-port SRAM cell. Wide range operation from 1.2Vdd to 0.4V is more flexible by user. There are many low power designs for low power reduction, such like share WBL, CLK gating, and power gating are used in this design. The new share WBL structure can reduce active power about 60% compare with conventional.

The chip is already tape out on Sep. 1 by CIC. This cell can operate under wide operate under wide operating voltage (VDD=1.4V~0.5V) that can cover all process and temperature variation. By post-layout simulation result, this 8K 2R2W multi-port SRAM can operate at 475 MHz at VDD=0.9V, TT corner and 25C and it also can operate at 150 MHz at VDD=0.6V, TT corner and 25C. The power consumption of read operation and write operation in VDD=0.9V, TT corner and 25C are 0.115(uW/t) bit-cell and 0.0692 (uW/t) bit-cell. The VDDmin is 0.5V in TT corner and 25C.

Chapter 4
Low-Power Register File Designs and New Bit-Cell Structure

4.1 Introduction

Lower VDD_{MIN} operation can achieve orders of magnitude low power consumption compare to convention super-threshold operation. In these year, near threshold is a new region for energy reduction than sub threshold region. Although operation in sub-threshold can reduce many orders than super-threshold, long operation time is needed. By this reason, it average total energy consumption in sub threshold region.

Nowadays, such as medical devices, portable devices, sensor networks and wireless body area network (WBAN) where performance is not constrained. Register file play a very important role in many process or SoC application. Not like the SRAM, register file need high bandwidth, high operation, and very robustness. In order to gain more bandwidth, multi-port structure is added. Increase port number can enlarge more bandwidth but area and power are overhead increasing at the same time. These kinds design can easily find in Intel or AMD CPU core [4.1] [4.2]. Not only power dissipation, more port operation in the same time degraded the cell SNM margin and longer access time. In this chapter, a low power multibank architecture for simultaneous access with collision detecting technology is proposed. Timing sharing for double data Read/Write can reduce CLK rate, and do a twice operation in one cycle. Multi thread is suit for register file switch when access is stall, by pipe line structure can gain more access efficiency. Combine with this new low power and high speed technology, a new near threshold supply voltage base register file is proposed. This register file can be applied to superscalar architecture or VLIW (Very Long Instruction Word) DSP.

The rest of this paper is organized as follow. An overview of recent low power register file cell is shown in section 4.2. Section 4.3 describes the register file architecture of bank based. Section 4.4 presents timing sharing technology to reduce port number, area and power consumption. Section 4.5 shows multi-thread switch structure and use in pipe line base. Finally, summary are given in section 4.6.

4.2 Previous of Low Power Register-File Design

4.2.1 Power Reduction [4.3][4.4]

Power gating has been widely used to reduce sub threshold leakage. However, the efficiency of power gating degrades very fast with technology scaling, and we can see the data plot in Fig.4.1. This is due to the gate leakage of circuits specific to power gating, such as storage elements and output interface circuits with a data-retention capability. A new scheme called supply switching with ground collapse is proposed to control both gate and subthreshold leakage in nanometer-scale CMOS circuits.

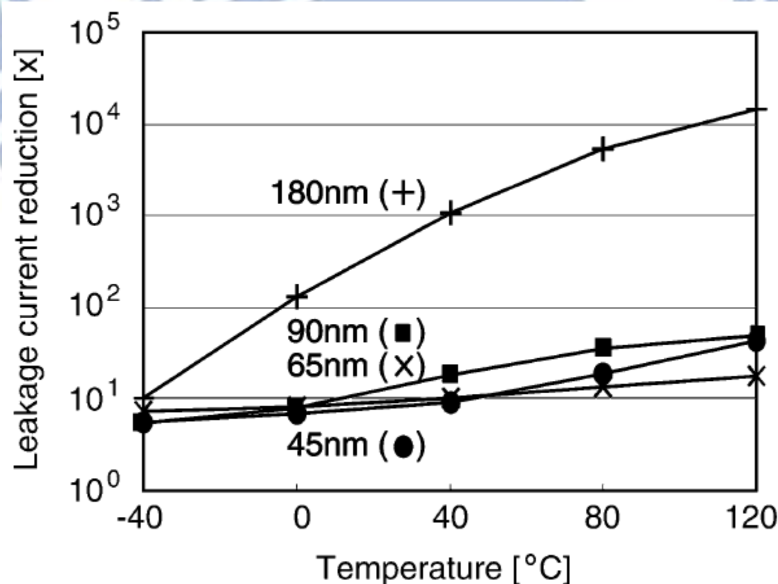


Fig. 4.1 Efficiency of power gating circuits

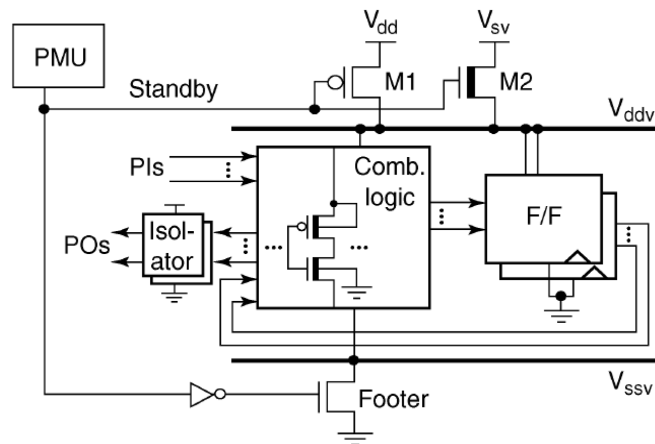


Fig. 4.2 Supply switching with ground collapse

Using SSGC, the component of gate leakage in storage elements is reduced by dropping the supply voltage, while power gating largely eliminates sub threshold leakage in the combinational circuits. Fig. 4.2 shows the SSGC concept. When the circuit is in active mode, the normal supply voltage is applied through supply control switches and the footer is turned on. When the PMU detects1 that the circuit is in standby state, it steers the supply control switches so that the standby supply voltage is applied to the circuit. At the same time, the footer is turned off and sub threshold leakage from the combinational logic is eliminated.

4.2.2 Banked Register File Architecture

Multi-banked structure is proposed in these years, with a low wiring coupling effect, area reduction and low power goodness. The drawbacks of using multi-port structure such like complex circuit, larger bit-cell and heat problem (high speed operation). By using bank structure in advanced process technology, a high speed and low power register is come to true. Besides lowering supply voltage for power issue is more available in multi-banked register file architecture compare to conventional register file structure. Not like conventional a multi-port with large bank, a small bank structure can reduce the BL loading significantly especially in low power operation mode.

Conflict issues should be considerate tightly in convention multi-port register file design. For data robust, bank-based structure has to do a priority port decision for port conflict problem. It prohibits that two ports write data in the same cell or write/Read in at the same moment. If this phenomenon happened, the cell node will in unknown statue. If no take a conflict circuit design to solve it, the register file function will not correct.

In [4.5], a small 1R1W SRAM cell with multi-bank structure is proposed. With 8-read, 8-write port, 64-Kbit, 32-bit word-length SRAM design with multi-bank architecture is reported. Using a 2-stage-pileline, a multi-stage-sensing scheme and a 2-port SRAM cell, high speed and high stability access is achieved simultaneously.

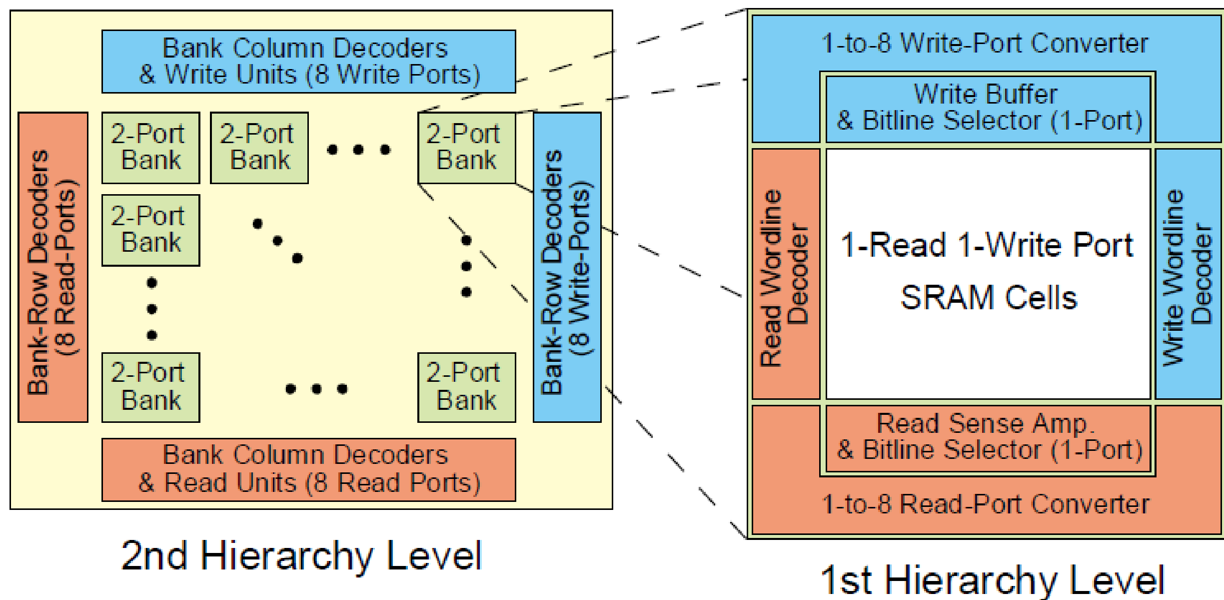


Fig. 4.3 16-port SRAM architecture with 2-port banks and distributed crossbar

In this paper bank-based multi-port architecture called hierarchical multi-port memory architecture, HMA [4.6], has been developed, to drastically reduce the area consumption by using 2-port banks and a distributed crossbar switch for large-port-number capability. At the same time, memory-access time and power dissipation are also reduced substantially. Access-conflicts, which may happen when a bank-based architecture is adopted, are avoidable by an access scheduling which takes

account of the bank structure.

Fig. 4.3 shows a block diagram of the applied 16 port HMA-memory architecture with distributed crossbar. The bank modules of the 1st hierarchy level consist of a 2-port SRAM core into which the 1-to-8 read-port and 1-to-8 write port converters of the distributed crossbar are integrated.

Not like the conventional large sensing, although it can do very robust, area is many large. The read access path of the 2-port SRAM core is shown in Fig. 4.4. We use a 2-stage sensing scheme for reading data from 2-port SRAM cells within an accessed bank. Local bit-lines are connected to only 8 cells (1st stage) and 4 local clusters are connected to the global bit-lines (2nd stage).

Fig. 4.5 shows the schematic diagram of a part of the bank-internal 1-to-8 read-port converter, which adopts dynamic CMOS technology with DOMINO logic, and the repeater structure for bank columns and rows until the output latches.

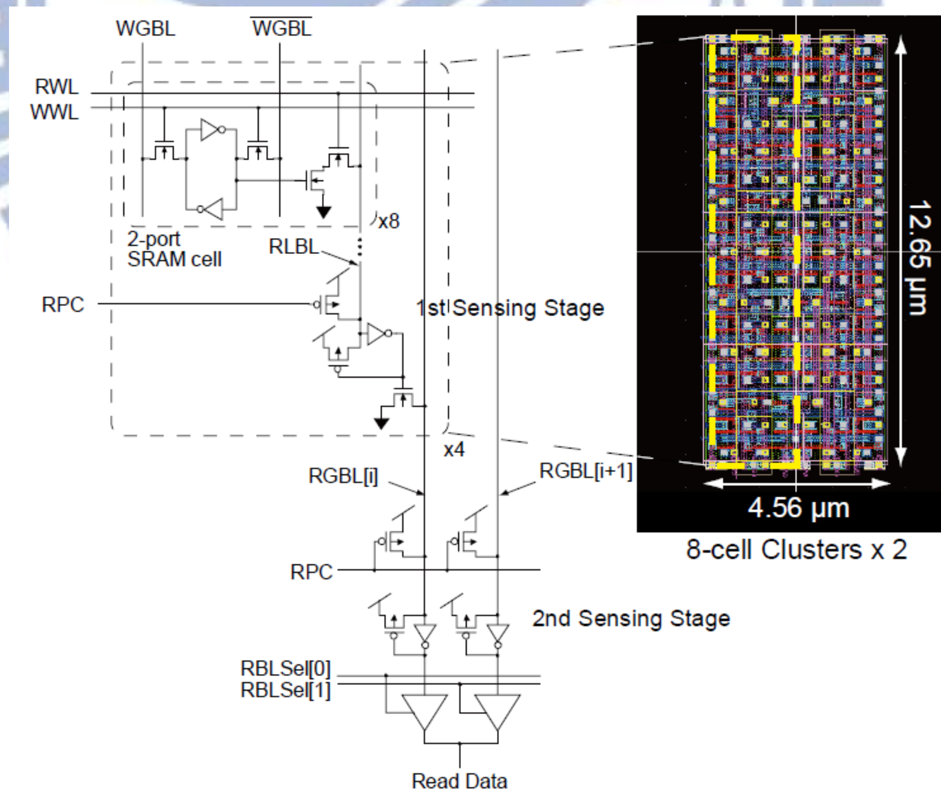


Fig. 4.4 Read-data path of the 2-port SRAM within each bank

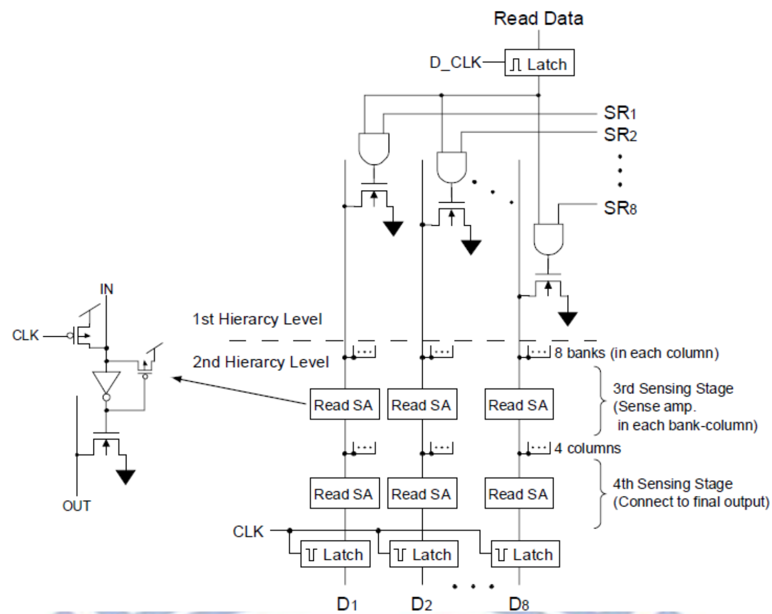


Fig. 4.5 The bank-internal 1-to-8 read-port converter,

4.2.3 Tri-state Register File Design

In this paper [4.7], the author proposed a novel integrated circuit and architectural level technique to reduce power consumption of register files in high performance microprocessors. Add a new state between "ON" or "OFF", which is named "Drowsy state" (Fig. 4.6). In order to achieve low leakage data-retention, if RF cell is not in work state, the gated ground technology is used. Simulation results on 32-nm process show 12.7-14.1% power reduction for ROB (Reorder Buffer)-based microprocessors and 12.4-17.9% power reduction for checkpoint-based microprocessors, respectively, with less than 5% impact on excess time.

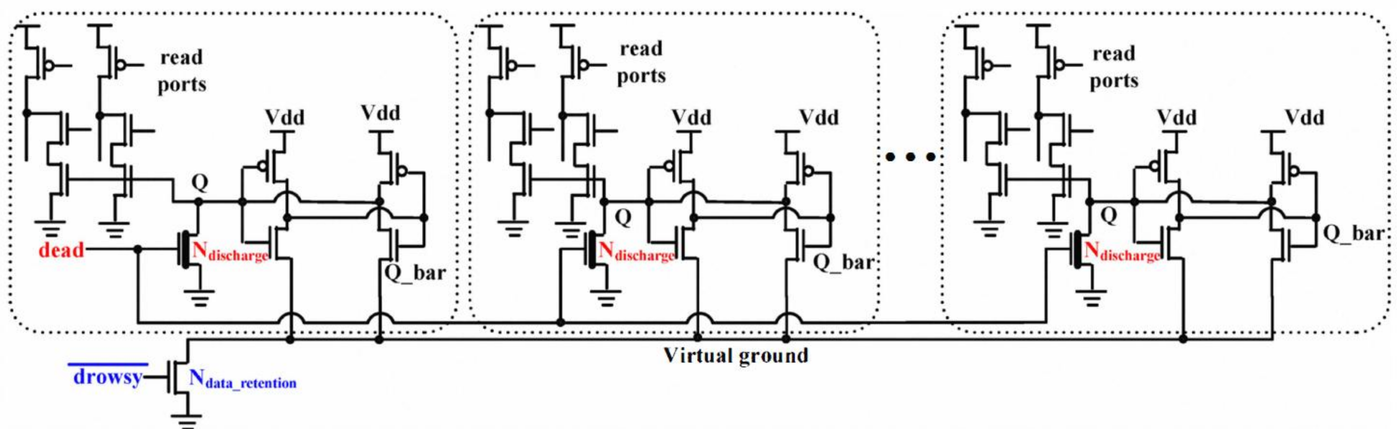


Fig. 4.6 Schematic design of novel register files

4.3 Multi-thread Register File Design

Current processors have reached their maximum operating frequency, and performance improvements must be sought in better organization of the computation. One area for improvements is the tolerance of latency of data caused e.g. by a memory or I/O access, which is usually handled by context switching and executing computation threads that have data available in processors that support multithreading. A new method for improvement SPARC v8 processor performance and the impact of the architectural by thread control is show in [4.8]. Other quantifying thread vulnerability for multicore architecture is discussed in [4.9], a special thread vulnerability factor metric is proposed for solving transient errors form a software perspective.

To implement a power-efficient Chip Multi-Threading architecture this maximizes overall throughput performance for commercial workloads. The target performance is achieved by exploiting high bandwidth rather than high frequency, thereby reducing hardware complexity and power. Combine high Thread-Level Parallelism with multiple and independent processes running concurrently. In paper [4.10] [4.11], the author introduces a new processor in SPARC which has 64 cores and 32 Thread.

In this process structural, each core executable instructions from up to four threads are selected each cycle on a round-robin basis. Fine-grain multithreading is implemented to eliminate thread-switch overhead. When any thread is stalled, the other threads issue instructions in its turn. This effectively hides the pipeline latency for the stalled thread, better utilizing available resources, and increasing overall throughput performance at a relatively modest cost. Adding the logic and registers needed to maintain the state of the four threads increased core area by only about 20%. Fig. 4.7 illustrates how multithreading effectively and efficiently hides latency in a

typical benchmark workload, based on measured results for 32-thread.

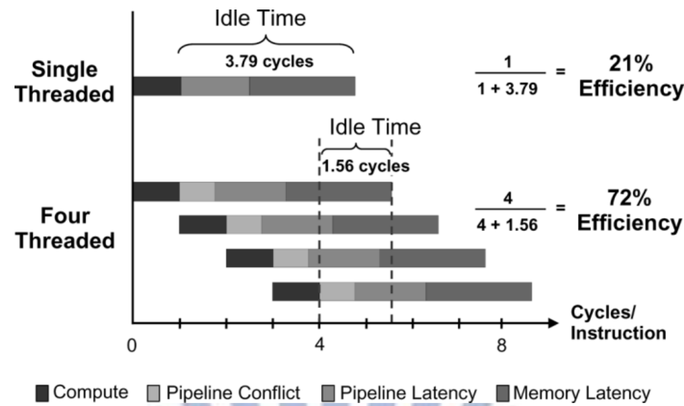


Fig. 4.7 CMT pipeline efficiency with 32 threads

4.3.1 Multi-thread Application Design

Multithreading application can see in [4.12]. In this paper showed that the designing the support for multithreading in pipe line structure. There are some basic operation units for multi-threading CPU can work successful. Thread management control thread switch, and reduce the core in status. This structure is no additional overhead related to embedded memory blocks because previous-unused storage is allocated to distinct register file for multi threads.

In Fig. 4.8, a picture pointed out the data pass through step by step. After program counters are instantiated, a thread status register is also included to indicate the active or inactive state of each thread. Finally, the basic five-stage pipeline includes the necessary logic for data forwarding and introduction of load-use stalls.

a) Multi program counters

b) Thread identifier

c) Thread status register

d) Multiple register file

e) Register identifier

f) Write to the processor register file

g) Instruction for thread management

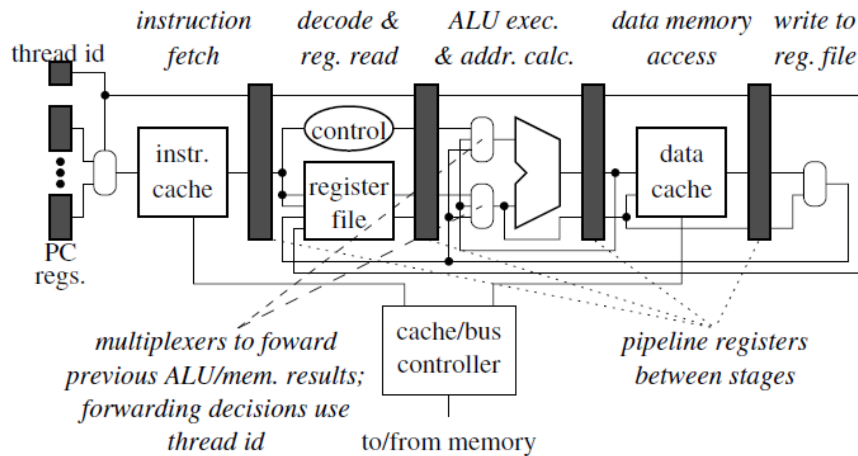


Fig. 4.8 Representation of pipeline and enhancements for multithreading

4.3.2 The Parity Protected Multi-thread Register File

In this paper [4.13] an integer and floating-point register files of the 90–nm generation Itanium Microprocessor are described. A pulsed, shared word line technique enables a 22 ported integer array with only 12 word lines per register.

The register file implements temporal multi-threading by multiplexing the read and write ports to two storage nodes enabling registers to write both foreground and background threads to the same register at the same time. Thread switching completes in one cycle.

Because with two thread for switching in RF cell, there is one more thread address in decoder. Two additional signals, thread and threadbar, are routed in the third layer completing the list of control signals. Two different decode techniques are used. In the IRF, with fewer timing and area constraints than the FRF decode, two pulsed-evaluated decoders are connected in domino fashion. Fig. 4.9

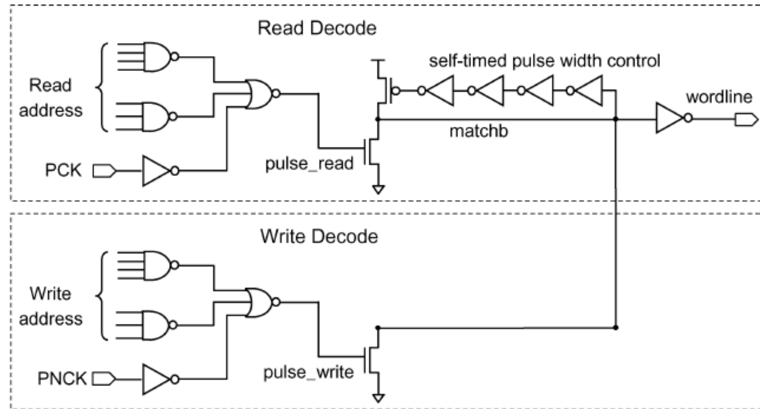


Fig. 4.9 FRF double-pumped pulse clock generator circuit diagram

Read/Write Operation

In this RF cell read & write are also adaptive single-end BL. The memory cell of the IRF incorporates 12 read ports and 10 write ports. Single ended read performance is maintained with a two-level bit line structure saving wires and reducing delay. Both the write of a "0" and "1" are accomplished through an nFET pass gate. SPICE simulations of both writes are demonstrated in Fig. 4.10. Writing a "1" requires special attention as the write data from the bit line suffers a threshold voltage drop across the nFET. A zero value on the storage node b0 is floated during writes by tri-stating the feedback with the signal write l which is asserted low during a write. The trip point of the back inverter is also skewed to allow node nb0 to flip low faster as rises.

