國 立 交 通 大 學

電子工程學系 電子研究所碩士班

碩士論文

低功率 8T 靜態隨機存取記憶體和次臨界多埠暫存器的設計與實現

Design and Implementation of Low Power 8T SRAM and Sub-threshold Multi-Port Register File

研究生 ： 楊仕祺

指導教授 ： 黃 威 教授

中 華 民 國 九 十 八 年 六 月

低功率 8T 靜態隨機存取記憶體和次臨界多埠暫存器的設計與實現

Design and Implementation of Low Power 8T SRAM and Sub-threshold Multi-Port Register File

研究生：楊仕祺　　Student : Shyh-Chyi Yang

指導教授：黃威 教授　　Advisor : Prof. Wei Hwang

國立交通大學

電子工程學系電子研究所

碩士論文

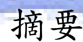
A Thesis
Submitted to Department of Electronics Engineering & Institute of Electronics
College of Electrical Engineering and Computer Engineering
National Chiao Tung University
In partial Fulfillment of the Requirements
For the Degree of
Master
In
Electronics Engineering
June 2008
Hsinchu, Taiwan, Republic of China

中 華 民 國 九 十 八 年 六 月

低功率 8T 靜態隨機存取記憶體和次臨界多埠暫存器的設計與實現

學生: 楊仕祺　　　　　　　　　　　　　　指導教授: 黃威教授

國立交通大學電子工程學系電子研究所



摘要

嵌入式記憶體在現今高效能低功率晶片佔了重要的地位。傳統的 6T 靜態隨機存取記憶體在先進製程的技術下面臨了許多的挑戰。在此篇論文中，將會討論 NBTI/PBTI 效應對於靜態隨機存取記憶體的影響，並且提出了一個能夠降低此效應影響的架構和方法，而降低幅度高達 32%-48%。另一個去解決傳統 6T 靜態隨機存取記憶體問題的方法是設計一個新的儲存架構。這篇論文呈現了一個新的 8T 架構，相較於傳統的設計，這個設計的資料的穩定性有 1.74 倍提升。此外，這篇論文提出一個新的介面電路能夠讓 8T 相容於傳統 6T 靜態隨機存取記憶體周邊電路。

另一項重要的靜態隨機存取記憶體的應用是多埠暫存器設計。在傳統的設計上，為了達到多埠的功能，設計者去會去增加儲存單元的埠數目。然而，在先進製程下，這個設計也將面臨許多的問題。因此一個可以應用在超長指令集架構的數位訊號處理器中並且操作在 1 伏到 0.25 伏的多埠暫存器設計在此篇論文呈現，其設計的架構是以多組區塊來達成多埠的實現。為了能讓暫存器操作在次臨界電壓，像是雙 Vt 的儲存單元設計，負電壓寫入機制，改進的讀取機制等，使之可以操作在低電壓下。這個可以支援同時 4 讀 4 寫的暫存器系以 UMC 90nm CMOS 製程設計，操作電壓在 0.25 伏特時，22.3-22.9 微瓦功率消耗。

Design and Implementation of Low Power 8T SRAM and Sub-threshold Multi-Port Register File

Student : Shyh-Chyi Yang Advisor : Prof. Wei Hwang

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

ABSTRACT

Embedded memory plays a significant role in high performance and low power VLSI technology. Stability and area of traditional 6T SRAM is difficult to scale down in future process due to the serious PVT variation and other effect, such as reliability issue: NBTI and PBTI. In this thesis, detailed analysis of timing control degradation caused by NBTI and PBTI on SRAM is presented. Furthermore, NBTI/PBTI tolerant design for nanoscale CMOS SRAM is also presented, which reduces 32%-48% degradation. Another method to address drawbacks of 6T SRAM cell is to design another new bit-cell, presented in this thesis. This new bit-cell eliminates read disturb and half-select disturb of 6T bit-cell and has 1.75X read SNM improvement when compared to the conventional 6T SRAM cell. An interface circuit design lets the unique structure of new 8T bit-cell combine the peripheral circuit of 6T SRAM without declining performance.

Another important SRAM design is multi-port SRAM-based register file. Similarly, multi-port bit-cell in nano-scale process or ultra low voltage works fail. As a result, a micro-watt multi-port register file with wide operating voltage range for micro-power applications is presented. Multibank architecture for simultaneous access with collision detecting technique is proposed. The architecture can be applied to VLIW DSP, and has been analyzed under wide operating voltage range between 1V to 0.25V with varies process corner. Negative voltage write scheme ensures successful write in deep sub-threshold region. Also, an improved read buffer footer and controllable pre-charge in read scheme are designed. A 4W/4R 16KB register file is implemented in UMC 90nm CMOS technology. The simulation results show that the maximum active power of multi-port register file can achieve near 22.3-22.9uW at 485 KHz under 0.25V.

誌　　謝

可以完成這篇論文，我想我要感覺非常多的人。首先，我的指導教授黃威提供了我研究的資源及環境，讓我能夠充份發揮，並且提供重要的指導與方向，讓我學習到在碩士班應有的處理問題態度。

接著是指導我的學長楊皓義，平時他提供了許多的觀點讓我學習，遇到困難時，也會廢寢忘食的指導我以度過難關。此外，也感謝張銘宏、黃柏蒼、謝維致學長們適時的幫助與討論。當然也感謝實驗室同學在枯燥的碩士生活增加不少樂趣。由於碩二有接智原科技的先進製程 SRAM 的計劃，在此也特別感謝莊景德教授的指導，及 MSCS 實驗室的林志宇同學和杜明賢學長和智原科技的同仁，讓我學到非常多有別於學界的業界資訊。

最後，感謝家人對我的鼓勵與支持，有了他們，才能讓我無後顧之憂的完成碩士學業。

Contents

Chapter 1 Introduction……………………….…………………………1

1.1 Background ……………………………………..……….........1

1.2 Motivation ……………………………………………...………..1

1.3 Thesis Organization…………………………………………2

Chapter 2 Overview of recent low-power SRAM……………………........... 3

2.1 Introduction…………………………………………………3

2.2 Power dissipation……………………………………...………3

2.2.1 Dynamic dissipation………………………………………………...3

2.2.2 Leakage dissipation…………………………………………………4

2.2.2.1 Junction leakage……………………………………………………5

2.2.2.2 Gate-induced drain leakage…………………………………………5

2.2.2.3 Gate direct tunneling leakage……………………………………5

2.2.2.4 Sub-threshold leakage…………………………………………7

2.2.3 Short circuit dissipation…………………………………………..8

2.2.4 Total power…………………………………………………9

2.3 Traditional 6T SRAM……………………………………… .10

2.3.1 6T bit-cell operation………………………………………………10

2.3.2 6T bit-cell stability……………………………………………

… . . 1 2

2.3.3 6T bit-cell in nano-scale process……………………………. …16

2.4 Recent low power SRAM technique………………………….....17

2.4.1 Single-end 8T SRAM bit-cell………………………………..17

2.4.2 Sub-threshold 10T SRAM bit-cell……………………………...19

2.4.3 Circuit technique in SRAM peripheral circuit…………………21

2.5 Summary……………………………………………………..27

Chapter 3 Timing Control Degradation and NBTI/PBTI Tolerant Design for Write-Replica Circuit in Nanoscale CMOS SRAM….......28

3.1 Background……………………………………………………..28

3.2 NBTI and PBTI…………………………...……………...……….29

3.3 Simulation environment…………………………..……...……..31

3.4 Impact on Write-replica timing control……………………….....33

3.4.1 Write delay of array and write-replica cell……………………...33

3.4.2 WWL pulse width……37
3.4.3 Write window……39
3.4.4 Write cycle time and write energy……41
3.5 NBTI/PBTI tolerant write-replica scheme……43
3.6 Summary……50

Chapter 4 High performance and low VDDmin of new 8T SRAM cell design……52
4.1 Introduction……52
4.2 Cell structure and basic operation of cell……53
4.2.1 Precharge/Standby mode……54
4.2.2 Read mode……54
4.2.3 Write mode……55
4.3 Cell Stability……57
4.4 Architecture of array……58
4.5 Interface circuit with negative voltage write scheme……62
4.5.1 Precharge/Standby mode……64
4.5.2 Read mode……65
4.5.3 Write mode……65
4.6 Simulation results……71
4.7 Summary……71

Chapter 5 Sub-threshold multi-port register file……73
5.1 Introduction……73
5.2 Overview of sub-threshold register file design and banked register file architecture……74
5.2.1 Sub-threshold register file design……74
5.2.2 Banked register file architecture……78
5.3 Architecture of low power multi-port register file……79
5.4 Design Concept……83
5.4.1 Dual Vt 8T Bit- Cell with High-Vt Passgate……83
5.4.2 Negative Voltage Write Scheme……86
5.4.3 Improved Read Buffer Footer, Controllable Pre-charge Scheme and Read Replica Circuit……90
5.4.4 Improved Output Latch……95
5.5 Design Implementation……96
5.6 Simulation Result……96
5.7 Summary……100

Chapter 6 **Conclusion……………………………………………………...101**

Bibliography **……………………………………………………………103**

Vita **…………………………………………………..……..112**

List of Figures

2.1 Circuit diagram of inverter……………………………………………………………………4
2.2 Leakage current in NMOS transistor……………………………………………………..5
2.3 Components of tunneling current……………………………………….........................7
2.4 Gate leakage current vs. gate oxide thickness……………………………………….......7
2.5 Gate leakage current vs. gate voltage………………………………………………………7
2.6 Traditional SRAM architecture (a) SRAM architecture; (b) Column configuration……..8
2.7 Circuit diagram of conventional 6T bit-cell…………………………………………..12
2.8 Read operation………………………………………………………………………………..12
2.9 Write operation…………………………………………………………………………….12
2.10 Definition of hold noise margin……………………………………………..………….13
2.11 Butterfly curve of hold noise margin…………………………………..…….……….14
2.12 Definition of read noise margin……………………………………………………….15
2.13 Butterfly curve of read and hold noise margin……………………………………….15
2.14 Setup for finding WTP………………………………………………………….………….16
2.15 Write margin of a SRAM cell, determined by WTP…………………………........…………16
2.16 A standard 8T bit-cell structure……………………………………………………...........18
2.17 (a) 10T SRAM bit-cell (b) Read mode……………………………………….………19
2.18 Structure of 10T bit-cell in [2.24]……………………………………….…..……………19
2.19 Another structure of 10T bit-cell which improves the leakage issue on read port..............20
2.20 Virtual GND replica circuit generator in [2.25]……………………………………………21
2.21 Schmitt trigger based Sub-threshold bit-cell……………………..…………………21
2.22 Lowering the WL voltage improves read noise margin, and the Gate…………………...21
2.23 By floating bit-cell supply voltage, write margin can increase……………………..…….22
2.24 Another layout style - Reaction diffusion bit-cell…………………………….…………..23
2.25 Several techniques to reduce power in stand-by mode……………………..……………..24
2.26 Hierarchical architecture can lower power consumption by reducing the local bit-line capacitance…………………………………………………………………….…………….24
2.27 An improved replica circuit generating the exact timing for word-line can reduce SRAM power...25
2.28 By using several replica cells in one replica column can against serious PVT variation which happens in replica column……………………………………………………….26
2.29 MTCMOS technique: improve performance and reduce power in low voltage operating……………………..…………………………………….……………………26
3.1 (a) Stress mode (b) Recovery mode…………………………………………………….27
3.2 Dynamic behavior in V_T……………………………………………………………………27
3.3 NBTI/PBTI induced V_T drifts vs. stressed time for 32nm poly-gate and high-k metal-gate

devices (V_T drifts based on AC RD model and calibrated with published data)....31
3.4 Critical paths of SRAM with Write-replica timing control..................................32
3.5 8T-SRAM cell schematic...34
3.6 Cell Write delay (Time-to-Write): (a) with NBTI in poly-gate CMOS; (b) with NBTI and PBTI in High-κ metal-gate CMOS...36
3.7 Increase of WWL pulse width due to degradation in logic control circuits: (a) with NBTI in poly-gate CMOS; (b) with NBTI and PBTI in High-κ metal-gate CMOS...............39
3.8 (a) Write Window of a "fresh" SRAM determined b rising edge of WD pulse and trailing edge of WWL pulse; (b) and (c) Write Window of a SRAM after stressing determined entirely by WWL pulse (in(b)) or entirely by WD pulse (in (c))..............................40
3.9 Increase of Write Window width: (a) with NBTI in poly-gate CMOS; (b) with NBTI and PBTI in high-k metal-gate CMOS..41
3.10 Increase of Write cycle time: (a) with NBTI in poly-gate CMOS; (b) with NBTI and PBTI in high-k metal-gate CMOS...43
3.11 Proposed circuits to mitigate NBTI/PBTI degradation: (a) Write-replica cell with power switch; (b) buffer with power switch...44
3.12 Proposed multi-bank SRAM architecture..45
3.13 Degradation mitigation by proposed NBTI/PBTI tolerant scheme (a) increase of WWL pulse width; (b) increase of Write Window width; (c) increase of Write cycle time.......48
3.14 Difference (Margin) between Write Window width and cell Write delay (Difference = Write Window width - cell Write delay) for proposed scheme across process corners. Positive difference (margin) indicates Write operation would be successful. (a) Case 1;(b) Case....46
4.1 Minimum-area comparison between 6T and 8T cells......................................52
4.2 Schematic of the new 8T cell...53
4.3 Precharge and standby mode..54
4.4 Read mode..55
4.4 Write mode (a)Write 0; (b)Write 1...57
4.5 TI (Texas Instruments)' 7T SRAM bit-cell..57
4.6 Read noise margin comparison...58
4.7 Monte Carlo simulation, times=1000, VDD=0.4V and sigma=3.............................58
4.8 Write margin...58
4.9 Worse noise margin in half-select write mode (a) Case 1; (b) Case 2; (c) Noise margin comparison...60
4.10 Array architecture (a) VVSS in row based; (b) VVSS in column base........................61
4.11 Power consumption comparison between row based architecture and column based architecture...62
4.12 Interface circuit architecture...63
4.13 Interface circuit – precharge/stand by mode..65

4.14 Interface circuit – read mode……………………………………………………….……….65
4.15 Interface circuit – write "1" mode…………………………………………….……….…….66
4.16 Negative voltage write scheme in TSMC (a) Waveform. (b) Detailed circuit................68
4.17 Negative voltage write scheme in IBM [4.4] (a) Waveform (b) Circuit diagram...............68
4.18 Interface circuit – write "0" mode (a) Phase I (b) Phase II (c) Waveform………………70
5.1 Conventional multi-port SRAM based register file bit-cell……………………….……..75
5.2 Another structure of RF bit-cell in sub-threshold region. [5.14-15] (a) Cell number on one bit-line is small (b) Two bit-cells share one bit-line…………………………………….76
5.3 A RF bit-cell with exceptional stability can be applied to special application and this RF bit-cell also can work in sub-threshold region……………………………….………..77
5.4 Mux-based selected logic is the robust design in sub-threshold region…………………...77
5.5 Proposed register file in VLIW DSP…………………………………………………...79
5.6 Architecture of the TMS320C64x family of DSPs. The C6x is an eight-issue traditional VLIW processor. (Courtesy Texas Instruments.)…………………………..………….80
5.7 Sub-decoder with power gating………………………………………….……………..81
5.8 Switch circuit with collision detecting logic…………………………………….…………..82
5.9 Read disturb in traditional 6T bit-cell……………………………………….………………83
5.10 Properties of dual Vt 8T bit-cell (a) Power consumption in Write half select mode; (b) Write half select noise margin of dual Vt 8T bit-cell compared with conv. 8T bit-cell….84
5.11 Monte Carlo simulation, sigma=3, VDD=0.5V, times=1000…………………………….85
5.12 Write margin of 8T cell with Hvt passgate and normal Vt passgate……………………..86
5.13 Write scheme comparison. (a) Boosting WL voltage (b) Negative bl voltage…………….87
5.14 Negative voltage generator with local BL sensing logic.(a) Circuit diagram; (b) Bit-line write voltage vs. VDD……………………………………………………………………..89
5.15 (a) Capacity of MOS capacitance degraded in sub-threshold region (b) Write scheme with local BL sensing logic reduces required area of MOS capacitance in deep sub-threshold region with PVT variation…………………………………..…………………………90
5.16 Read architecture and IREAD tracing replica circuit……………………………………..91
5.17 PMOS of read buffer footer in [5.9] generates leakage to increase access time in worst pattern…………………………………………………………………..………………93
5.18 Read buffer footer with PMOS increases array read access time in Fig. 5.17. (a) 0.18V – 1V (b) 16-256 cells/BL…………………………………………………...…………..94
5.19 Controllable pre-charge scheme reduces both power and access time in this read architecture ……………………………………………………………………………..94
5.20 Improved output latch can ensure data stability……………………………….….….…96
5.21 Layout photograph of the proposed sub-threshold multi-port register File…………….… 97
5.22 Access time significantly varies across process and temperature variation…………..….99
5.23 The power comparison between this work and conventional design…………….……...99

List of Tables

3.1 SRAM write frequency and target WWL selected frequrency……………………………33
3.2 Stressed probability distribution……………………………………...……………34
4.1 True table of interface circuit signal………………………………………….……64
4.2 Timing information in SS corner -40°C…………………………………...…..71
4.3 Timing information in SS corner 125°C………………………………...…….71
5.1 The simulation results………………………………………………….…………..98

Chapter 1 Introduction

1.1 Background

According to Moore's law, the number of transistors that can be placed inexpensively on an integrated circuit doubles approximately every 18 months. That also means processing speed and memory capacity improve at exponential rates as well. In modern CMOS technology, dynamic power is not the most significant in power consumption anymore. Due to the process technology, architecture and application, leakage power becomes crucial to total power consumption.

Embedded memory plays a key role in high performance and low power VLSI technology today. Memory module will occupy over 90% area of a high performance chip in recent 10 years. As a result, memory usually becomes a bottleneck of chip performance. In other words, embedded memory dominates the area, power and performance of chip. In addition, one of high performance embedded memory design is Static Random Access Memory (SRAM).

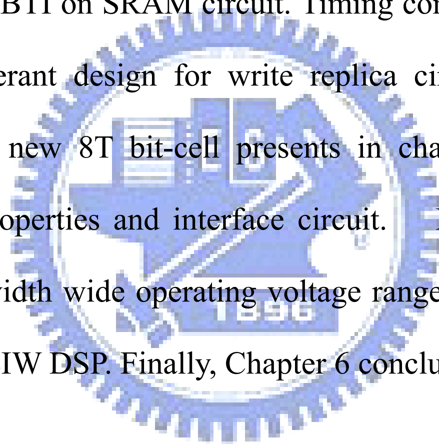
1.2 Motivation

Nowadays, in order to produce high performance and low power chip design in advanced process technology, more and more companies design custom SRAM, for particular application by self instead of making use of general SRAM compiler. On the other hand, multi-port SRAM-based register file design is also important. Due to the PVT variation in future process, the circuit level design of register file is not the same as the traditional register file. In addition, the traditional register file can't work

with wide-operating-voltage range or in ultra low voltage which is important for mobile device. Consequently, a custom SRAM/register file demanded by each product is necessary.

1.3 Thesis Organization

The rest of this thesis is organized as following. Chapter 2 presents the traditional SRAM architecture and low power SRAM design in recent years. This chapter includes overview of power dissipation, properties of traditional 6T SRAM, low power SRAM bit-cell, and circuit techniques in recent years. Chapter 3 shows the influence of NBTI and PBTI on SRAM circuit. Timing control degradation in SRAM and an NBTI/PBTI tolerant design for write replica circuit in nanoscale CMOS SRAM is presented. A new 8T bit-cell presents in chapter 4, including detailed analysis of new 8T's properties and interface circuit. In chapter 5, a micro-watt multi-port register file width wide operating voltage range is proposed which can be applied to low power VLIW DSP. Finally, Chapter 6 concludes all this work.



Chapter 2 Overview of Recent Low-power SRAM Design

2.1 Introduction

This chapter represents a study of power dissipation of CMOS circuit and recent low power SRAM design. Power dissipation, including dynamic dissipation, leakage dissipation, and short circuit dissipation, is presented in Section 2.2. Other sections show the low-power SRAM design in recent years. To begin with, operation, importance, and architecture of introduction of traditional 6T SRAM is presented. In addition, conventional 8T bit-cell, other new 10T bit-cell and register file bit-cell for the low power concern and other interesting design also presented.

2.2 Power dissipation

2.2.1 Dynamic dissipation

Fig. 2.1 shows a CMOS inverter, the average dynamic power dissipation can be obtained by summarizing the average dynamic power of NMOS and PMOS. Assuming that the input Vin is a square wave having a period *T* and that the rise and fall times of the input are much less than the repetition period, the dynamic power is given by

$$P_D = \frac{1}{T}\int_0^{T/2} i_N(t)V_{out}dt + \frac{1}{T}\int_{T/2}^{T} i_P(t)(V_{DD} - V_{out})dt \tag{2.1}$$

Where C_L is the load capacitance, f is the operating frequency, and the i_N is relative to load capacitance.

$$P_D = fC_L V^2{}_{DD} \tag{2.2}$$

Moreover, power dissipation is in data dependence form, i.e. Power dissipation

depends on the switching probability, thus, dynamic power can be expressed as

$$P_D = \alpha \cdot f C_L V^2{}_{DD} \qquad (2.3)$$

From (2.3), dynamic power dissipation of a logic gate is proportional to switching frequency, capacitance of load capacitance, square of the supply voltage, and operation frequency.

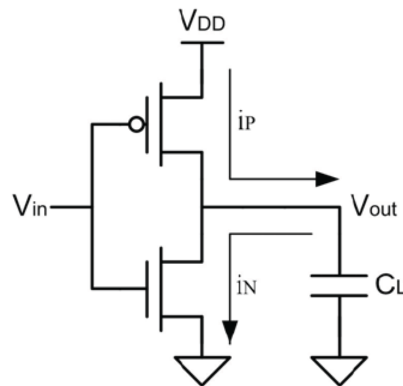


Fig.2.1 Circuit diagram of inverter

2.2.2 Leakage dissipation

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}) are composed of leakage current in a CMOS transistor, as illustrated in Figure 2.2 [2.1-3]. Each source of leakage current will be described in the followings.

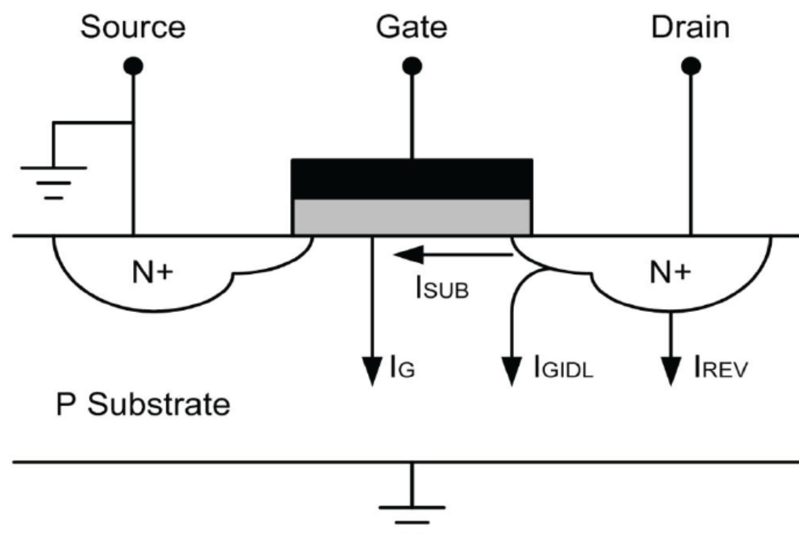


Fig. 2.2: Leakage current in NMOS transistor.

2.2.2.1 Junction leakage

While the transistor off, junction leakage occurs from the source or drain of transistor to the substrate through the reverse-biased diodes, I_{REV}, as illustrated in Fig. 2.2. A reverse-biased pn junction leakage includes: minority carrier diffusion/drift near the edge of the depletion region; the other is due to electron-hole pair generation in the depletion region of the reverse-biased junction. The area of the drain diffusion and the leakage current density influence junction leakage current, which is determined by the doping concentration. Junction leakage components from both the source-drain diodes and the well diodes are generally negligible with respect to the other three leakage components.