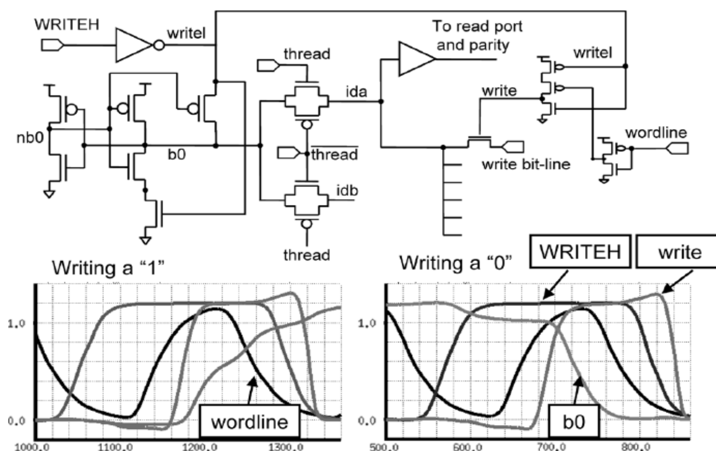


Fig. 4.10 SPICE simulations of both writes "0" and "1"

4.3.3 Thread Switching

To support dual-threaded execution, each register bit-cell incorporates two identical storage cells see Fig. 4.11. Four transmission gates determine the thread selection, where each storage cell is exclusively accessible by either thread. This switching of the I/O to the storage cells removes the third storage node needed using copy-in-place schemes and replaces the complex self-timed signaling to the storage nodes required in such schemes to switch threads with a simple thread id signal (thread) and its inversion (thread bar). Switching threads in this way can induce charge-sharing noise between the storage cells b0/b1 and the internal bit lines ida/idb, which have considerable capacitance load due to the large number of ports. idb, which have considerable capacitance load due to the large number of ports.

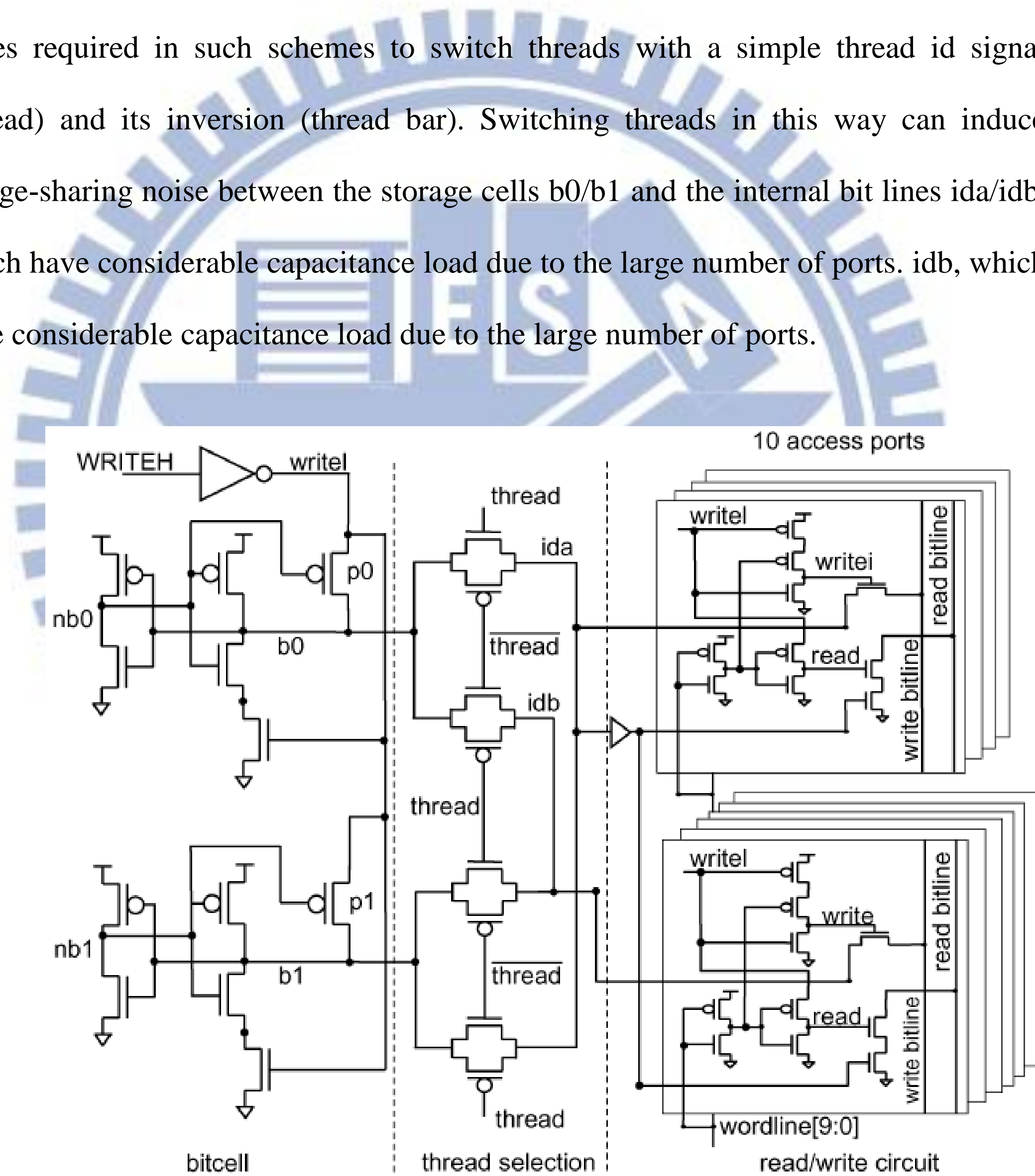


Fig. 4.11 Schematic of two threads switch cell

Others thread switch technology proposed by IBM Power 7 microprocessor [4.14] [4.15] design. In Fig. 4.13 shows the double-bit 6R 4W VRF cell with the associated timing diagram. One addressable location holds two sub-cells. Subcell“0” and subcell“1” store thread 0/1 data and thread 2/3 data, respectively. The six read ports are divided into two groups of three operands, with each set belonging to one thread. The thread select signal that selects sub-cell 0 or 1 is done inside the cell via the two-way built-in multiplexer. The multiplexing inside the cell is done via two OR–AND–INVERT operations. Read ports 0–2 are selected via the first OAI22, and read ports 3–5 via the second one. A duplication of the register files, to support up to four threads, is avoided with this configuration of double-bit cell and cell internal MUX. OR–AND–INVERT in this circuit design can reduce transistors number by Bollinger function. After simplification the logic, transistors counts can down to eight.

The picture is show in the Fig.4.12 and one sub-cell with three read port for parallel data read out operation.

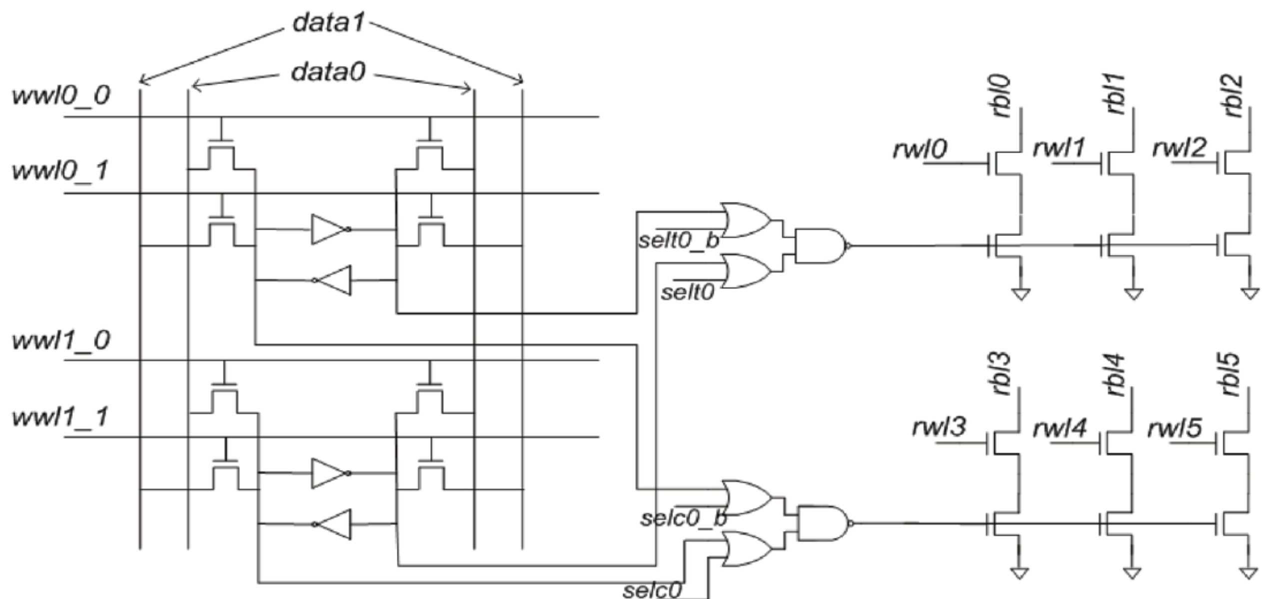


Fig. 4.12 The vector register file cell unit

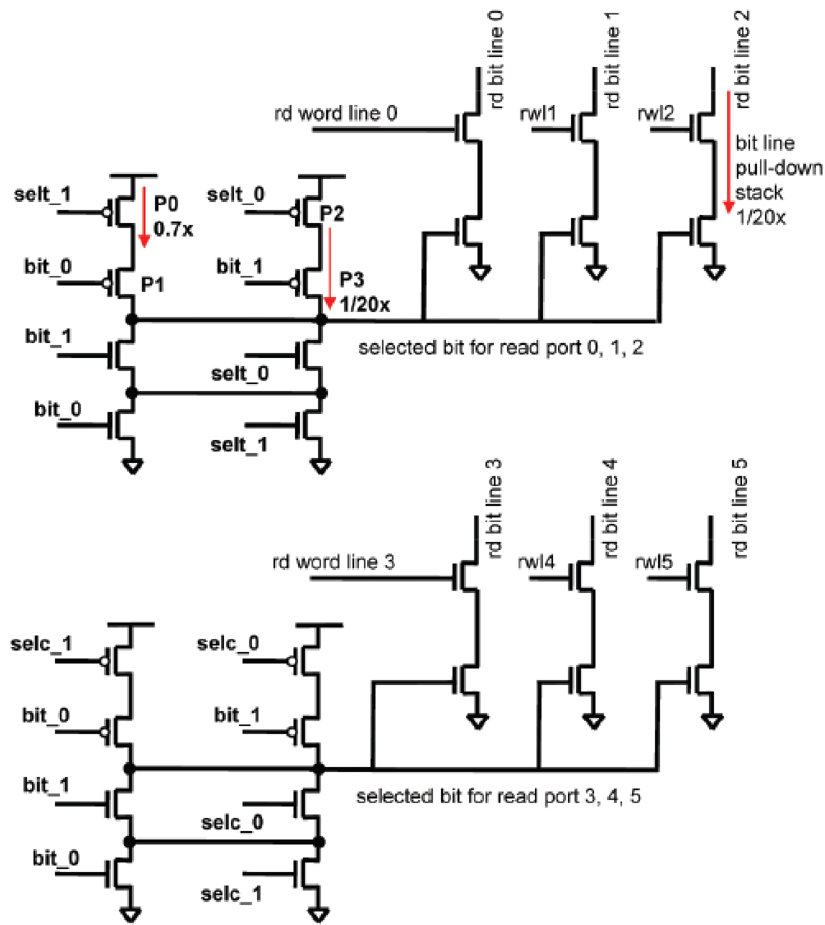


Fig. 4.13 VRF cell capable of holding 2 bits and inter-bit thread decoder

4.4 Timing Sharing Technology

4.4.1 Pervious Work

In this section, I will introduce many methods for reduce power consumption by timing control. Normally register file operation in high speed frequency, and CLK transition very faster bring huge power consumption. Not only power dissipation, a robust high frequency is not easy to implement. In these years, a double pump or many kinds of timing sharing technology are proposed. By low operation frequency, and can do a twice read/write can be approach. For example, the work (Fig. 4.14) is used by DDR first, and used in register for timing control by Kuo's work [4.16]. In the proposed

timing sharing access scheme, local Read and Write ports can be accessed twice in a clock. In other words, a clock is divided into two time slots, and an access operation can be finished in one slot. The waveform is showed in fig.4.13, after one replica pulse finished, a short delay another replica pulse is turn on in the same CLK period.

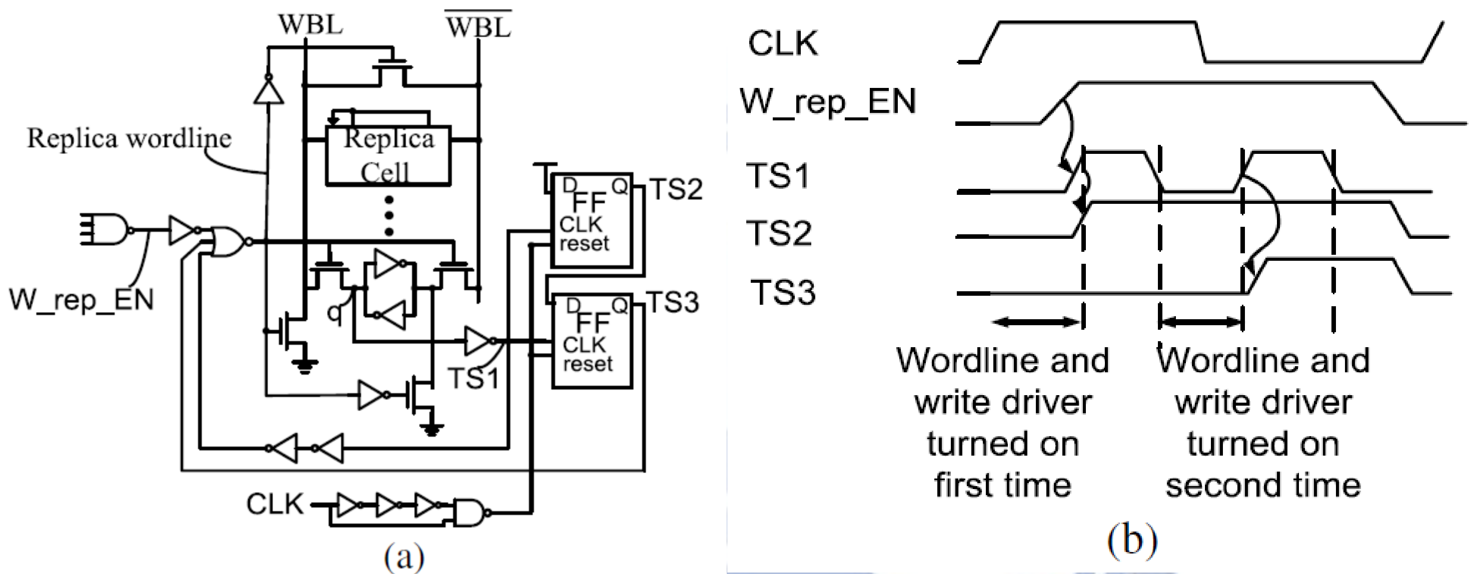


Fig. 4.14 (a) Write replica circuit (b) Signal waveforms of this circuit

After Kuo's work, IBM also proposed a new double pump technology for lower power and reduced R/W port number [4.17]. In order to generate a two pulse, IBM used a CLK chopper to generate two different CLK delay, and separate control the W/R replica pulse operation. But in this work not like the Kuo's work (WL pulse is step by step trigger); timing control between two slots must be very carefully. If two pulses overlay together, W/R error will happen. These problems also refer in this paper, and give a robust timing delay. Fig. 4.15 is the structure double pump proposed by IBM register file.

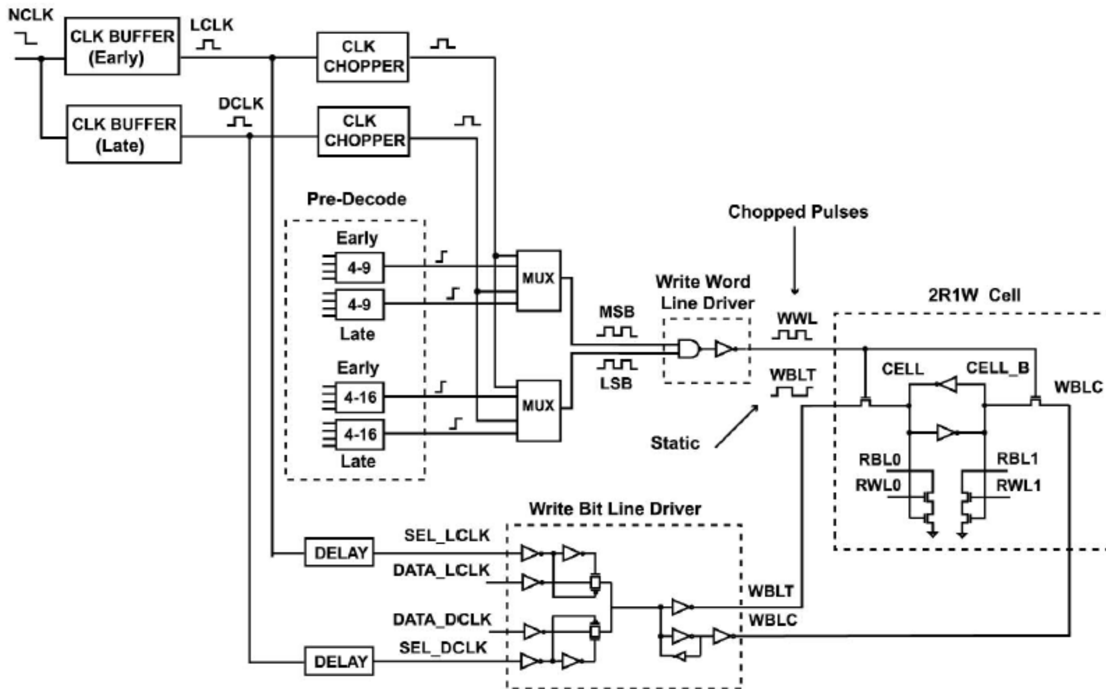


Fig. 4.15 Circuit cross-section of double-pumped write path

4.4.2 Conflict Issues

In conventional register file design, the bit cell operation is like dual port SRAM cell. Like the normal dual port SRAM, when multi-port operation at same time will produce write/ read conflict in half select bit cell. This effect make the register file must have to add another conflict circuit for W/R free. Some technology showed in [18] [19], by use read after write or write first then read go on ……etc. Complex circuit detect for this problem will not appear, and give the register file very robustness. In order to solve this problem, a new cell unit or more safe detect circuit is proposed. Because this already discussed in the chapter 3, in this section will no discussed again.

4.5 This Work

4.5.1 Bank Structure

Applying for mobile or wireless product, low power register file is

needed in these years. So I try to do a low power register and supply multi-thread instruction is proposed. This design not only can do wide range operation but also add data switch technology. In this low power register file design, I make a two bank structure. Use two banks to double the R/W ports form 2R2W extent to 4R4W. The Fig.4.16 is show this design architecture and port number. Each bank with 1K bit, and if four ports no conflicts that can operate in the same time. By adapting multi-bank structure, the cell ports count can reduce to 1/2. Area and power consumption can drop down significantly and only loss a little bank conflict detect time. In this chip, a two stage conflict detect include. 1st level is input pins to inter port priority detect, and 2nd level is bank address conflict detect.

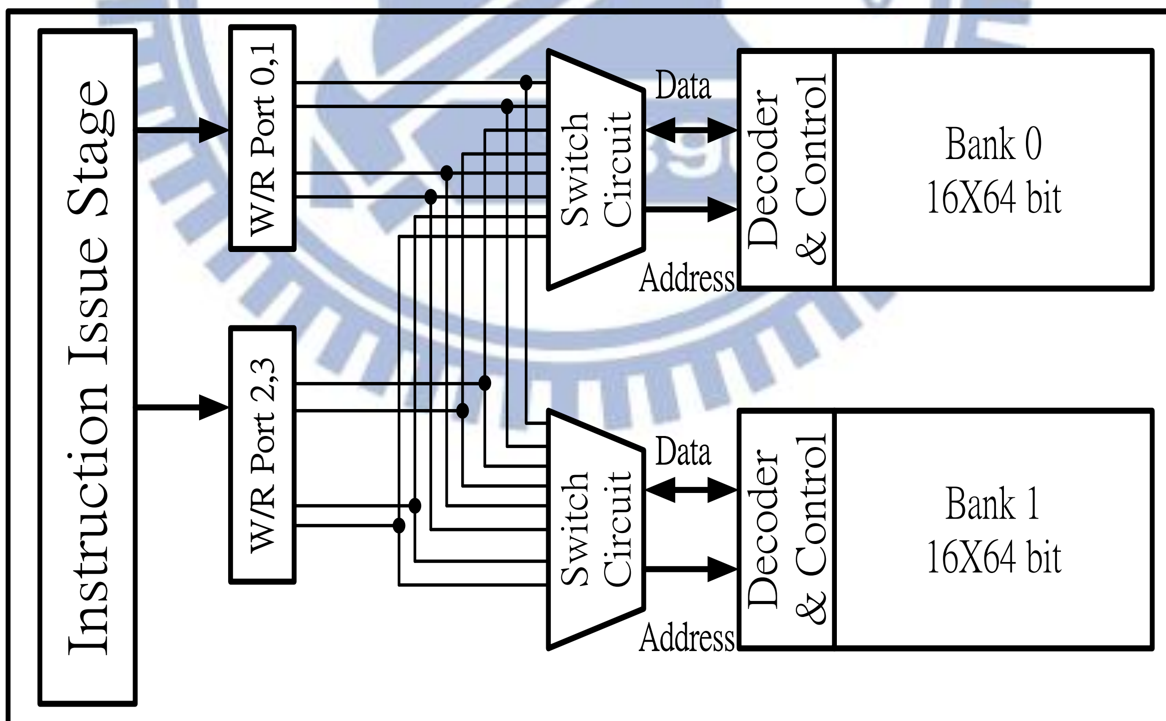


Fig. 4.16 4R4W Multi-Bank structure

➢ 1st Level

Detect WEN/REN signal between the negative and rising edge. There are two bits for enable port address, one is select which bank will select, another one is decision turn on "A" port or "B" port. After finish detected, a true signal can store in DFF and wait for the CLK rising edge to trigger it. The priority is set that (Port0 > Port1> Port2 > Port3). If conflict happen, the high priority port will still turn on. The Fig. 4.17 shows this step operation, and data transmit path.

Contrarily, if two port address point to the same intra-port, the second priority will turn off and give out a conflict signal. By this design, conflict W/R can never happen in register file work. The specific waveform is presented in Fig.4.18. P0_WEN is represent input port WEN signal, and O_P0_WEN signal is stand for true trigger WEN signal after detecting. Because there are two slot in one cycle (double pump technology), I call slot is "S0" and second slot is "S1". S0 &S1 are all needed for conflict detect, so the circuit is doubled.

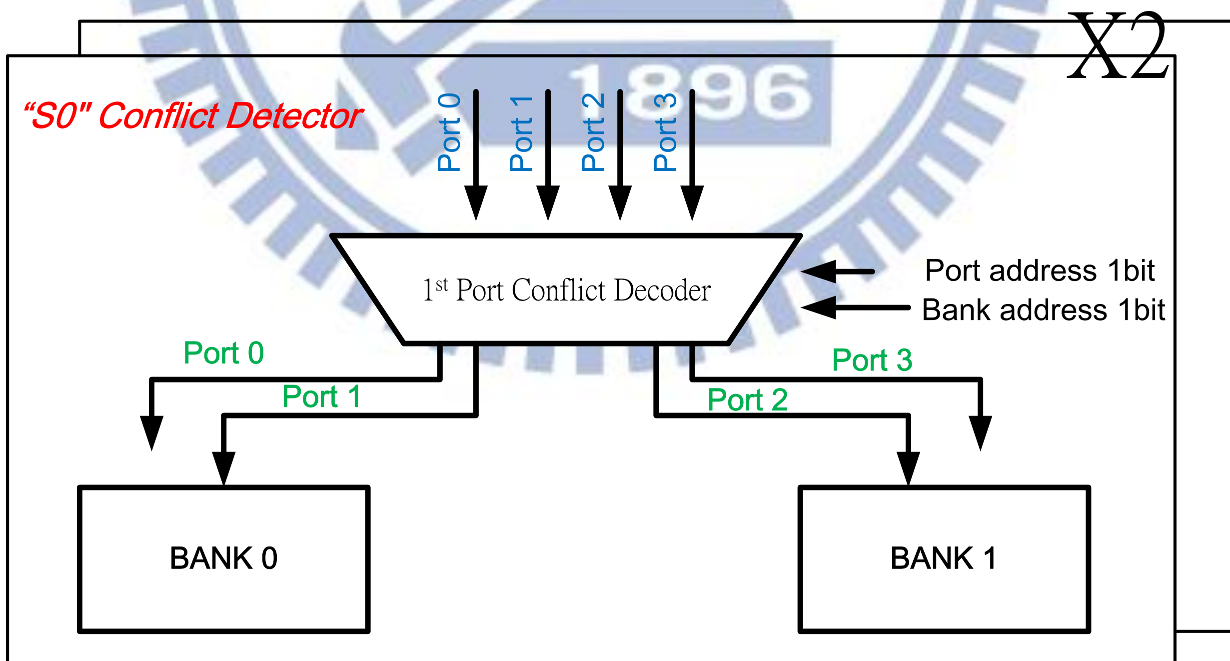


Fig. 4.17 1st Data In/Out conflict detector

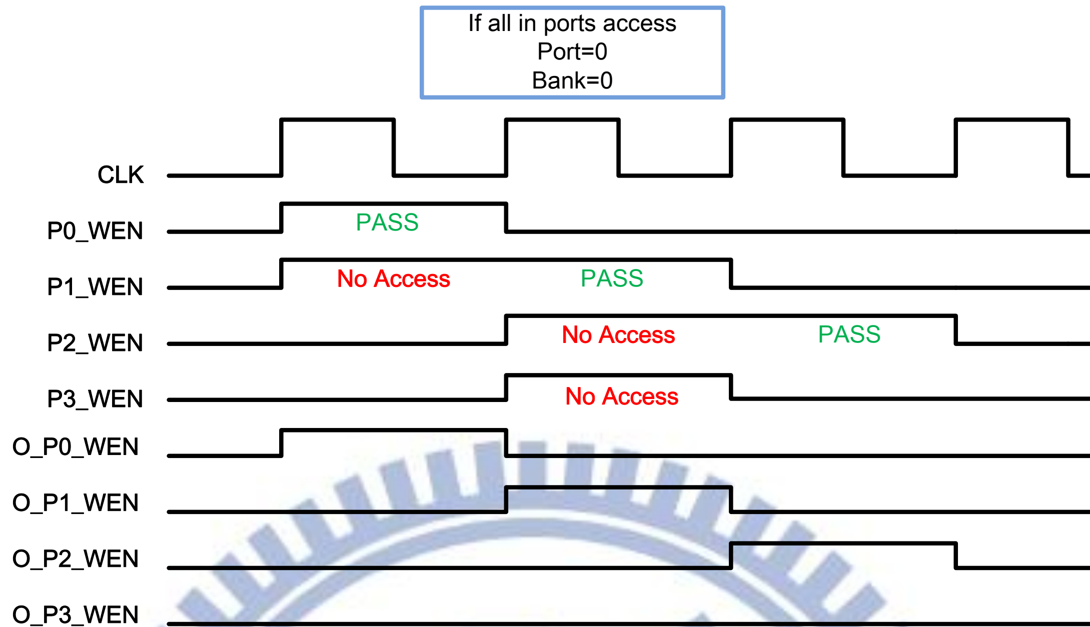


Fig. 4.18 1st Conflict WEN/REN detect waveform

- **2nd Level**

In this part, address conflict is detected. A new data slot conflict switch (DSCS) technology is proposed. Combine with double data rate and one/two operation in one cycle. If data is need high speed, double W/R operation is selected; contrariwise, one operation in one cycle mode can safe more power consumption. High speed and low power mode can switch by user easily.

In this stage, an intra-bank address detected which start by CLK trigger and W/R conflict detect circuit start working in this design. Normally, Read is non-disturbed data and write will change date forever. I set Read priority is always higher than write. Every time CLK rising edge trigger wen signal is go on test mode and REN signal is like conventional data path no need detect. Fig. 4.19 is show the conflict circuit in 2nd conflict detect and DSCS control logic for solve conflict problem. By data switch technology, it can solve many time for waiting data write redo and give the register file more conflict free.

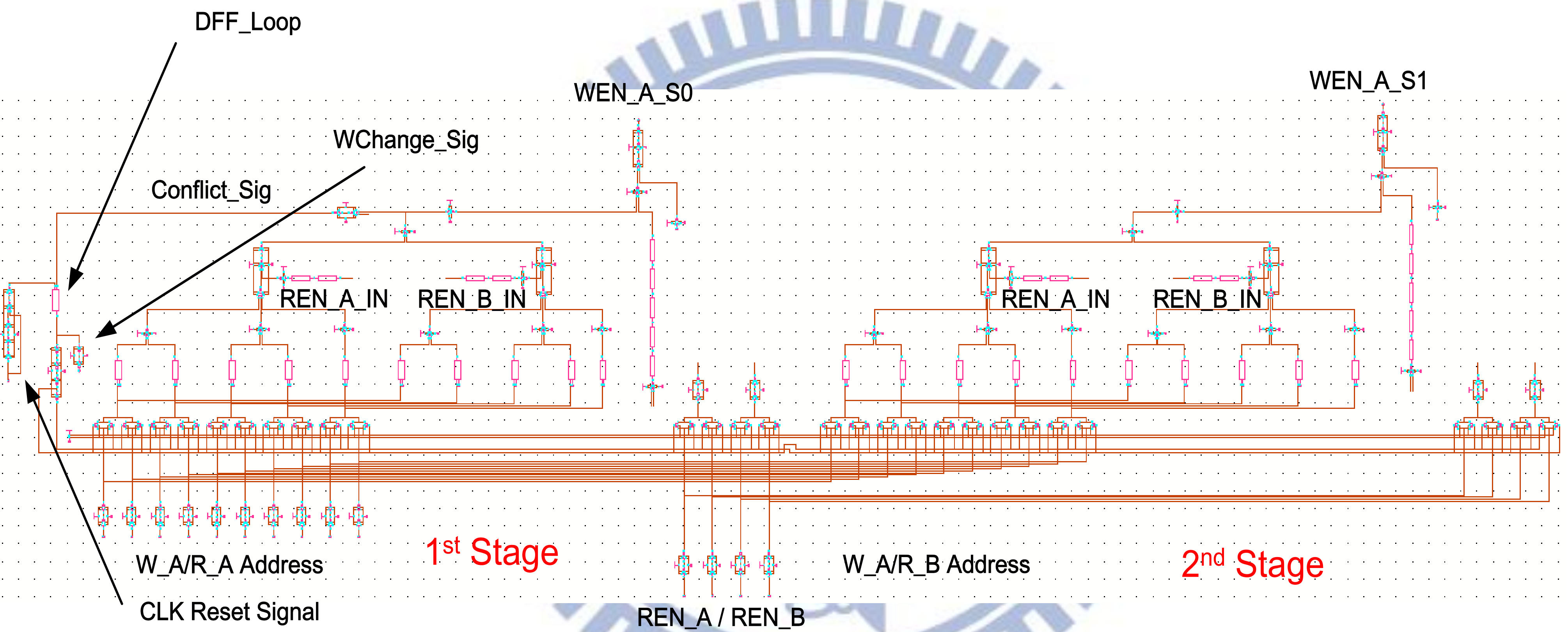


Fig. 4.19 4R4W Multi-Bank structure

Fig.4.20 is a finite state machine in this DSCS circuit. When CLK rising edge signal come, a reset pulse is generated for reset the "WChange" single to 0. The reset pulse is output by inverter delay and AND logic. 1st is set to detect the S0 conflict state, it has to compare to A and B port read address (state "S0"). If one of two addresses is not same each other, a conflict signal will keep "0". And still output the WEN signal after this address conflict detect. In other words; if conflicts happen, a signal will trigger the switch DFF_Loop and switch the input address from S0 to S1. The state is S1 in this finite state machine. If no conflict anymore, the switch operation is finished (state "S2"). Worst case is in "S3", in this state is showed that conflict -> switch -> conflict remain. In this case, no WEN signal will output in this cycle.

By this technology, although data is conflict at first, it keep more than 50% write successfully probability. Not like the conventional design, when conflict coming, this state will not do anything. Timing is save many much compare with conventional design.

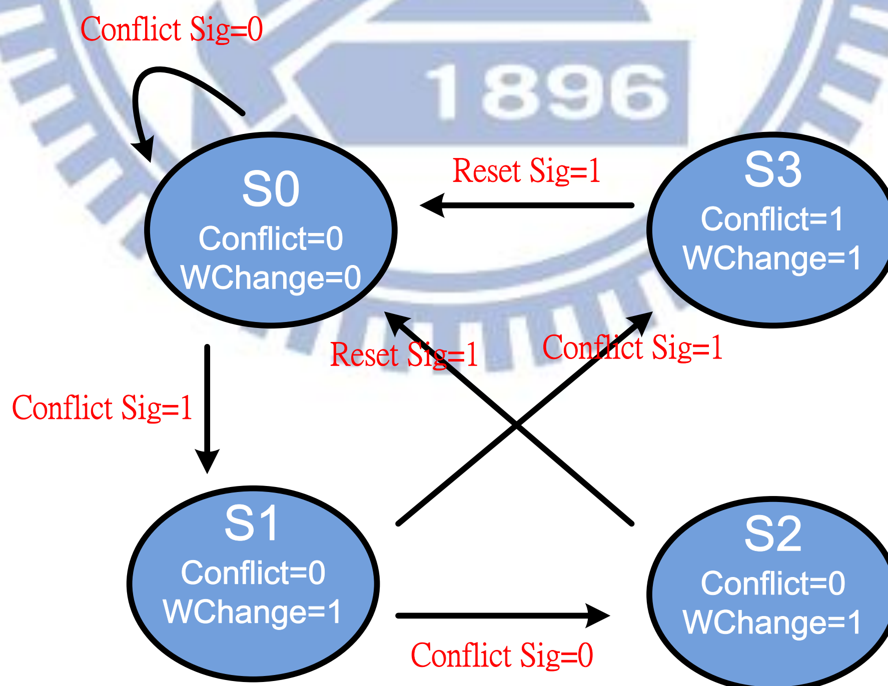
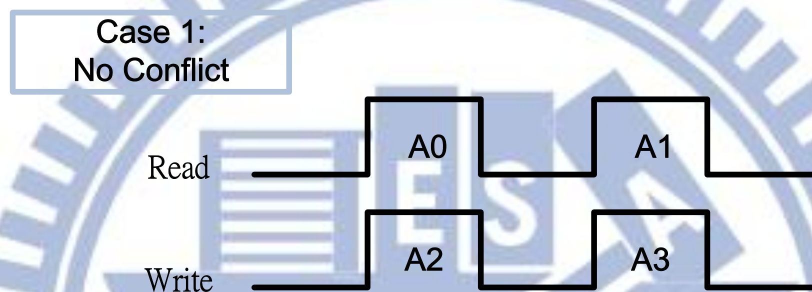


Fig. 4.20 Conflict finite state machine

4.5.2. Read/Write Slot Controller

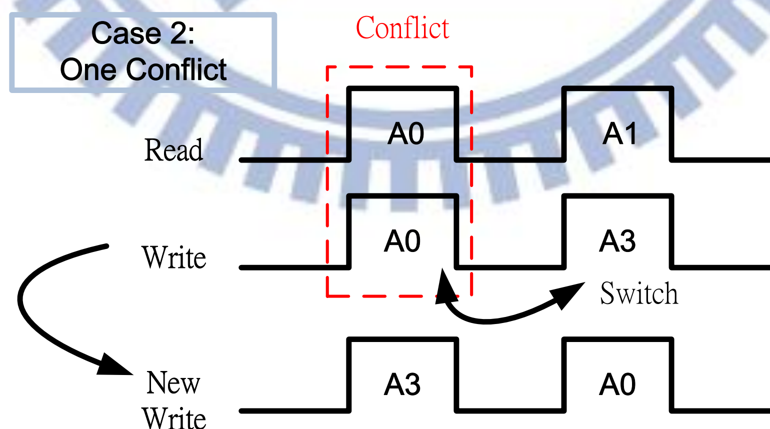
In this section, I will discuss the control slot operation mechanism. With double pump and multi-port characteristic, more special case should be considered. This circuit must be very robust and include all operation case. There are five case will happen in Write/Read operation, I will show that below.

Case 1:



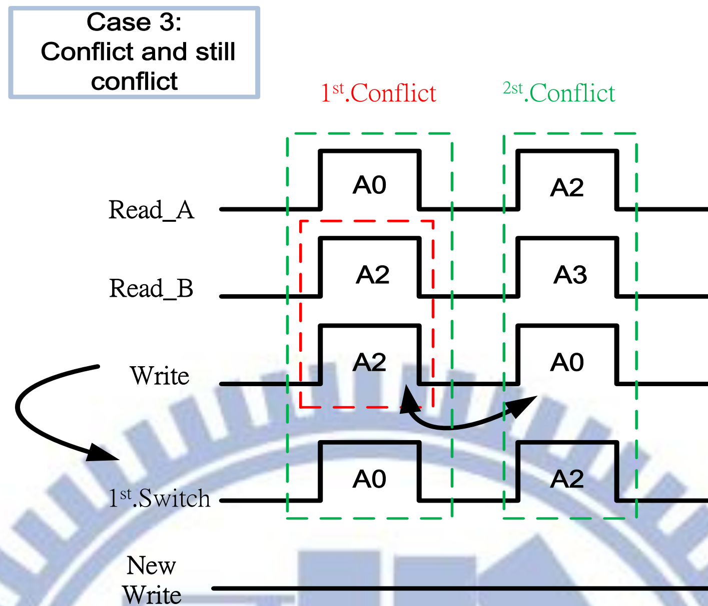
Case 1 is showed that no conflict happed situation. Read port and write port is point to different address. Conflict detect is free, after conflict detect operation WEN signal will output to next stage.

Case 2:



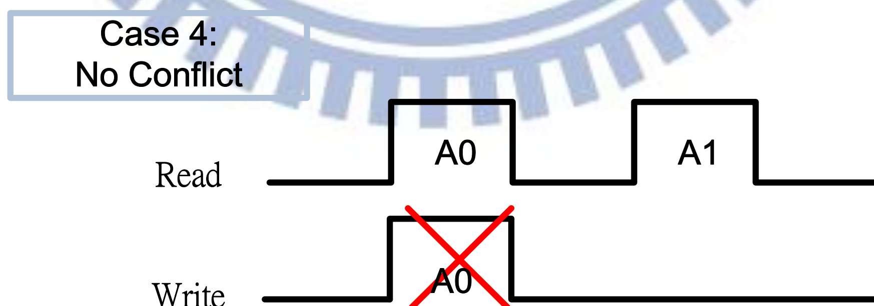
In this case, conflict is happened at first CLK trigger. After detect the address, slot changed by conflict signal. Finally, success twice writes data in the cell in one cycle is accomplished.

Case 3:



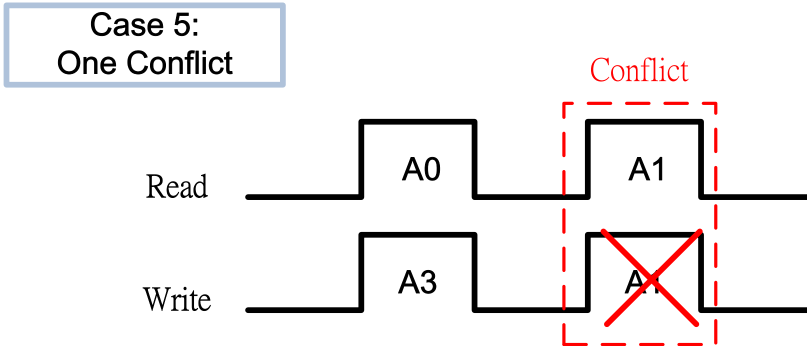
Case 3 is the worst case in these operations. In this case, switch operation still can do anything. At beginning, conflict happened, so a conflict signal is trigger next stage. After switching, address remains conflict to another port. In this case, WEN signal will not turn on any more.

Case 4:



Case 4 is show the S1 is not turn on, and conflict happened at beginning. The address all point to "A0", so in this cycle WEN signal is disappear.

Case 5:



Case 5 is a case with no conflict at first. Conflict happened at second stage, in this situation WEN_S1 is disappearing. Because first slot is not conflict, in this case has not switch operation.

4.5.3 Switch Data Circuit

After conflict detect, if WChange rise to high, this signal will trigger data switch. S1 will pass through in the first slot and S0 comes later. Fig.4.21 showed the D_IO switch output control circuit, and also reset at every CLK trigger.

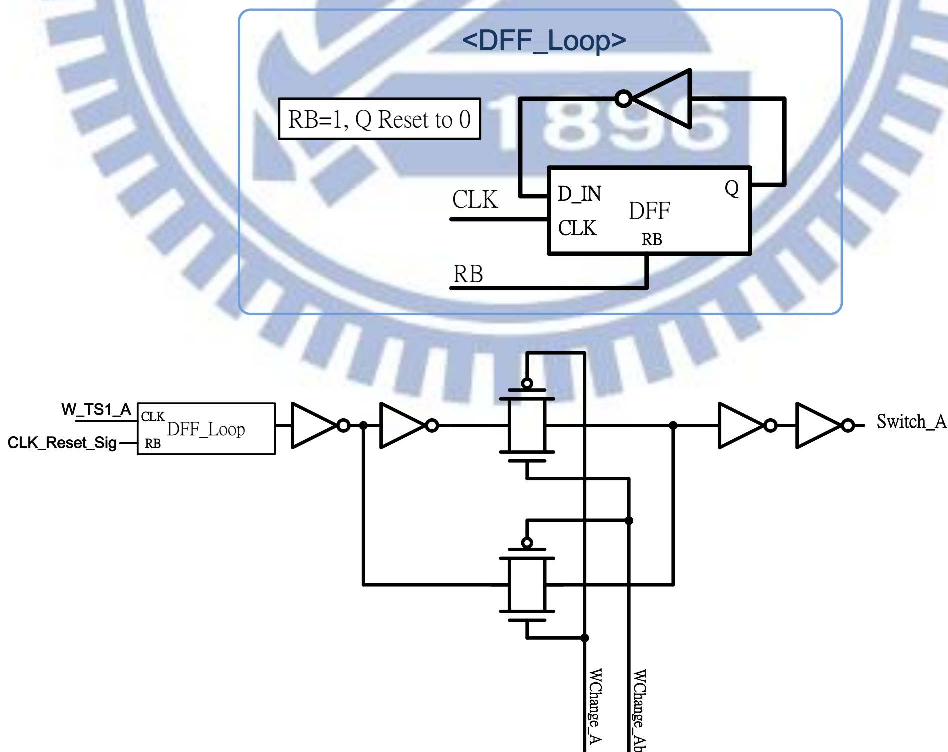


Fig. 4.21 S0/S1 switch control circuit

4.5.4 Thread Switch Control

In this section, I show my register file multi-thread control design. In this design, I refer the IBM's work (one bit cell with two subcell). IBM's work is showed in fig.4.22 and every bit cell with two threads switch decoder. Shorter delay can accomplish by using this design, but area and power is more than conventional SRAM long BL sensing design. In order to do a low power and wide voltage range operation register file. A new cell is proposed, this cell I will show detail in next chapert5. Not like IBM's bitcell, I combine with my new "share RBL structure" and each column share only one thread decoder. In this new design, power can down to more than 50% compare with IBM's work and area consumption is about 70% lower than before.

Not only area and power reduction are very significantly, thread decoder read out is more flexible in my work. In IBM's work, three read ports are already set for subcell 0 or subcell 1. Six read port can't point to the thread 0 at the same time. In other word, 4read port can select which thread I want to read out in this design. Fig. 4.23 is my RBL structure. One bit cell has 4 RBL, and share each other with neighbor cell. Subcell 0 and subcell1 have itself RBL which can read out data at the same time. In read cycle, "A" port or "B" port data will pass through by RBL then into the thread switch circuit (Fig. 4.24) for decoder. When decoder is finished, RBL data will output to latch.

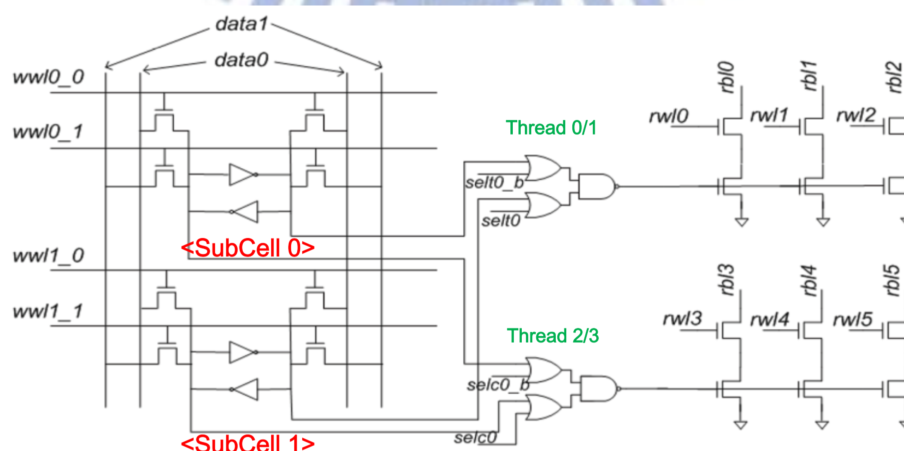


Fig. 4.22 Power7 thread decoder control circuit

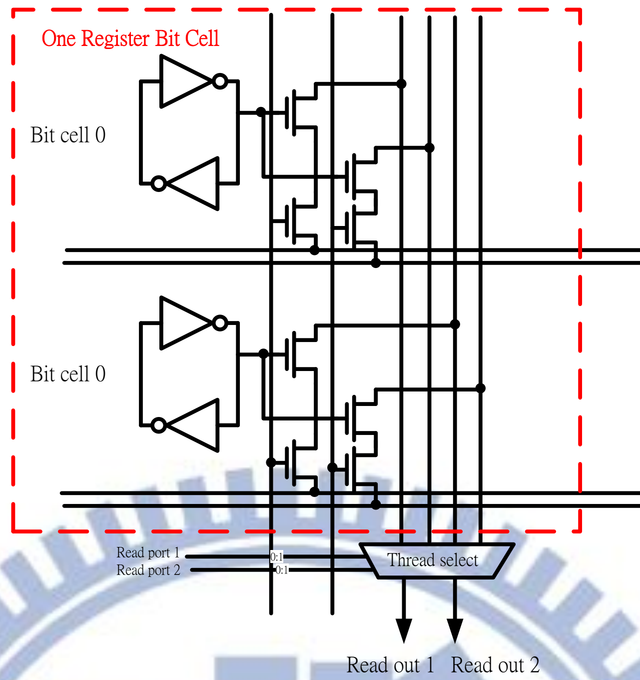


Fig. 4.23 Bit cell read out architecture

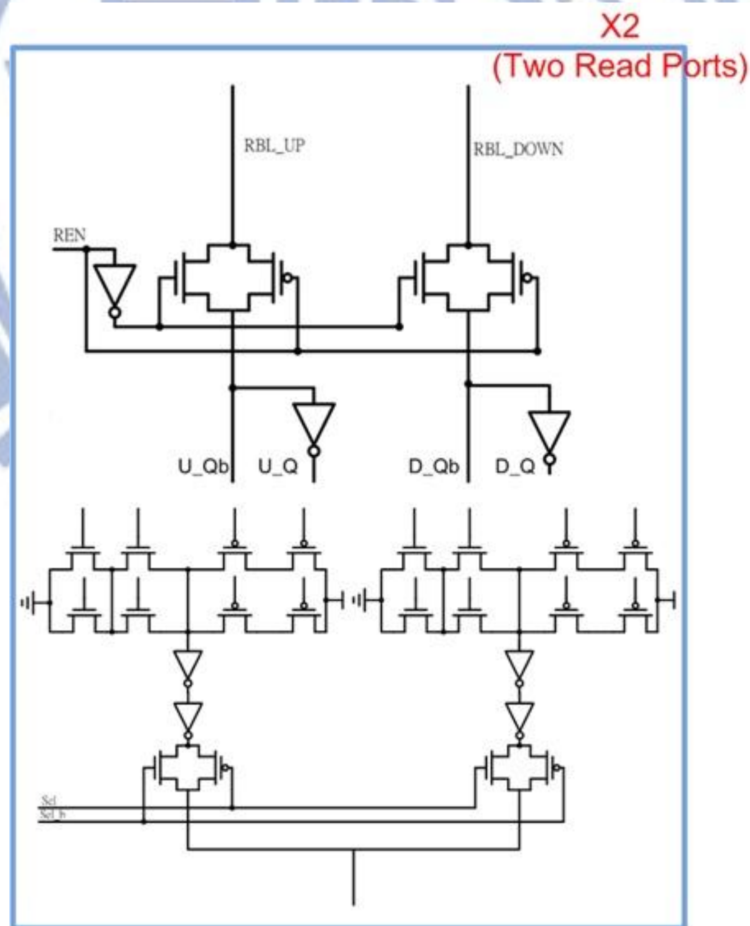


Fig. 4.24 A/B port threads switch control

4.5.5 Double Pump Operation

Base on Kuo's work, a more robust and lower power design is proposed in this section. The slot control circuit is show in fig.4.25, and operation waveform is in fig.4.26. This circuit will check second pulse is on or off. When "WEN_S1" signal is 0, it means one write only this cycle. After TS1 is trigger then TS2 is follow it rising up (which is write finish signal).In this low power design, SA is floating when SA pulse is no trigger. Power gating makes leakage reduction significantly in sleep mode. For register more robust, write followed by read control is added. Simulation data show that read time is more than write time (In next chapter). If write pulse is too fast and second write has a probability disturbed to first read slot. In this reason, I set second write pulse trigger always waiting for read first slot finished. This design is very help for data write/read disturbed free and no any timing degraded.