2.2.2.2 Gate-induced drain leakage

Gate-induced drain leakage (GIDL), which is I_{GIDL} in Fig. 2.2, arises in the high electric field under the gate/drain overlap region. GIDL occurs at large V_{DB} and generates carriers into the substrate and drain from surface traps or band-to-band tunneling. Thinner oxide, higher supply voltage, and lightly doped drain structures increase GIDL current.

2.2.2.3 Gate direct tunneling leakage

Gate direct tunneling current is due to the tunneling of an electron/hole from the bulk silicon through the gate oxide potential barrier into the gate [2.4-5]. Reduction of gate oxide thickness results in the increase in the field across the oxide. The high electric field coupled with low oxide thickness results in tunneling of electrons from

substrate to gate and also from gate to substrate through the gate oxide, resulting in the gate leakage. In nanometer-scale CMOS technologies, where ultra-thin gate oxide thickness takes place for effective gate control, gate leakage becomes appreciable and dominates the total leakage dissipation [2.6].

Figure 2.3 shows the components of tunneling current in a scaled NMOS transistor. They are classified in to three categories:

1. Edge direct tunneling components between the gate and the source-drain extension overlap region (I_{gdo} and I_{gso}).
2. Gate-to-channel current (I_{gc}), part of which goes to the source (I_{gds}), and the rest goes to the drain (I_{gcd}).
3. Gate-to-substrate leakage current (I_{gb}).

Therefore, the gate leakage (I_G) can be divided into three major components:

1. Gate-to-source ($I_{gs}= I_{gso}+I_{gcs}$).
2. Gate-to-drain ($I_{gd}= I_{gdo}+ I_{gcd}$).
3. Gate-to-substrate (I_{gb}).

The magnitude of the gate leakage current increases exponentially with the gate oxide thickness T_{OX} and the gate-to-source voltage V_{GS}, as shown in Fig. 2.4 and Fig. 2.5, respectively [2.7].

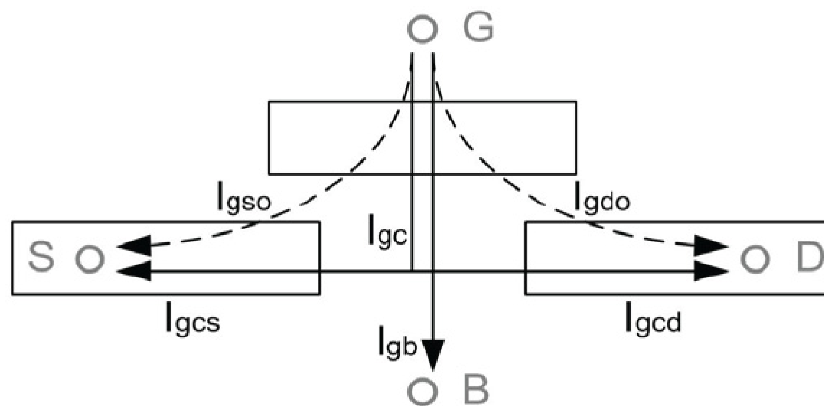


Fig.2.3: Components of tunneling current. [2.6]

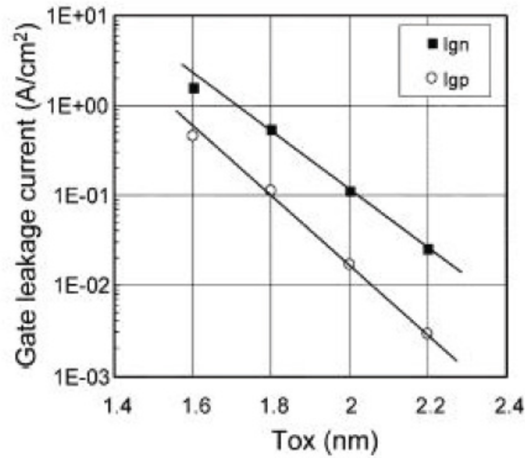


Figure 2.4: Gate leakage current vs. gate oxide thickness.

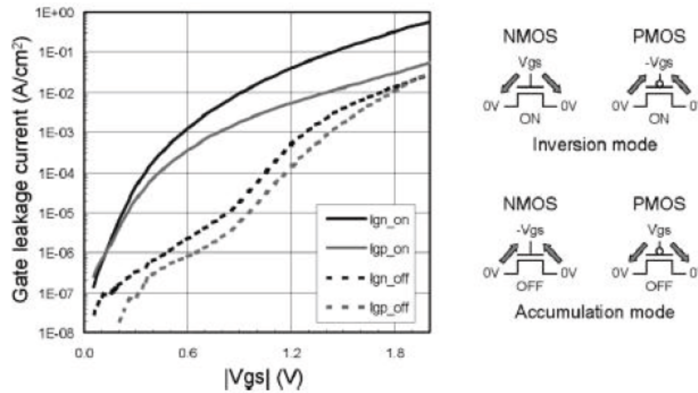


Figure 2.5: Gate leakage current vs. gate voltage.

2.2.2.4 Sub-threshold leakage

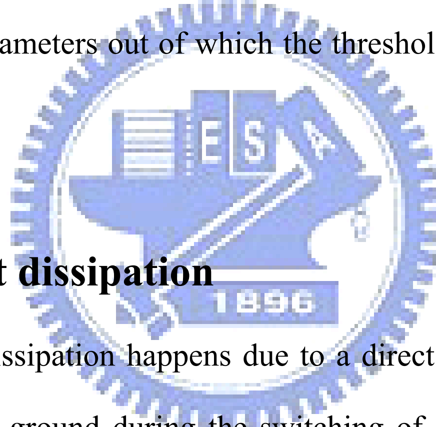
Sub-threshold or weak inversion conduction current between source and drain of a MOS transistor occurs when gate voltage is below the threshold voltage level. Unlike the strong inversion region in which the drift current dominates, the sub-threshold conduction is due to the diffusion current of the minority carriers in the channel for a MOS device. For instance, in an inverter with a low input voltage and high output voltage, for the NMOS transistor, even V_{GS} is 0V, there is still a current flowing in the channel of the off NMOS transistor due to the V_{DD} potential of the V_{DS}.

Sub-threshold leakage current (I_{SUB}) becomes apparent as CMOS technologies

enter the submicron era [2.8]. I_{SUB} can be expressed based on the following:

$$I_{SUB} = \frac{W}{L}\mu v_{th}^{2} C_{sth} e^{\frac{V_{GS}-V_{T}+\eta V_{DS}}{nV_{th}}} (1-e^{\frac{-V_{DS}}{V_{th}}}), n = 1+\frac{C_{sth}}{C_{ox}} \quad (2.4)$$

where W and L denote the transistor width and length, μ denotes the carrier mobility, $V_{th} = kT/q$ denotes the thermal voltage at temperature T, $C_{sth} = C_{dep} + C_{it}$ denotes the summation of the depletion region capacitance and the interface trap capacitance both per unit area of the MOS gate, and η is the drain-induced barrier lowering (DIBL) coefficient. n is the slope shape factor and is calculated as: Where C_{ox} denotes the gate input capacitance per unit area of the MOS gate. Thus, the magnitude of the sub-threshold leakage current is a function of the temperature, supply voltage, device size, and the process parameters out of which the threshold voltage plays a dominant role.



2.2.3 Short circuit dissipation

Short circuit power dissipation happens due to a direct path current flowing from the power supply to the ground during the switching of a static CMOS gate. Short circuit dissipation can be expressed as:

$$P_{SC} = I_{mean} V_{DD} \quad (2.5)$$

where I_{mean} is the mean value of the short circuit current, and I_{mean} is modeled as [2.9]:

$$P_{SC} = \frac{1}{12}\frac{\beta}{V_{DD}}(V_{DD}-2V_{T})^{3}\frac{\tau}{T} \quad (2.6)$$

where β is the gain factor of a transistor, τ is the input rise/fall time. Although this is a simplified model, it reveals the fact that short circuit dissipation is affected by supply voltage, threshold voltage, rise/fall time, and operation frequency. As a result, by lowering supply voltage, increasing threshold voltage, and minimizing input rise/fall time, short-circuit power can be decreased.

2.2.4 Total power

By summarizing power above, total power can be described as following.

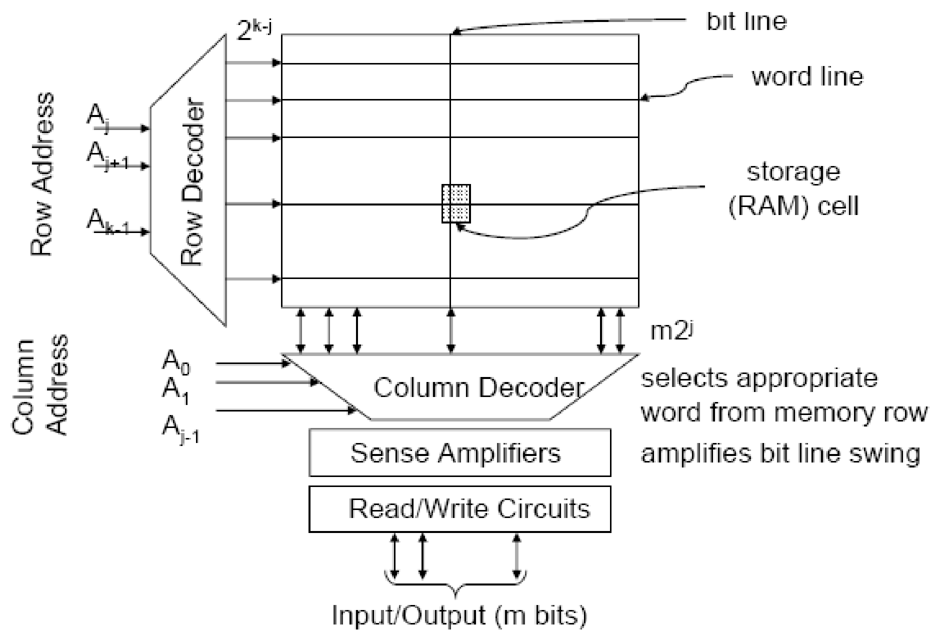
$$P_{Total} = P_D + P_{leakage} + P_{SC} = \alpha f \cdot C_L V_{DD}^2 + I_{Leak} V_{DD} + I_{SC} V_{DD} \tag{2.7}$$

From (2.7), supply voltage dominates total power consumption. Lowering the supply is the most effective method to reduce power consumption.

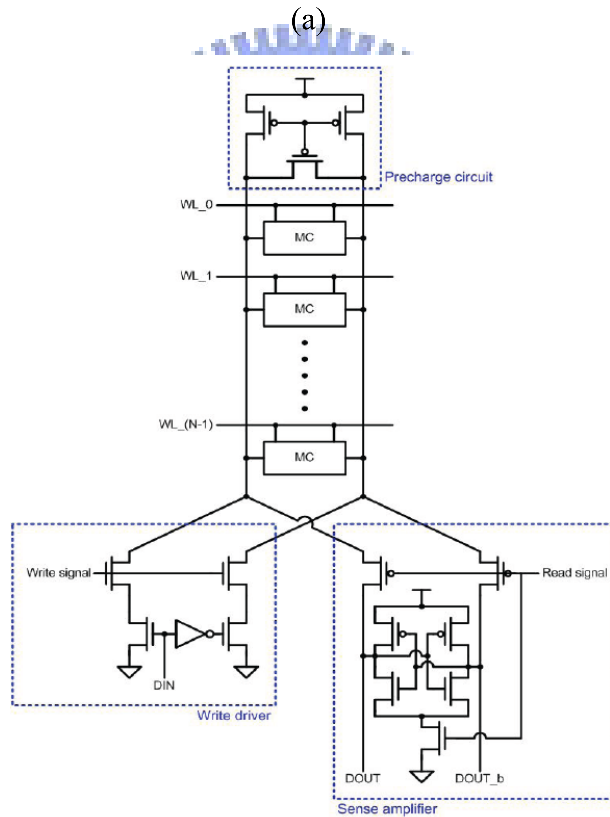
2.3 Traditional 6T SRAM

The issue of SRAM design is more and more popular in recent years. Why? SRAMs comprise a significant percentage of the total area (90%) and total power for many digital chips [2.10]. SRAM leakage can dominate total chip leakage, and switching highly capacitive bit-line and word-line is costly in terms of energy. Besides, the process variation is more and more badly. Process variation will make the SRAM bit-cell or peripheral circuit unstable or damage. If the SRAM doesn't work, we can say that the chip can't work, too.

Fig. 2.6 (a) shows the structure of SRAM. We can see the SRAM cell, decoder, sensing amplifier, select circuit and buffer. Fig. 2.6 (b) shows SRAM column configuration. The precharge circuit is used to pull the bit-line to high voltage level and equalize bit-line pair before operation. Each column must also contain write drivers and read sensing circuits. Write drivers pull the bit-line or its complement low during write operation. The sense amplifier shown is a commonly used latch type sense amplifier. When the sense amplifier is activated, the cross-coupled inverter pair pulls one output low and the other high through regenerative feedback.



(a)



(b)

Fig. 2.6: Traditional SRAM architecture [2.11] (a) SRAM architecture; (b) Column configuration.

2.3.1 6T bit-cell operation

Fig. 2.7 shows the schematic of the 6T SRAM cell commonly used in practice. The cell uses a single word-line and both true and complementary bit-lines. The cell contains a pair of cross-coupled inverters for data storage and an access transistor for each bit-line.

For read operation, BL and BLb are first precharged to high level. The WL is then turned on, and one of the bit-lines will be pulled down by the cell. For example, in Figure 2.8, Q=0 and Q_b=1, BL will therefore be pulled down by transistors MAL-MNL, while BLb stays high. A differential signal is generated on the bit-line pair, and the sense amplifier at the read output end will detect this small signal and transforms it into full swing voltage. For write operation, one bit-line is driven low and the other high. The word-line is then turned on, and data on bit-lines will transfer to the cell stored node. For example, in Figure 2.9, Q=0, Qb=1, BL=1, and BL_b=0, Qb will be pulled to low, and Q will rises to high.

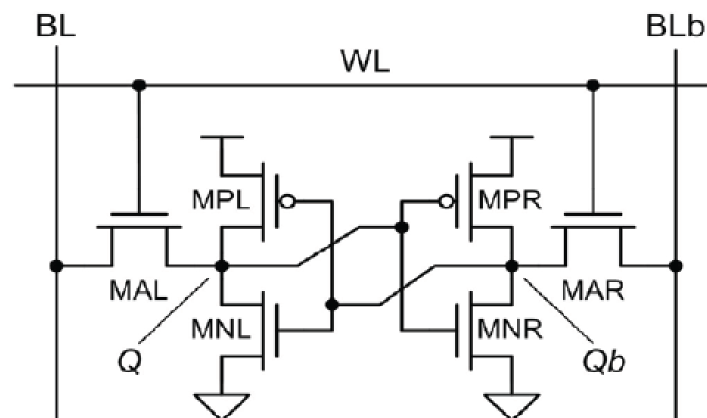


Fig. 2.7: Circuit diagram of conventional 6T bit-cell.

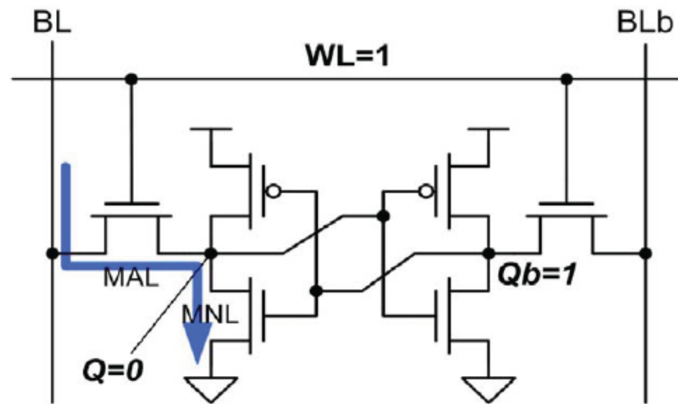


Fig. 2.8: Read operation.

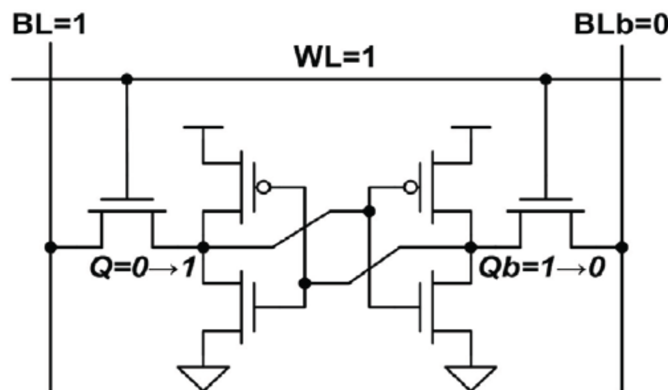


Fig. 2.9: Write operation.

2.3.2 6T bit-cell stability

When the bit-cell is holding data, the WL is low so that NMOS access transistors (MAL and MAR) are off. The cross-coupled inverters must maintain bi-stable operating points in order to properly hold data. The best measure of the ability of the cross-coupled inverters to maintain their state is the static noise margin (SNM) [2.12]. The Hold SNM is defined as the maximum value of DC voltage noise that can be tolerated by the SRAM cell without changing the stored bit when the access transistors are off. Figure 2.10 shows the standard setup for modeling Hold SNM. DC noise sources V_N are introduced at each of the internal nodes in the bit-cell. Cell stability changes as V_N increases. Figure 2.11 [2.13], known as the butterfly curve, is

the most common way of representing the SNM graphically. The butterfly curve plots the voltage transfer characteristic (VTC) of Inverter R and the inverse VTC of Inverter L. Inverter R and Inverter L are shown in Figure 2.42. The SNM is defined as the length of the side of the largest square that can be embedded inside the lobes of the butterfly curve. When the value of V_N increases, the VTCs move horizontally and/or vertically. When the value of V_N is equal to the value of SNM, the VTCs meet at only two points. Further noise flips the cell content.

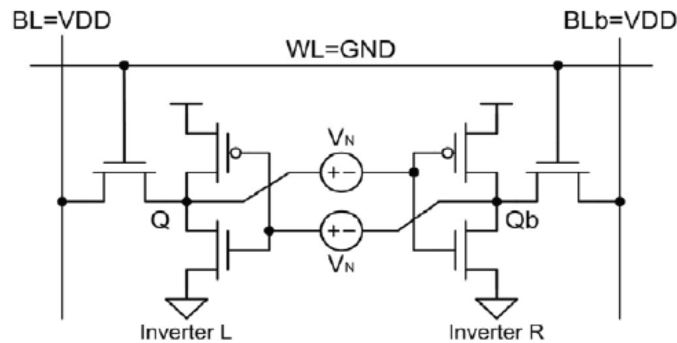


Fig. 2.10: Definition of hold noise margin [2.12].

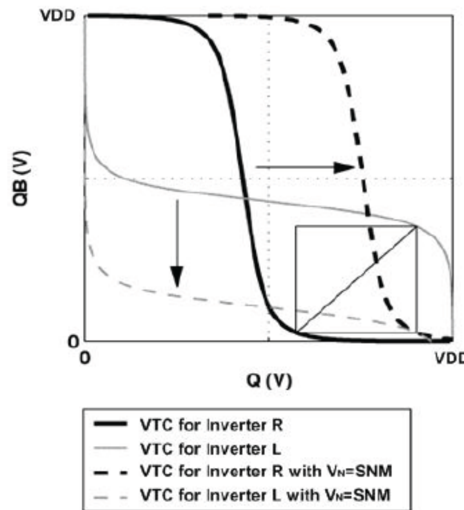


Fig. 2.11: Butterfly curve of hold noise margin[2.13].

The most common method to measure read stability is the Read SNM. SNM is

defined in the previous subsection, but the setup for Read SNM is different from Hold SNM. Figure 2.12 shows the standard setup for modeling Read SNM. WL is on for read access; BL and BLb are set to V_{DD} to indicate the initial value of bitlines are precharged to high.

In a conventional 6T cell, Read SNM is worse than Hold SNM. During read, the cell begins with the WL being turned on, with the bitlines initially high. This causes the low node within the cell to rise due to the voltage dividing effect across the access transistors and the pull down transistors. If this node voltage becomes close to the threshold of the pull down devices, process variations combined with noise coupling may flip the state of the cell. Fig. 2.13 [2.13] shows example of butterfly curves during hold and read, revealing the degradation in SNM during read.

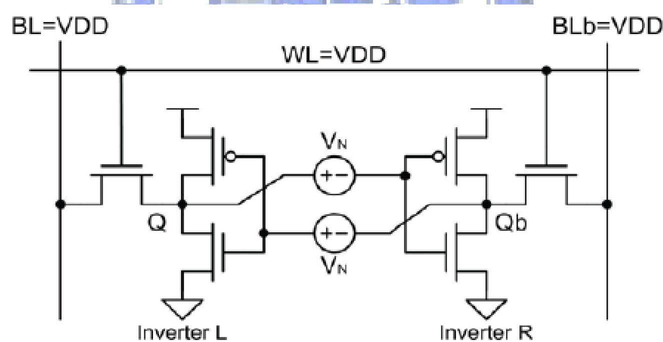


Fig. 2.12: Definition of read noise margin.

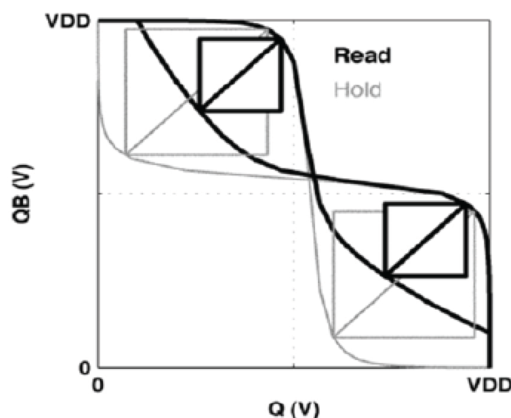


Fig. 2.13: Butterfly curve of read and hold noise margin [2.13].

Although there are many definition of write margin, a common way for characterizing write ability write margin (WM) or write trip point (WTP) [2.14-15, 17]. WTP defines the maximum voltage on the bit-line needed to flip the cell content. Figure 2.14 shows the conceptual setup to measure WTP of 6T SRAM cell. Figure 2.15 shows a corresponding example of finding WTP [2.16]. As the bit-line voltage is lowered to a certain level, the cell content is flipped, indicating a successful write. Larger WTP means smaller voltage must be lowered below V_{DD} for successful write, indicating it is easier to write into the cell. If the WTP becomes negative, it means that it is not possible to write into the cell. To sum up, a higher WTP represents better write ability.

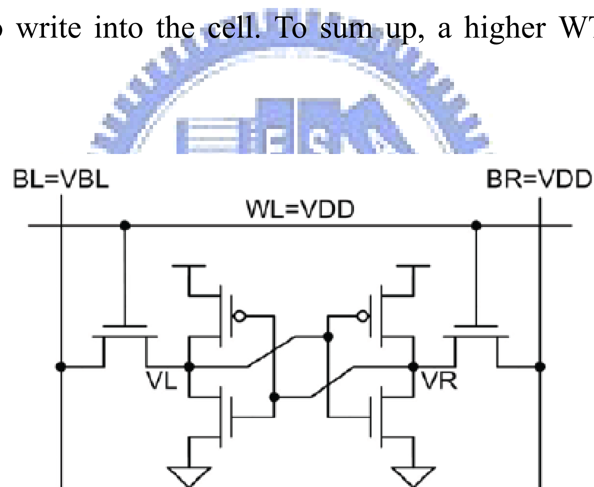


Fig. 2.14: Setup for finding WTP

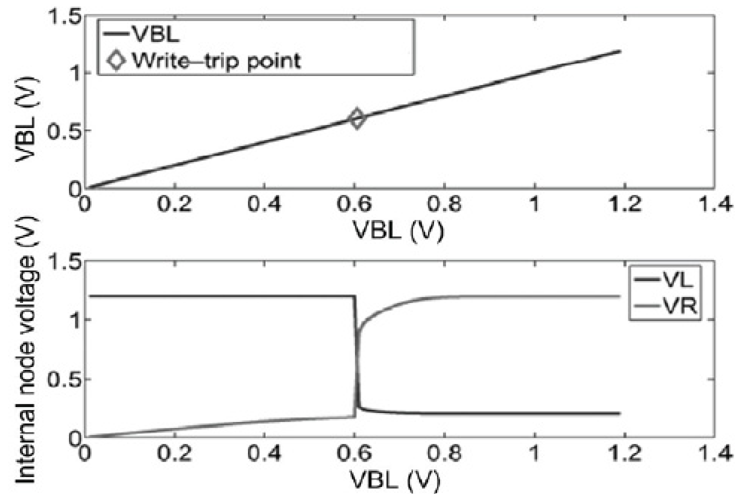


Fig. 2.15: Write margin of a SRAM cell, determined by WTP [2.16].