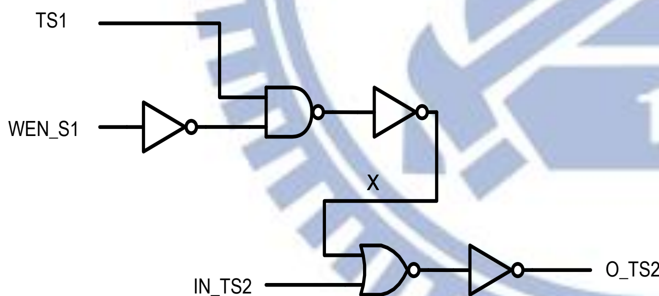


Fig. 4.25 Slot detected circuit

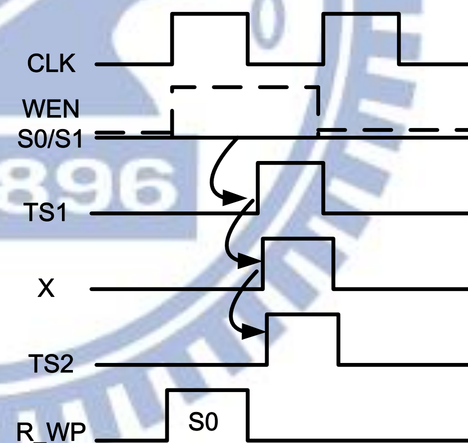


Fig. 4.26 Slot control waveform

Fig.4.27 is replica design in this chip. Double pump pulse is generating by replica and twice trigger in one cycle. Because each bank has four ports, one port need itself port replica bit. For read test, give the worst case that only the top read dummy cell store "0", others are store "1". In this case, RBL can drop down by select RWL bit cell. In order to

test the worst write time, I put the right cell in the bottom of column. Write path combine with column capacitive loading and simulation Write”0” & “1” operation at the same time. And W_OK signal will trigger when worst case write is finished. The replica operation waveform is showed in fig.28, R_WP signal is rise up when CLK in ing edge, and R_WP pulse is turn off when R_W_OK pulse up and close R_WP. Waiting for a short time then replica read path again.

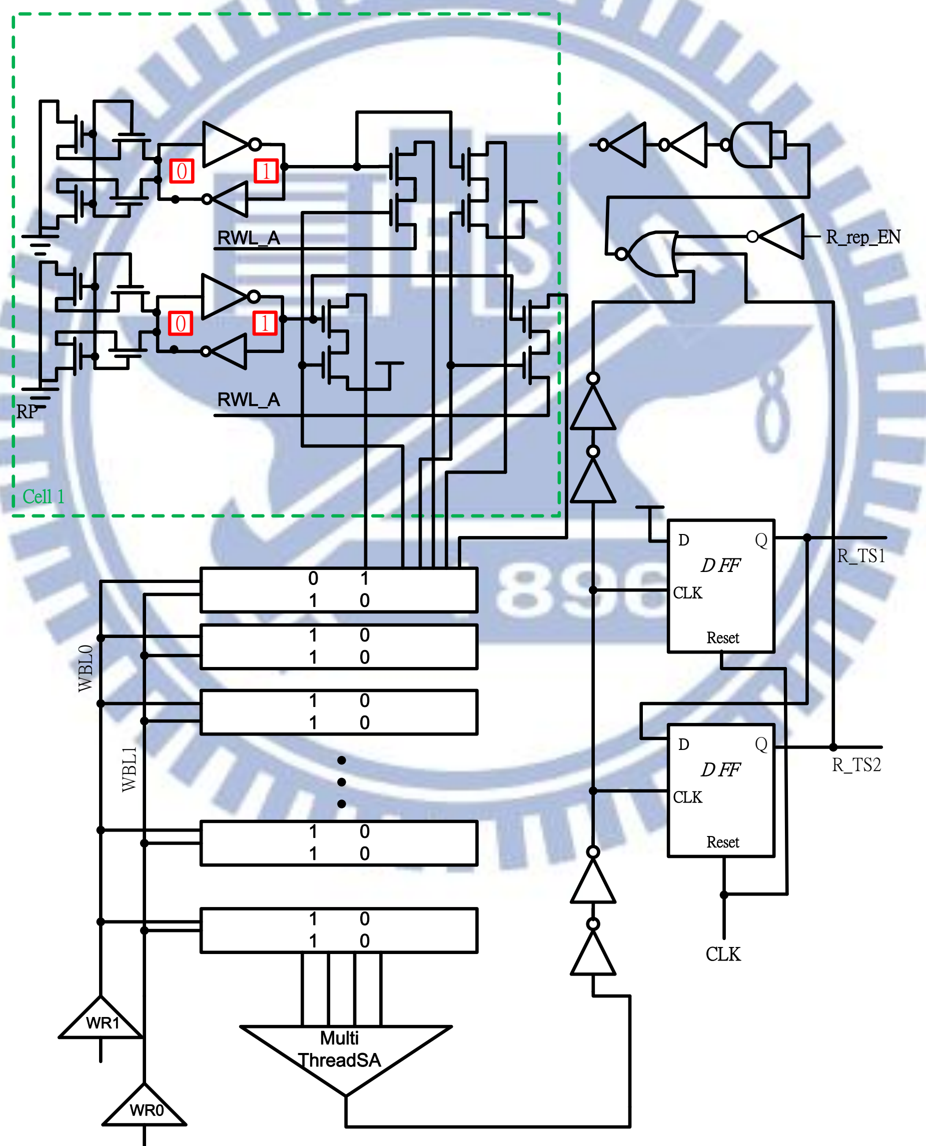


Fig. 4.27 Register files replica circuit

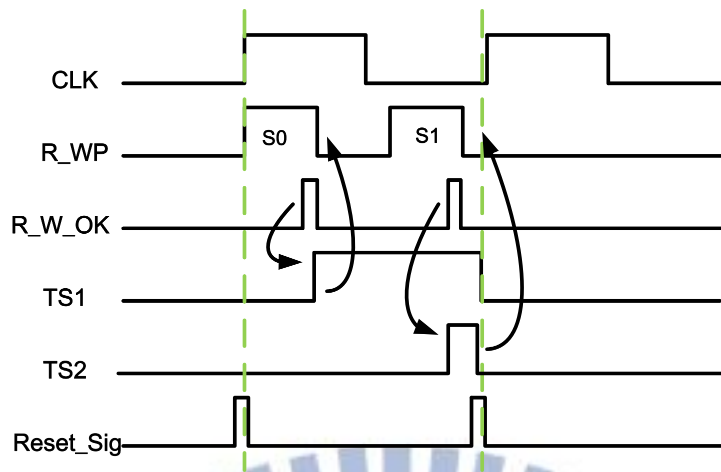


Fig. 4.28 Register files slot control

4.6 Summary

In this chapter, first we introduced pervious register design and conventional dual port conflict issues in write/read mode. Low power design such as multi-bank and power gating non-used bit cell are showed in Chap 4.2. In Chap. 4.3, multi threads cell design is discussed. There are many technologies for thread switch, not only circuit level but also computer architecture. Multi-threads supply high bandwidth between data transmission, this design is very suit for video or high performance CPU. Double pump design and time sharing technology are discussed in Chap. 4.4. By using lower CLK frequency, power consumption reduce is very significantly. Different technology to generate twice pulse on one cycle can lower the port counts to half, and save more operating time. In Chap. 4.5 is my design work. A new share RBL structure is proposed for power reduction and Data slot switch can gain more write successful probability. RBL base multi-thread decoder can reduce area about 50~60% than IBM's work, and not reduce too much performance. One/twice slot switch control is proposed for power reduction and give cell more robustness.

Chapter 5

Low VDD_{MIN} Multi-thread 4R4W Register File Design in TSMC 40nm CMOS Process

5.1 Introduction

I design a low power 2K 13T 4R4W register file with multi-thread switch in this chapter. There are many technologies for improving register file performance, power consumption and data retention I discussed in Chap 3, Chap 4 and this chapter. The floor plan, pin count, pin definition and specification of this 4R4W register file will also be introduced. In Chap 5.2 I will show the cell layout and this register file spec. Share WBL and RBL structure and thread read out structure. Chap 5.3 shows SNM simulation compare with conventional 8T, dual port, and this design. In this section, iso-area concept is included for more fare comparison. Chap 5.4 shows assist circuit design, such as negative VVSS for low VCC write assist and cut off feedback loop when write operation. In order to do floating problem free, extra Y_Cut NMOS is added in this unit cell. In Chap 5.4, Design implementation & Test-waveform function of proposed 4R4W register file is discussed. Whole chip floor plan will show in this section. Finally, post simulation and analysis are based on TSMC 40nm TN40G process. Power consumption and bank layout are show in Chap5.5. The technology file is supply by CIC, and finishes it in secrecy Lab.

5.2 4R4W Register File Structure

5.2.1 2R2W Register File Unit Cell & Layout View

Fig. 5.1 shows the 2R2W register file cell I proposed. In this cell has two subcell, each one store thread0/1 and thread 2/3. The cell is composed of 28 N/PMOS, and use

share WBL and share RBL technology to reduce transistor counts. In this design, there have 8 rows signals and 6 columns signal to operation this 2R2W register file cell. RWL_A and RWL_B control is share with up/down cell. When write operation, Xsel signal will select up or down which is need to write data in. In read mode, subcell "0" and subce11 "1" data will read out at the same time. Negative assist technology is added in the MNA VVSS for write improvement. This technology only turn on in write "1" mode, otherwise it will reduce write "0" performance. Compare with Chap 3 I proposed 2R2W multi-port SRAM, conventional read buffer is changed to new structure RBL sharing structure. RBL sharing side in this picture is not sow because the plot is already too large. I will discuss this new structure in next section.

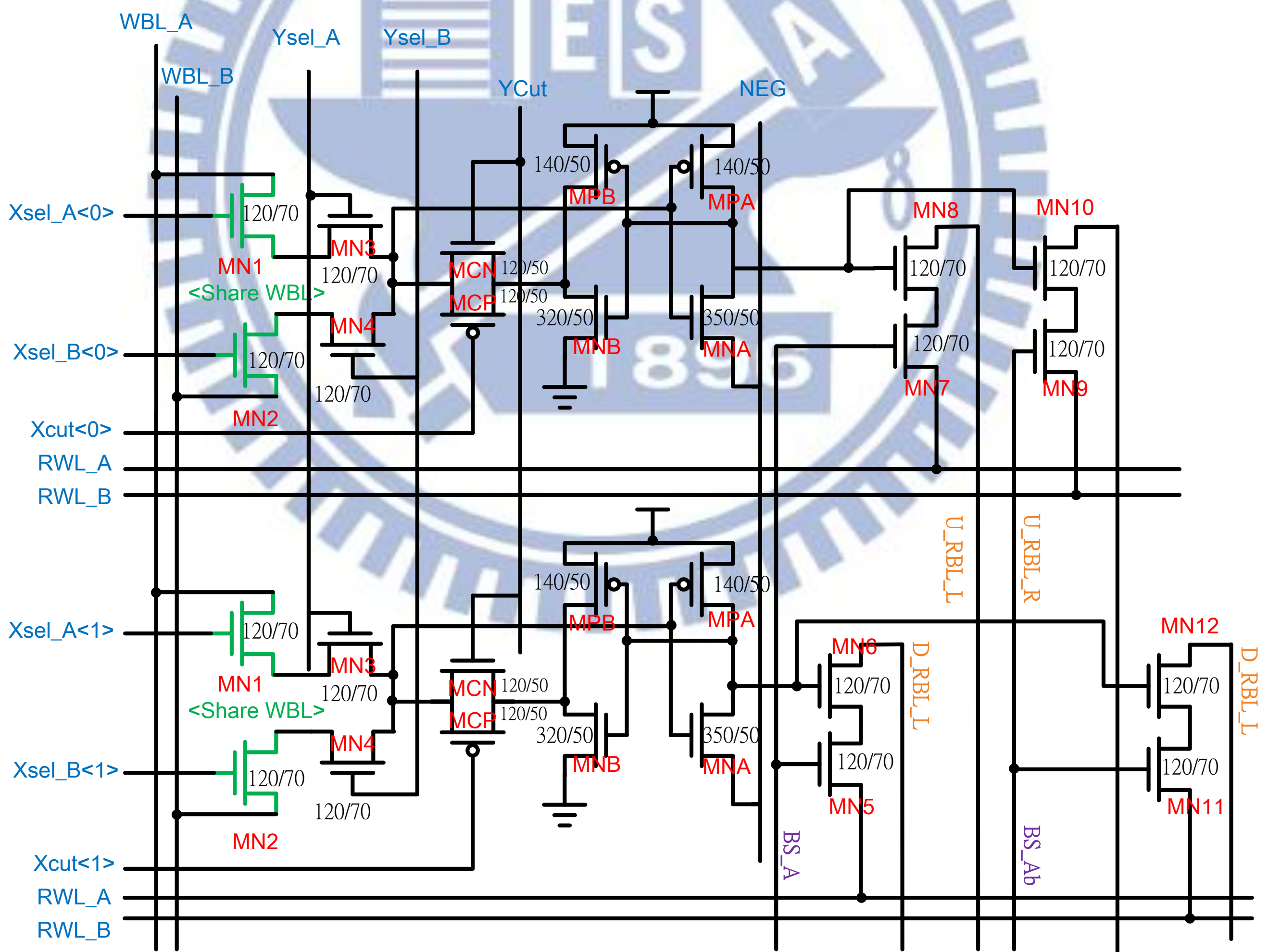


Fig. 5.1 The 2R2W register file unit cell

In layout view, metal layer increases from metal 3 to metal 5 and no area damage. Fig. 5.2 is metal used in this cell and layout view is showed in fig. 5.3. Metal layer is change for Ysel & Y_Cut signal from M2 to M4, and row base signal metal is change from M3 to M5. M1 & M3 is become to inter layer connection. Fig.5.3 shows two neighbor bit-cell layout share RBL structure, and sub-cell architecture

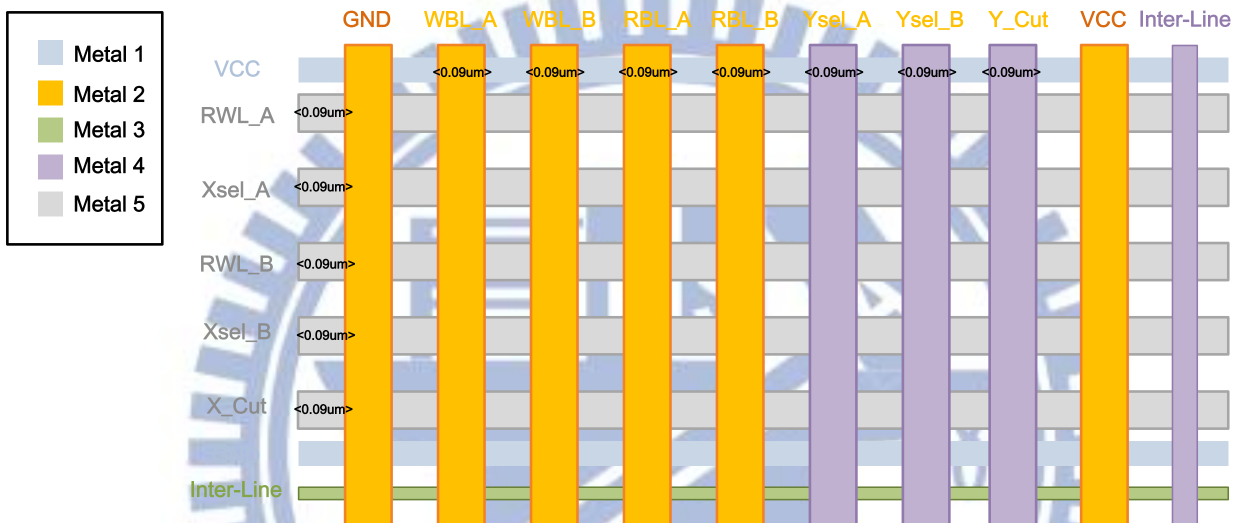


Fig. 5.2 The metal layer implement of 2R2W register file unit cell

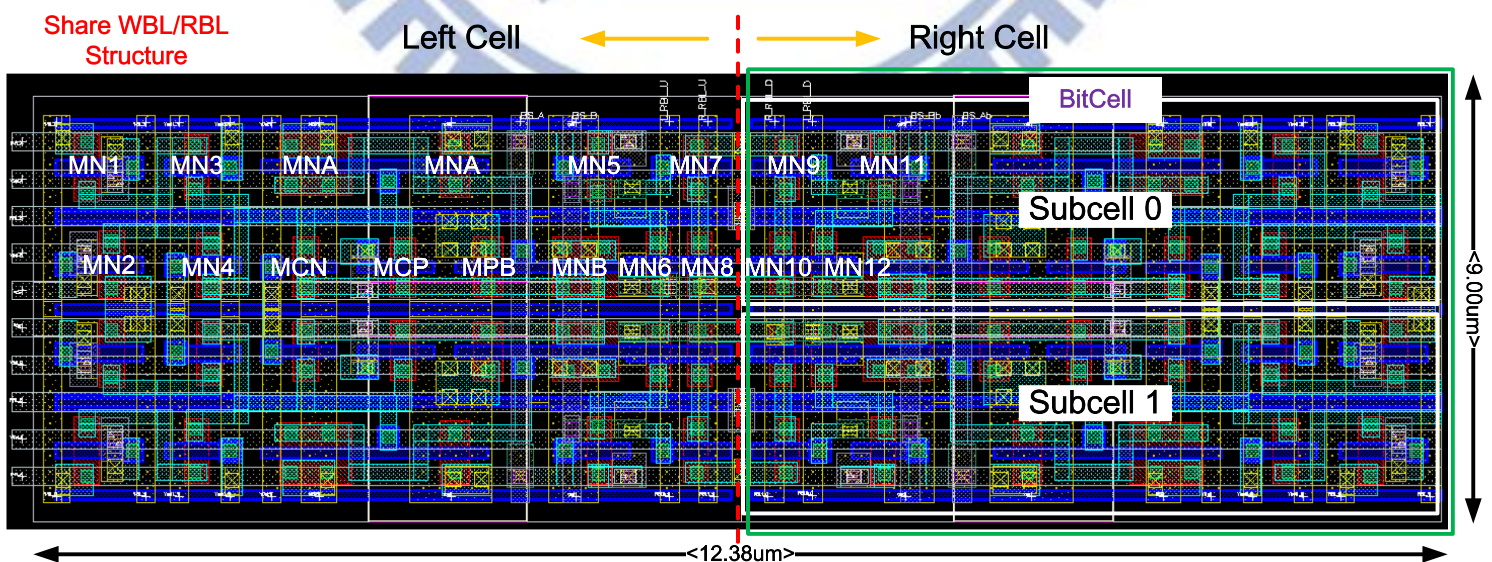


Fig. 5.3 Layout view of 2R2W register file

5.2.2 Share WBL Structure

Like 2R2W multi-port single end write operation (Chap. 3), so in this section I will not repeat again operation. An only one problem is that it needs extra one WBL in the two side in register file array. So in this design, we have to do a new switch circuit design for WBL port switch in bit-interleaving structure.

5.2.3 Share RBL Structure

For lower power consumption, a new structure read buffer is proposed. In this way, port numbers can reduce to 1/2 and no have dummy read problem. It is very suit for lower power device, and it can increase more battery life. Dummy read is normally see in the conventional dual port or 8T SRAM cell used in bit-interleaving structure. More than 30% power reduction by using conventional 8T dual port compare with this new read-out structure. Not only this method can do dummy read free, RWL keeps high in standby can also reduce leakage power consumption.

Fig. 5.4 is a dummy read happened in conventional SRAM design. Normally register file read frequency is more than write. So this technology is very used for this design, and only need extra 2 column select signal to control it (Left or right signal). By the way, the dummy read power consumption will enlarge along with bit-interleaving bit.

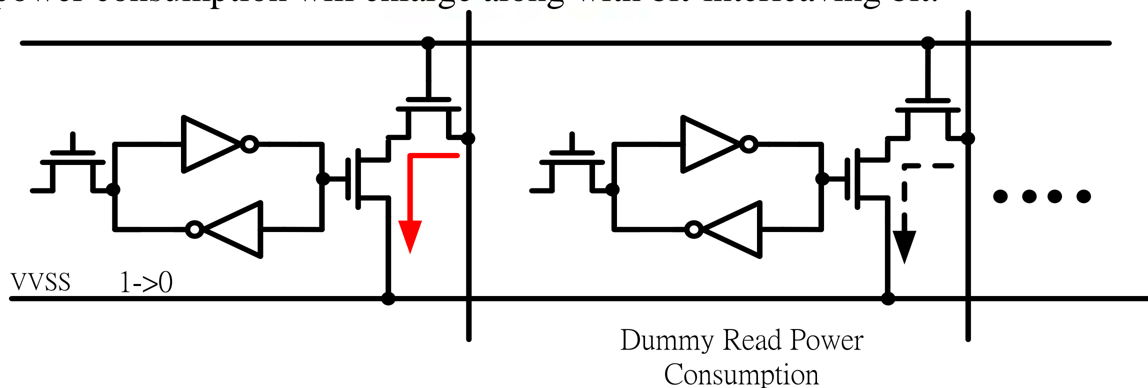


Fig. 5.4 Conventional dummy read operation in half select cell

Fig. 5.5 shows my design of share RBL structure. The green color is the share with left bit cell and black is stand for right bit cell read buffer. There are four ports in bit cell (for thread switch detect need), then two A & B ports output. Fig.5.6 is my read buffer design in read operation mode. A one side NMOS open in at the same time, it will not worry about dummy read power consumption anymore.

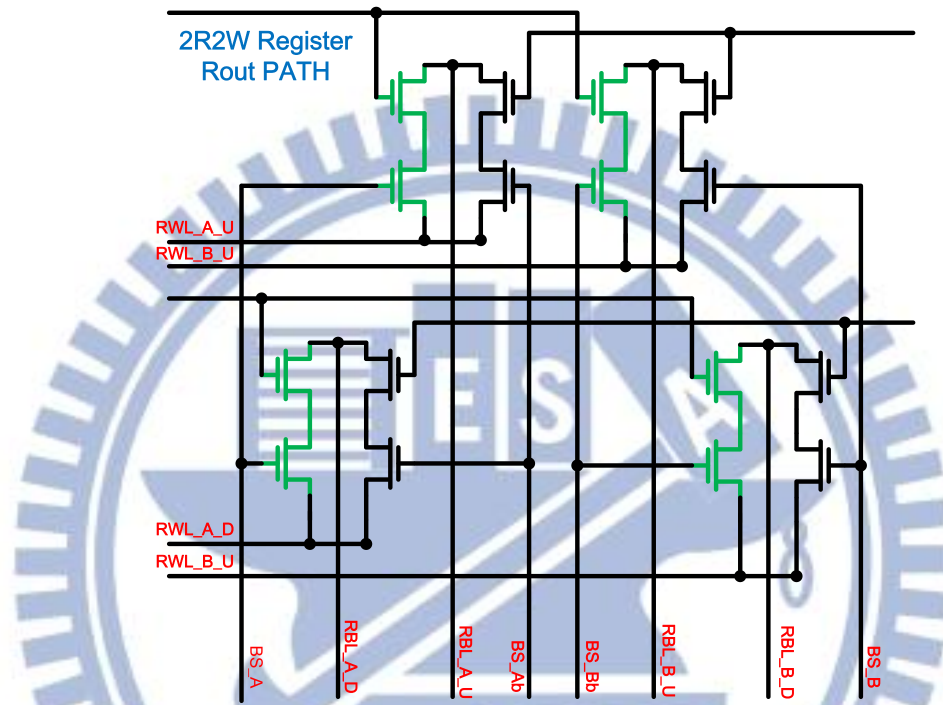


Fig. 5.5 Share RBL structure in 4R4W register file

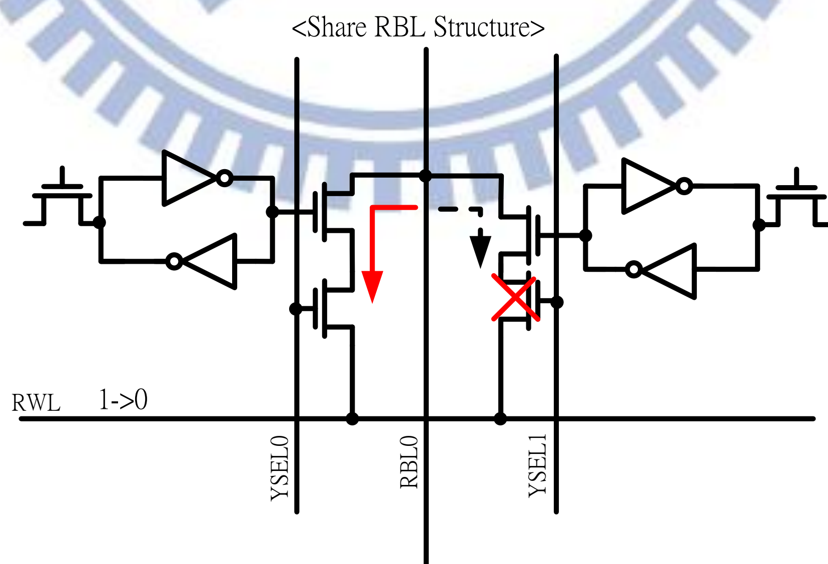


Fig. 5.6 Share RBL structure no dummy read power consumption

5.3 Register File Assist Technology

5.3.1 Negative VVSS Design

In this cell design, write “1” is not easy than write “0’. In order to solve this problem, a negative VVSS in used for improving write “1” case. The negative VVSS circuit is show in the fig. 5.7. At first, wen signal is input then the negative start detect this moment is write “0” or write “1”. If A and B port all choose write “0”, VVSS will not change statue and connected to the ground. In other word, either one of two port select write “1” mode, the negative circuit is start working. The capacitance sizes choose should be very carefully. If size is too small, the write assist is not enough; otherwise, if the size is too large, it will distribute others standby cell in the same column. Table 5.1 is the NEG circuit work function and Fig. 5.99 shows negative VVSS level by post simulation result in different voltage. The second Cap. Only turn on when two port is write “1”. In this moment the WBL loading is larger than single write “1”, so it needs more capacitance to assist write “1”. Fig. 5.60 shows negative VVSS variation in different corner in 0.6V supply voltage.

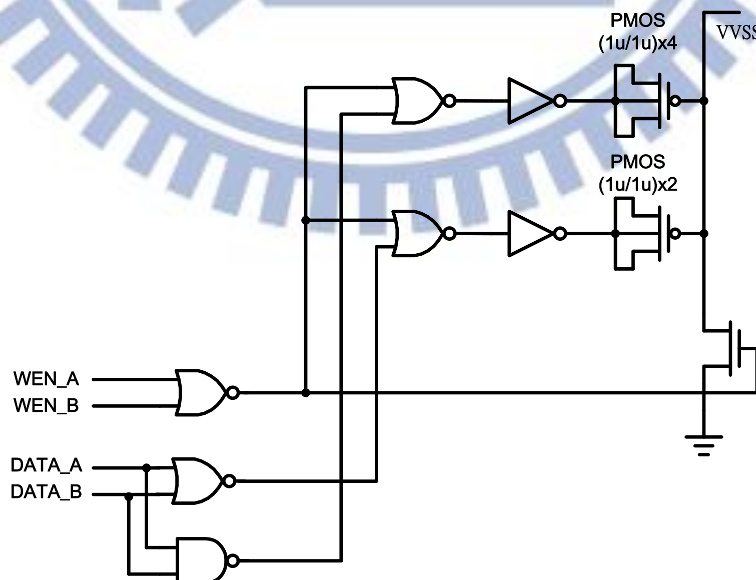


Fig. 5.7 Share RBL structure in 4R4W register file

	WEN = 0	WEN = 1	WEN = 1	WEN = 1	WEN = 1
Data A	**0**	**0**	**0**	**1**	**1**
Data B	**0**	**0**	**1**	**0**	**1**
Cap. 1	**OFF**	**OFF**	**ON**	**ON**	**ON**
Cap. 2	**OFF**	**OFF**	**OFF**	**OFF**	**ON**

Table 5.1 Negative VVSS circuit function

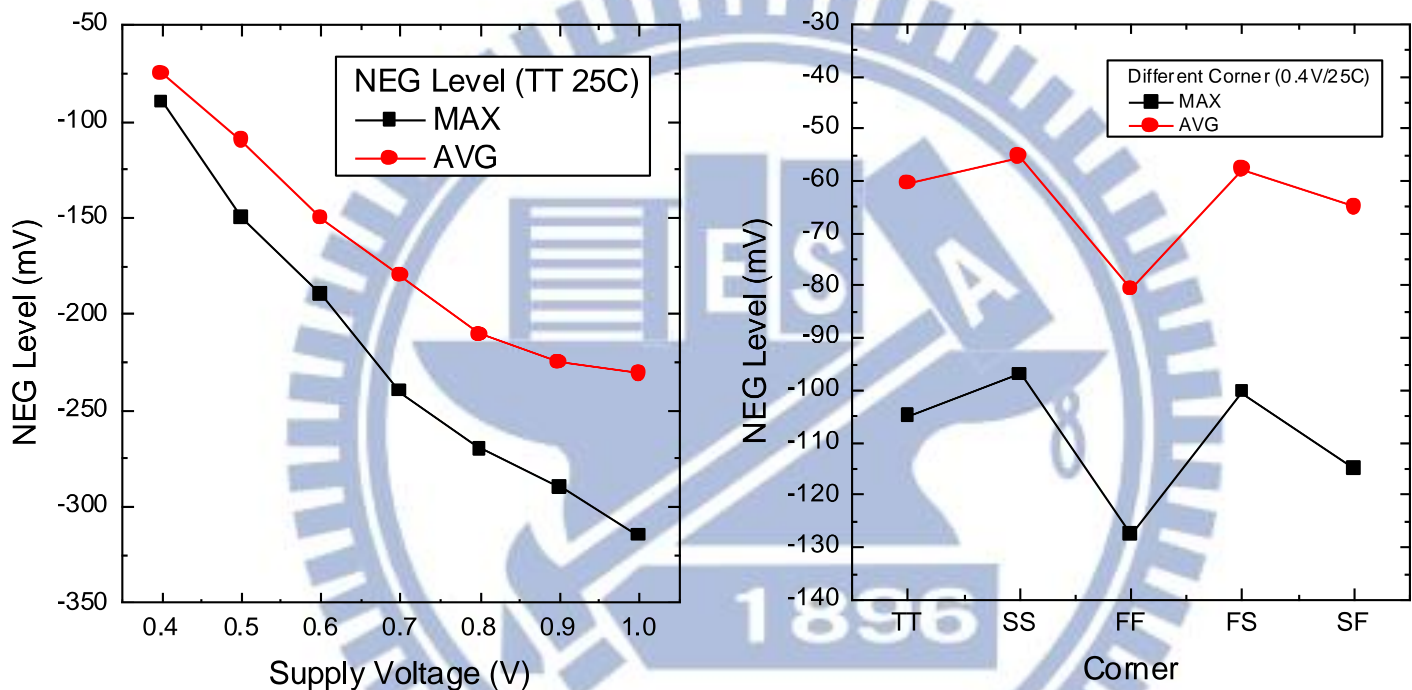


Fig. 5.8 (a) Negative level in different voltage (b) Negative level in different corner

5.3.2 Single-end Write Cut-off & Y_Cut for Floating Issues Free

In this section, write mode is like 2R2W multi-port SRAM cell. Single-end write is not like differential write can do small signal sensing, but driver power consumption, periphery circuits and cell size is reduced many much. Normally, write “1” is a big problem when circuit operative under low supply voltage. In this design, cut off feedback loop can help write “1” easier, but in the same low based will happen floating issue.

5.4 Implementation of Multi-thread 4R4W RF

5.4.1 4R4W Register File Floor Plane

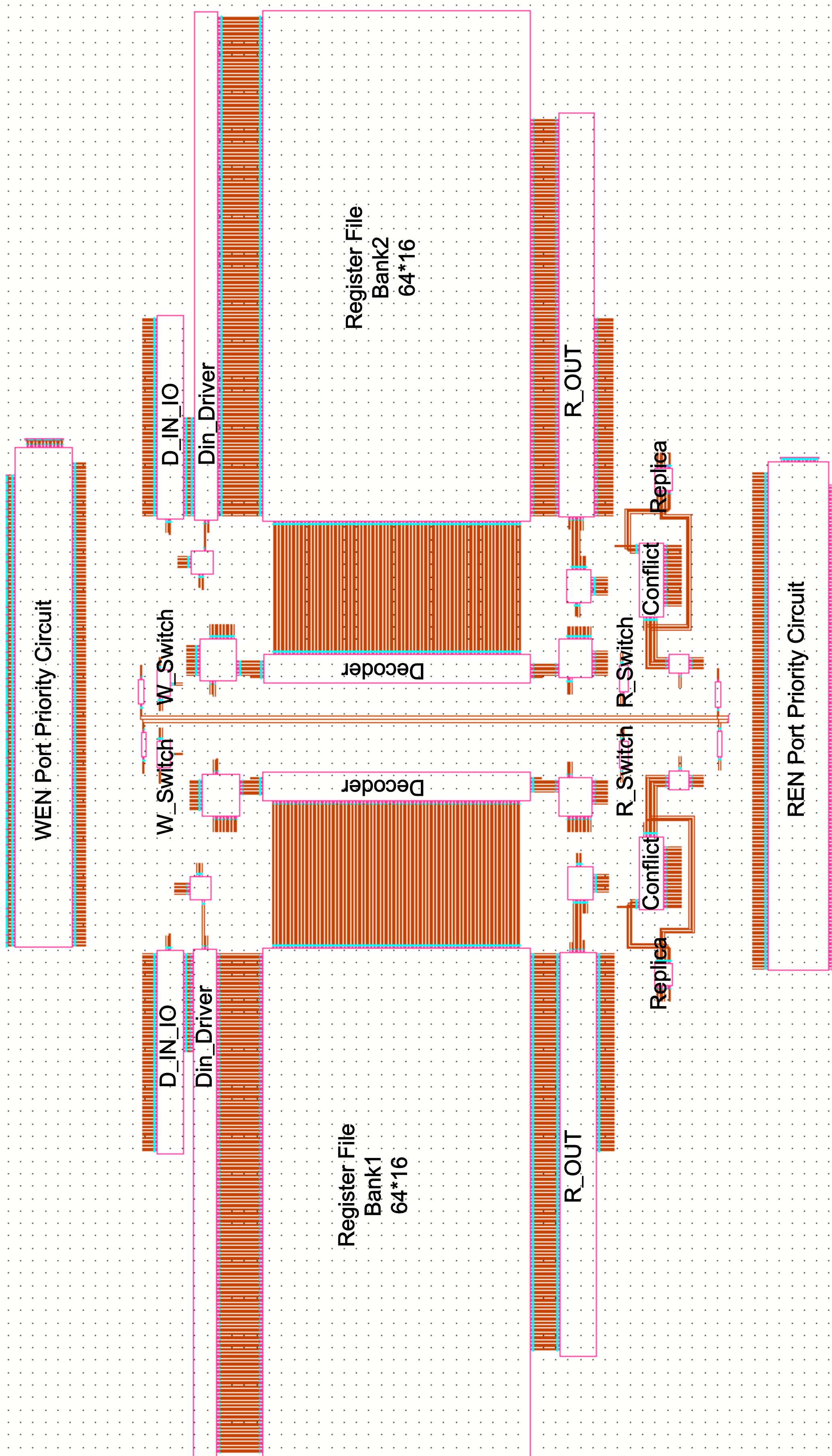


Fig.5.13 is a schematic view of 4R4W multi-threading register file design. Including two 1k bits bank and hierarchy level conflict detect circuit, no conflict problem will happen. Each bank have itself control circuit such as R/W decoder, replica, R/W driver and switch detect circuit. CEN0 & CEN1 signal is set to bank each other. If CEN signal is keep low, all of bank is in sleep mode. Periphery circuit is in power gating for leakage power reduction. Compare with conventional always on circuit, power saving is very huge. In this figure 5.13, write control circuit is on the top side and read control is on the bottom side. Replica is place in the left side and right side for detect the worst case. WL pulse is including not only replica sensing delay but also long inverter delay chain. By this design, it supplies a very robust WL pulse for write/read successful. Read and write slot 0/1 switch circuit is different, read is no disturbed the cell node so I set it priority is always higher than write; otherwise, write will flip the storage node and change data so it need pass thought the conflict detect circuit.

Macro Size	2K bits (1024*2*1)
Process Technology	TSMC 40nm General purpose CMOS process
Data-width	32 bit
Read Address (each port)	9 bit*2 (two slot)
Write Address (each port)	9 bit*2 (two slot)
Bit-Interleaving	2 bit
Read / Write Port	4R4W
Multi-Threading	4 Thread
Voltage range	0.4V~1.2V
Cell size (Bank)	213.2um x 111.81 um = 23.83 um^2
Access time @ 0.4V TT 25	17.39 KHz (Double slots)
Cycle time @0.4V TT 25	16.68 KHz
Read power	98.63 nW/t (per bit c-ell, Double read)
Write power	112.3 nW/t (per bit c-ell, Double write)

Table 5.2 The specification of proposed

5.4.2 Design Implementation & Test-flow of Proposed 4R4W Register File

Fig. 5.10 shows my test function waveform. At first, single write “0” operation is in 1st CLK cycle. In second CLK cycle, write the bottom bit and read the bit which is write in first cycle. In this cycle, test signal read and double write function is showed. Write bit is all set one nearest side and one furthest side, and try to find the critical path. In next cycle, two read operation at this moment. In fourth cycle, try to test W/R conflict circuit. Finally, the data switch test is in the 5th CLK cycle.

After using simulation test pattern can trace the worst case in this chip. If function is not working, waveform is different. Test CLK frequency is adaptive 50 MHz in this work. The write simulation post-simulation waveform can see in Fig. 5.11. Fig5.12 shows read mode waveform. In 5th CLK cycle, a WChange signal is raised up and changed S0 & S1 order.