2.3.3 6T bit-cell in nano-scale process

In .35, .18, .13um process, the 6T SRAM is main structure of embedded memory. Due to several disadvantages [2.18], traditional 6T bit-cell is difficult to scale down in future process technology.

First, the read and half-select disturb. Due to the structure of 6T bit-cell, both storage nodes of a 6T SRAM cell connect to the BL pairs after its WL turns on during read. Since the access pass-transistor NMOS and the pull-down NMOS form a voltage divider, the cell “0” storage node rises during read. If the disturb voltage is larger than the trip voltage of another inverter in the same bit-cell, the data lose. Besides, in an interleaving SRAM structure, while WL selecting during a read or write operation, the half-selected cells in the same row are also do a read operation. This situation makes the half-select cell face the same issue as the selected cell.

The second is the conflicting read/write requirements. For addressing read disturb, it is necessary to design a strong pull-down NMOS and a weak access pass-transistor.

The ratio is called beta ratio. However, in order to improve write ability, it is desirable to have a strong access pass-transistor NMOS and weak load-transistor PMOS, as gamma ratio. Large PVT variation and local Vt mismatch in future process make 6T bit-cell disturb more serious especially in future process technology. Traditional 6T bit-cell is highly susceptible to PVT variation, as a result, large beta ratio and gamma ratio ensure the cell stability but increase cell area and power a lot.

Finale is the high VDDmin: The SNM of 6T bit-cell becomes worse as supply voltage scaled down. Sizing issue in circuit design is only available in high voltage supply, so 6T bit-cell's VDDmin is limited to high value (e.g. > 0.8 V at 65 nm).

2.4 Recent low power SRAM technique

2.4.1 Single-end 8T SRAM bit-cell

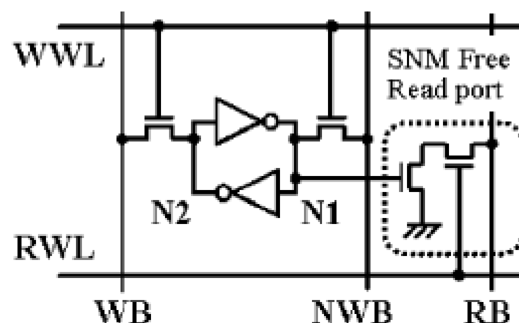
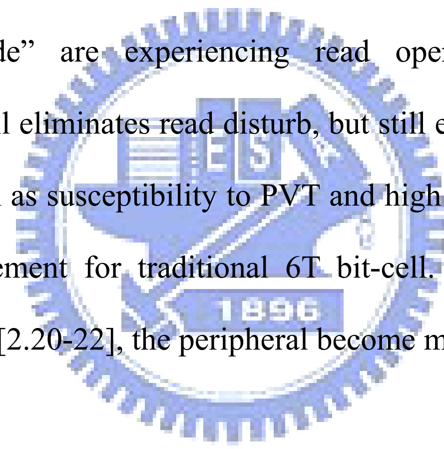


Fig. 2.16: A standard 8T bit-cell structure [2.19].

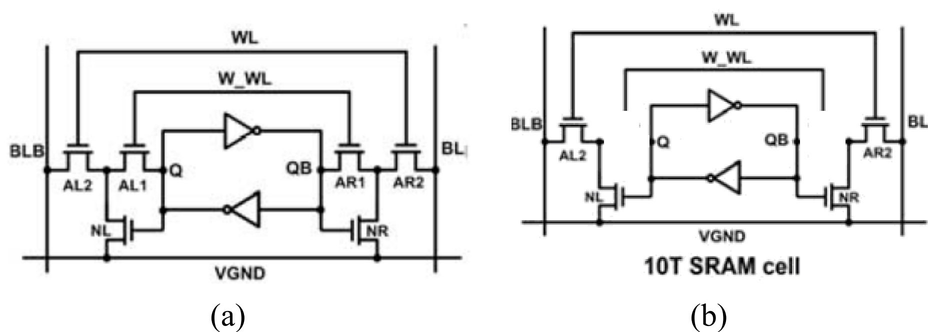
In order to address read disturb in traditional 6T bit-cell, 8T bit-cell can provide a solution. The main difference is that there are two extra NMOS. The two extra NMOS construct a read stack which connecting to a read bit-line. That means the RB won't influence the stored data, N1. As a result, 8T bit-cell eliminates read and half-read disturb issue. 8T Bit-cell has a good read margin which is almost the same as hold margin, and the small beta ratio which makes the cell area down.

However, when one WWL is pulled up, all of the access transistors of cells in the same row will turn on, too. The bit-line (WB) and bit-line-bar (NWB) of unselected column will also be precharged to high level, so stored node will influenced by BL. The fluctuation of N1 node voltage affects the read operation. When N1 is zero, the NMOS should be off so that the RB-line can maintain the high voltage. However, the read disturb voltage will turn NMOS on weakly, the leakage lets the RB-line voltage go down, which is called "misreading."[2.19]

Similar to above description, in the bit-interleaving architecture, 8T bit-cell in "half-select write mode" are experiencing read operation of a 6T bit-cell. Consequently, 8T bit-cell eliminates read disturb, but still exists half-write disturb. As a result, drawbacks such as susceptibility to PVT and high VDDmin make 8T bit-cell can't become a replacement for traditional 6T bit-cell. Although there are some solutions for 8T bit-cell [2.20-22], the peripheral become more complicated.



2.4.2 Sub-threshold 10T SRAM bit-cell



(a) (b)

Fig. 2.17: (a) 10T SRAM bit-cell [2.23] (b) Read mode

The difference between 10T SRAM bit-cell and 6T SRAM bit-cell is that there are extra M_{AL2}, M_{AR2}, M_{NL} and M_{NR} [2.23]. The WL signal is from a row decoder, but the

W_WL is from a column decoder. Fig. 2.17 (b) shows the read mode. Read disturb is eliminated. During Write, only both pass-gates are open in selected cell, the Write-Half-Select-Disturb for other half-selected bit-cells are eliminated, too.

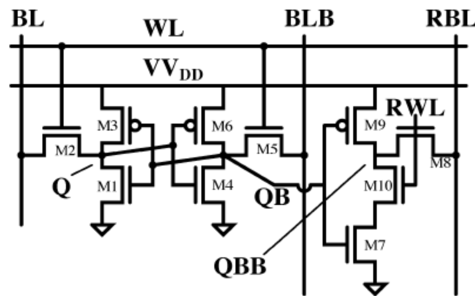


Fig. 2.18: Structure of 10T bit-cell in [2.24]

In Fig. 2.18, another structure of 10T bit-cell is presented in [2.24]. This bit-cell uses 4 transistors for read to improve read issue. The method tries to reduce the leakage power, but there is still a leakage issue (data dependence) and the bit-line only connects less than 256 bit-cells instead of 1024 bit-cells. The authors of [2.25] proposed a new structure to improve leakage issue. There are three transistors which are controlled by RWL, as shown in Fig. 2.19. The key idea is let the leakage always from supply source. Besides, the passgate of bit-cell uses RSCE (reverse short channel effect) to increase drive ability. Also, it uses VGND replica circuit to make sure the sensing amplifier work, as shown in Fig. 2.20.

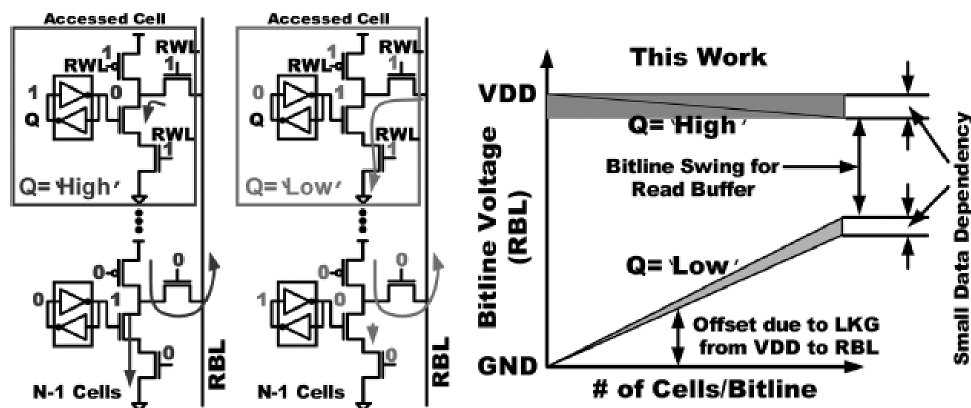


Fig. 2.19: Another structure of 10T bit-cell which improves the leakage issue on read

port. [2.25]

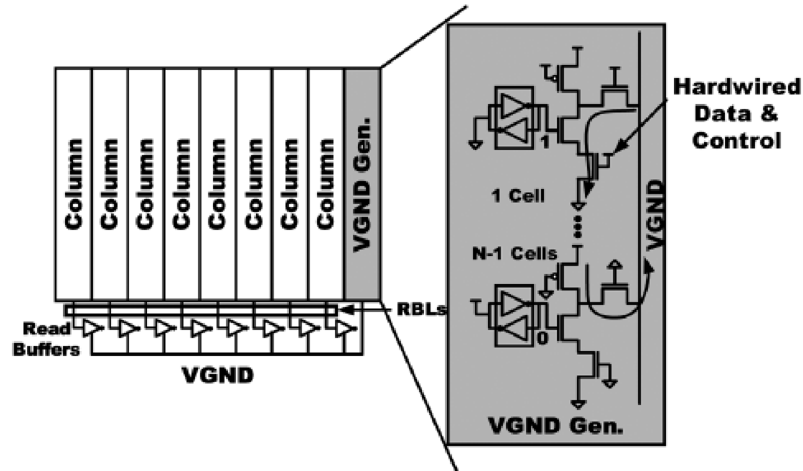


Fig. 2.20: Virtual GND replica circuit generator in [2.25]

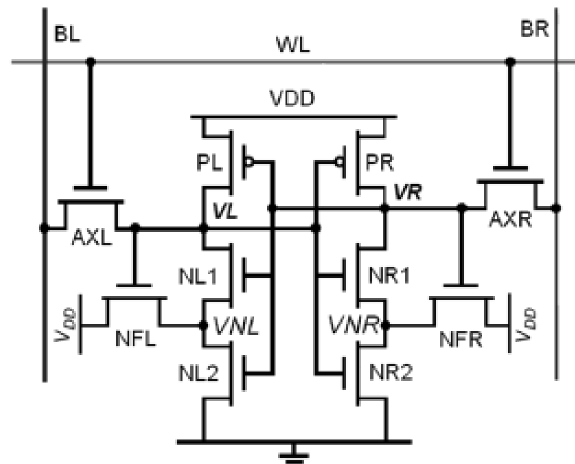


Fig. 2.21: Schmitt trigger based Sub-threshold bit-cell [2.35]

Besides, [2.35] designs a Schmitt trigger based sub-threshold SRAM cell, Fig. 2.21. The design is quite good no matter in stability, structure or area. The drawbacks of the cell are low speed access time and require fine process technology. All of the measurements in this paper are from 1V to 0.4V, although paper announced that the cell can work in 0.16V. However, there is no any information about the cell has ability against serious process variation.

2.4.3 Circuit technique in SRAM peripheral circuit

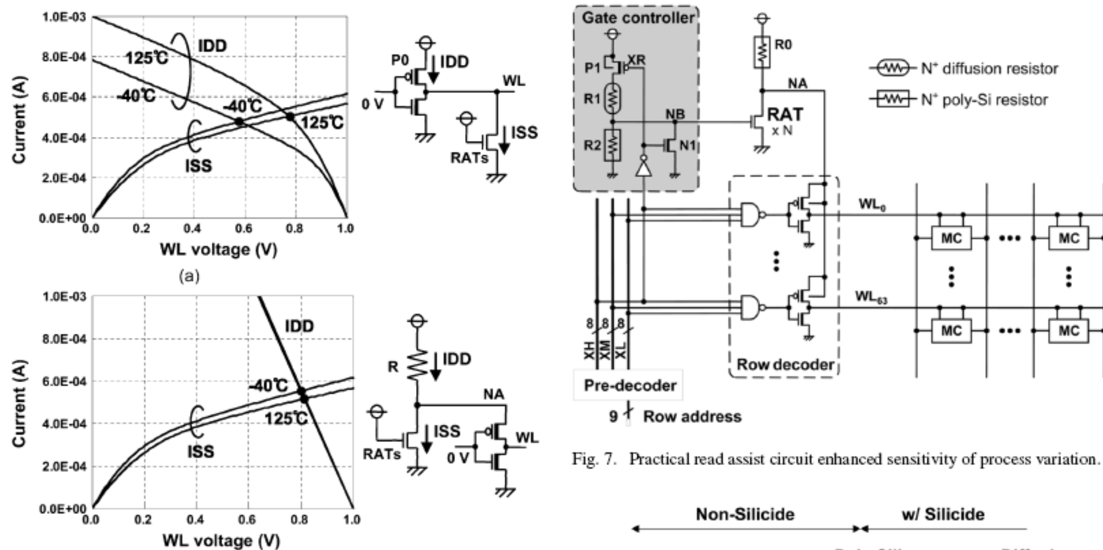


Fig. 2.22: Lowering the WL voltage improves read noise margin, and the Gate controller improves PVT variation impact. [2.26]

In [2.26], this design is in 45-nm technology and it use some assist circuit to improve read and write ability of conventional 6T cell, Fig. 2.22. Its key point is to decrease the voltage of word-line to improve the read margin. Another issue is to increase immunity against temperature. Fig. 2.22 shows the assist circuit. The design uses resistance instead of PMOS to compose the circuit because the temperature has larger influence on PMOS than on resistance. There are R1 and R2 in gray block. The key point is to use different voltage NB by R1 and R2, and the feedback loop to adjust the WL voltage.

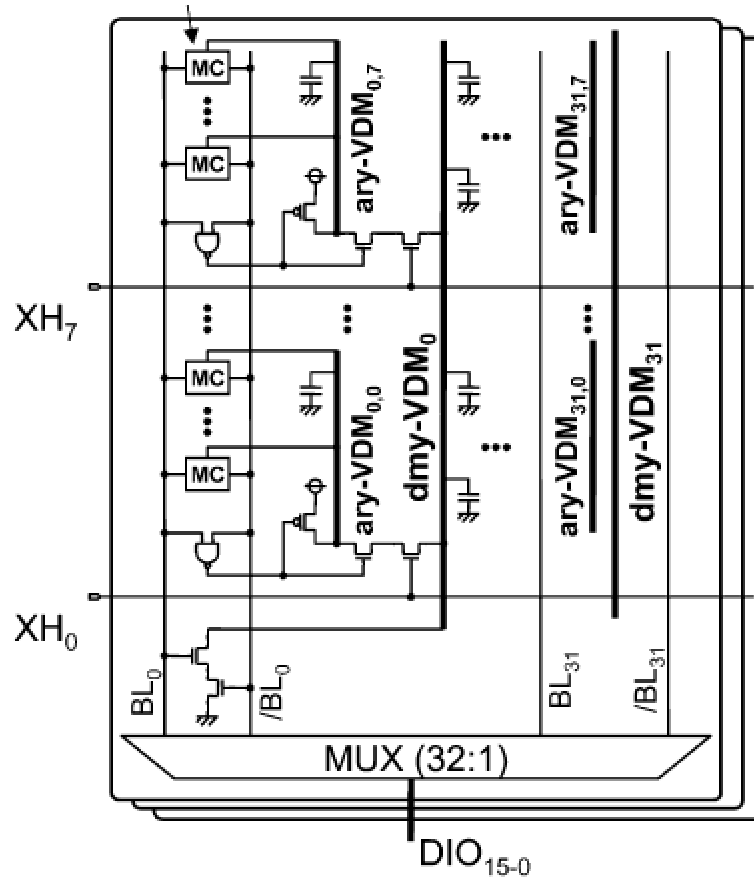


Fig. 2.23: By floating bit-cell supply voltage, write margin can increase. [2.26]

A good way to improve WM is to decrease the supply voltage of cell. The authors in [2.26] use floating technique to achieve this objective. They use 8 cells as a group because it is the optimization results by testing, Fig. 2.23. These assist circuits help the 6T cell work perfectly under 45-nm technology, but the drawback is to increase the complexity of peripheral circuit. However, the 6T cell has the smallest area compared to other cell structure and the lowest power consumption.

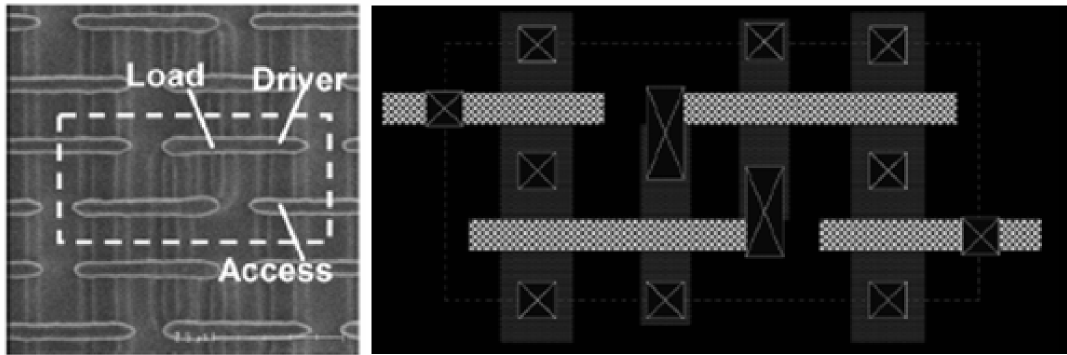


Fig. 2.24: Another layout style - Reaction diffusion bit-cell. [2.28]

RD cell [2.27-28] (Rectangular Diffusion Cell, Fig. 2.23) may be a method to solve variation in future process. It means that the beta ratio equals 1. Although beta ratio should be large enough to keep the read margin, sometimes there are some different lengths or widths occur in this layout of cell due to the process variation. For example, the two sides of passgate don't have the same width while there is a process offset. In future process, this issue becomes more serious and worse. Because the length and width is smaller than before, the contact area can't scale down with that. It needs to change the contact layout style in order to prevent PMOS and passgate's diffusion damage from contact due to process variation.

The RD cell has another advantage: the electric beta ratio in RD cell is better than that of conventional cell. Driver ability of passgate transistor to pull-down transistor is defined as electric beta ratio. In low voltage operation, due to reverse narrow channel effect, the electric beta ratio decreases. It means the conventional cell have some problem in read stability. But RD cell will not happen. Also, using this style, the area is smaller than conventional cell because of its large beta ratio.

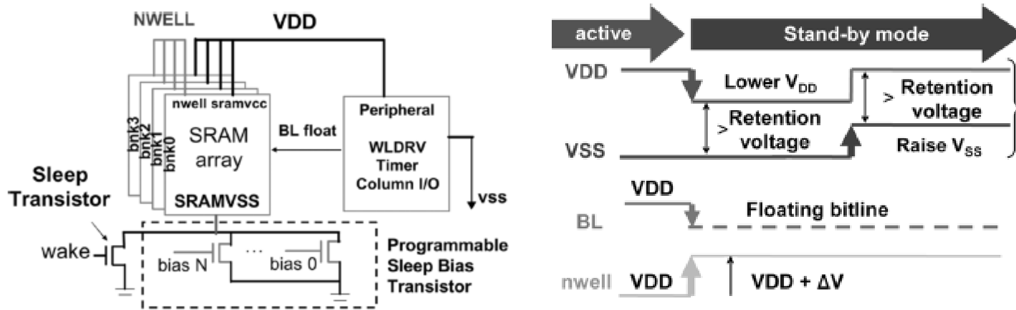


Fig.2.25: Several techniques to reduce power in stand-by mode [2.29]

In Fig. 2.25 [2.29], authors use “sleep transistor” to reduce the power. If the cell bank has no write or read, the cell bank will be operated in sleep mode. Lowering the supply voltage, power gating, eliminating the bit-line precharge, and adjusting body bias reduce the leakage consumption.

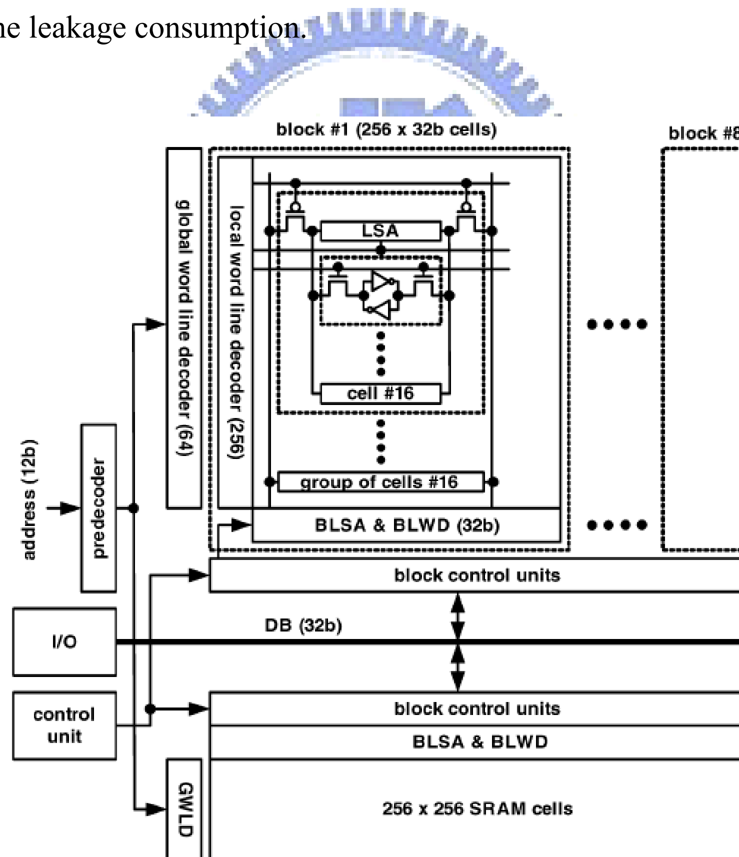


Fig. 2.26: Hierarchical architecture can lower power consumption by reducing the local bit-line capacitance. [2.30]

In [2.30], it uses hierarchical architecture to design the SRAM. If the SRAM’s

capacity is large, we often use one bit-line to connect 1024 or 512 bit cells. As a result, the capacitance on bit-line is very huge. Only one cell can do read or write operation in one bit-line. It is necessary to pull down or pull up the voltage of bit-line in every read or write cycle. If the bit-line design uses hierarchical structure, which only connects 16 cells, can reduce the power significantly, Fig. 2.26. In addition to hierarchical, it also uses word-line pulse to reduce the voltage difference. This design can lower the power effectively, but it needs to use many local amplifier and buffer to keep the correct the signal.