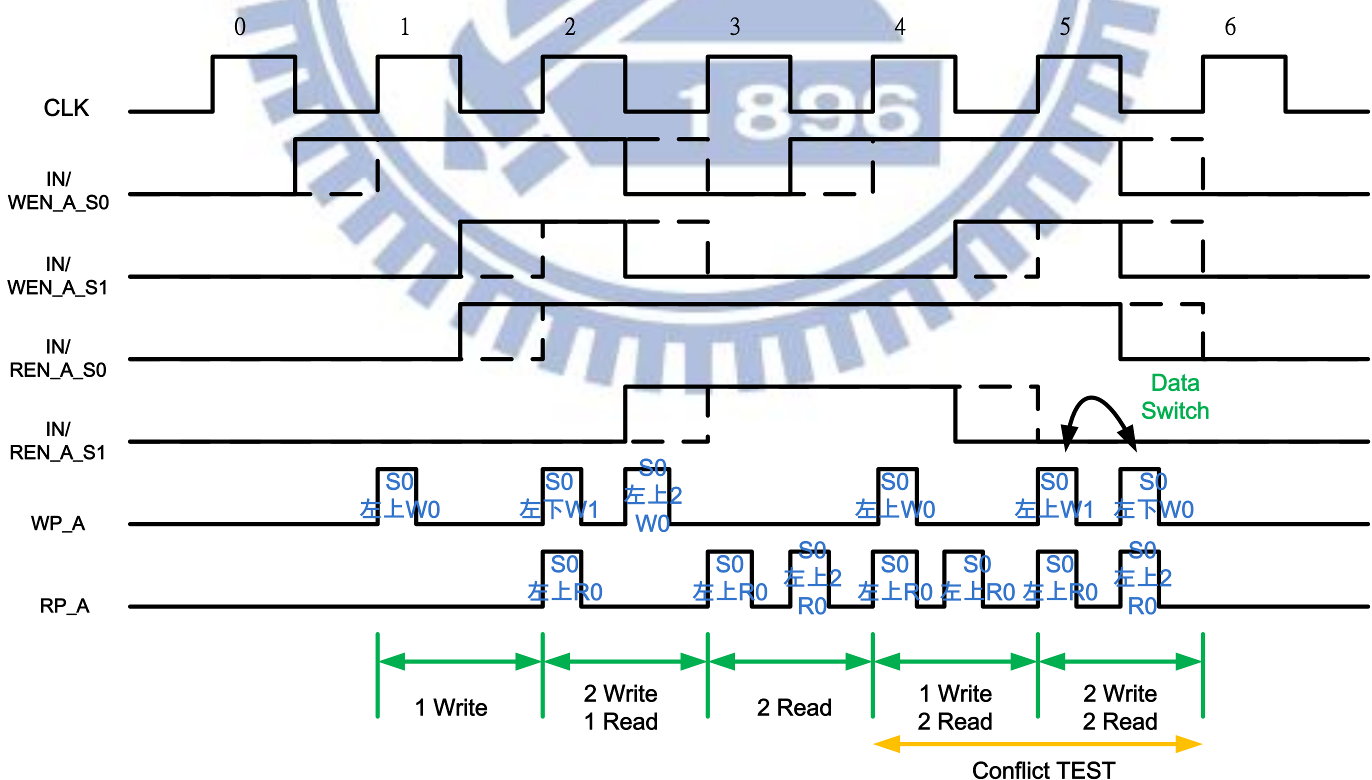


Fig. 5.10 Test pattern and simulation waveform result

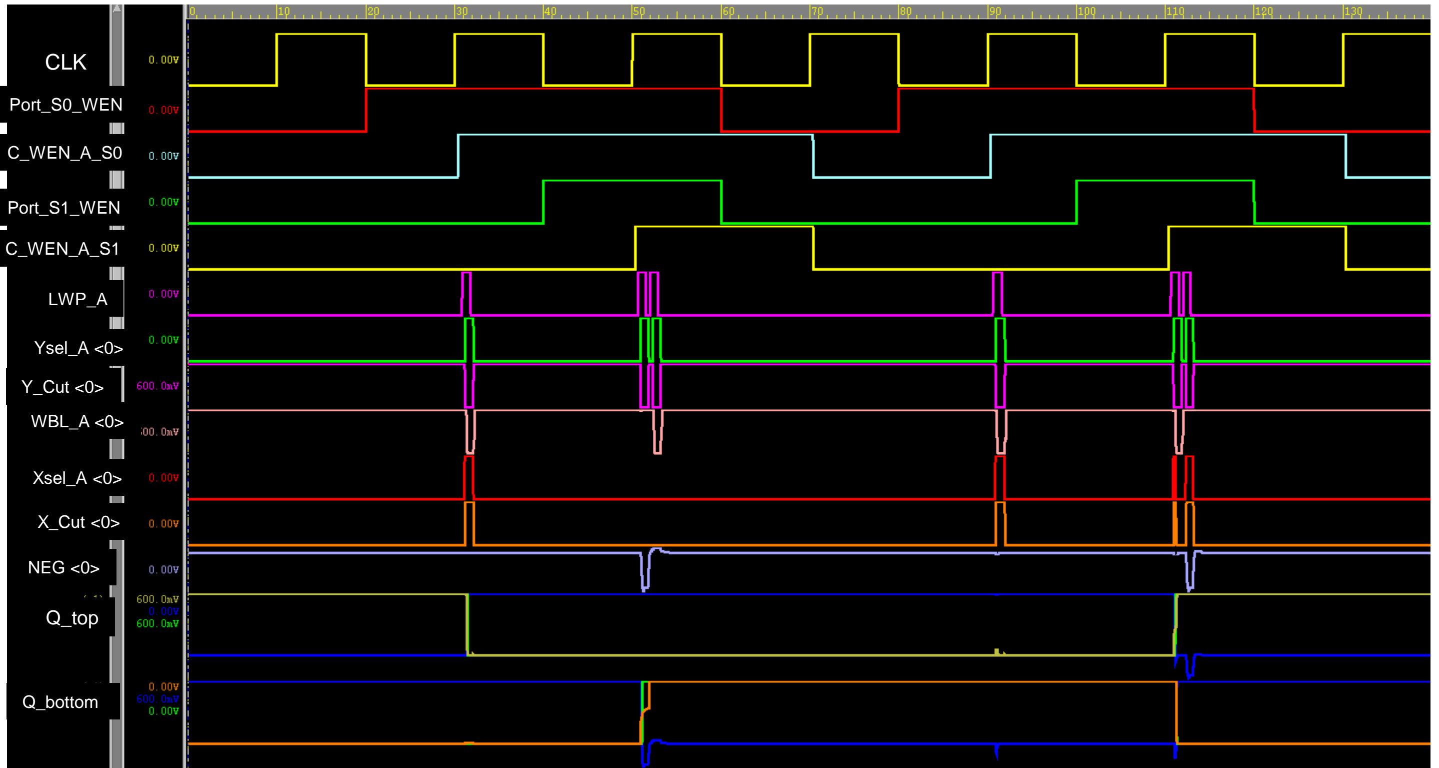


Fig. 5.11 Write post-layout simulation waveform result

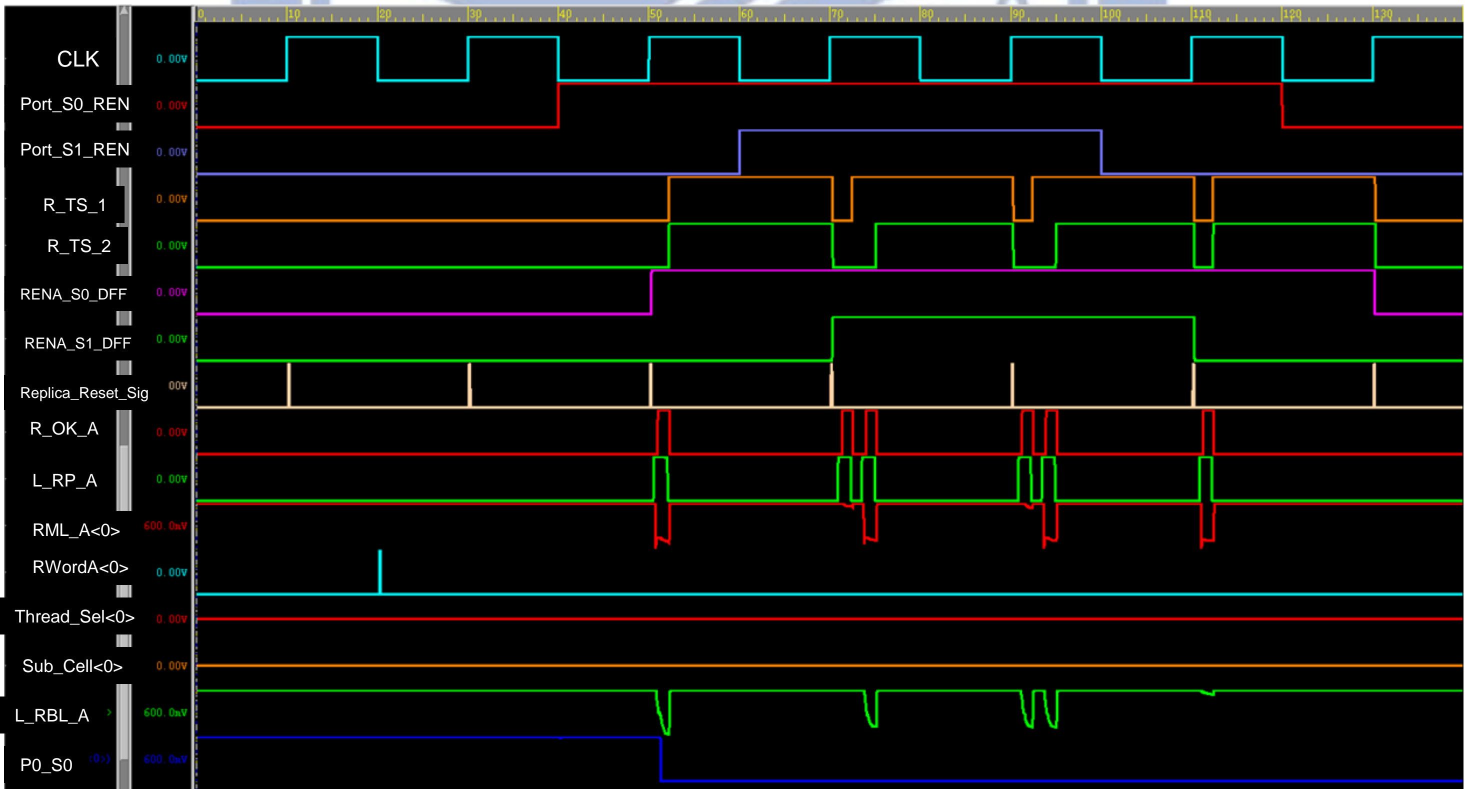


Fig. 5.12 Read post-layout simulation waveform result

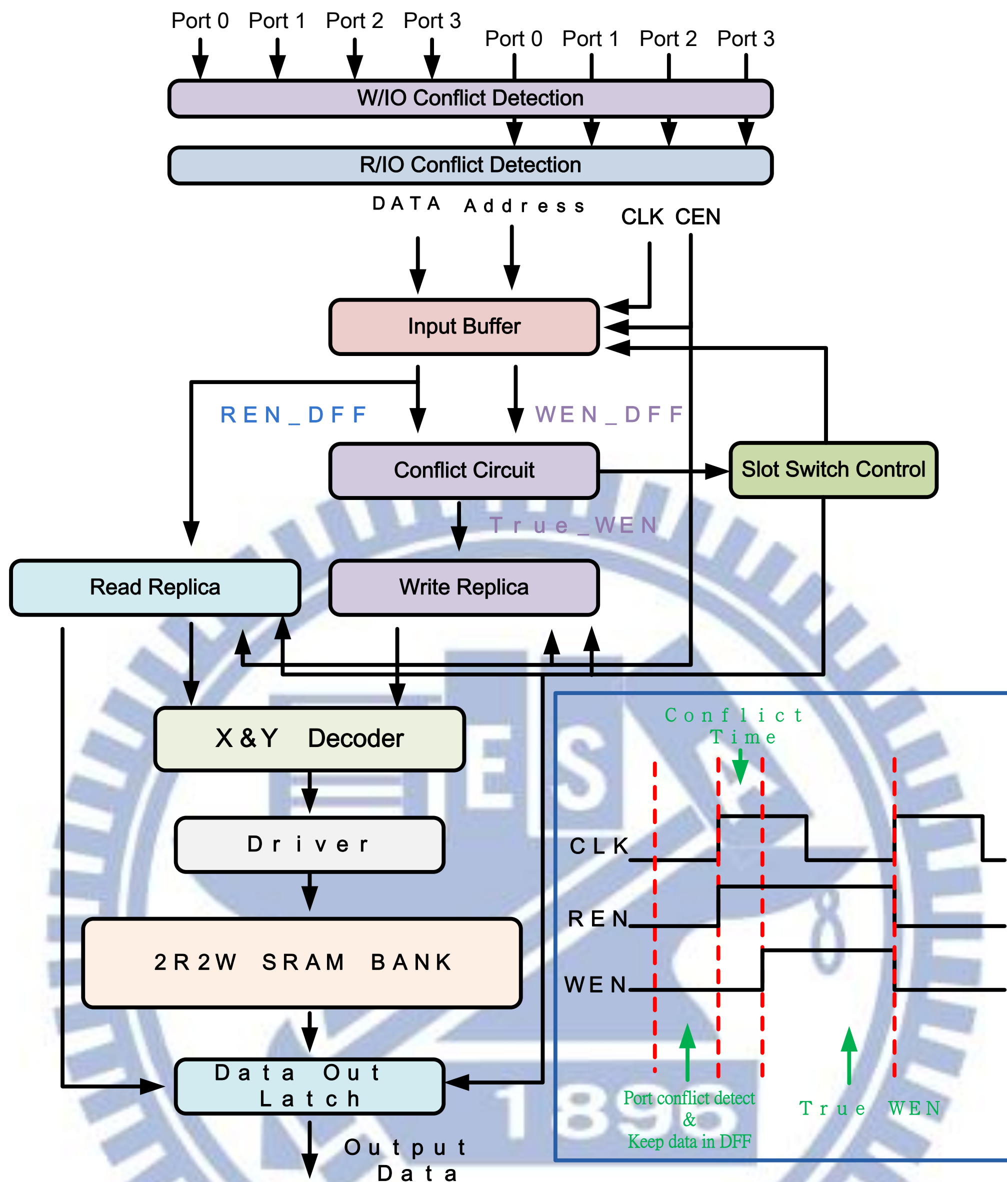


Fig. 5.13 Data transmission path in this 4R4W multi-threading register file

Fig.5.11 shows the post-layout simulation in this chip design, test pattern can refer to Fig. 5.10. Below this picture, read simulation waveform is showed in Fig. 5.12. Fig.5.13 shows data transmission path in this 4R4W multi-threading register file design. There are two level conflict detect circuit level I have introduces in chapter 4.5 and chapter 5.2. Data input have to give the signal before the CLK rising edge and finished first level port priority conflict detect. Then CLK rising edge trigger DFF, start second detect level address detect. WEN signal is turn on if no conflict problem issues. This signal passes to replica circuit and generates a WWL pulse to control write driver.

REN is like conventional read design and also have to pass port conflict detect circuit.

5.5 Post-layout Simulation Result

Fig. 5.14 shows a bank layout view of the proposed 4R4W multi-thread register file bank. The proposed 4R4W 2K register file is fabricated using TSMC 40nm general purpose process. The area of bit-cell is 3.095um x 1.8um = 5.571um^2 and the bank size is 213.2um x 111.81um = 23.837 mm^2. Below is all of improved technology of this design Table 5.3. These technologies are introduced in the Chap. 4 &Chap. 5, and post-simulation result is showed in next section.

These technologies I have introduce the design concept and logic circuit in Chap.3, Chap.4 and Chap.5. Low power, area reduction and wide range operation register file is present in this section.

NO.	Technology
1.	**Share RBL structure**
2.	**Share WBL structure**
3.	**Power gating sensing circuit**
4.	**Negative VVSS for write assist**
5.	**Transmission gate cut off write assist**
6.	**Data slot switch control**
7.	**Column base Multi-thread switch**
8.	**Double pump timing sharing**
9.	**Y_Cut NMOS for floating issue free**
10.	**RWL VVSS reduce leakage current**
11.	**Bit-interleaving structure**
12.	**Hierarchical conflict detect structure**
13.	**Wide range operation 0.4~1.2V**

Table 5.3 Improve technology of 4R4W register file design

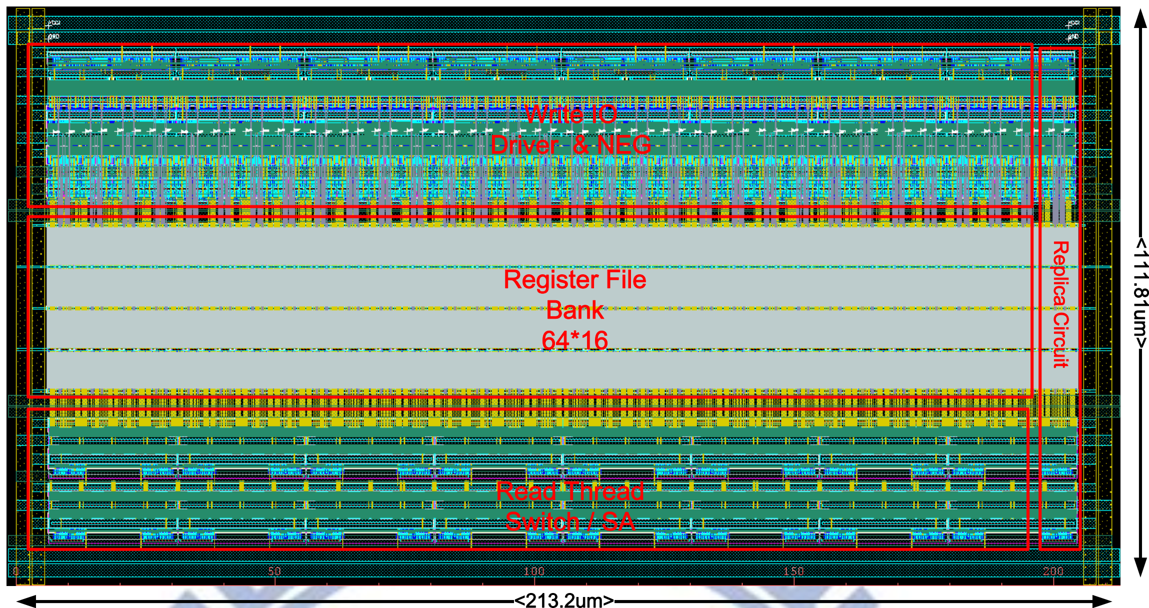


Fig. 5.14 4R4W multi-thread register file bank layout view

5.5.1 Performance

Base on post-layout simulation result, this proposed 2K 4R4W register file array can operate at 220 MHz in VDD=0.9V, TT corner &25C. When operation in Vmin (VDD=0.4V), the cycle time also can operate at 6 MHz, TT corner & 25C. Although two slot operate is need more time for timing control and waiting data, it also save 20% time compare with conventional two cycle can finish it. The timing reduction is show in the below.

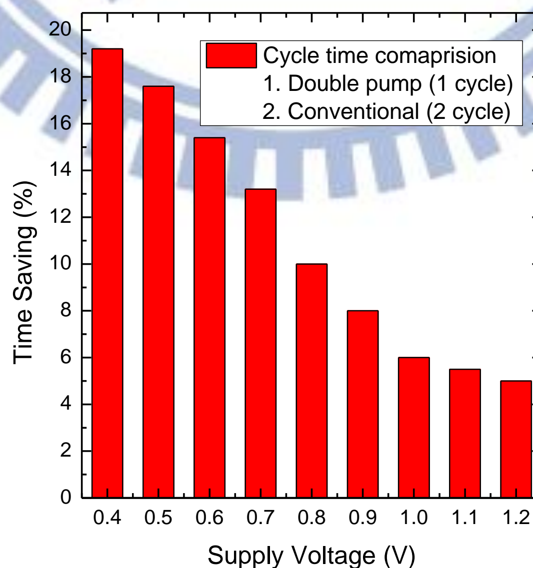


Fig. 5.15 Improve technologies of 4R4W register file design

In this post-layout simulation, the cycle time is dominated by Read 0 operation. Read 0 speed is limited by read buffer discharge share read bit line and multi-thread decoder operation. This problem is become more significantly in low power operation (VDD lower than 0.5V). In order to reduce read 0 time in this cell design, short channel is used for increasing more read current and used a short RBL structure (Fig.5.16).

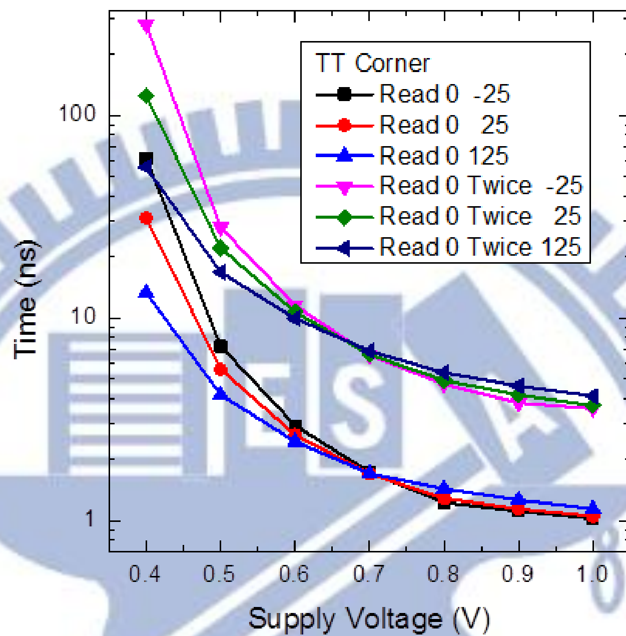


Fig. 5.16 Read “0” performance for post-layout simulation

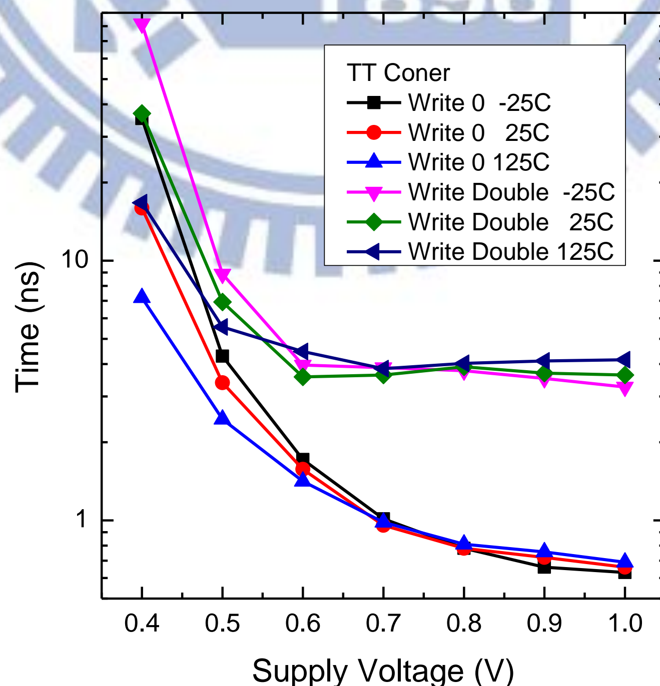


Fig. 5.17 Write “0” performance for post-layout simulation

Fig 5.18 is one/double read simulation result in different operation voltage. By this picture 125C with higher speed in high supply and slower in low voltage. Fig 5.17 shows write "0" time simulation, in this cell writes "0" is easier than write "1". NMOS access transistor is good for write "0" operation.

Fig 5.18 is write "1" simulation result in this design. Write "1" in this design is faster than write "0", because cut off feedback and negative VVSS write "1" assist is used. The WEN address conflict detect time is included in this data, and one write and double write timing more closed in low voltage operation. This effect is conflict circuit with XOR delay raise significantly in low power mode.

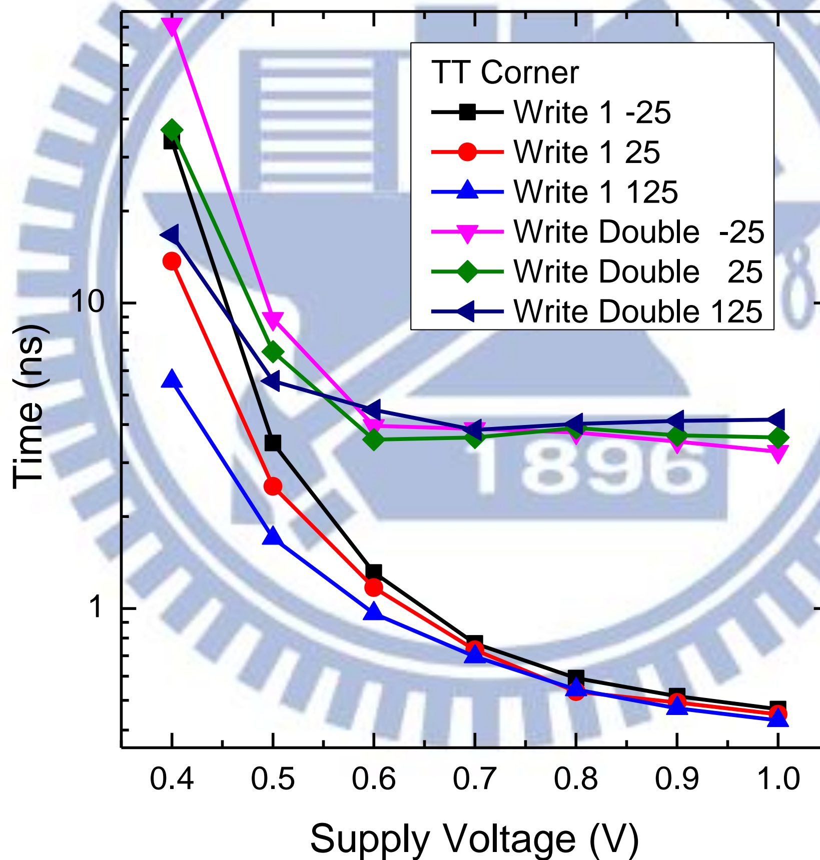


Fig. 5.18 Write "1" performance for post-layout simulation

Fig 5.19 (a) shows address conflict detect circuit operate in wide range and (b) is delay of write worst case compare with read "0" access time. This delay is similar in super threshold voltage operation, but split to each other in sub threshold region.

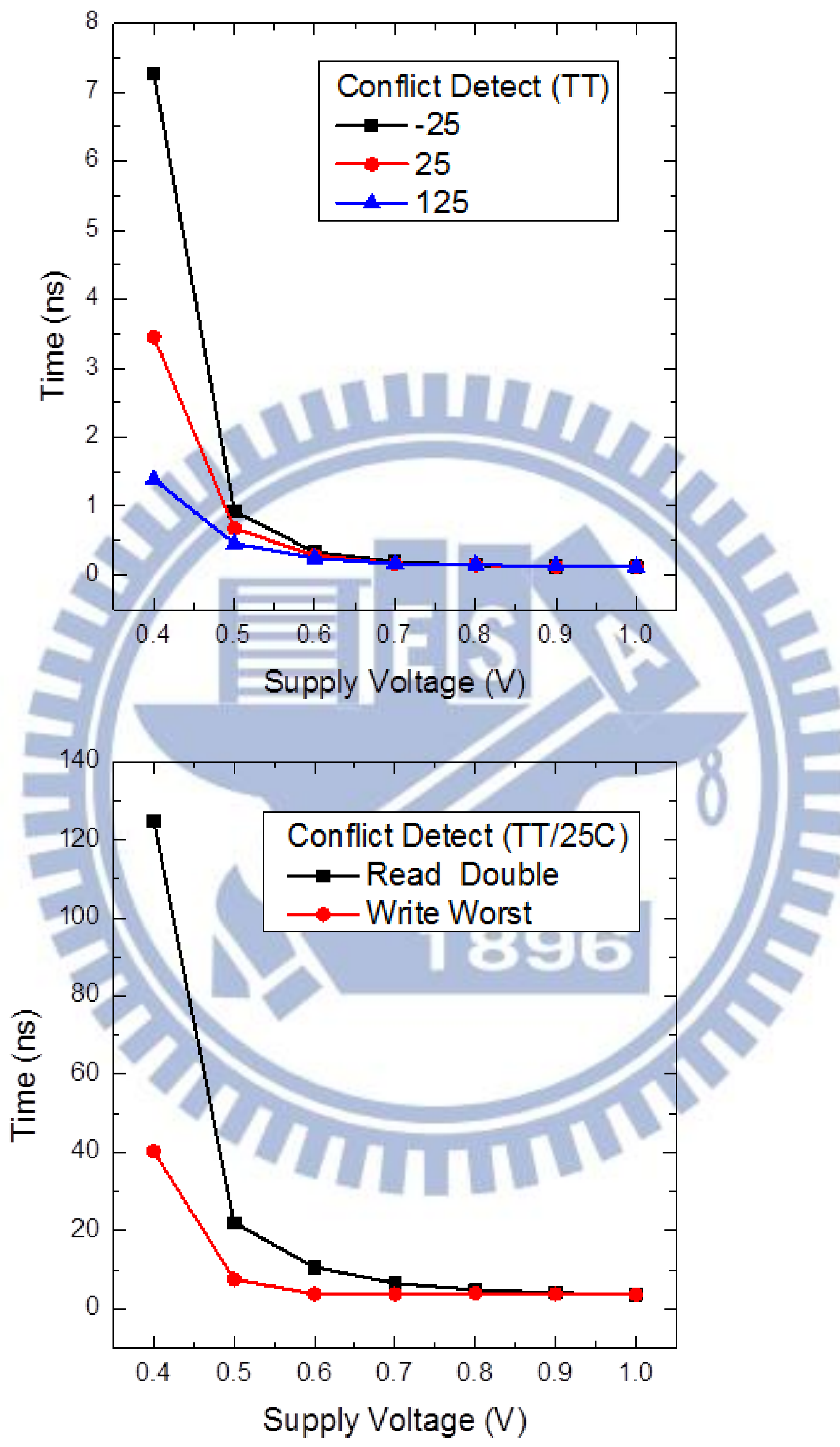


Fig. 5.19 (a) Address conflict detect circuit delay under wide range

(b) Delay of write worst case compare with read "0" access time.

5.5.2 Power Consumption

Fig 5.20 shows the power consumption of read/write operation and standby mode in different supply voltage and one/two slots operation in one cycle. Power in write “0” is lower than others, because in this case negative VVSS circuit will not turn on and save many power dissipation. Another reason is that RBL pass through thread switch is complex and read two sub-cells at the same time. This data is based on two bank are turn on and do the same operation in the one cycle.

Consider lowest power timing product, the minima energy point is operation under VDD=0.65V (Fig 5.21). Although this cell can scaling the supply voltage down to 0.4V all corner pass, read delay too long and raise its energy consumption. Read Ion/off is become very weak under sub threshold voltage, it domain the all 4R4W register file cycle time.

.

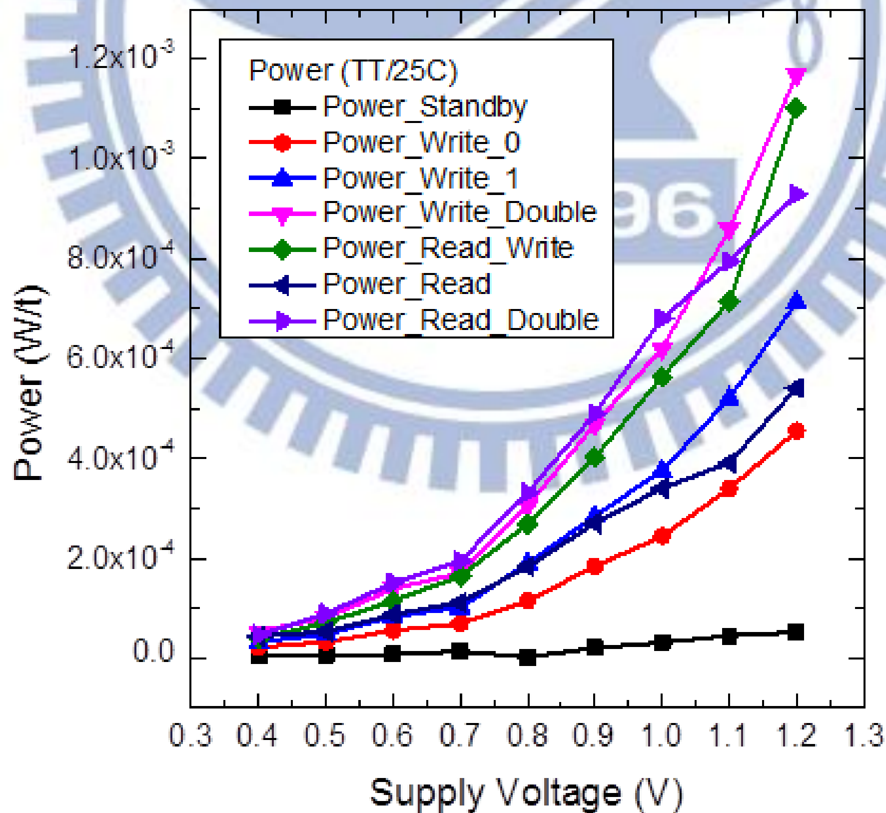


Fig. 5.20 Power consumption with different supply voltage

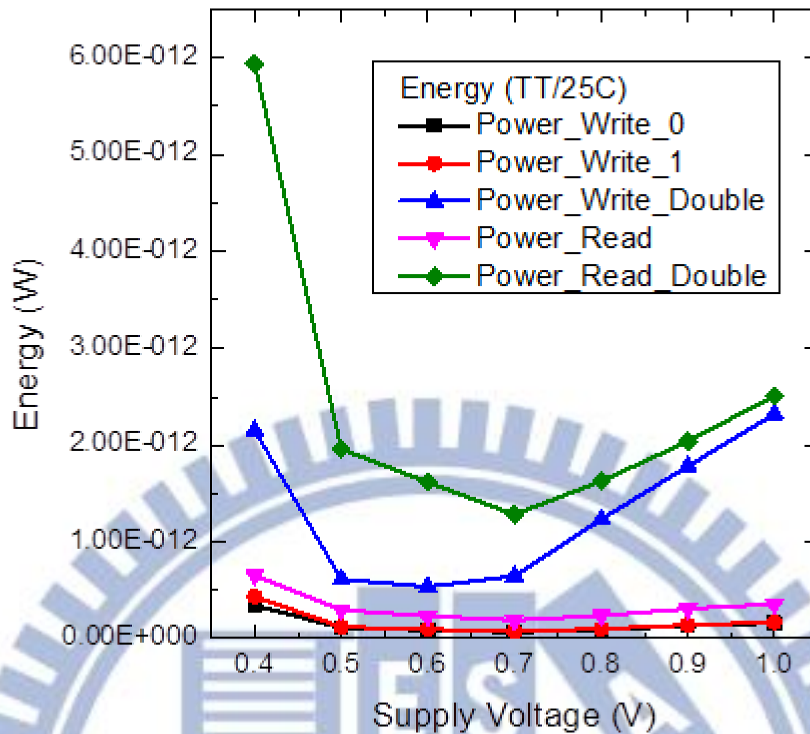


Fig. 5.21 Energy consumption with different supply voltage

In this design, no dummy read in others unselect bit-interleaving in the same row. Compare with conventional design, read power reduction can reduce more than 50%. Power consumption reduction is relative with bit-interleaving number and data store in the bit-cell. If the data is “1”, conventional read buffer NMOS is cut off. No dummy read current; otherwise, if data is “0”, dummy read current occur in half select cell. Fig. 5.22 shows simulation data with different bit-interleaving number and all node store “0” worst case.

In standby mode, conventional 8T SRAM read buffer is connect to ground. Leakage current will pass through two stack NMOS, and extra power consumption generate by this leakage path. RWL keep high when no read operation, compare with conventional 8T SRAM read buffer power saving about 28% in standby mode (Fig. 5.23). Keep RWL on VVSS also cannot take care of RBL sensing failed by RBL leakage induced.

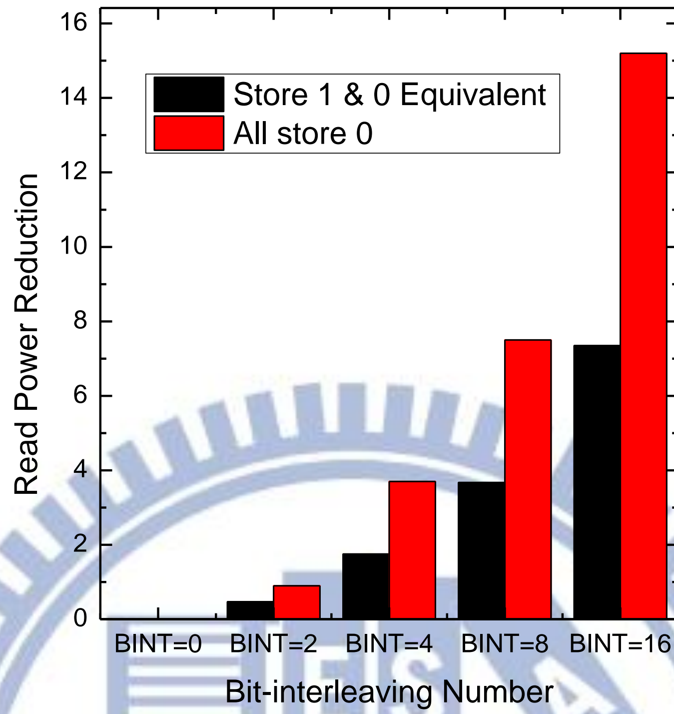


Fig. 5.22 Read power consumption with different bit-interleaving structure

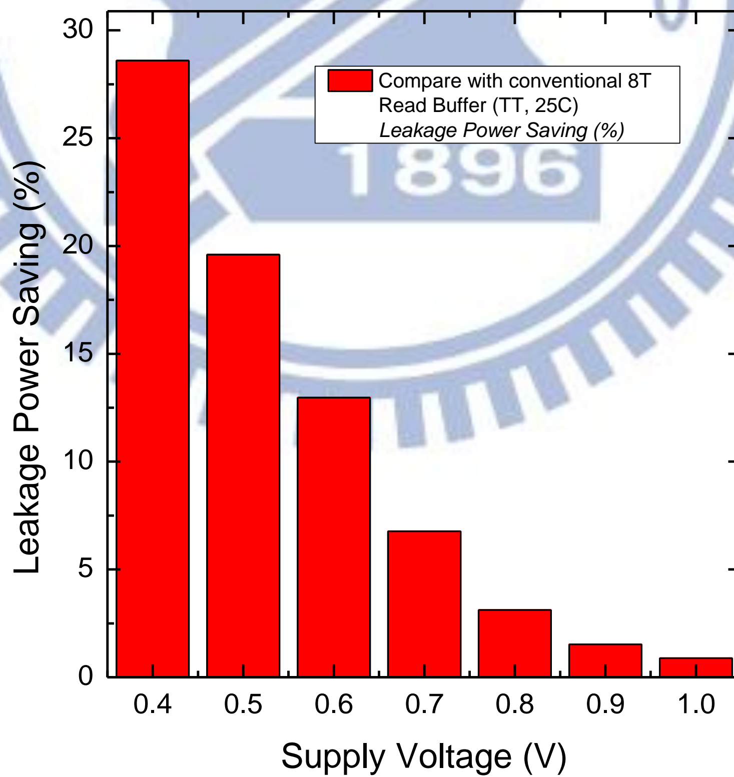


Fig. 5.23 Leakage power saving different voltage

5.5.3 Iso-Area SNM Simulation and Comparison

Comparing with the conventional dual-port (DP) 8T and single-ended (SE) 8T, the proposed 2R2W multi-port bit-cell and 4R4W multi-thread bit-cell area consumption are showed on table 5.4. The cell layout use TSMC 40nm technology, in this design dummy poly is needed between poly and poly. In this reason, this work bit cell is not too larger compare with convention 8T and dual 8T design.

	DP 8T	**SE 8T**	**This Work 2R2W**	**This Work 4R4W**
Normalized Bit-cell area	1	0.84	1.43	2.86
Total bit-cell area	N	0.84N	1.43N	2.86N
Array efficiency	70%	50%	73%	81%
Peripheral circuit area	0.43N	0.91N	0.52N	0.67N
Total sub-array area	1.43N	1.75N	1.95N	3.53N

Table 5.4 These works compare with conventional SRAM bit-cell

Stability issues are very important in register file design, and SNM is the easy way to check the cell robustness. 10000 times Monte Carlo simulation at 0.6V TT corner are showed below. Fig 5.24 shows hold static noise margin (HSNM) compare with this 2R2W and 4R4R new cells. In this design, Con. 8T and dual-8T are all change to same area to be fairer. Fig 5.25 shows write trip point (WTP) in these cells, and in my design, not only HSNM is better but also write trip point is higher than DP 8T and SE 8T design. So in this simulation can show this work is very robust in near threshold even work under subthreshold region.

Write operation in this work is single-end write, so write “0” and write “1” should be consider independently. Write one is the worst case in this cell because NMOS is no perfect 1, and the node will rise up to VDD-V_{TN}.

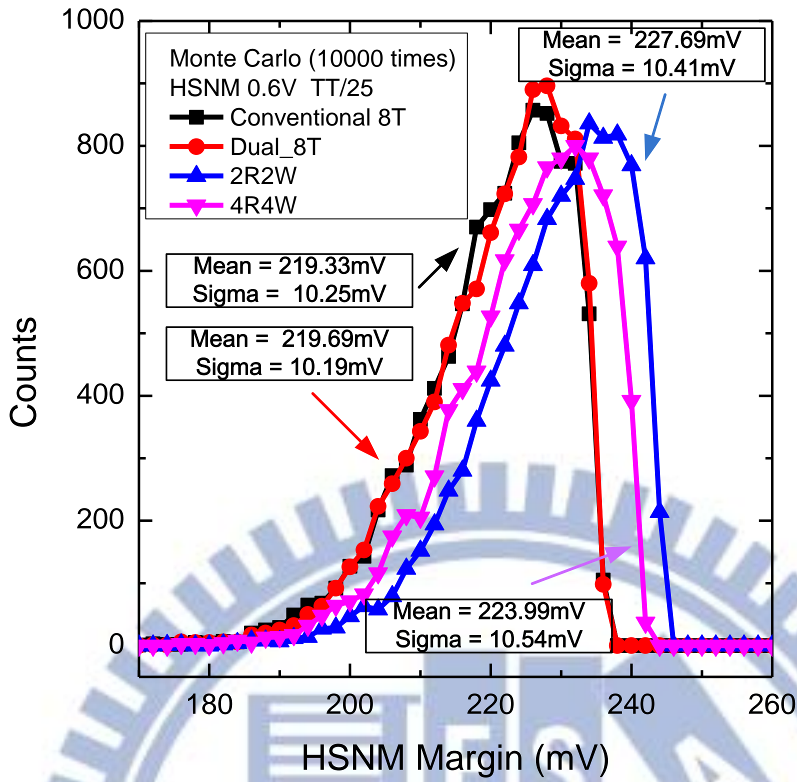


Fig. 5.24 Hold static noise margin simulation

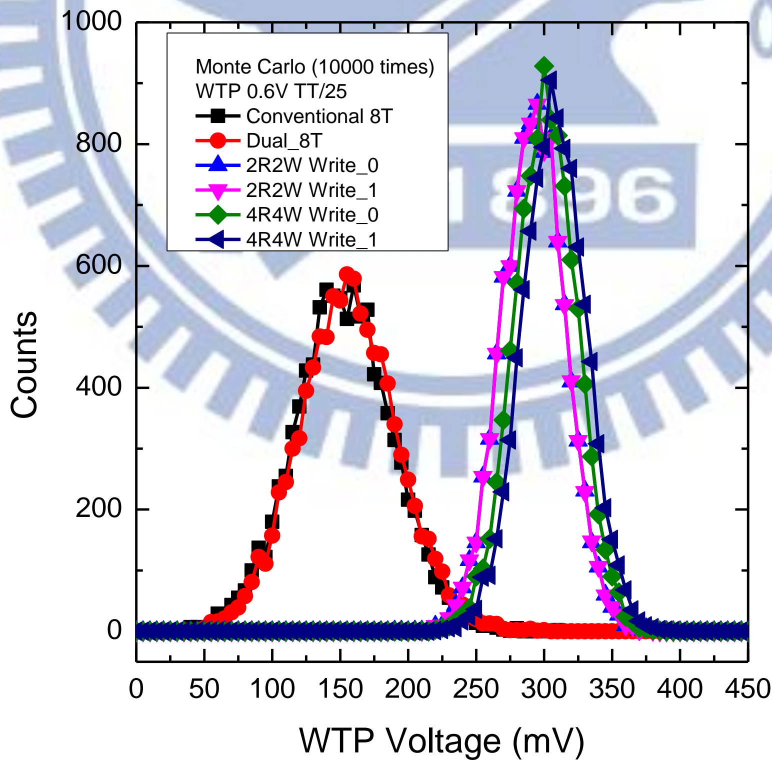


Fig. 5.25 Write trip point simulation

Table 5.5 shows this work compare with pervious works. By this table, our design power is very low. Wide range operation from 0.4V~ 1.2V and low area overhead is better than others work.

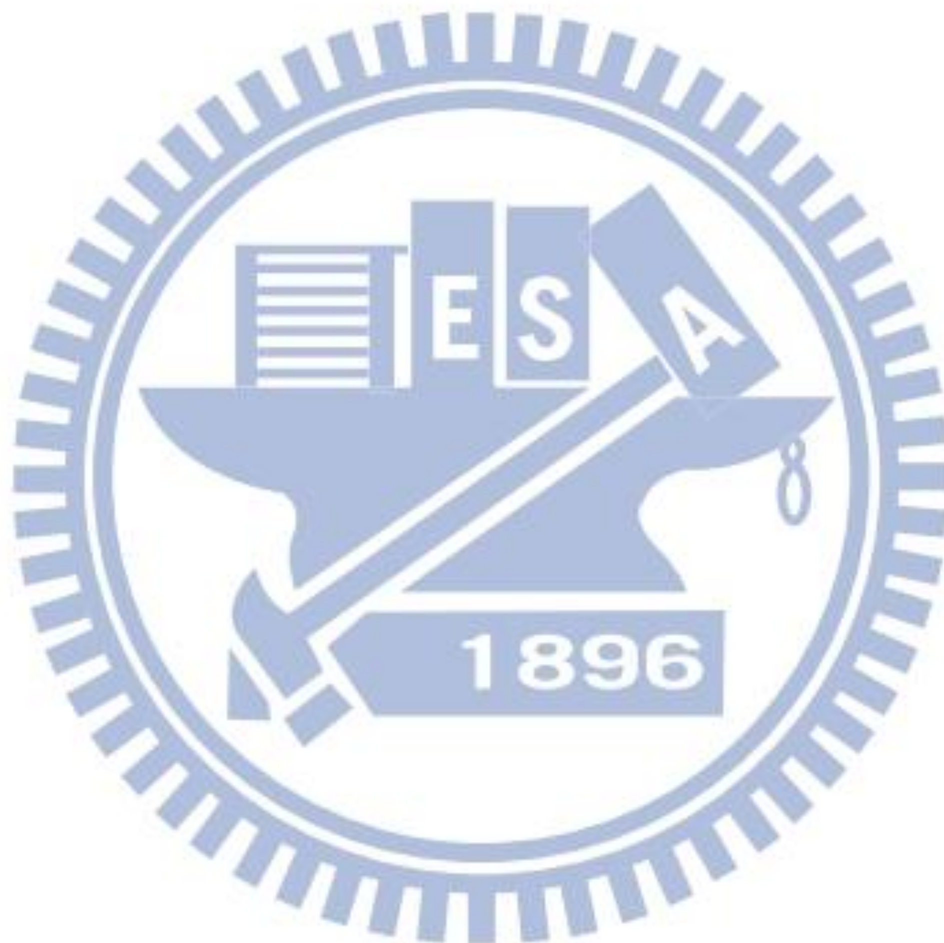
Item	JSSC [1]	ISSCC [2]	ISSCC [3]	ASICON [4]	VLSI CAD [5]	VLSI CAD [6]	4R4W Register file
Technology	90nm	45nm SOI	45nm SOI	65nm	90nm	90nm	40nm
Company	SPARC	IBM	IBM	Fudan	LPMD	LPMD	LPMD
Port Numb.	3R2W	2R1W	2R1W*2	4R2W	4R4W	4R4W	4R4W
Transistor number	35	10	10*2	20	16	8	13
Thread	4	4	No	None	2	None	4
Operation Voltage	1.2V	0.7~1.1V	0.7~0.9V	1.2V	0.5V	0.5V	0.4~1.2V
Freq. / Voltage	1.2Ghz (1.2V)	1.4GHz (0.77V)	2.76GHz (0.9V)	1.2GHz (1.2V)	48MHz (0.5V)	48MHz (0.5V)	43MHz (0.5V)
RF. Size	2Kb	4Kb	11.2Kb	1K	1K	4K	2K
Max Power	75mW	9.375mW	28mW	1.15mW	0.195mW	0.823mW	0.116mW

Table 5.5 Compare with pervious works and this work

5.6 Summary

A wide range 0.4V~1.2V 4R4W multi-thread register file is present in this chapter. Supply four thread structures and low power technology, it very suit for low power portable device, mobile phone or notebook. Multi-thread structure supply high bandwidth and pipe line application. Double pump technology use low CLK frequency can generate double read/write, low CLK frequency and less port number

make power and area reduction. A new structure share read bitline and data switch save power more than 50%, by this new technology which give register file more robust. This work is based on TSMC 40nm general purpose technology process and post-layout simulation by 2K 4R4W multi-port register file design.



Chapter 6
Conclusion & Future Work
6.1 Conclusions

Low power design is the popular design in these years, more and more portable devices need low power IC technology for battery life enlarge. Voltage scaling is a one of method to reduce energy in digital circuit, power can down by the $P=CVdd^2$ rule. Not only low power consumption, but also high resolution video needs high bandwidth for data transmission. Thus, the conventional single-port 6T SRAM bit-cell is no longer suitable in this case. Conventional 8T dual-port is proposed for this reason, but it also has many problems such as conflict problem and half select disturbed issues.

In previous SRAM circuit design, conventional 8T dual-port SRAM bit-call is the best solution for high bandwidth device or non-synchronism circuit design. Due to advance process scaling, the cell stability and read/write ability are degraded due to global and local process variation. The conventional dual 8T SRAM is like 6T SRAM, it is not suitable in low voltage region due to read disturb problem and half select disturb problem.

In this thesis, a robust wide range and low power 2R2W multi-port SRAM bit-cell is proposed. For power saving, new share WBL structure design and single-end read/write structure is used. By using this structure, write driver can reduce to 1/2. Power reduction in this design is very significantly. This bit-cell is very robust that no have half select problem and bit-interleaving supported. Negative VVSS and cut off feedback are improving write ability in low voltage operation. Hence, the proposed 13T 2R2W multi-port SRAM can do a very wide range operation (0.5V~1.4V).

Register file is the most important memory in process, it must be very robustness. Parallel and multi-thread can do pipe-line to save time if one of thread is in stall. Power

hungry is often happened in the high speed processor, by this reason a low power design in portable device is very important. For cell robust consideration, conflict problem is not allowable and many methods are present for solved the W/R conflict problem.

In this work, a new low power dissipation and conflict free technology is proposed. Double pump combines with multi-thread design can supply high bandwidth for media parallel operation. Data conflict switch technology can gain more write success chance than conventional conflict circuit design. Active power reduction is very obviously that new share RBL structure can eliminate dummy read power consumption. Standby power reduction is reduction by keep RWL to VVSS, leakage reduction is significantly in lower supply voltage. Multi-bank structure reduces area and power dissipation, and only need a priority circuit for access decision. Therefore, it can achieve low power and high robustness when operation under low supply voltage.

6.2 Future Work

The proposed 13T register file bit cell can significantly improve write/read ability by using multi-threshold structure. To use low Vt NMOS device can raise the voltage level (V_{DD}-V_{TN}) when write “1” operation, and body effect of two stack NMOS can also keep V_{TN} not too low. Cross couple inverter use high Vt to reduce leakage power and more robustness. Voltage detect circuit included for controlling negative VVSS on /off, post-layout simulation shows that only in low voltage need negative VVSS assist. For achieving more robust in low voltage operation, a PVT monitor is need to detect and give information to replica circuit.

References

Chapter 1

[1.1] Bo Zhai, D. Blaauw, D. Sylvester, K. Flautner, "***Theoretical and practical limits of dynamic voltage scaling,***" Design Automation Conference, 2004. Proceedings. 41st, pp.868-873, 7-11 July 2004

[1.2] Wang, A. Chandrakasan, "***A 180mV FFT processor using subthreshold circuit techniques,***" 2004 IEEE International Solid-State Circuits Conference (ISSCC), pp. 292- 529 Vol.1, 15-19 Feb. 2004

[1.3] N. Lindert, T. Sugii, S. Tang, Hu Chenming, "***Dynamic threshold pass-transistor logic for improved delay at lower power supply voltages,***" IEEE Journal of Solid-State Circuits, vol.34, no.1, pp.85-89, Jan 1999

[1.4] G. Chen, D. Sylvester, D. Blaauw, T. Mudge, "***Yield-Driven Near-Threshold SRAM Design,***" Very Large Scale Integration (VLSI) Systems, IEEE Transactions on , vol.18, no.11, pp.1590-1598, Nov. 2010

[1.5] S. Mukhopadhyay, H. Mahmoodi, and K. Roy, "***Modeling of failure probability and statistical design of SRAM array for yield enhancement in nanoscaled CMOS,***" IEEE Trans. Comput.–Aided Design (CAD) integer. Cricuit Syst.,vol. 24, no. 12, Dec. 2005, pp. 1859–1880.

[1.6] J. P. Kulkarni and K. Roy "***Ultralow-voltage Process-variation-Tolerant Schmitt-Trigger-Based SRAM Design,***" IEEE Trans. VLSI System 2011.