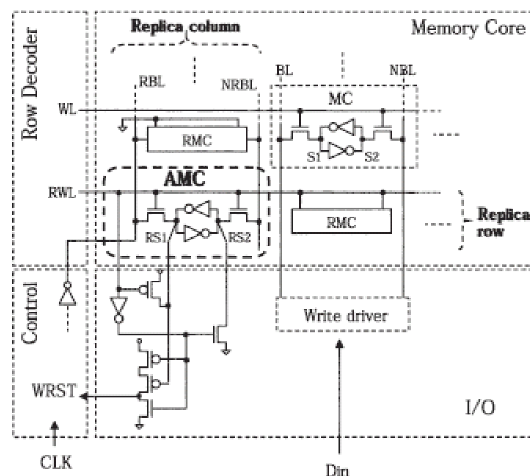


Fig. 2.27: An improved replica circuit generating the exact timing for word-line can reduce SRAM power [2.31]

In [2.31], this paper talks the conventional replica circuit and the circuits can work fail against variation. Because of variation, some read/write operation in cell is faster or slower than normal cell. An error may happen while there are fast replica cells and slow array cells in one SRAM. It use asymmetrical cell to make sure the sufficient operation time for array cell, Fig. 2.27.

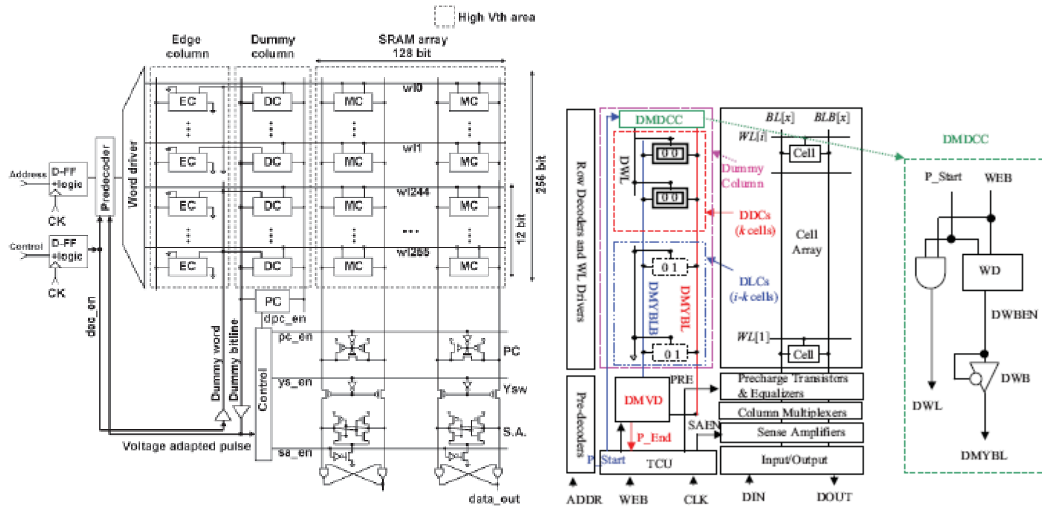


Fig. 2.28: By using several replica cells in one replica column can against serious PVT variation which happens in replica column. [2.32-33]

In [2.32-33], the authors use many replica cells to do the operation together and found that it is more stable to against temperature and process variation. Because sometimes the array has good yield and there is a serious variation in one of replica cells, the SRAM still can work correctly.

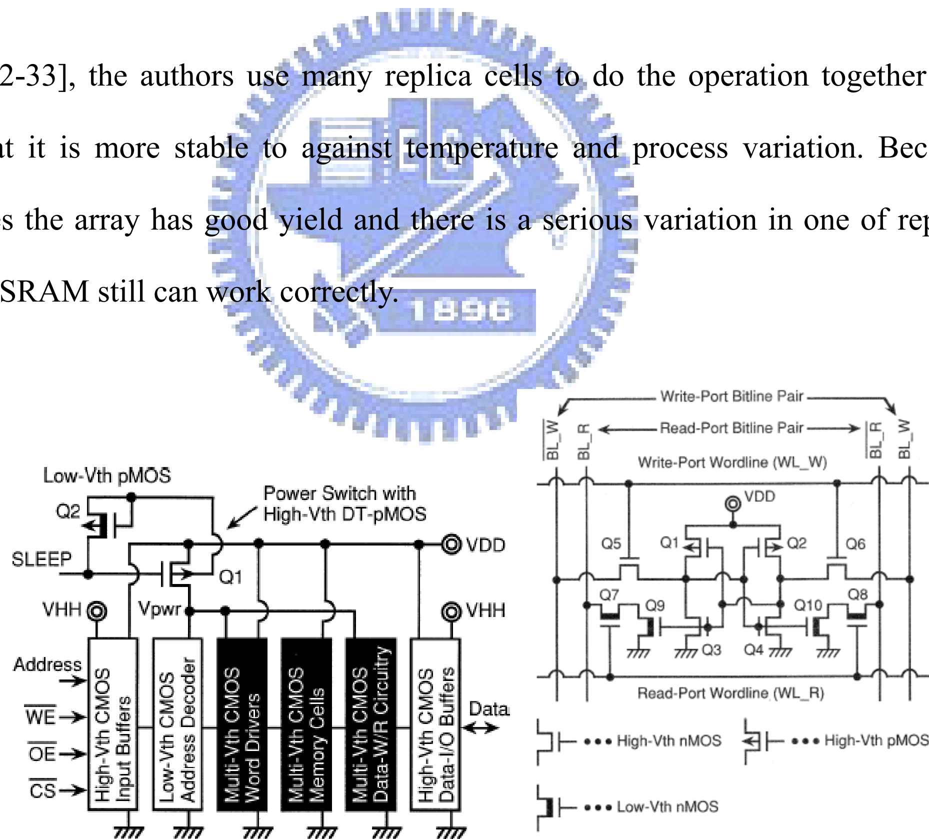


Fig. 2.29: MTCMOS technique: improve performance and reduce power in low voltage operating. [2.34]

In [2.34], MTCMOS technique is applied to low-power and high speed SRAM. Use

low Vt logic in critical path to speed up and high Vt logic in other block to reduce leakage power [2.34]. Besides, it also use multi-voltage to design the SRAM, some blocks need high speed by using high voltage and some blocks use low voltage to reduce power consumption, as shown in Fig. 2.29.

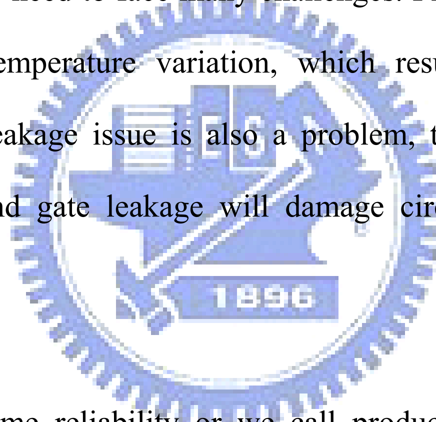
2.5 Summary

This chapter introduces traditional 6T SRAM structure, operation and other new SRAM design for low power in recent years. Due to the properties of advanced process technology, traditional 6T SRAM contrarily become a bottleneck in chip performance. As a result, lots of probable substitutes for traditional 6T SRAM or novel assist circuits are presented in recent years. Besides, ultra low power application such as mobile device, wireless sensor and medical device might require embedded memory working successfully with ultra low/low supply voltage. Obviously, 6T SRAM can't satisfy the demands. Following chapters will present the influence caused by advanced process on SRAM, a replacement of traditional 6T bit-cell and a multi-port SRAM based register file with wide operating voltage range.

Chapter 3 Timing Control Degradation and NBTI/PBTI Tolerant Design for Write-Replica Circuit in Nanoscale CMOS SRAM

3.1 Background

Static Random Access Memory plays a key role in high performance and low power VLSI technology due to SRAM occupies most part of chip area. In nano-scaled SRAM design, designers need to face many challenges. First is the variation, such as process, voltage and temperature variation, which result in circuit fail or bad performance. Second, leakage issue is also a problem, too. As the process scaled down, large junction and gate leakage will damage circuit function and increase power consumption.

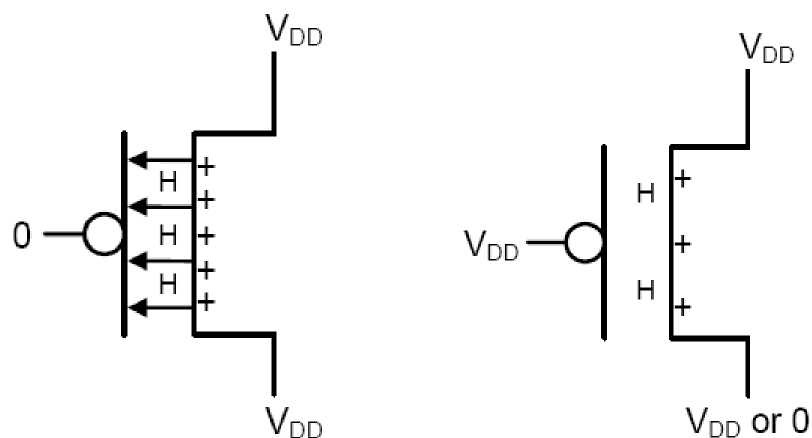


Besides, usage lifetime reliability or we call production reliability, is also an important issue. One of lifetime reliability is NBTI (negative bias temperature Instanbility) and PBTI (positive bias temperature Instanbility) effect, which is becoming more and more serious in advanced technology. The main influence of NBTI and PBTI on circuit design is that the threshold voltage (Vt) changes with usage time. Similarly, these effects also influence SRAM circuit. SRAM is composed of peripheral circuit and cell array. Peripheral is a kind of digital circuit and the bit-cell is latch type device. NBTI and PBTI make the logic has abnormal trip-point voltage result in logic timing mismatch and make the bit-cell has different static noise margin and write margin in data dependence form since SRAM might works a long time

without turning off because it need to keep the data. What's more, it is common that only a part circuit function in a huge SRAM macro every cycle. These similar and parallel signal paths in SRAM might have different timing issue after NBTI stressing. In other words, timing degradation is dependent on SRAM architecture, operation and signal path. In this topic, we focus on write timing issue on a prototype SRAM. This SRAM has replica circuit to control timing for ensuring the write operation against fail. All the simulation and analysis are based on 32nm prediction model which is published by Arizona University [3.1].

3.2 NBTI and PBTI

This section shows the detailed properties and causing reason of NBTI and PBTI. For a PMOS transistor, when gate voltage equals to zero, positive interface traps are accumulating over stressing time with hydrogen diffusion toward the gate. Actually, these traps are broken bonds and they increase threshold voltage of transistor. The stressed value increases with stressing time and saturate at higher value under higher voltage stressed. On the contrary, if the "stress" is removed, the PMOS can be recovery, Fig. 3.1. The dynamic behavior also is shown in the prior publish data, as shown in fig. 3.2 [3.2]



(a) (b)

Fig.3.1: (a) Stress mode (b) Recovery mode

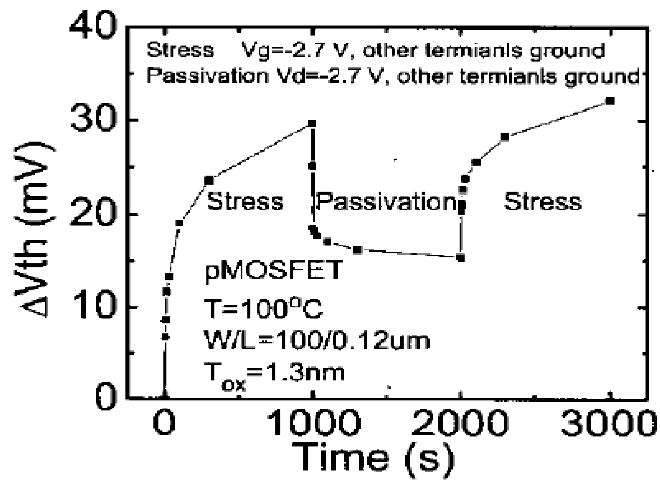


Fig. 3.2: Dynamic behavior in Vt. [3.2]

The above exception is rather simple. The physical level reason is shown in following. Negative bias temperature instability is a result of continuous trap generation in Si-SiO2 interface of PMOS transistor. Undesirable Si dangling bonds exit due to structure mismatch at Si-SiO2 interface. These dangling bonds act as charged interfacical traps of NBTI [3.6].

As a result, different signal stressed probabilities determine the threshold voltage of transistors. NBTI influence PMOS threshold voltage not only in conventional poly gate CMOS technology but also in high-k metal gate process technology which is one choice of substitute for poly-gate CMOS technology [3.7]. On the other hands, PBTI influence NMOS threshold voltage only in high-k metal gate technology. PBTI in high-k metal gate technology is the most serious.

The threshold voltage drift of transistor can be expressed by AC reaction diffusion model.

$$\Delta V_{TH}(t) \fallingdotseq K_{AC} \times t^{n} \fallingdotseq \alpha(S,f) \times K_{DC} \times t^{n} \qquad (1)\ [3.3]$$

However, in simulation and analysis, we can't directly use the AC model to do simulation. In this equation, K_{DC} is technology constant which is determined by technology process. Alpha is a prefactor which is a function of stress probability (S) and stress frequency (f). In fact, only stress probability is the most important and there are known prefactors for different stress probability. As a result, we can calculate each transistor of threshold voltage drift value by its stress probability so that we can analysis the whole circuit. Calculation of prefactor and the analysis method are from the prior published data. The following line graph shows the information of NBTI and PBTI. The VT drift due to NBTI and PBTI are based on AC RD model and calibrated with published data, in Fig. 3.3 [3.4-5]. The horizontal axis represents the stress time, whereas the vertical axis represents the increase value of threshold voltage. Here is three model stacked up with each other. PBTI is a serious problem. It dramatically goes up and has the least time to be saturated.

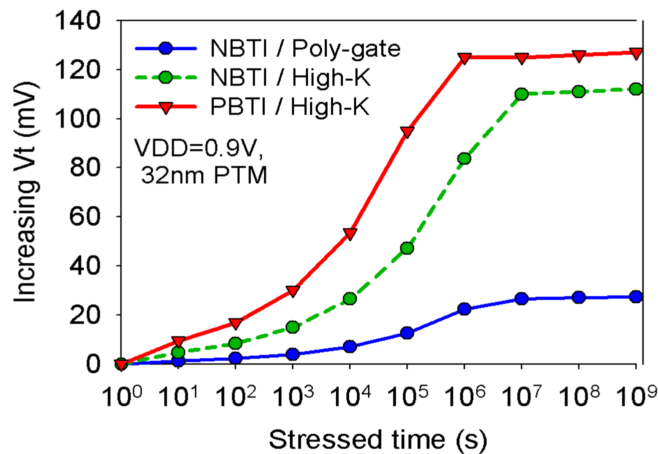


Fig 3.3: NBTI/PBTI induced V_T drifts vs. stressed time for 32nm poly-gate and high-k metal-gate devices (V_T drifts based on AC RD model and calibrated with published data).

3.3 Simulation environment

In this section, simulation environment is presented. A prototype SRAM is prepared. The architecture of this SRAM is 128 rows by 64 columns. This SRAM has a replica circuit to control the timing. Replica circuit can ensure WL pulse is wide enough for successful write or read operation, and it turns off WL as soon as possible after data are written into selected cell or read data. The beneficial provided by replica cell is that reduced power consumption and minimize the half-select disturb [3.8]. The operating frequency of this SRAM is 2.25GHz with 0.9V supply voltage.

The critical path of write operation is shown in Fig. 3.4.

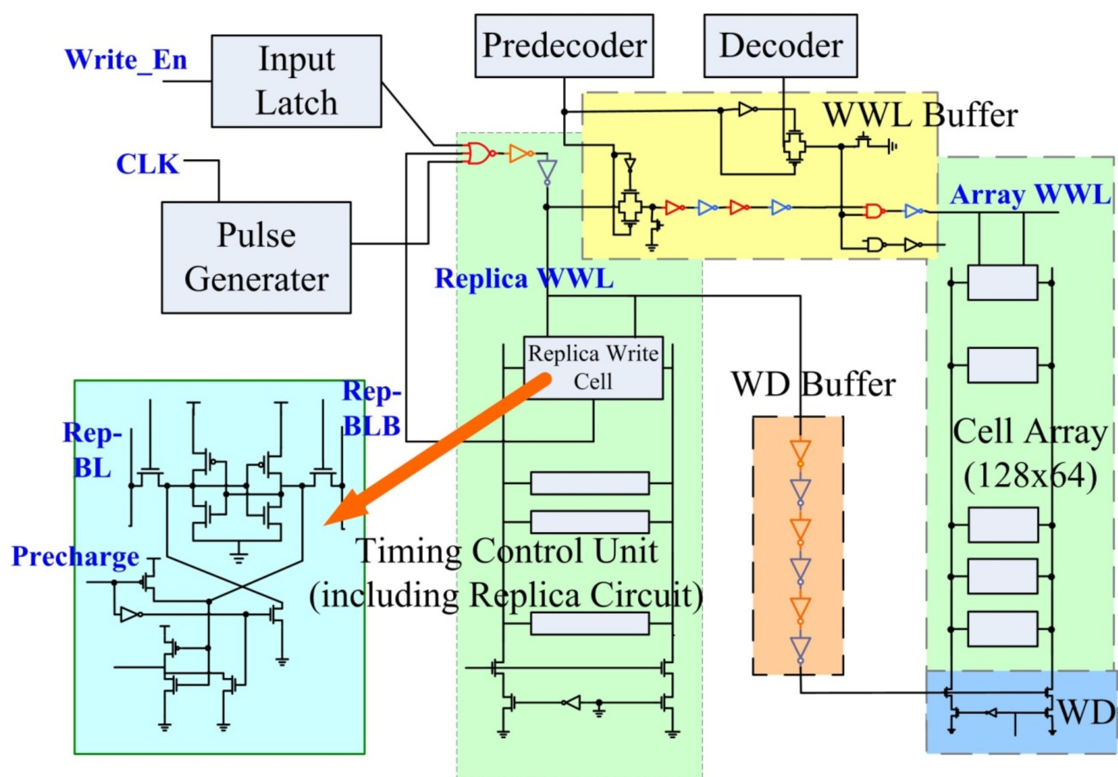


Fig 3.4: Critical paths of SRAM with Write-replica timing control.

In the beginning, the pulse generator using the CLK rising edge to generate the precharge signal for bit-line and read bit-line. The replica circuit resets the write replica cell during the precharge phase. If there is a write operation, the replica WWL is turned on, so is the replica write driver. In the meanwhile, the selected WWL and write driver is turned on through the long path. After detecting change in stored data,

the replica circuit turns the replica WWL and Write driver off. The selected WWL and Write driver in array are turned off though the long path once again. Then a write operation is finished.

In the following analysis, we classify SRAM write operation into 3 cases. Case 1 represents the case where the SRAM seldom performs Write operations. Case 2 represents the case where the SRAM has high probability of performing Write operation, but the considered Write Word-Line (WWL) is seldom selected. Case 3 represents the case that the SRAM has high probability of performing Write operation, and the considered Write Word-Line (WWL) is always selected. These 3 cases are summarized in Table 3.1.

Table3.1: SRAM write frequency and target WWL selected frequrency.

Case	SRAM Write frequency	Target WWL
1	Seldom write	Seldom selected
2	Frequently write	Seldom selected
3	Frequently write	always selected

3.4 Impact on Write-Replica Timing Control

In this section, detailed analysis about NBTI and PBTI impact on SRAM circuit. The analysis includes write time of cell, WWL pulse width, Write Window, Write cycle time and energy consumption. First is the bit-cell.

3.4.1 Write Delay of Array and Write Replica Cell

Following figure shows the circuit diagram of conventional 8T bit-cell. NBTI and

PBTI induce threshold voltage drifts and mismatch between cell transistors of its cross-coupled inverter pairs. The same color on the transistors means that the stress probability is the same. Because it is a cross coupled inverter structure, the two stress probability of blue and red are complement each of the other. Therefore, the cell write delay varies under different signal probability of stored data. They can be summarizes into 5 condition, as shown in this table.

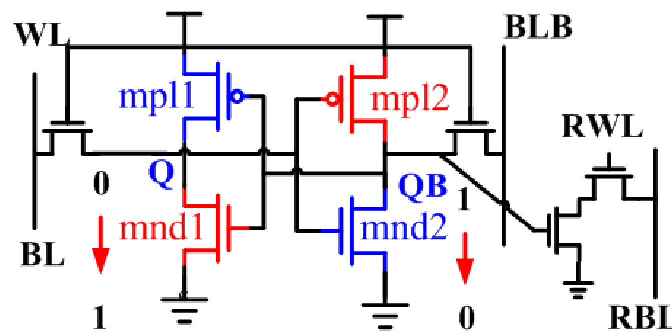


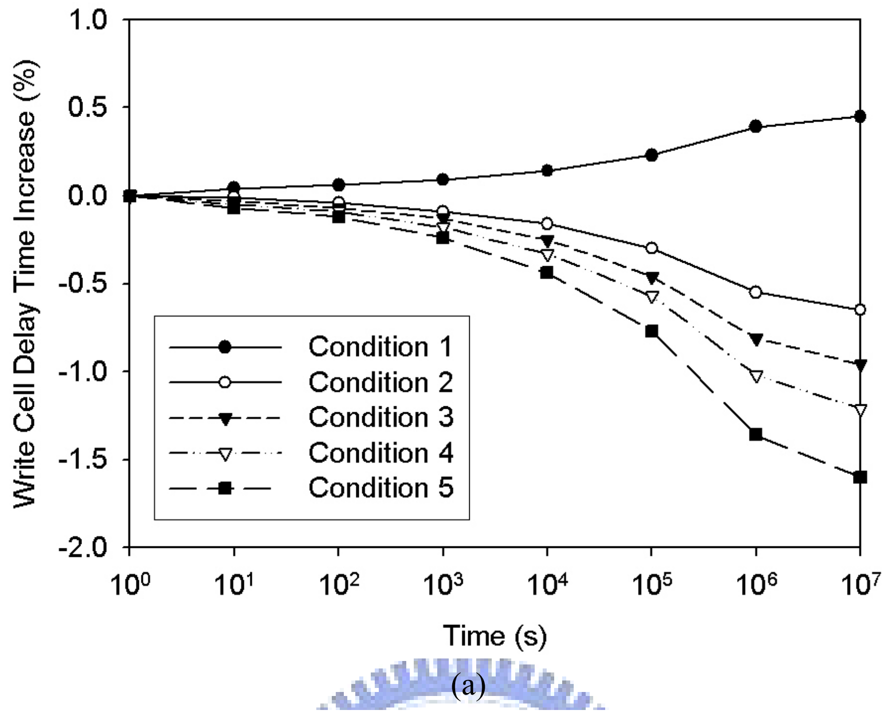
Fig.3.5: 8T-SRAM cell schematic

Table 3.2: Stressed probability distribution.