[1.7] J. Shrivas, S. Akashe, "***Impact of Design Parameter on SRAM Bit Cell,***" *2012 Second International Conference on Advanced Computing & Communication Technologies*, pp.353-356, 7-8 Jan. 2012

[1.8] M.E. Sinangil, N. Verma, A.P. Chandrakasan, "***A Reconfigurable 8T Ultra-Dynamic Voltage Scalable (U-DVS) SRAM in 65 nm CMOS,***" IEEE Journal of Solid-State Circuits, vol.44, no.11, pp.3163-3173, Nov. 2009

Chapter 2

[2.1] ITRS. (2011). Process Integration, Devices, and Structures (PIDS) Plus New MASTER Model Table [Online].

http://www.itrs.net/Links/2011ITRS/home2011.htm

[2.2] Andrew B. Kahng, "***Product Futures,***" *Design & Test of Computers,* IEEE , vol.28, no.6, pp.88-89, Nov.-Dec. 2011

[2.3] Peter Kuoyuan Hsu, Yukit Tang, Derek Tao, Ming-Chieh Huang, Min-Jer Wang, CH Wu, Quincy Lee, "**A SRAM cell array with adaptive leakage reduction scheme for data retention in 28nm high-k metal-gate CMOS,**" *VLSI Circuits (VLSIC), 2012 Symposium* , pp.62-63, 13-15 June 2012

[2.4] F. Fallah, M. Pedram, "**Standby and Active Leakage Current Control and Minimization in CMOS VLSI Circuits,**" *IEICE Trans. Electron*, vol. E88-C, no. 4, April 2005, pp. 509-519.

[2.5] Kaushik Roy, S. Mukhopadhyay, and H. Mahomoodi-Meimand, "**Leakage Current Mechanisms and Leakage Reduction Techniques in Deep-Submicrometer CMOS Circuits,**" *Proceedings of the IEEE*, vol. 91, no. 2, February 2003, pp. 305-327.

[2.6] Valmiki Mukherjee, Saraju P. Mohanty, Elias Kougianos, Rahul Allawadhi, Ramakrishna Velagapudi,"**Gate leakage current analysis in READ/WRITE/ IDLE states of a SRAM cell,**" *Region 5 Conference, 2006 IEEE* , pp.196-200, 7-9 April 2006

[2.7] Semiconductor Industry Association, International Technology Roadmap for Semiconductors, 2003 edition, http://public.itrs.net.

[2.8] J. Shrivas, S. Akashe, "**Impact of Design Parameter on SRAM Bit Cell,**" *Advanced Computing & Communication Technologies (ACCT), 2012 Second International Conference* , pp.353-356, 7-8 Jan. 2012

[2.9] Xingsheng Wang, G. Roy, O. Saxod, A. Bajolet, A. Juge, A. Asenov, "**Simulation Study of Dominant Statistical Variability Sources in 32-nm High- /Metal Gate CMOS,**" *Electron Device Letters, IEEE* , vol.33, no.5, pp.643-645, May 2012

[2.10] Ming-Long Fan, Vita Pi-Ho Hu, Yin-Nien Chen, Pin Su, Ching-Te Chuang, "**Impacts of Random Telegraph Noise on FinFET devices, 6T SRAM cell, and logic circuits,**" *Reliability Physics Symposium (IRPS), 2012 IEEE International*, pp.CR.1.1-CR.1.6, 15-19 April 2012

[2.11] K. Endo, S. O'uchi, Y. Ishikawa, Yongxun Liu, T. Matsukawa, K. Sakamoto, J. Tsukada, H. Yamauchi, M. Masahara, "**A Correlative Analysis Between Characteristics of FinFETs and SRAM Performance,**" *Electron Devices, IEEE Transactions* , vol.59, no.5, pp.1345-1352, May 2012

[2.12] D. Eckerbert, P. Larsson-Edefors, "**Interconnect-driven short-circuit power modeling,**" *Digital Systems, Design, 2001. Proceedings. Euromicro*

Symposium , pp.414-421, 2001

[2.13] H. J. M. Veendrick. "**Short-circuit dissipation of static CMOS circuitry and its impact on the design of buffer circuits.**" *IEEE Journal of Solid-state Circuits*, pages 468-473, Aug. 1984.

[2.14] N. Hedenstiema, K. o. Jeppson. "**CMOS circuit speed and buffer optimization.**" *IEEE Transactions on Computer- AidedDesign*, CAD-6(2):27@-281, March 1987

[2.15] N. Verma, and A. P. Chandrakasan, "**A 256 kb 65 nm 8T sub-threshold SRAM employing Sense-amplifier Redundancy,**" *IEEE Journal of Solid-State Circuits*, vol. 43, no. 1, pp. 141–14, Jan. 2008.

[2.16] Y. Nakagome, M. Horiguchi, T. Kawahara, K. Itoh, "**Review and future prospects of low-voltage RAM circuits,**" *IBM Journal of Research and Development*, vol.47, no.5.6, pp.525-552, Sept. 2003

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

[2.18] Naveen Verma, Joyce Kwong, and Anantha P. Chandrakasan, "***Nanometer MOSFET Variation in Minimum Energy Subthreshold Circuits***," IEEE Trans. on Electron Device, January 2008, pp. 163-174.

[2.19] Kevin Zhang, "***F1: Embedded Memory Design for Nano-Scale VLSI Systems,***" *Solid-State Circuits Conference, ISSCC 2008.* IEEE International Digest of Technical Papers., pp.650-651, 3-7 Feb. 2008

[2.20] A. Alvandpour, R.K. Krishnamurthy, K. Soumyanath, S.Y. Borkar, "***A sub-130-nm conditional keeper technique,***" *IEEE Journal of Solid-State Circuits,* vol.37, no.5, pp.633-638, May 2002

[2.21] R.G.D. Jeyasingh, N. Bhat, B. Amrutur, "***Adaptive Keeper Design for Dynamic Logic Circuits Using Rate Sensing Technique***," IEEE Transactions on Very Large Scale Integration (VLSI) Systems, vol.19, no.2, pp.295-304, Feb. 2011

[2.22] Y. Lih, N. Tzartzanis, W. W. Walker, "***A Leakage Current Replica Keeper for Dynamic Circuits***," IEEE Journal of Solid-State Circuits, vol.42, no.1, pp.48-55, Jan. 2007

[2.23] A. Raychowdhury, B. Geuskens, J. Kulkarni, J. Tschanz, K. Bowman, T.

Karnik, Shih-Lien Lu; V. De, M.M. Khellah, "***PVT-and-aging adaptive wordline boosting for 8T SRAM power reduction***," 2010 IEEE International Solid-State Circuits Conference Digest of Technical Papers (ISSCC), pp.352-353, 7-11 Feb. 2010

[2.24] J. Kulkarni, B. Geuskens, T. Karnik, M. Khellah, J. Tschanz, V. De, "***Capacitive-coupling wordline boosting with self-induced VCC collapse for write VMIN reduction in 22-nm 8T SRAM***," 2012 IEEE International Solid-State Circuits Conference Digest of Technical Papers (ISSCC), pp.234-236, 19-23 Feb. 2012

[2.25] Ming-Dou Ker, Shih-Lun Chen, Chia-Shen Tsai, "***Design of charge pump circuit with consideration of gate-oxide reliability in low-voltage CMOS processes***," IEEE Journal of Solid-State Circuits, vol.41, no.5, pp. 1100-1107, May 2006

[2.26] Xueqiang Wang, Dong Wu, Fengying Qiao, Peng Zhu, Kan Li, Liyang Pan, Runde Zhou, "***A high efficiency CMOS charge pump for low voltage operation***," IEEE 8th International Conference on ASICON '09., pp.320-323, 20-23 Oct. 2009

[2.27] K. Nii, Y. Masuda, M. Yabuuchi, Y. Tsukamoto, S. Ohbayashi, S. Imaoka, M. Igarashi, K. Tomita, N. Tsuboi, H. Makino, K. Ishibashi, H. Shinohara, "***A 65 nm Ultra-High-Density Dual-Port SRAM with 0.71um 8T-Cell for SoC***," Symposium on Digest of VLSI Circuits Technical Papers., pp.130-131. 2006.

[2.28] T. Suzuki, H. Yamauchi, Y. Yamagami, K. Satomi, H. Akamatsu, "***A Stable 2-Port SRAM Cell Design Against Simultaneously Read/Write-Disturbed Accesses***," IEEE Journal of Solid-State Circuits, vol.43, no.9, pp.2109-2119, Sept. 2008

[2.29] Ping Liu, J. Wang, M. Phan, M. Garg, R. Zhang, A. Cassier, L. Chua-Eoan, B. Andreev, S. Weyland, S. Ekbote, M. Han, J. Fischer, G.C.-F. Yeap, Ping-Wei Wang, "***A dual core oxide 8T SRAM cell with low Vccmin and dual voltage supplies in 45nm triple gate oxide and multi Vt CMOS for very high performance yet low leakage mobile SoC applications***," 2010 Symposium on VLSI Technology, pp.135-136, 15-17 June 2010

[2.30] J. Singh, D.S. Aswar, S.P. Mohanty, D.K. Pradhan, "***A 2-port 6T SRAM bitcell design with multi-port capabilities at reduced area overhead***," 11th International Symposium on Quality Electronic Design, pp.131-138, 22-24 March 2010

[2.31] Jui-Jen Wu, Yen-Huei Chen, Meng-Fan Chang, Po-Wei Chou, Chien-Yuan Chen, Hung-Jen Liao, Ming-Bin Chen, Yuan-Hua Chu, Wen-Chin Wu, H. Yamauchi, "***A Large δ Vt/VDD Tolerant Zigzag 8T SRAM With Area-Efficient Decoupled Differential Sensing and Fast Write-Back Scheme***," IEEE Journal of Solid-State Circuits, vol.46, no.4, pp.815-827, April 2011

[2.32] N. Verma, A.P. Chandrakasan, "**A 65nm 8T Sub-Vt SRAM Employing Sense-Amplifier Redundancy,**" Digest of Technical Papers. IEEE International Solid-State Circuits Conference (ISSCC), pp.328-606, 11-15 Feb. 2007

[2.33] I.J. Chang, J.J. Kim, S.P. Park, K. Roy, "***A 32kb 10T Subthreshold SRAM Array with Bit-Interleaving and Differential Read Scheme in 90nm CMOS***," Digest of Technical Papers. IEEE International Solid-State Circuits Conference (ISSCC), pp.388-622, 3-7 Feb. 2008

[2.34] Zhiyu Liu, V. Kursun, "***Characterization of a Novel Nine-Transistor SRAM Cell***," IEEE Transactions on Very Large Scale Integration Systems, vol.16, no.4, pp.488-492, April 2008

[2.35] S.K. Jain, P. Agarwal, "***A low leakage and SNM free SRAM cell design in deep sub- micron CMOS technology***," VLSI Design, 2006. 19th International Conference on Held jointly with 5th International Conference on Embedded Systems and Design., pp. 4 pp., 3-7 Jan. 2006

[2.36] Jinhui Chen, L.T. Clark, Tai-Hua Chen, "***An Ultra-Low-Power Memory With a Subthreshold Power Supply Voltage***," IEEE Journal of Solid-State Circuits, vol.41, no.10, pp.2344-2353, Oct. 2006

[2.37] Chen Tai-Hua, Chen Jinhui, L.T. Clark, J.E. Knudsen, G. Samson , "***Ultra-Low Power Radiation Hardened by Design Memory Circuits***," IEEE Transactions on Nuclear Science, vol.54, no.6, pp.2004-2011, Dec. 2007

[2.38] M. Haghi, J. Draper, "***The 90 nm Double-DICE storage element to reduce Single-Event upsets***," 52nd IEEE International Midwest Symposium on Circuits and Systems, pp.463-466, 2-5 Aug. 2009

[2.39] A. S.Leon, K. W. Tam, J. L. Shin, D. Weisner, F. Schumacher, "***A Power-Efficient High-Throughput 32-Thread SPARC Processor***," IEEE Journal of Solid-State Circuits, vol.42, no.1, pp.7-16, Jan. 2007

[2.40] C. Johnson, D.H. Allen, J. Brown, S. Vanderwiel, R. Hoover, H. Achilles,

C.-Y. Cher, G.A. May, H. Franke, J. Xenedis, C. Basso, "***A wire-speed powerTM processor: 2.3GHz 45nm SOI with 16 cores and 64 threads***," 2010 IEEE International Solid-State Circuits Conference Digest of Technical Papers, pp.104-105, 7-11 Feb. 2010

[2.41] G.S. Ditlow, R.K. Montoye, S.N. Storino, S.M. Dance, S. Ehrenreich, B.M. Fleischer, T.W. Fox, K.M. Holmes, J. Mihara, Y. Nakamura, S. Onishi, R. Shearer, D. Wendel, Leland Chang , "***A 4R2W register file for a 2.3GHz wire-speed POWER™ processor with double-pumped write operation***," 2011 IEEE International Solid-State Circuits Conference Digest of Technical Papers, pp.256-258, 20-24 Feb. 2011

[2.42] E.S. Fetzer, M. Gibson, A. Klein, N. Calick, Chengyu Zhu, E. Busta, B. Mohammad, "**A fully bypassed six-issue integer data path and register file on the Itanium-2 microprocessor,**" IEEE Journal of Solid-State Circuits, vol.37, no.11, pp. 1433- 1440, Nov 2002

[2.43] D. Wendel, R. Kalla, R. Cargoni, J. Clables, J. Friedrich, R. Frech, J. Kahle, B. Sinharoy, W. Starke, S. Taylor, S. Weitzel, S.G. Chu, S. Islam, V. Zyuban, "***The implementation of POWER7TM: A highly parallel and scalable multi-core high-end server processor***," 2010 IEEE International Solid-State Circuits Conference Digest of Technical Papers (ISSCC), vol., no., pp.102-103, 7-11 Feb. 2010

Chapter 3

[3.1] K. Nii, Y. Tsukamoto, M. Yabuuchi, Y. Masuda, S. Imaoka, K. Usui, S. Ohbayashi, H. Makino, H. Shinohara, "**Synchronous Ultra-High-Density 2RW Dual-Port 8T-SRAM With Circumvention of Simultaneous Common-Row-Access,**" *Solid-State Circuits, IEEE Journal* , vol.44, no.3, pp.977-986, March 2009

[3.2] Y. Ishii, H. Fujiwara, K. Nii, H. Chigasaki, O. Kuromiya, T. Saiki, A. Miyanishi, Y. Kihara, "**A 28-nm dual-port SRAM macro with active bitline equalizing circuitry against write disturb issue,**" *VLSI Circuits (VLSIC), 2010 IEEE Symposium* , pp.99-100, 16-18 June 2010

[3.3] Y. Ishii, H. Fujiwara, S. Tanaka, Y. Tsukamoto, K. Nii, Y. Kihara, K. Yanagisawa, "**A 28 nm Dual-Port SRAM Macro With Screening Circuitry Against Write-Read Disturb Failure Issues,**" *Solid-State Circuits, IEEE Journal* , vol.46, no.11, pp.2535-2544, Nov. 2011

[3.4] T. Suzuki, H. Yamauchi, Y. Yamagami, K. Satomi, H. Akamatsu, "***A Stable 2-Port SRAM Cell Design Against Simultaneously Read/Write-Disturbed Accesses,***" *Solid-State Circuits, IEEE Journal* , vol.43, no.9, pp.2109-2119, Sept. 2008

[3.5] Y. Ishii, Y. Tsukamoto, K. Nii, H. Fujiwara, M. Yabuuchi, K. Tanaka, S. Tanaka, Y. Shimazaki, "**A 28nm 360ps-access-time two-port SRAM with a time-sharing scheme to circumvent read disturbs,**" *Solid-State Circuits Conference Digest of Technical Papers (ISSCC), 2012 IEEE International* , pp.236-238, 19-23 Feb.

[3.6] Y. Tsukamoto, T. Kida, T. Yamaki, Y. Ishii, K. Nii, K. Tanaka, S. Tanaka, Y. Kihara, "**Dynamic stability in minimum operating voltage Vmin for single-port and dual-port SRAMs,**" *Custom Integrated Circuits Conference (CICC), 2011 IEEE* , pp.1-4, 19-21 Sept. 2011

Chapter 4

[4.1] H. McIntyre, S. Arekapudi, E. Busta, T. Fischer, M. Golden, A. Horiuchi, T. Meneghini, S. Naffziger, J. Vinh, "**Design of the Two-Core x86-64 AMD “Bulldozer” Module in 32 nm SOI CMOS,**" *Solid-State Circuits, IEEE Journal* , vol.47, no.1, pp.164-176, Jan. 2012

[4.2] R. Riedlinger, R. Arnold, L. Biro, B. Bowhill, J. Crop, K. Duda, E.S. Fetzer, O. Franza, T. Grutkowski, C. Little, C. Morganti, G. Moyer, A. Munch, M. Nagarajan, C. Parks, C. Poirier, B. Repasky, E. Roytman, T. Singh, M.W. Stefaniw, "**A 32 nm, 3.1 Billion Transistor, 12 Wide Issue ItaniumR Processor for Mission-Critical Servers,**" *Solid-State Circuits, IEEE Journal* , vol.47, no.1, pp.177-193, Jan. 2012

[4.3] Youngsoo Shin, Sewan Heo, Hyung-Ock Kim, Jung Yun Choi, "**Supply Switching With Ground Collapse: Simultaneous Control of Subthreshold and Gate Leakage Current in Nanometer-Scale CMOS Circuits,**" *IEEE Transactions on Very Large Scale Integration Systems* , vol.15, no.7, pp.758-766, July 2007

[4.4] Hyung-Ock Kim, Bong Hyun Lee, Jong-Tae Kim, Jung Yun Choi, Kyu-Myung Choi, Youngsoo Shin, "**Supply Switching With Ground Collapse for Low-Leakage Register Files in 65-nm CMOS,**" *IEEE Transactions on Very Large Scale Integration Systems*, vol.18, no.3, pp.505-509, March 2010

[4.5] K. Johguchi, Y. Mukuda, S. Izumi, H.J. Mattausch, T. Koide, "**A 0.6-Tbps, 16-port SRAM design with 2-stage- pipeline and multi-stage-sensing scheme,**" *33rd European Solid State Circuits Conference*, ,pp.320-323, 11-13 Sept. 2007

[4.6] Koh Johguchi, Ken-ichi Aoyama, Tetsuya Sueyoshi, Hans Jurgen Mattausch, Tetsushi Koide, Moto Maeda, Tetsuo Hironaka, Kazuya Tanigawa, "**Multi-Bank Register File for Increased Performance of Highly-Parallel Processors,**" *Proceedings of the 32nd European Solid-State Circuits Conference*, pp.154-157, Sept. 2006

[4.7] Na Gong, Geng Tang, Jinhui Wang, R. Sridhar, "**Low power tri-state register files design for modern out-of-order processors,**" *SOC Conference (SOCC), 2011 IEEE International* , pp.323-328, 26-28 Sept. 2011

[4.8] M. Danek, L. Kafka, L. Kohout, J. Sykora, "**Instruction set extensions for multi-threading in LEON3,**" *Design and Diagnostics of Electronic Circuits and Systems (DDECS), 2010 IEEE 13th International Symposium* , pp.237-242, 14-16 April 2010

[4.9] I. Oz, H.R. Topcuoglu, M. Kandemir, O. Tosun, "**Quantifying Thread Vulnerability for Multicore Architectures,**" *Parallel, Distributed and Network-Based Processing (PDP), 2011 19th Euromicro International Conference* , pp.32-39, 9-11 Feb. 2011

[4.10] A. S. Leon, K. W. Tam, J. L. Shin, D. Weisner, F. Schumacher, "**A Power-Efficient High-Throughput 32-Thread SPARC Processor,**" *Solid-State Circuits, IEEE Journal* , vol.42, no.1, pp.7-16, Jan. 2007

[4.11] C. Johnson, D.H. Allen, J. Brown, S. Vanderwiel, R. Hoover, H. Achilles, C.-Y. Cher, G.A. May, H. Franke, J. Xenedis, C. Basso, "**A wire-speed powerTM processor: 2.3GHz 45nm SOI with 16 cores and 64 threads,**" *2010 IEEE International Solid-State Circuits Conference Digest of Technical Papers*, pp.104-105, 7-11 Feb. 2010

[4.12] Naraig Manjikian; , "**Implementation of Hardware Multithreading in a Pipelined Processor,**" *Circuits and Systems, 2006 IEEE North-East Workshop on* , vol., no., pp.145-148, June 2006

[4.13] E.S. Fetzer, D. Dahle, C. Little, K. Safford , "**The Parity protected, multithreaded register files on the 90-nm itanium microprocessor,**" *IEEE*

Journal of Solid-State Circuits, vol.41, no.1, pp. 246- 255, Jan. 2006

[4.14] D. F. Wendel, J. Barth, D. M. Dreps, S. Islam, J. Pille, J. A. Tierno, "**IBM POWER7 processor circuit design,**" *IBM Journal of Research and Development* , vol.55, no.3, pp.1:1-1:8, May-June 2011

[4.15] V. Zyuban, J. Friedrich, C. J. Gonzalez, R. Rao, M. D. Brown, M. M. Ziegler, H. Jacobson, S. Islam, S. Chu, P. Kartschoke, G. Fiorenza, M. Boersma, J. A. Culp, "**Power optimization methodology for the IBM POWER7 microprocessor,**" *IBM Journal of Research and Development* , vol.55, no.3, pp.7:1-7:9, May-June 2011

[4.16] U-Chan Kuo, Hao-I Yang, Wei Hwang, " ***A Sub-threshold Multi-Port Register File,***" 19th VLSI Design/CAD Symposium, Taiwan, Aug. 2008.

[4.17] G.S. Ditlow, R.K. Montoye, S.N. Storino, S.M. Dance, S. Ehrenreich, B.M. Fleischer, T.W. Fox, K.M. Holmes, J. Mihara, Y. Nakamura, S. Onishi, R. Shearer, D. Wendel, Leland Chang , "**A 4R2W register file for a 2.3GHz wire-speed POWER7 processor with double-pumped write operation,**" *Solid-State Circuits Conference Digest of Technical Papers (ISSCC), 2011 IEEE International* , pp.256-258, 20-24 Feb. 2011

Chapter 5

[5.1] A.S. Leon, K. W.Tam, J.L. Shin, D. Weisner, F. Schumacher, "***A Power-Efficient High-Throughput 32-Thread SPARC Processor,***" IEEE Journal of Solid-State Circuits, vol.42, no.1, pp.7-16, Jan. 2007

[5.2] C. Johnson, D.H. Allen, J. Brown, S. Vanderwiel, R. Hoover, H. Achilles, C.-Y. Cher, G.A. May, H. Franke, J. Xenedis, C. Basso, "***A wire-speed power™ processor: 2.3GHz 45nm SOI with 16 cores and 64 threads,***" 2010 IEEE International Solid-State Circuits Conference Digest of Technical Papers (ISSCC), vol., no., pp.104-105, 7-11 Feb. 2010

[5.3] G.S. Ditlow, R.K. Montoye, S.N. Storino, S.M. Dance, S. Ehrenreich, B.M. Fleischer, T.W. Fox, K.M. Holmes, J. Mihara, Y. Nakamura, S. Onishi, R. Shearer, D. Wendel, Leland Chang; , "***A 4R2W register file for a 2.3GHz wire-speed POWER™ processor with double-pumped write operation,***" 2011 IEEE International Solid-State Circuits Conference Digest of Technical Papers (ISSCC), pp.256-258, 20-24 Feb. 2011

[5.4] U-Chan Kuo, Hao-I Yang, Wei Hwang, " ***A Sub-threshold Multi-Port Register File,***" 19th VLSI Design/CAD Symposium, Taiwan, Aug. 2008

[5.5] Shyh-Chyi Yang, Hao-I Yang, Wei Hwang "***Thermal Management with In-Situ Process-Temperature Sensor for TSV 3D-ICs,***" 20th VLSI Design/CAD

Symposium, Taiwan, Aug. 2009.
[5.6] Jun Han, Xingxing Zhang, Baoyu Xiong, Zhiyi Yu, Xiaoyang Zeng, "***A control scheme for a 65nm 32×32b 4-read 2-write register file,***" 2011 IEEE 9th International Conference on ASIC , 739-742, 25-28 Oct. 2011