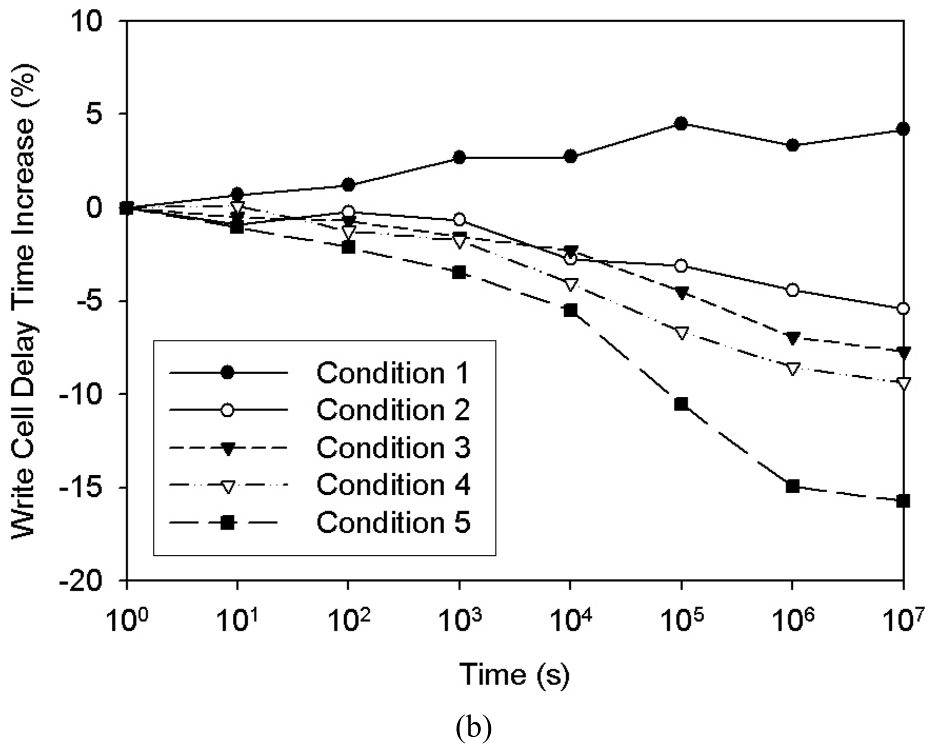
Cond.	mpl1/mnd2	mpl2/mnd1	Write delay
1	P=1	P=0	Increase
2	P=0.75	P=0.25	Decrease
3	P=0.5	P=0.5	Decrease
4	P=0.25	P=0.25	Decrease
5	P=1	P=1	Decrease

Fig. 3.5 shows the schematic of an 8-T SRAM bit-cell [3.10]. NBTI and PBTI induce V_T drifts and mismatch between cell transistors of its cross-coupled inverter pairs. Therefore, Write delay (or Time-to-Write) of a SRAM cell varies under different signal (stress) probability of stored data. Table 3.2 summarizes 5 different signal (stress) probability conditions, and Fig. 3.6(a) and 3.6(b) show Write delays of

SRAM cells in PTM 32 nm poly-gate and high-k metal-gate CMOS technology respectively. They show that Write delay of SRAM cell improves in most conditions except Condition 1. This is because when a cell is affected by NBTI and the cell signal (stress) probability is not 100% (0%), both PFET loading transistors become weaker. A weaker holding PFET helps the initial discharging of the “logic 1” storage node through the access NFET, while a weaker pull-up PFET impedes the subsequent pull-up of the “logic 0” storage node. Since the initial discharging of the “logic 1” storage node tends to be the dominating factor for Write operation, the Write Margin (WM) and Time-to-Write improve with both PFET weakened. However, when the cell signal (stress) probability is 100% (0%) (as in Condition 1), only one PFET loading transistor becomes weaker. For the worst case pattern, the PFET holding the original “logic 1” storage node is not stressed/weakened, so the pull down of the “logic 1” storage node is not getting easier. The PFET corresponding to the original “logic 0” storage node, however, would be fully stressed/weakened, and thus slowing down the charging of its storage node to “logic 1” during Write operation. As a result, the WM and Time-to-Write degrade.



(a)



(b)

Fig. 3.6: Cell Write delay (Time-to-Write): (a) with NBTI in poly-gate CMOS; (b) with NBTI and PBTI in High-κ metal-gate CMOS

The Write-replica cell is also impacted by NBTI and PBTI. The replica cell is reset (with the right cell storage node conditioned to “1”, see Fig. 3.4) during

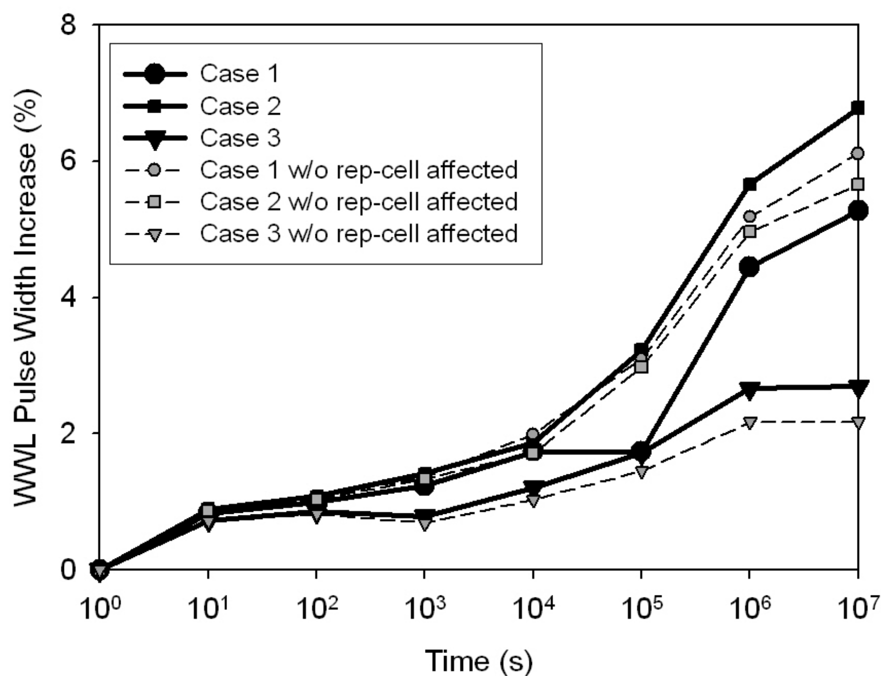
precharge phase of every clock cycle regardless of the operation as long as the array block is active. When a Write operation is performed, the writing of the replica cell is *always* performed by *pulling down the right cell storage node*. As such, if the SRAM seldom performs Write operations, the right Holding PFET in the replica cell is always stressed (hence weakened). Thus Write delay (Time-to-Write) of the replica cell improves when a Write operation is actually performed later. On the other hand, when the SRAM performs Write operations frequently, the replica cell would be "written" frequently, and the signal (stress) probability would be distributed more evenly between the right and left Holding PFET. As a result, the right Holding PFET would be less stressed, and the replica cell Write delay would increases.

3.4.2 WWL pulse width

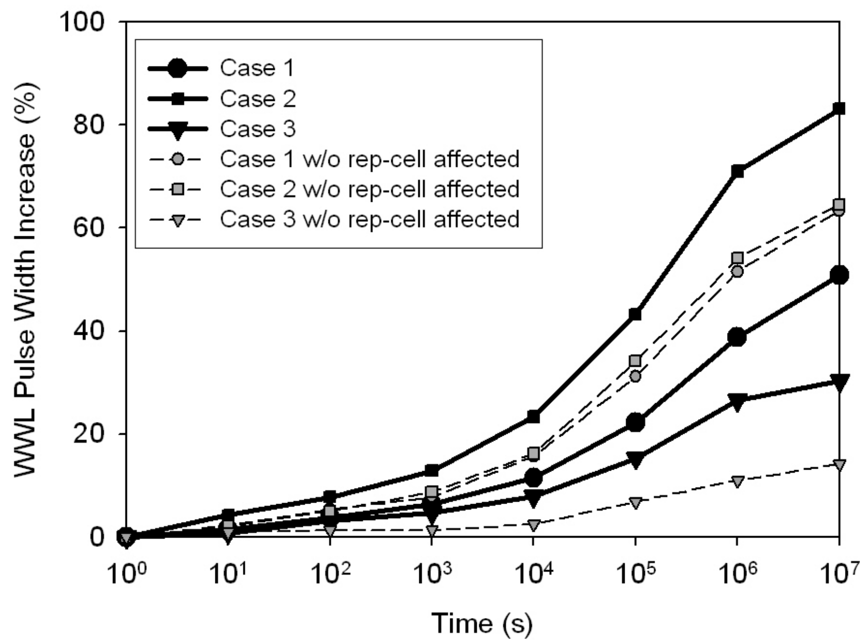
After long and sustained usage time, signal transferring delay of logic control circuit is degrades by NBTI and PBTI, and the pulse width of the selected WWL increases (Fig. 3.7). Because high-k metal gate devices are more sensitive to NBTI and PBTI, the WWL pulse width increase is more significant/severe in SRAM with high-k metal gate devices (Fig. 3.7(b)). Moreover, when a logic gate is stressed by a particular input signal, it becomes more difficult to maintain or transfer the output corresponding to the input signal. For example, an inverter with its input at "1" for sustained long period would exert NBTI stress on the pull-down NFET, causing increase of the NFET VT and thus making it more difficult to transfer "0" to the output or maintain the output at "0". Hence, if the considered WWL is seldom selected, the inverter/driver chain becomes less capable to transfer and maintain the "Logic 0" output signal. Consequently, the selected WWL turns on earlier and turns off slower, and WWL pulse width increases more in Case 1 and Case 2. Furthermore, if the SRAM seldom performs Write operation (as in Case 1), the Write delay of array

cell increases while the Write delay of Write-replica cell decreases. In contrast, if the SRAM experiences frequent Write operation (as in Case 2 and Case 3), the Write delay of array cell decreases, while the Write delay for the Write-replica cell increases. As a result, the WWL pulse width of Case 2 is wider than Case 1, although their WWL pulse widths would be close/closer if the Write- replica cell is assumed to experience no stress.

The larger/wider WWL pulse width causes more severe half-selected disturb during Write operation. It also results in larger Write power and more serious gate leakages in cell access NFETs across the selected WWL. In addition, it significantly degrades the cycle time of a high performance SRAM.



(a)



(b)

Fig.3.7: Increase of WWL pulse width due to degradation in logic control circuits: (a) with NBTI in poly-gate CMOS; (b) with NBTI and PBTI in High-κ metal-gate CMOS.

3.4.3 Write Window

Write Window is defined by the duration when both the selected WWL pulse and WD pulse are active. When the SRAM is "fresh", the selected WWL and WDs turn on and turn off in sync, and the Write Window is determined by the rising edge of WD pulse and the trailing edge of WWL pulse (Fig. 3.8(a)). Nevertheless, because the number of buffering stages and loading of WD and WWL driver are different, timing tracking between WD pulse and WWL pulse deteriorates after long usage time. The Write Window may become completely determined by WWL pulse alone (Fig. 3.8(b)) or by WD pulse alone (Fig. 3.8(c)), depending on the relative shift of the signal edges. The Write Window width determines whether the Write operation succeeds or not. If the Write Window is shorter than the Write delay of a cell, the Write operation fails. Although larger Write Window improves the Write operation success rates, it

degrades the cycle time and hence performance.

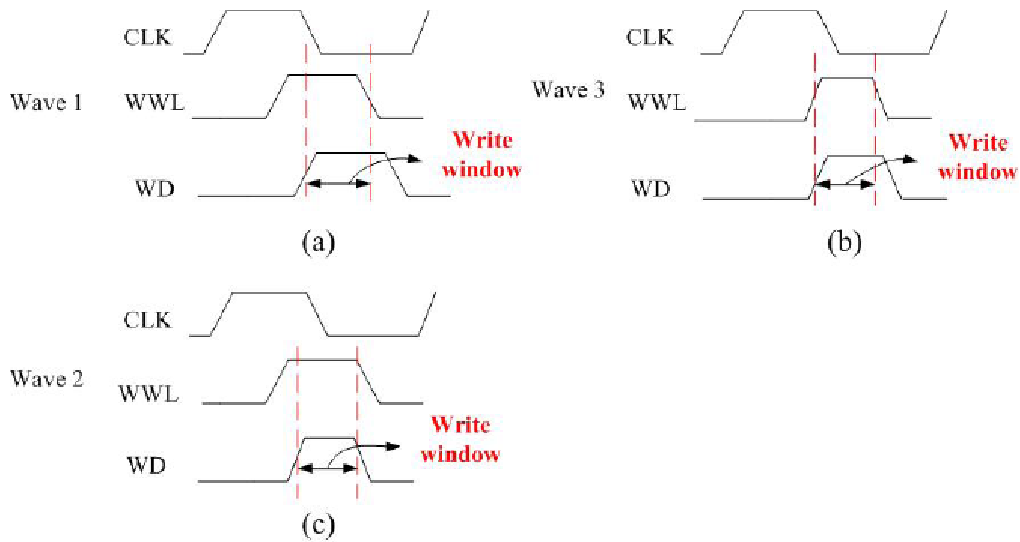
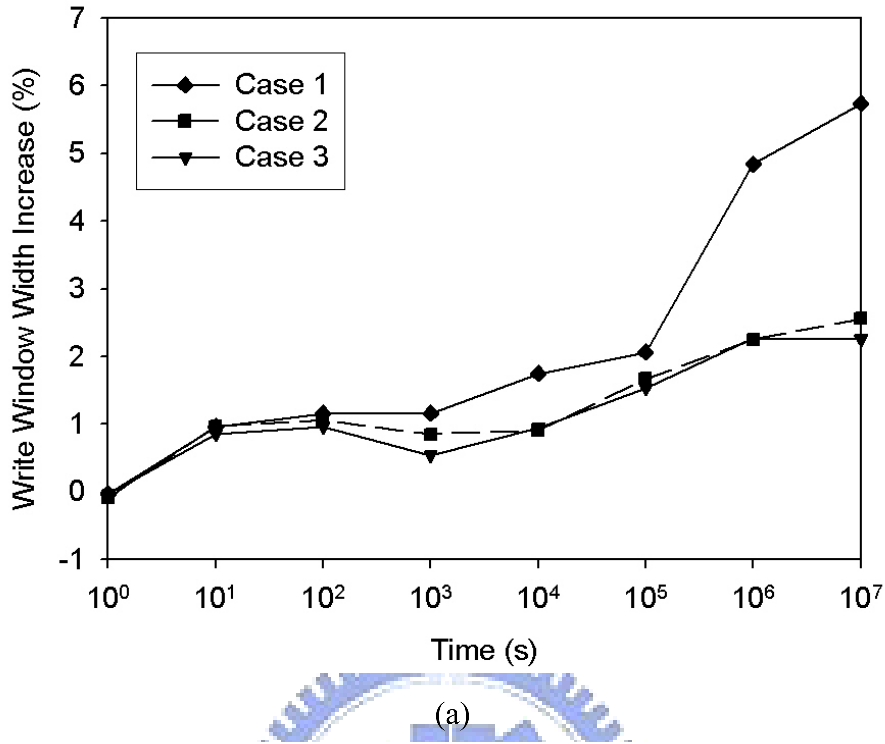
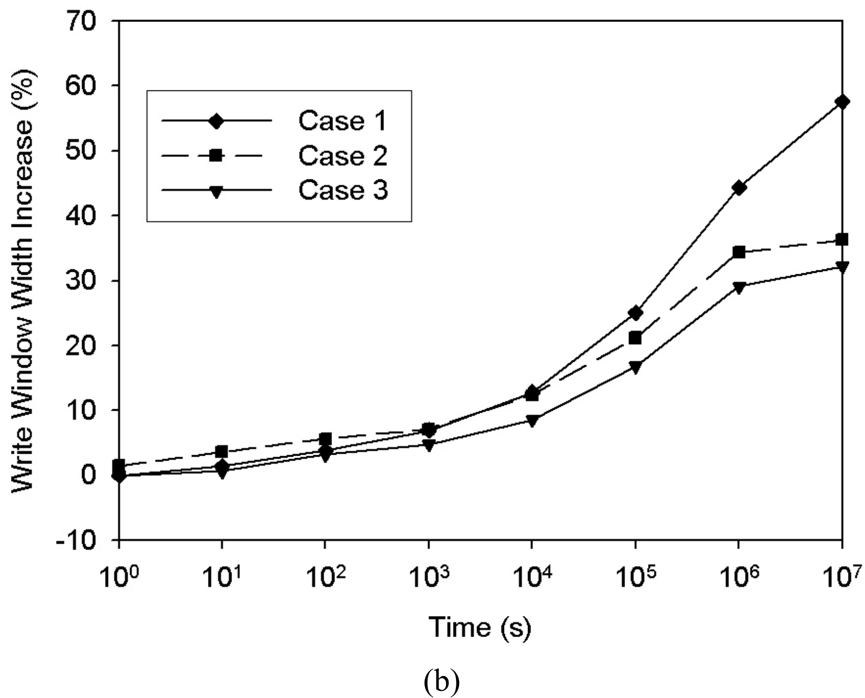


Fig. 3.8: (a) Write Window of a “fresh” SRAM determined b rising edge of WD pulse and trailing edge of WWL pulse; (b) and (c) Write Window of a SRAM after stressing determined entirely by WWL pulse (in(b)) or entirely by WD pulse (in (c)).

The pulse widths, in general, increase after stressing as explained previously. As shown in Fig. 3.9(a), the Write Window width becomes wider after stressing since the pulse widths of both WWL and WD increase. Comparing Fig. 3.6(a) and Fig. 3.9(a), we can find that the increase of Write Window width is larger than the increase of the SRAM cell Write delay in each case. It implies that SRAM Write operation would be successful post NBTI stress, and there is margin left for circuit techniques to reduce the Write Window width for post NBTI stress performance improvement. The same phenomenon can also be observed in SRAM with high-k metal-gate devices (Fig. 3.6(b) and Fig. 3.9(b)).



(a)

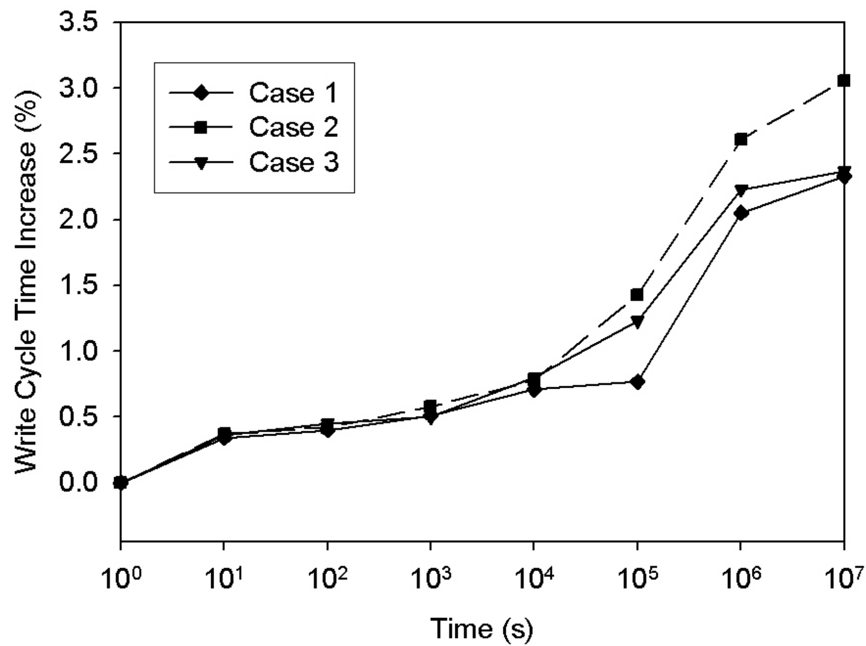


(b)

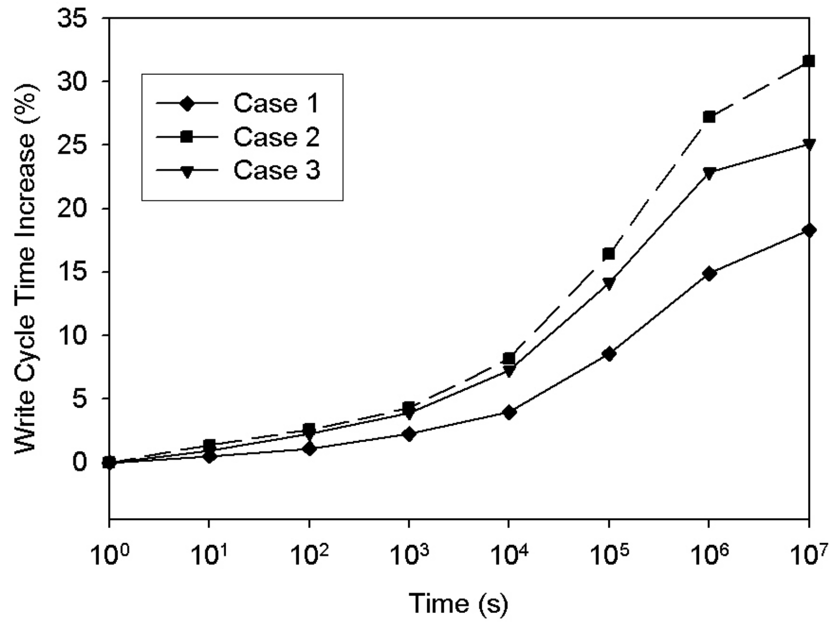
Fig: 3.9 Increase of Write Window width: (a) with NBTI in poly-gate CMOS; (b) with NBTI and PBTI in high-k metal-gate CMOS.

3.4.4 Write cycle and Write energy

The minimum required cycle time for Write operation is defined between CLK rising edge and the time that the last logic circuits (the selected WWL buffer or WDs) turn off. As shown in Fig. 3.10, the Write cycle time degrades after stressing for both poly-gate and high- k metal-gate technology. In the worst case, The Write cycle time of SRAM with high-k metal-gate devices degrades 30% after stressing compared with 3% degradation for SRAM with poly-gate devices. On the other hand, because VT drifts higher post NBTI/PBTI stress, the leakage currents and active power improve. However, due to the lengthened Write cycle time, the Write energy consumption increases.



(a)



(b)

Fig. 3.10: Increase of Write cycle time: (a) with NBTI in poly-gate CMOS; (b) with NBTI and PBTI in high-k metal-gate CMOS.

3.5 NBTI/PBTI tolerant Write replica Scheme

To mitigate the NBTI/PBTI induced Write performance degradation, an NBTI/PBTI tolerant Write-replica scheme is proposed. Our analysis indicates that the degradation of WWL buffers and WD buffers are dominating factors for Write performance and reliability deterioration. We also find that the precharge time increases less than 0.3% after 3 year stressing due to the small fan-out of the precharge signal path. Therefore, we focus on refining Timing Control Unit, WWL drivers, and WD buffers.

In order to mitigate the Write delay degradation of the Write-replica cell, a power switch is inserted between VVDD of the replica cell and the power supply line (VDD), as shown in Fig. 3.11(a). When the SRAM is not performing Write operation, node VVDD is connected to GND. Thus, PMOS and NMOS pairs of the replica cell are not

stressed in this condition, and NBTI/PBTI induced Write delay degradation of the replica cell is reduced. In addition, since the degradation of logic path delay induced by NBTI/PBTI is quite severe in WWL buffers and WD buffers, power switches are also added to the VVDD of these buffers, as shown in Fig. 3.11(b). These power switches connect VVDD of these buffers to GND when the SRAM is not performing Write operation. As such, there is no bias and hence no NBTI/PBTI stress across PMOS and NMOS in these buffers. Notice that in both cases (for replica-cell Write delay path, and for WWL/WD buffers), the zero-bias state not only removes the NBTI/PBTI stress, but also provides a period for "Recovery" phase to further reduce the VT drift.

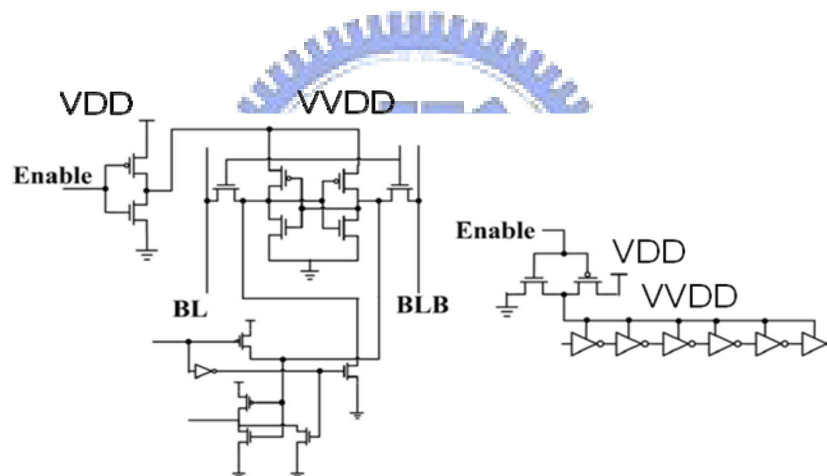


Fig. 3.11: Proposed circuits to mitigate NBTI/PBTI degradation: (a) Write-replica cell with power switch; (b) buffer with power switch.

To further mitigate NBTI/PBTI degradation, we partition the SRAM into multiple banks. Timing Unit, WWL buffers, and WD buffers are also localized (Fig. 3.12). The method is just like in [3.9]. In this proposed multi-bank architecture, when a bank is performing Write operations, VVDD of other banks' Timing Units and buffers can be connects to GND. Thus, Timing Units and WWL/WD buffers of other banks are put into "Recovery" phase. Moreover, the "active probability" of local Timing Units and WWL/WD buffers in the proposed scheme is less than the single bank architecture. In

other words, local Timing Units and WWL/WD buffers have more opportunities to stay in “Recovery” phase. However, if a bank frequently performs Write operations, it becomes a hot spot and its Write cycle time would dominate the performance of the SRAM. Although this condition seldom happens, its possibility exists to limit the SRAM performance. Hence, higher level mechanism (software or memory management) is needed to ensure more even probability of bank access.

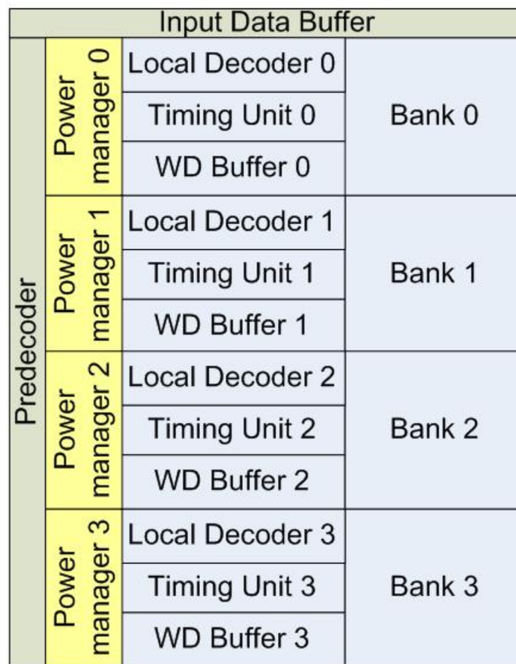
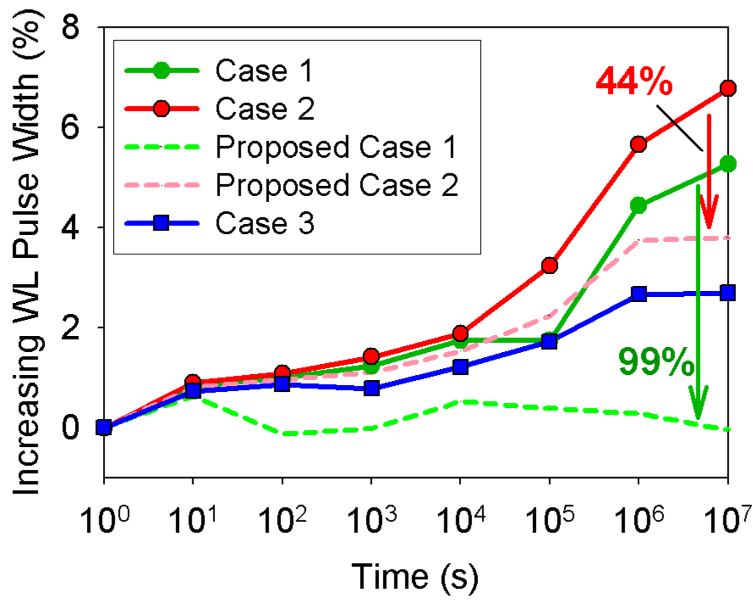
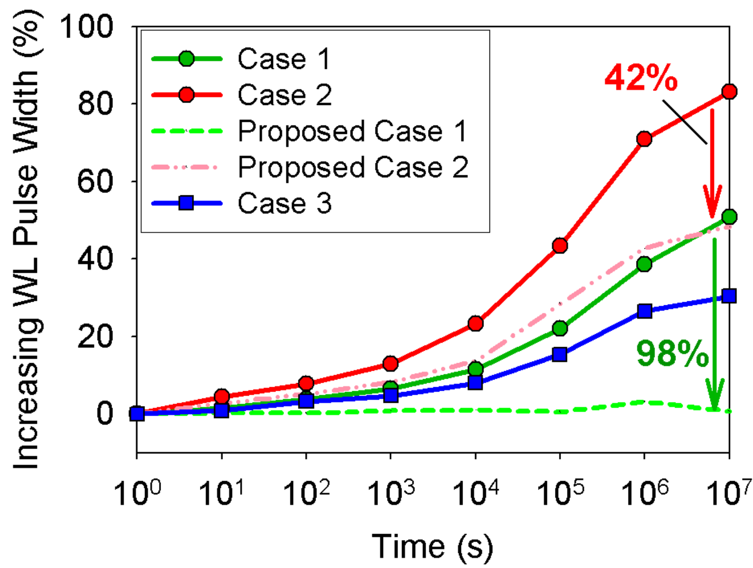


Fig. 3.12: Proposed multi-bank SRAM architecture.

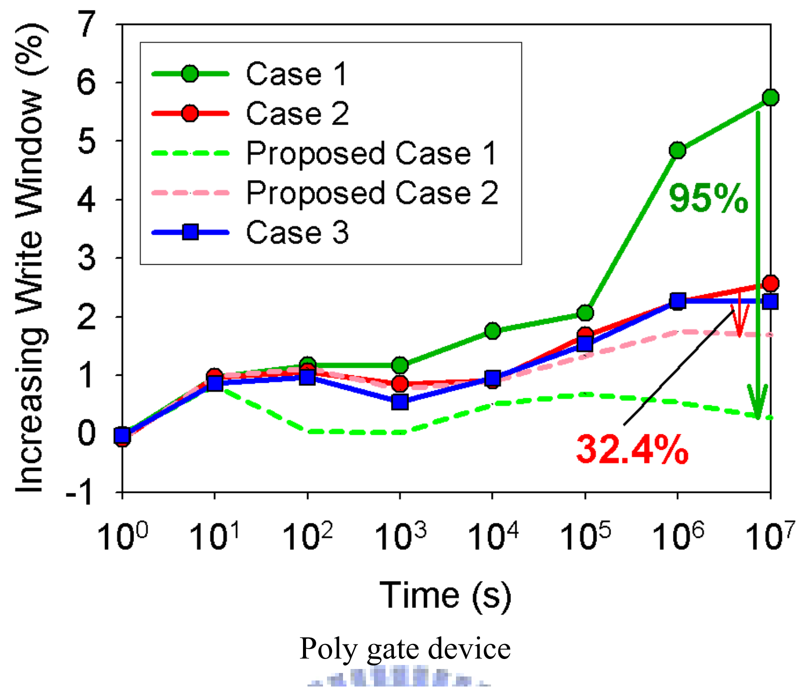


Poly gate device

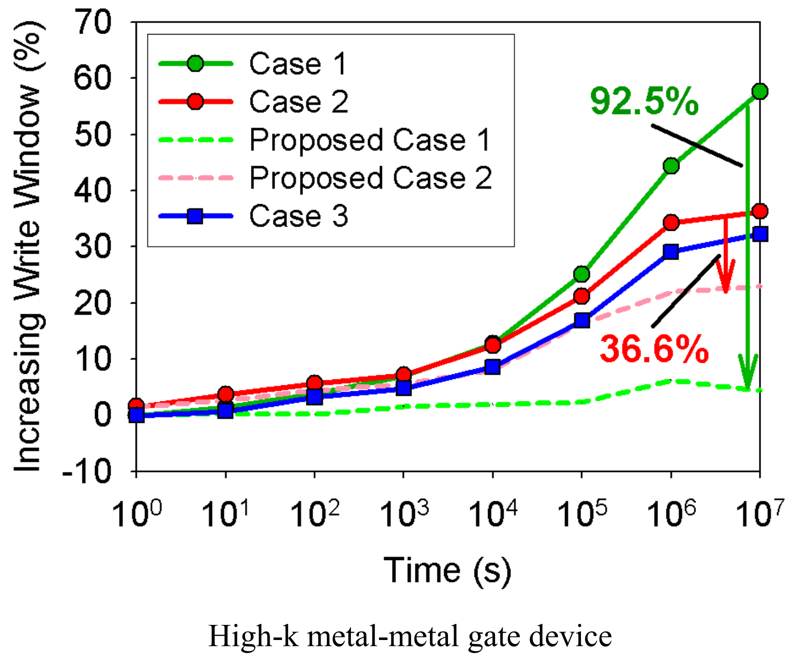


High-k metal-metal gate device

(a)

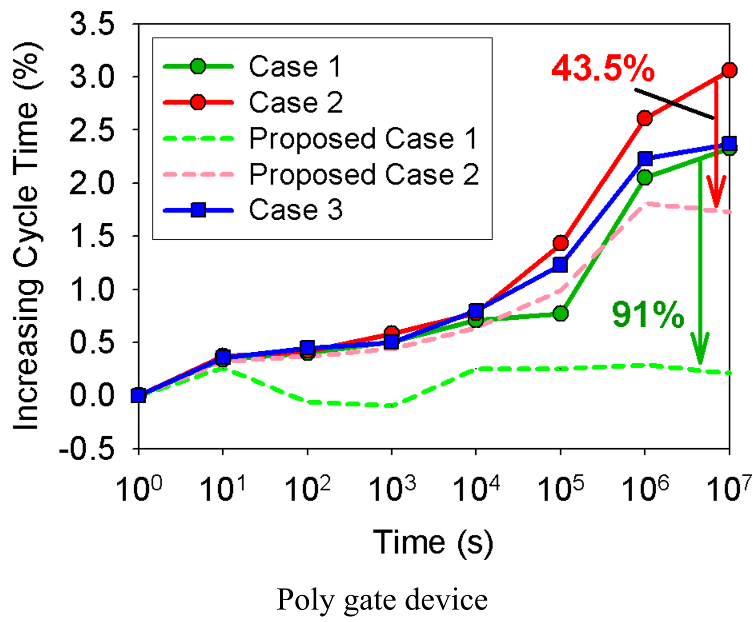


Poly gate device

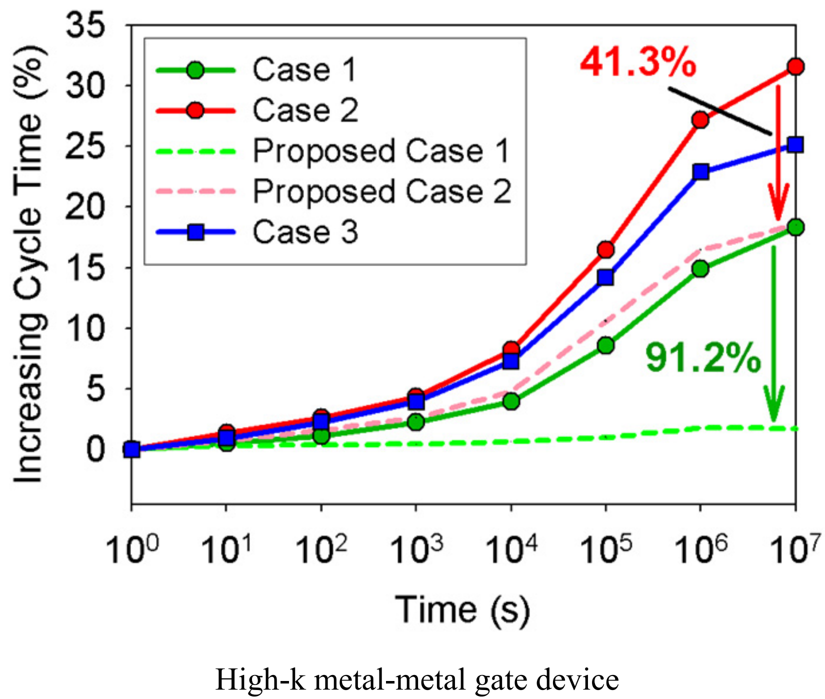


High-k metal-metal gate device

(b)



Poly gate device



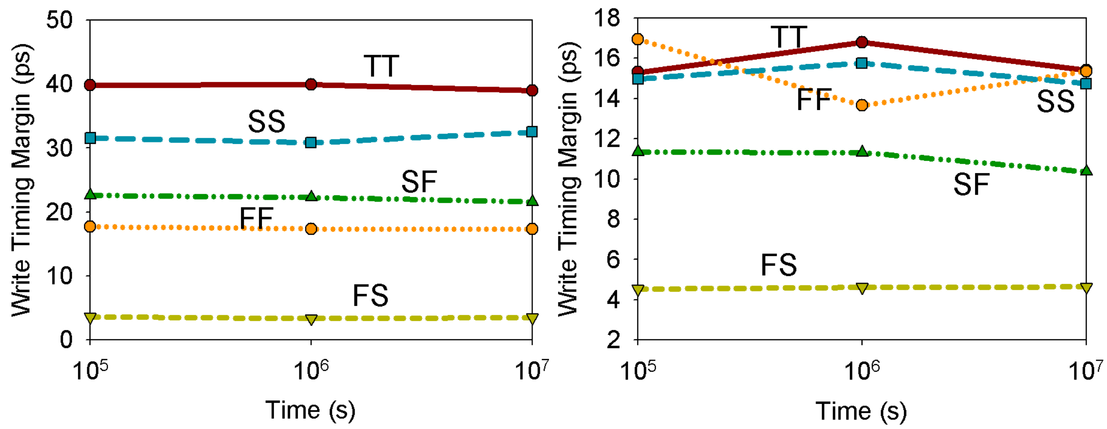
High-k metal-metal gate device

(c)

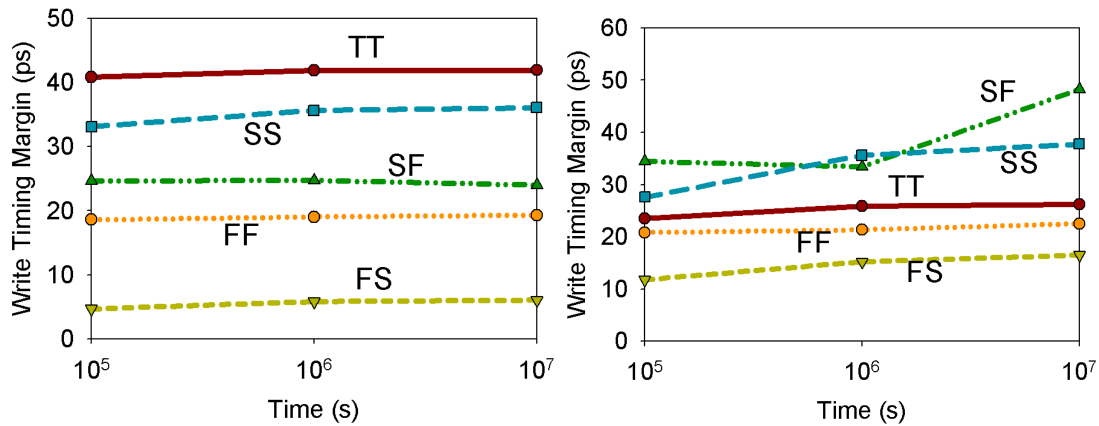
Fig 3.13: Degradation mitigation by proposed NBTI/PBTI tolerant scheme (a) increase of WWL pulse width; (b) increase of Write Window width; (c) increase of Write cycle time.

Fig. 3.13 shows simulation results of the proposed scheme compared with the conventional Write-replica scheme. The WWL pulse width increase of Case 1 and

Case 2 are significantly reduced by the proposed scheme (Fig. 3.13(a)). The maximum improvement is about 98% for Case 1 and 42% for Case 2. The increase in Write Window width is also significantly reduced. As shown in Fig. 3.13(b), there is almost no change in Write Window width for the proposed scheme in Case 1. Fig. 3.13(c) shows the marked reduction (41% to 91%) of Write cycle time degradation with the proposed scheme. Notice that while minimizing the Write Window width degradation (increase) is beneficial for performance and power, it may lead to Write fail if the width of the Write Window is narrower than the Write delay of a SRAM cell. Fig. 11 shows the difference (margin) between Write Window width and the cell Write delay across various process corners. It shows adequate margin across process corners, indicating the Write operations would be successful with the proposed scheme in both poly-gate and high-k metal-gate CMOS technology. Fig. 3.14 also shows that our scheme offers little improvement in Case 3. Case 3 represents the condition wherein a SRAM frequently performs Write operations and a particular address is always selected. Timing Unit, WWL buffers, and WDs are always active in Case 3, and their VVDD are always connected to VDD. Therefore, the proposed scheme has little effect on these circuits in Case 3. Notice that with the proposed scheme which significantly improves the Write cycle time for other cases, the Write cycle time would be limited by Case 3.



(a)



(b)

Fig 3.14: Difference (Margin) between Write Window width and cell Write delay (Difference = Write Window width - cell Write delay) for proposed scheme across process corners. Positive difference (margin) indicates Write operation would be successful. (a) Case 1;(b) Case 2.

3.6 Summary

This section presented a detailed analysis on NBTI/PBTI induced timing control degradation in SRAM Write-replica circuit based on PTM 32nm technology models. The widths of WWL pulse and Write Window, and the cell Write delay, were shown to increase after stressing, thus degrading the Write cycle time. An NBTI/PBTI tolerant Write-replica scheme with power switches inserted in the Timing Unit and WWL/WD buffers was proposed. The scheme minimized the stress time and provided

“Recovery” period for timing-critical devices and circuits to mitigate the degradation. The results showed that around 42%-98% reduction in WWL pulse width increase, 32%-92% reduction in Write Window width increase, and 41%-93% reduction in Write cycle time increase, could be achieved while maintaining adequate Write timing margin across process corners. A multi-band architecture was proposed to allow inactive bands to be put into zero-bias state to further minimize the stress time and maximize the “Recovery” phase. High-level mechanism would be needed to ensure more even bank access to leverage the multi-bank architecture.

Chapter 4 High performance and low VDDmin of new 8T SRAM cell design

4.1 Introduction

In nanoscale CMOS SRAM design, serious process, voltage and temperature variation let conventional 6T SRAM cell difficult to scale down with process. In [4.1], authors point out that area of 6T bit-cell will larger than conventional 8T bit-cell in future process due to the large beta and gamma ratio against process variation, as shown in Fig. 4.1. In other words, read disturb is more serious in advanced process technology especially lowering operating voltage. In recent years, low voltage circuit design for low power application is very common. Conventional 6T cell is difficult to work under low voltage. Furthermore, with advanced process technology, lowering VDDmin of 6T bit-cell is almost impossible. It is necessary to design a substitute for conventional 6T bit-cell for advance process or low voltage design.

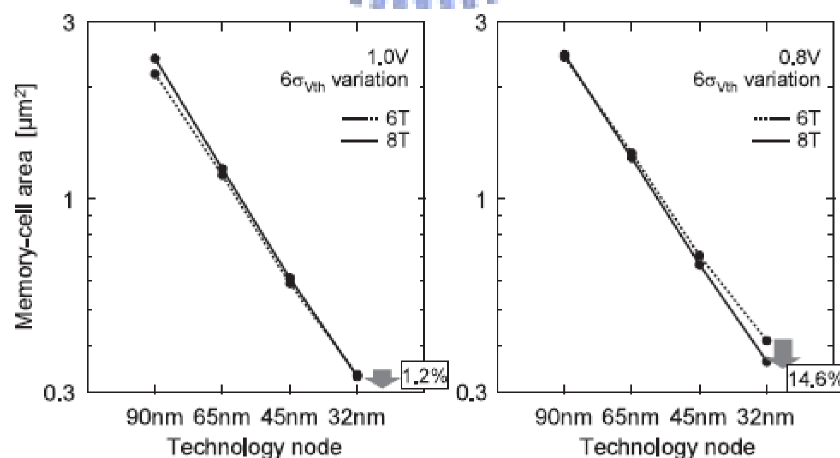


Fig 4.1: Minimum-area comparison between 6T and 8T cells [4.1]

In this chapter, a new 8T SRAM bit-cell for low power and high performance is

presented. Since this 8T bit-cell is a new type design, detailed bit-cell analysis such as stability and read/write operation are presented in this section. Also, the appropriate architecture of new 8T array and methods that allow the new 8T bit-cell working with conventional 6T SRAM peripheral circuit is shown in this chapter. Following simulation and analysis are based on UMC 65nm process technology. Besides, this project are discussed with and supported by professor Ching-Te Chuang of Digital VLSI Lab, Hao-I Yang of LPMD Lab, Jihi-Yu Lin and Ming-Hsien Tu of MSCS Lab. Also, this design will tape out in November, 2009 supported by ***Faraday Technology Corporation***.

4.2 Cell structure and basic operation of cell

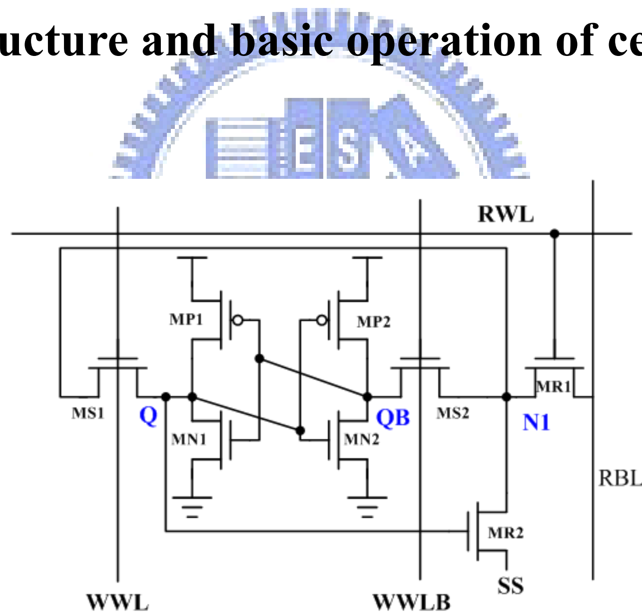


Fig.4.2: Schematic of the new 8T cell

Fig.4.2 shows the circuit diagram of the new 8T bit-cell. This bit-cell has inner layer and outer layer of passgates which are controlled by column based write word-line (WWL and WWLB) and row based read word-line (RWL). Read port is controlled by read word line and stored node of cell. SS is controllable node according to operation or leakage consideration. Besides, this cell only has one

bit-line for performing read or write operation.

4.2.1 Precharge/Standby mode

While cell performing precharge mode or standby mode, both RWL and WWL pair are turned off. Therefore, the three passgates in bit-cell are cut off for improving the stability of stored data. In precharge mode, the voltage level of RBL is logic 1, as shown in Fig. 4.3.

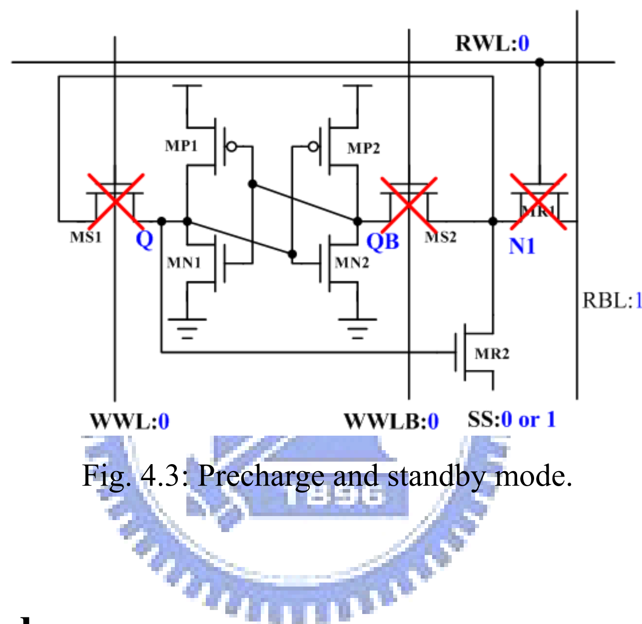


Fig. 4.3: Precharge and standby mode.

4.2.2 Read mode

RBL is precharged to a high voltage level first. SS of selected cell is connected to ground. Then, row-based RWL turns on, and Mr1 is open. The selected BL either stays at high voltage level or discharges through the outer layer pass-gate NMOS and the Mr2 according to the data stored in the cell storage node. Read operation is also illustrated in Fig. 4.4. Read operation is the same as the conventional 8T bit-cell. Read disturb and read half-select disturb is eliminated. SS of other unselect cell can be pulled up for leakage current concern.

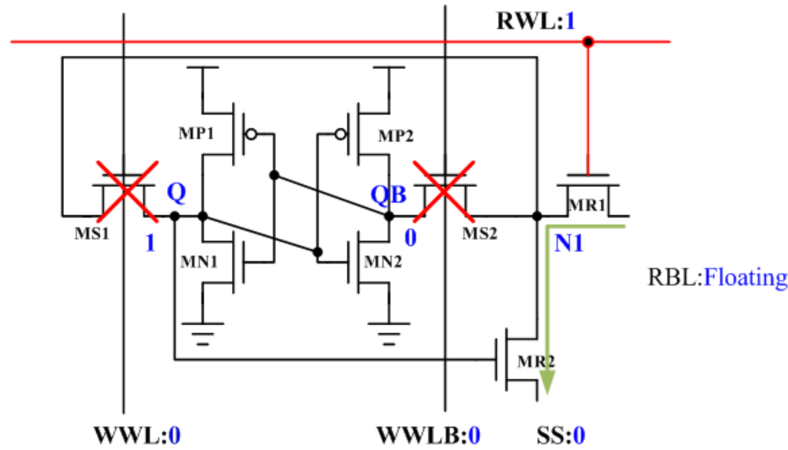


Fig. 4.4: Read mode.

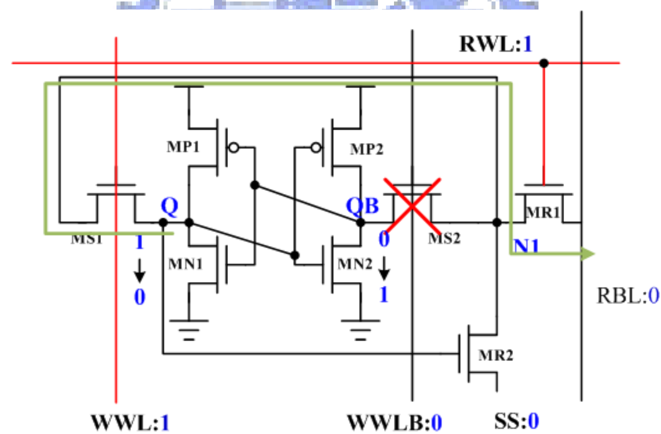
4.2.3 Write mode

RBL is precharged to a high voltage level first. SS of selected cell is connected to ground. Then, "Row-based" RWL and one of the "Column-based" WWL turn on, so the outer layer passgate NMOS and one of the inner layer passgate NMOS are open. The selected BL discharges to 0. The data ("0") on the selected BL is then written to the cell storage node Q if WWL is "High", as shown in Fig. 4.4 (a), and to the cell storage node QB if WWLB is "High", as shown in Fig. 4.4 (b).

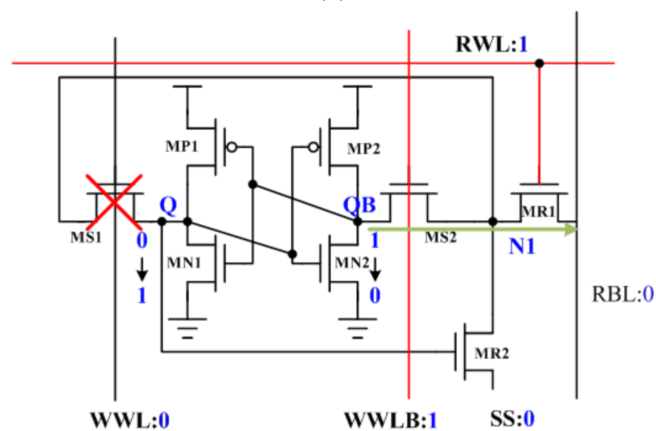
For traditional 6T SRAM during write operation, the input data is used to complementarily drive the selected bit-line pair. In the proposed scheme, the input data is used to complementarily control the "Column-based" write word-line pairs. Besides, this scheme also provides a solution for half-select issue. Conventional 8T bit-cell address read disturb issue, but still face write half-select disturb [4.3]. This is why traditional 8T bit-cell can't replace conventional 6T bit-cell. Based on the structure of this new 8T bit-cell, only the selected cell has a path to RBL since passgate is controlled by selected WWL and RWL. Since the read disturb and read/write half-select of 6T doesn't happen to this new this 8T bit-cell, this new 8T

bit-cell has several advantages, such as: Read and write noise margin are independent of transistor sizing; While lowering operating voltage, bit-cell stability is still good; Size modification for noise margin concern within each process generation becomes an easier job.

TI (Texas Instruments)' 7T SRAM cell, in Fig 4.5 [4.2], also has only one bit-line. However, due to the Vt loss in passing a logic 1 signal through passgate NMOS, the stack NMOS will make this 7T cell performing write "1" poor or even work fail in low voltage operation. There is no such problem in this new 8T bit-cell. RBL is pull down to perform "write 1" or "write 0." Consequently, equal capability in writing 1 or 0, and write half-select disturb makes this new 8T bit-cell perform write operation well in low voltage operating for low power issue.



(a)



(b)

Fig. 4.4: Write mode (a)Write 0; (b)Write 1.

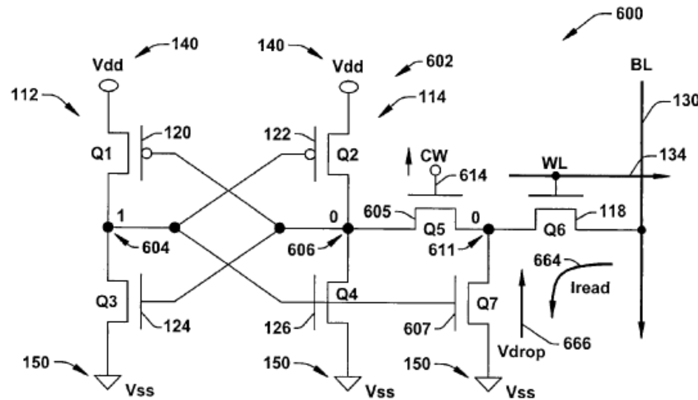


Fig.4.5: TI (Texas Instruments)' 7T SRAM bit-cell [4.2]

4.3 Cell stability

In Fig. 4.6, read noise margin of traditional 6T and new 8T bit-cell are shown. Read noise margin of "regular size 6T" is from UMC 0.62um^2 6T bit-cell. On the contrast, "same size 6T" means that all transistor sizes are the same. From Fig. 4.6, transistor sizing has influence on noise margin of traditional 6T bit-cell especially in high voltage range. However, transistor sizing doesn't have influence on this new 8T bit-cell. In other words, designer doesn't modify cell sizing for noise margin concern within each process generation. Besides, read noise margin of new 8T bit-cell has 1.75X improvement compared to tradition 6T bit-cell.

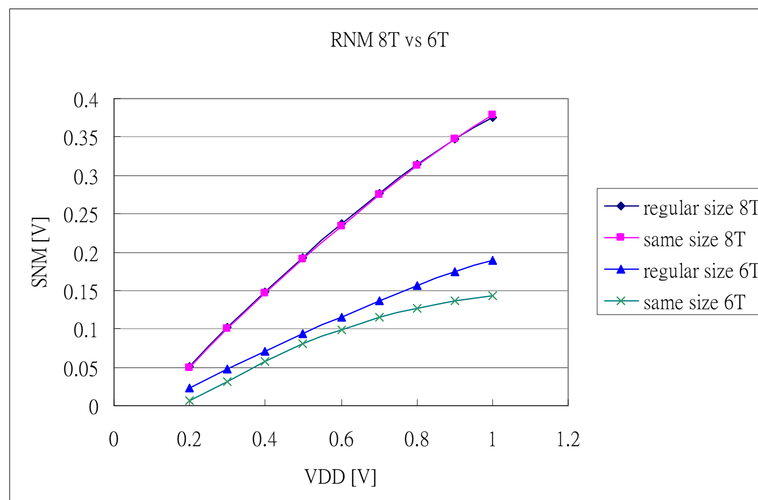


Fig. 4.6: Read noise margin comparison.

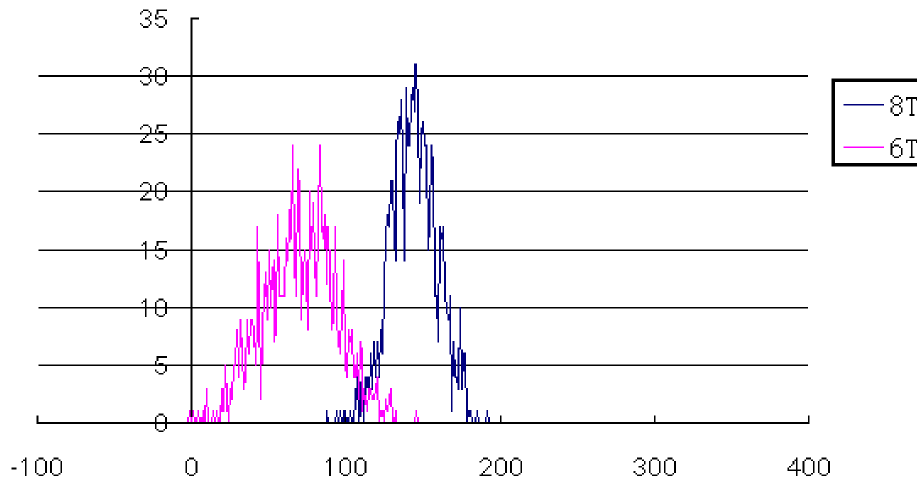


Fig. 4.7: Monte Carlo simulation, times=1000, VDD=0.4V and sigma=3.

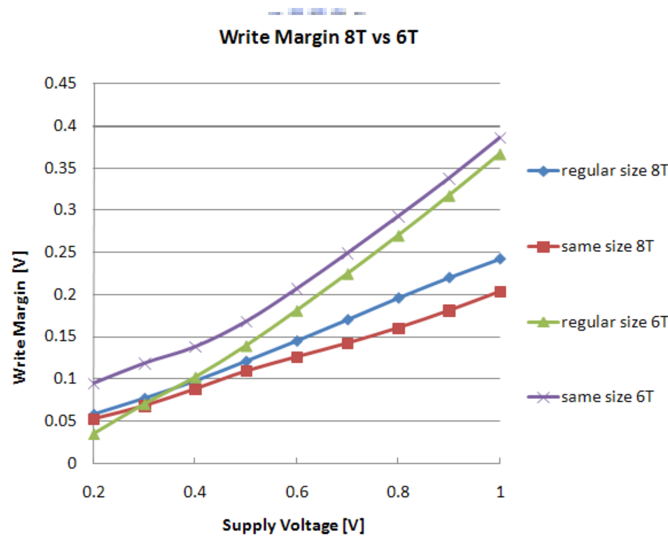


Fig. 4.8: Write margin.

Write margin in this section is defined that the maximum bit-line voltage can flip-out the stored data. From Fig. 4.8, write margin of new 8T bit-cell is worse than traditional 6T bit-cell due to the two layer passgate NMOS.

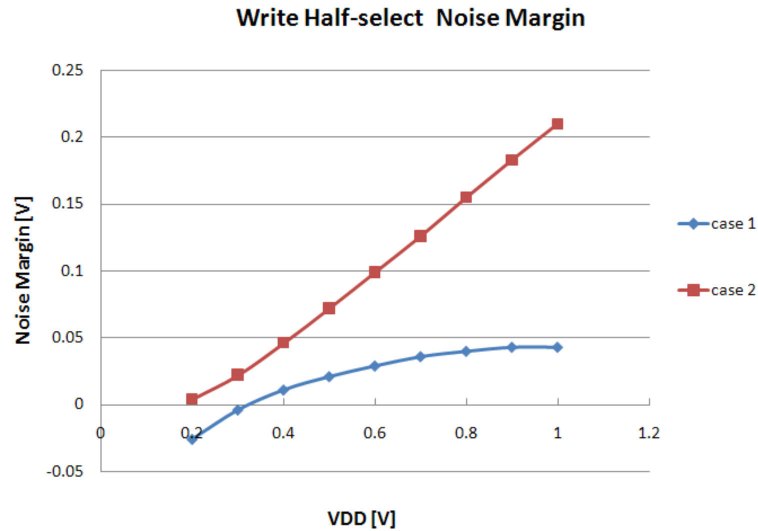
4.4 Architecture of array

According to the structure of new 8T bit-cell, there is stability issue in half-select

write mode. Fig.4.9 (a) (b) shows the worst case. The causing reason is that both Mr1 and one of inner layer passgate NMOS (Ms1 or Ms2) are open in half-select write mode. Voltage level of node SS determines the cell stability. In case 1, Fig. 4.9 (a), if SS voltage is logic 0, left stored node might be pull down to 0 through this open path. Similarly, if SS voltage is logic 1, right stored node might be pull up a little through the path, as shown in case 2, Fig. 4.9(b). From Fig. 4.9 (c), noise margin in case 1 is worse than case 2 due to the capability of transferring logic 1 passing serious NMOS is weak. Besides, this action in case 2 is just like 6T bit-cell in read mode. An appropriate voltage level of node SS for half-select cell is very important. What's more, it determines the array structure.

(a)

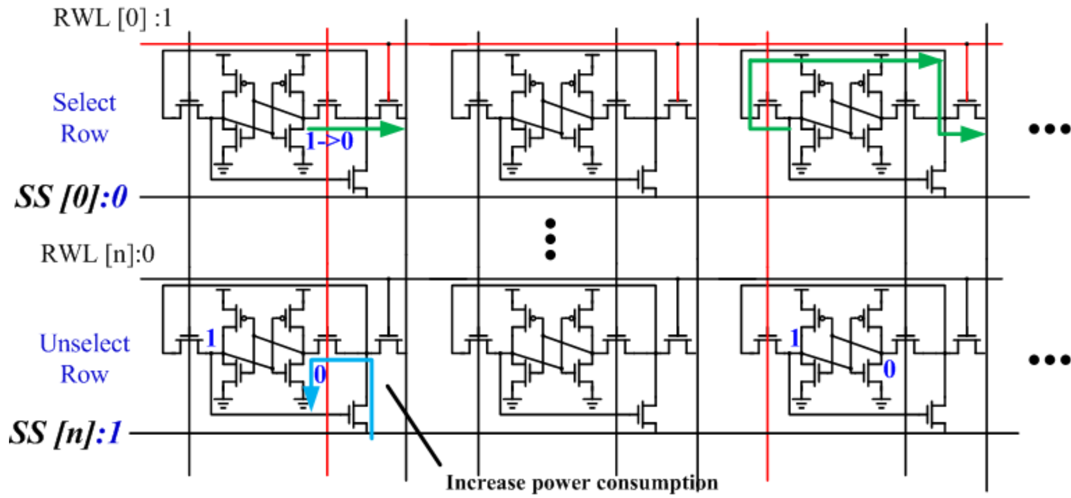
(b)



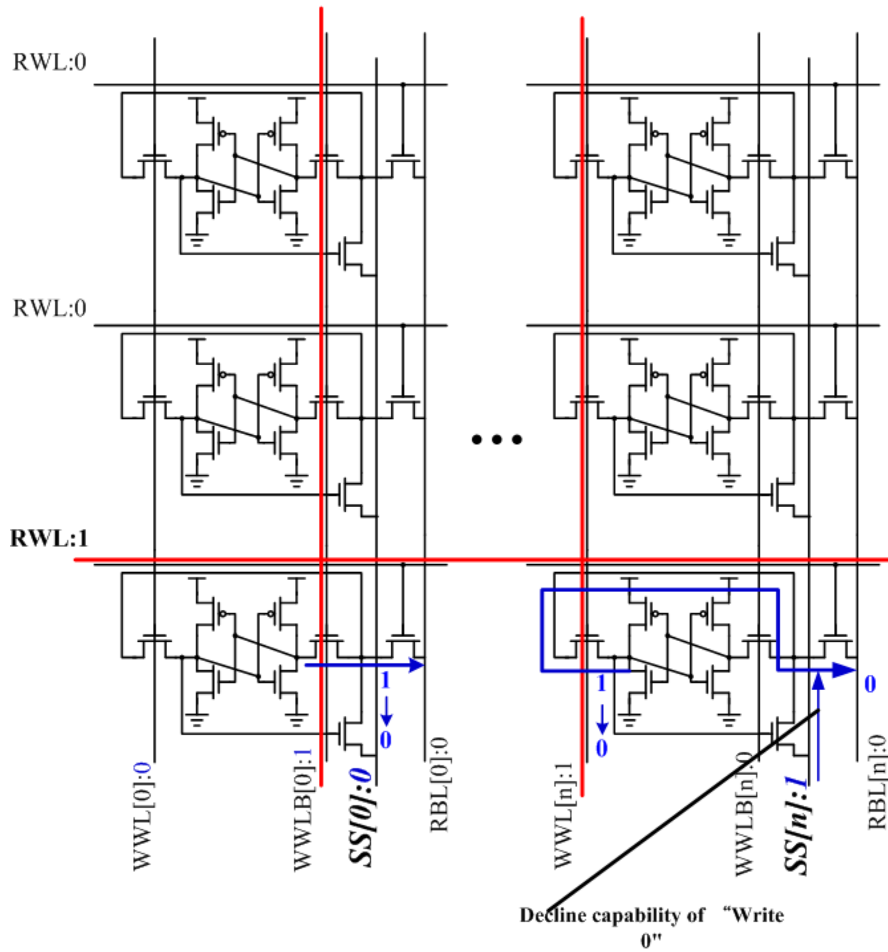
(c)

Fig. 4.9: Worse noise margin in half-select write mode (a) Case 1; (b) Case 2; (c) Noise margin comparison.

Architecture of SS which is the column based control or row based control will affect the array stability, peripheral circuit design and the power consumption. If the SS is row based, control design is not a difficult job. SS of unselected row are biased at logic 1, but SS of the selected row is biased at logic 0. The worse case of half-select mode in bit-cell is case 2. However, if all the half-select cells perform case 1, power consumption will be large.



(a)



(b)

Fig. 4.10: Array architecture (a) VVSS in row based; (b) VVSS in column base.

If SS is column based, as shown in Fig. 4.10 (b), the voltage level of SS and

WWL0 is the same. As a result, the half-select issue can be addressed completely in this architecture. The disadvantage of column based structure is that the control design is more complicated and the capability of write is decreased. In other hands, advantage is that there is no noise margin concern in half-select cell and the power consumption is much less than row based architecture, as shown in Fig. 4.11. From above analysis, the architecture should be the column based obviously.

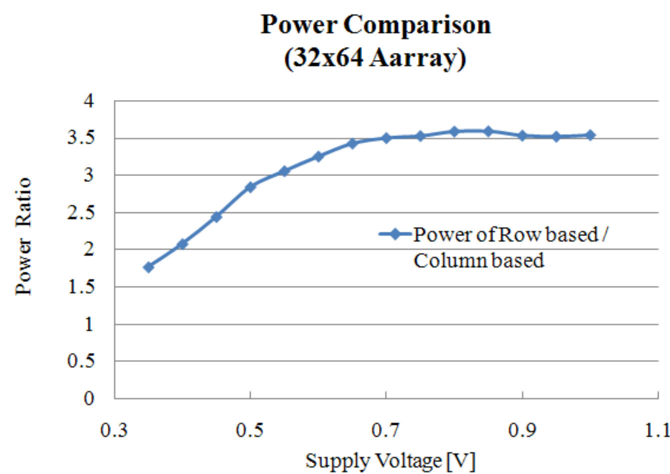


Fig. 4.11: Power consumption comparison between row based architecture and column based architecture.

4.5 Interface circuit with negative voltage write scheme

Column based WWL and SS signal line has large difference from the traditional 6T SRAM structure. However, interface circuit can address this issue and remain the original peripheral circuit. Fig. 4.12 shows the concept of interface circuit. Interface circuit is a bridge between traditional peripheral circuit and the new 8T bit-cell. For the interface circuit, there are 3 inputs such as write enable (WE) and bit-line pair (BL and BLB) and an output which is RBL and connects to sensing amplifier. In other

words, a traditional prototype SRAM without 6T cell array but with 8T cell array and interface circuit can still work. Besides, the performance is closer to original 6T SRAM in high voltage range. In a traditional prototype SRAM design, critical timing of row direction is longer than column direction. Since the interface circuit only increases critical timing of column direction, interface circuit has almost no influence on SRAM performance.

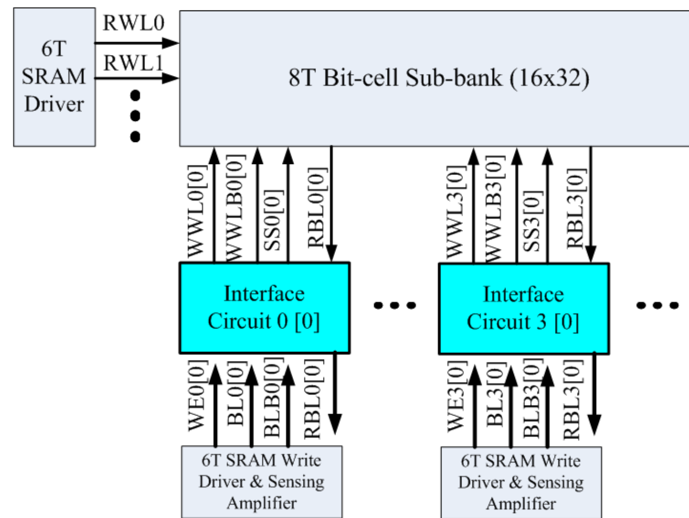


Fig. 4.12: Interface circuit

Table 4.1: True table of interface circuit.

	Precharge /standby	Write 1	Write 0	Read
BL0	High	High	Low	Floating
BLB0	High	Low	High	Floating
WE	Low	High	High	Low
WWL0	Low	Low	High	Low
WWL1	Low	High	Low	Low
RBL	High	Low	Low	Floating
SS	Low	Low	High	Low

Table 4.1 shows the true table of the interface circuit. In write 0, although SS and WWL0 are the same voltage level, SS is needed to pull up before WWL0 pull up for noise margin concern in half-select write mode. Following sub-sections will represent the detailed operation.

4.5.1 Precharge/standby mode

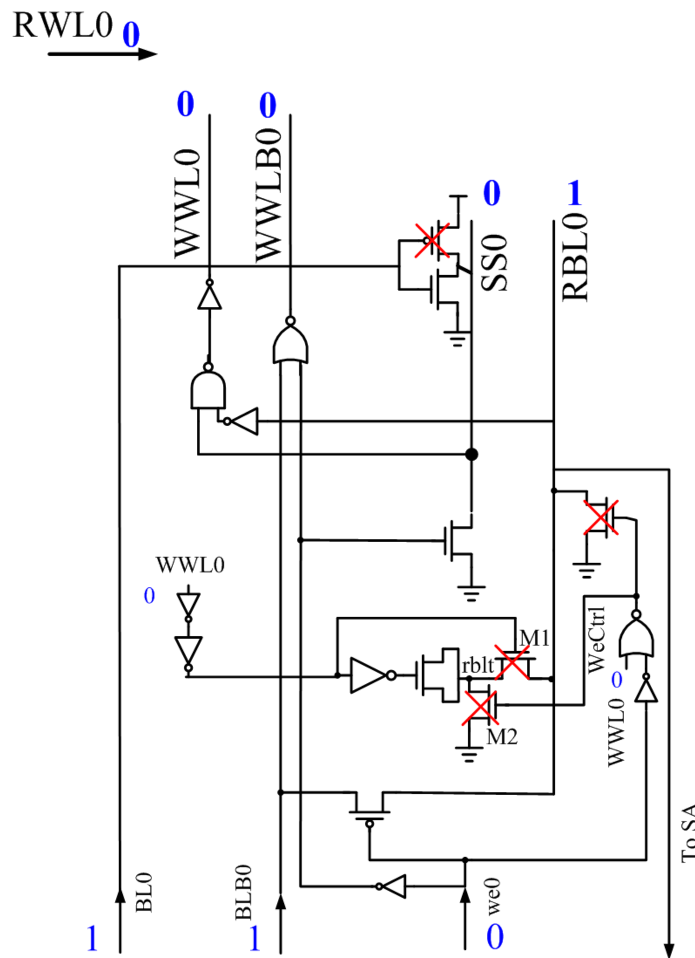


Fig. 4.13: Interface circuit – precharge/stand by mode

In precharge and stand by mode, BL0 and BLB0 are precharged to logic 1 in the prototype SRAM. RBL of interface circuit is also logic 1 by connecting to BLB through PMOS. About other output signals, such as WWL pairs and SS, they are pulled down to 0.

4.5.2 Read mode

Interface circuit can't determine when the operation is read or precharge due to interface only receives the 3 signals. As a result, the action of logic in interface circuit is the same as precharge mode. After BLB and RBL are precharged, BLB and RBL become floating. RBL either stays at high voltage level or discharges through the cell according to the stored data. Besides, RBL connects to sensing amplifier for single sensing.

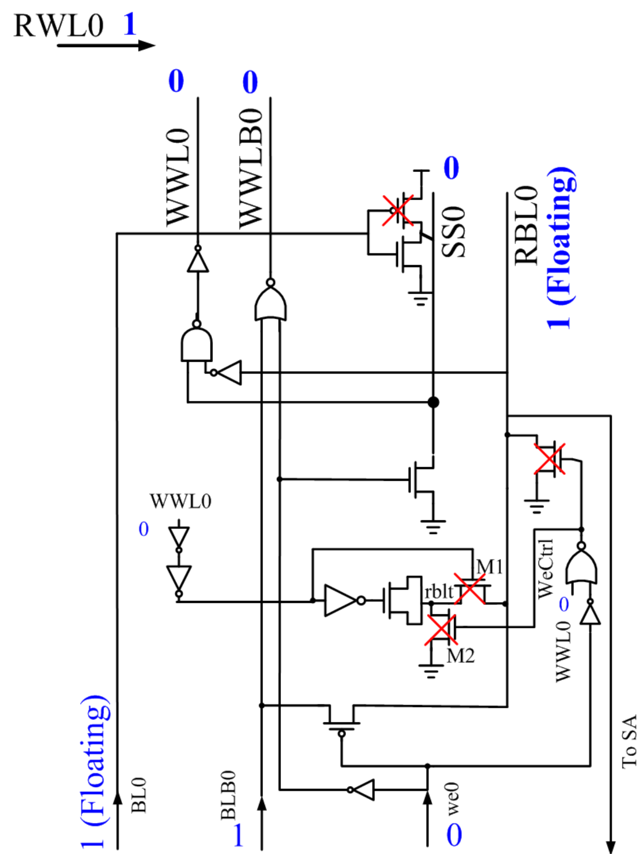


Fig. 4.14: Interface circuit – read mode

4.5.3 Write mode

In write 1, according to the table 4.1, BLB equals to logic 0 and WE equals to logic 1. Interface circuit pulls RBL to logic 0 and turns on WWL1 for writing "1." In write

0, Fig. 4.15, interface circuit pulls the SS to logic 1 and discharges the RBL in the first. After that, WWL0 is pulled up for performing write operation. The main reason is that while WWL pulled up, half-select cell has better noise margin during write. However, write operation might be fail if the worst local variation happen to bit-cell. For example, local variation makes Mr2 is strong and Mr1 is weak, “write 0” might be fail. To address this issue, negative voltage write scheme provide an excellent solution for this proposed 8T bit-cell.

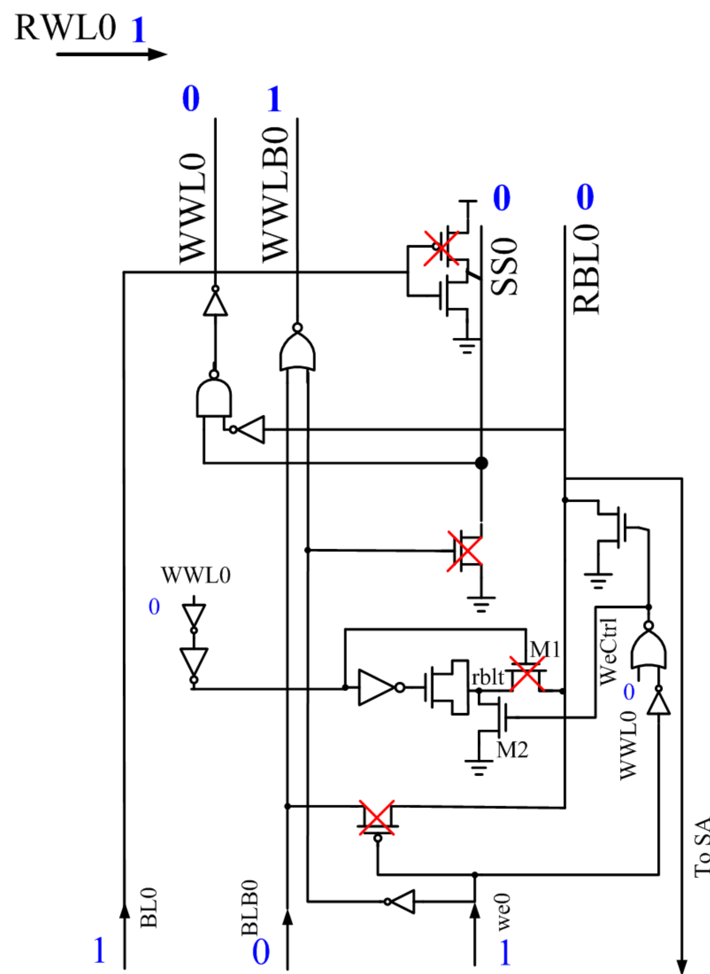
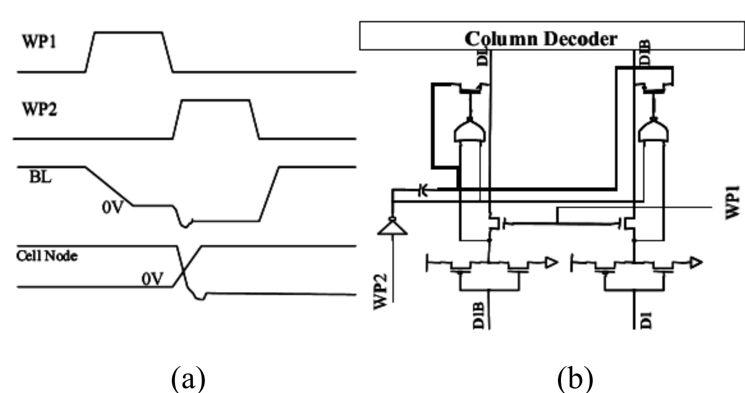


Fig. 4.15: Interface circuit – write “1” mode

In write 0, interface circuit generates negative voltage at RBL to ensure write successful against fail due to serious local variation. From recent papers, [4.3-4] negative voltage write scheme can ensure the write operation success. From Fig. 4.16

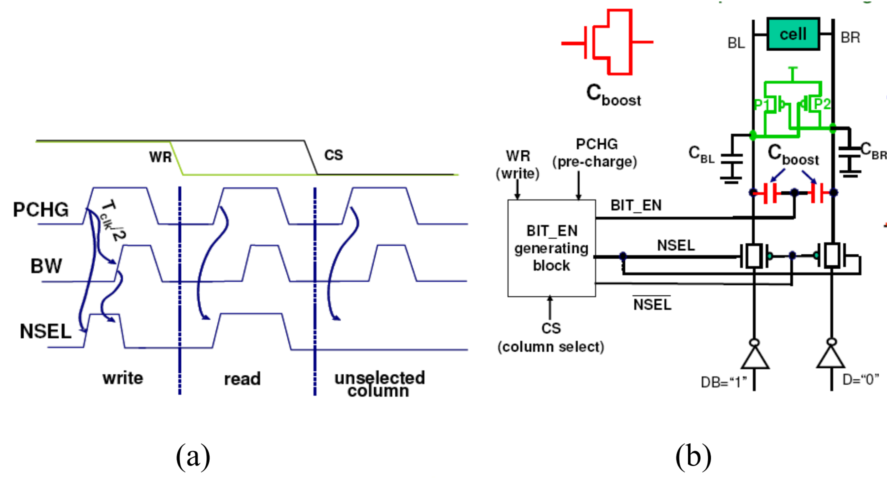
[4.3], write circuit generates the negative voltage on bit-line through two stages. First, bit-line is discharged to 0V during WP1 pulling up. Second, by using inverter and MOS capacitance generate negative voltage at bit-line during WP2 pulling up. Similarly, Fig. 4.17 shows another negative voltage write scheme, which is designed by IBM. By separating write cycle into two cycles, BW and NSEL, control circuit discharges bit-line in first-half cycle and drives C_{boost} generating negative voltage on bit-line in second-half cycle. According to authors of [4.4], C_{boost} won't increase read time a lot. In read cycle, because BIT_EN is pulled up first, the loading caused by C_{boost} on bit-line can be neglected. Consequently, generating negative voltage can be separated into two periods: First, discharging to 0V; Second, generating negative voltage by driving capacitance.

Timing issue of negative voltage generator can directly make use of properties of interface circuit. Fig. 4.18 explains the operation of "write 0." In phase I, interface circuit discharges the RBL to 0V, pulls up SS to high and resets the MOS capacitance before the WWL0 opening. In phase II, after WWL0 opening, interface circuit drives MOS capacitance generating negative voltage and opens the transistor M1 to complete the write operation. Meanwhile, the half-write select cell has good noise margin because SS is pulled up to appropriate level before WWL open.



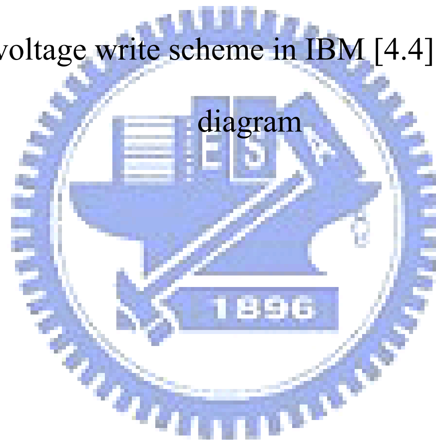
(a) (b)

Fig. 4.16: Negative voltage write scheme in TSMC [4.3]. (a) Waveform. (b) Detailed circuit.

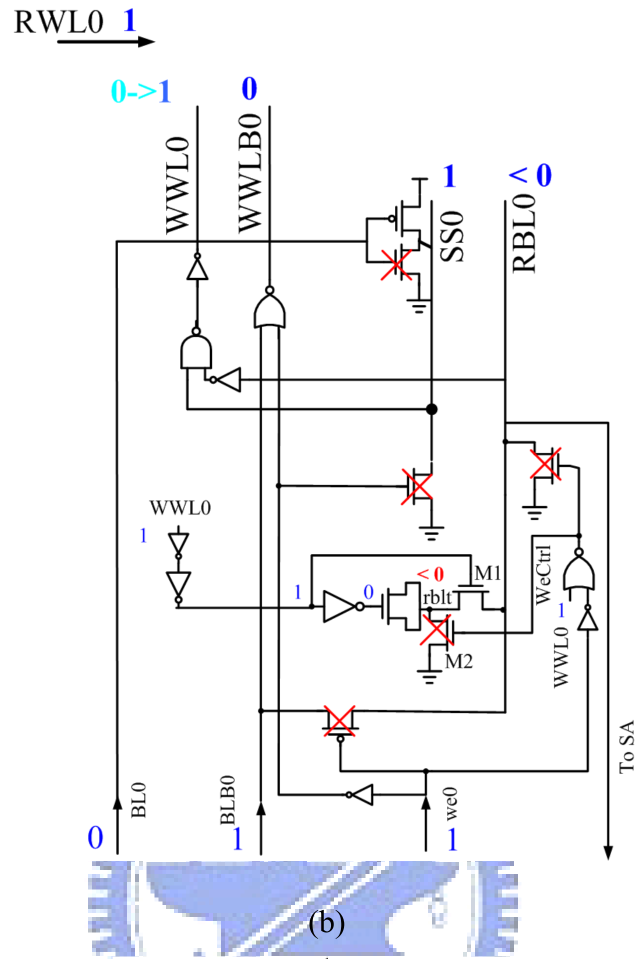


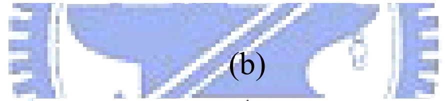
(a) (b)

Fig. 4.17: Negative voltage write scheme in IBM [4.4] (a) Waveform (b) Circuit diagram

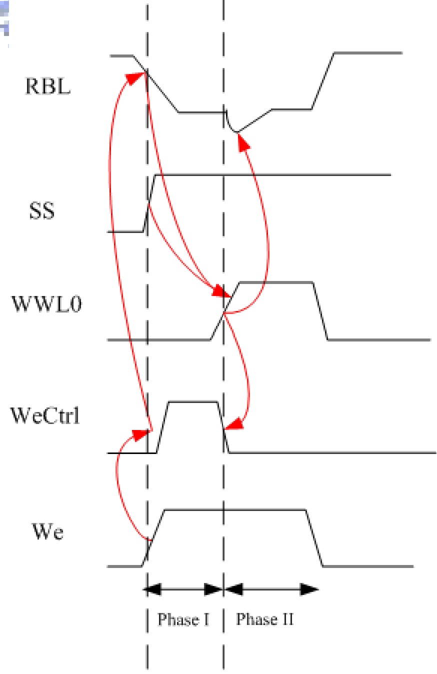


(a)





(b)



(c)

Fig. 4.18: Interface circuit – write “0” mode (a) Phase I (b) Phase II (c) Waveform

4.6 Simulation results

Table 4.2 and 4.3 show the timing comparison between traditional 6T bit-cell (UMC 0.62um^2 6T SRAM cell) and proposed 8T bit-cell. This corner is requested by ***Faraday Technology Corporation***. Read and write time of 8T bit-cell with interface circuit are closer to traditional 6T bit-cell. As a result, proposed 8T array with interface circuit become a probably replacement for traditional 6T bit-cell in future process technology.

Table 4.2: Timing information in SS corner -40°C

Voltage	Read in 6T	Write in 6T	**Write 0 in 8T**	Read in 8T
1	118ps	68ps	**81ps**	126ps
0.9	166ps	88.5ps	**100ps**	158ps
0.8	274ps	129ps	**133ps**	217ps
0.7	597ps	237ps	**203ps**	344ps

Table 4.3: Timing information in SS corner 125°C

Voltage	Read in 6T	Write in 6T	**Write 0 in 8T**	Read in 8T
1	126ps	74ps	**88ps**	132ps
0.9	162ps	92ps	**105ps**	155ps
0.8	229ps	122ps	**131ps**	193ps
0.7	375ps	177ps	**175ps**	261ps

4.7 Summary

This new type 8T-bit cell could be a possible replacement of traditional 6T/8T SRAM at 32/28 nm or below. This cell can achieve high-density design due to only one bit-line in cell. By controlling inner layer and outer passgate and cell structure, read-select disturb and read/write half-select disturb is eliminated. As a result, this cell can work at wide voltage operation region (1V-0.4V). What’s more, the size modification for noise margin concern within each process generation becomes an

easier job. Besides, write word-lines are controlled by the input data during Write operation. In contract with TI' 7T bit-cell, capability and effectiveness are equal in writing "0" and writing “1“. Therefore, robust write margin and performance is performed for low voltage operation. With interface circuit design, it is simple and convenient to apply this new 8T bit-cell into conventional SRAM architecture. Column based architecture of 8T array and interface circuit with negative voltage write scheme make the proposed 8T array have no any read/write disturb and half-select disturb and the similar performance compared to traditional 6T SRAM.

Chapter 5 Sub-threshold Multi-Port Register File

5.1 Introduction

Sub-threshold operation can achieve orders of magnitude low power consumption compared to conventional super-threshold operation. It can be used in applications such as medical devices, portable device, sensor networks and wireless body area network (WBAN) where performance is not constrained.

Register file is a key component of many processors or SoC applications. Not only its access time dominates the application speed but also its area and power occupies the most part of chip in high performance processor design. In order to achieve sufficient bandwidth, designers increase the port number on bit-cell in conventional register file design [5.1], such as multi-port SRAM-based register file. However, such approaches make the cell have larger area, worse noise margin, longer access time and limited operation voltage. To address these issues, many techniques were investigated to reduce the port number [5.2]. In this chapter, a low power multibank architecture for simultaneous access with collision detecting technique is proposed. The port number of cell can also be reduced. The architecture has been analyzed under wide operating voltage range between 1V to 0.25V. The proposed register file can be applied to the Superscalar architecture or VLIW (Very Long Instruction Word) DSP.

The rest of this paper is organized as follows: An overview of recent sub-threshold

register file is shown in section 5.2. Section 5.3 describes the architecture of low power multi-port register file. Section 5.4 presents circuit techniques to improve register file performance under wide voltage operating. Section 5.5 shows the layout and floor plan of proposed register file. Section 5.6 shows simulation results, and summary are given in Section 5.7.

5.2 Overview of sub-threshold register file design and banked register file architecture

5.2.1 Sub-threshold register file design

It is common using registers to store template data in chip. The system might request the multiple inputs or outputs to access data which resulting in multi-port design. If the chip is successful to work in sub-threshold region, the logic and latch have to face some problems. First, in logic design, the current drive strength in sub-threshold is so weak. Unfortunately, the leakage current doesn't scale down with supply voltage in the same region. The speed of logic is dominant by weak drive current and leakage current. If there is transistor stacked in the design, the speed becomes slower due to the poor drive current strength. Variation influence is quite large in sub-threshold region compared to high voltage region. The FNSP and SNFP corner, or local variation, make the logic path have different transmission time results in glitch or wrong signals.

Second, small ratio of Ion/Ioff not only makes the logic slow but also influences the fan-in/fan-out in SRAM array design. It determines how many bit-cells can connect to

one bit-line. If there are many than one port in a bit-cell, the large capacitance make longer access time. More importantly, the conventional 6T cell almost has no SNM in the ultra low voltage. It is necessary to design a new cell structure to replace the conventional 6T. Sensing amplifier is almost the same. In high voltage region, we don't need to discharge the bit-line rail to rail. In contrast, bit-line needs to be pull down to 0 to make the sensing amplifier work well. Besides, conventional latch type sensing amplifier faces the serious process variation influence. The above reasons make sub-threshold register difficult to design.

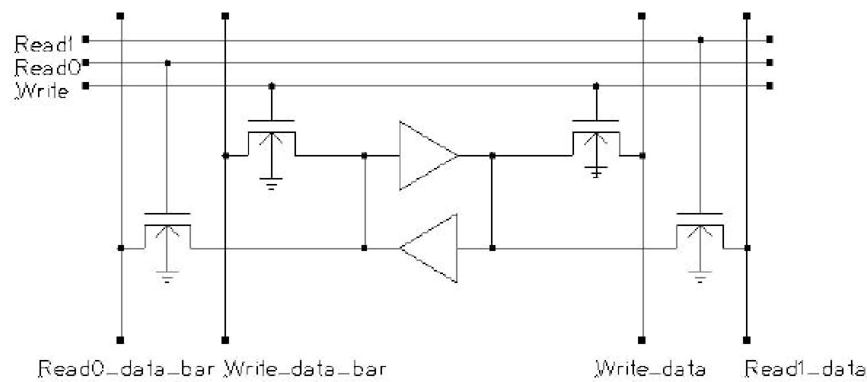


Fig. 5.1: Conventional multi-port SRAM based register file bit-cell [5.13]

Fig. 5.1 shows a conventional multi-port SRAM based register file bit-cell [5.13]. The peripheral circuits are easy to implement in high voltage operation. However, in sub-threshold region, the structure will damage the speed and stability in low voltage operation.

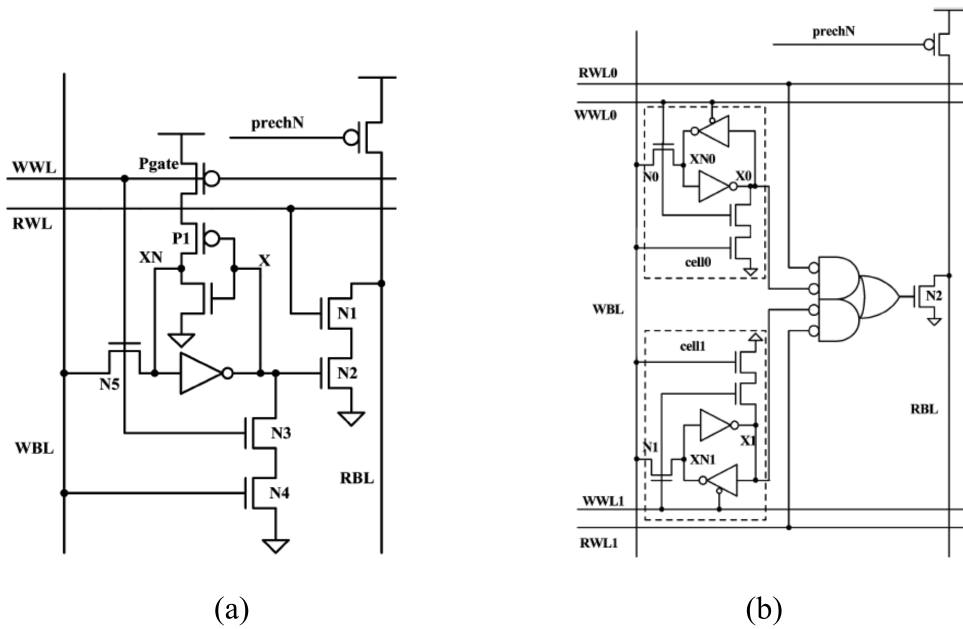


Fig. 5.2: Another structure of RF bit-cell in sub-threshold region. [5.14-15] (a) Cell number on one bit-line is small (b) Two bit-cells share one bit-line

The above left figure [5.14-15] shows another RF bit-cell which can work in sub-threshold region, 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 above right cell. It provides a solution to reduce the capacitance. However, the speed will degrade and cause large area in array.

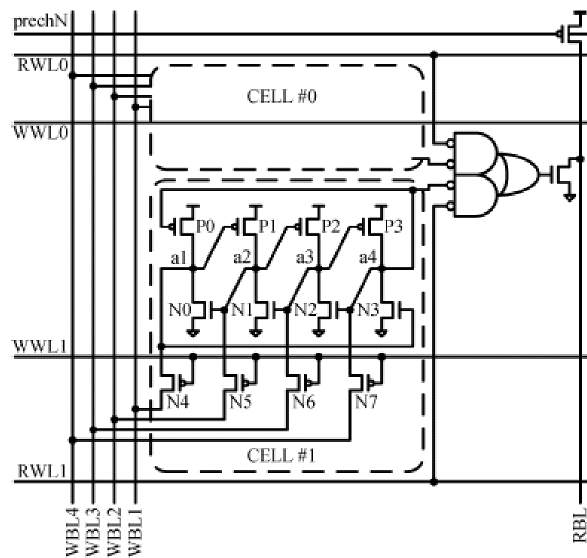


Fig. 5.3: A RF bit-cell with exceptional stability can be applied to special application and this RF bit-cell also can work in sub-threshold region. [5.15]

A likely design in Fig. 5.3 also uses the same combinational circuit to reduce the loading on RBL [5.15]. This bit-cell with large number of transistors also can work in sub-threshold region and can against the serious radiation damage for special application.

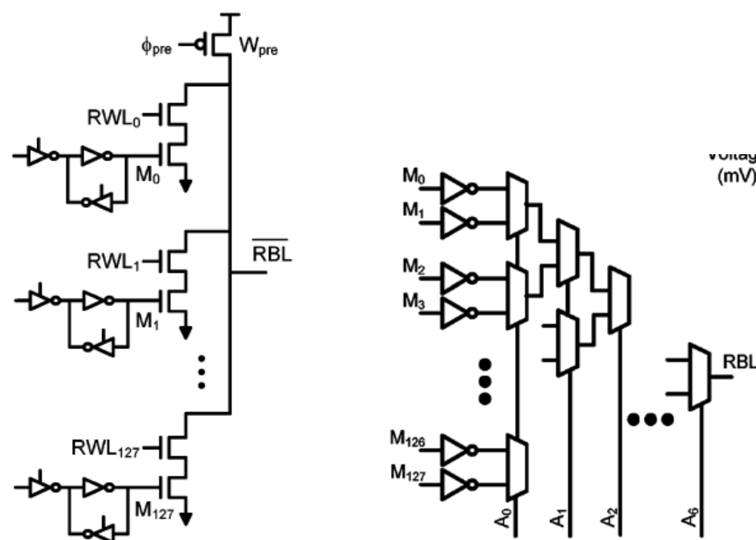
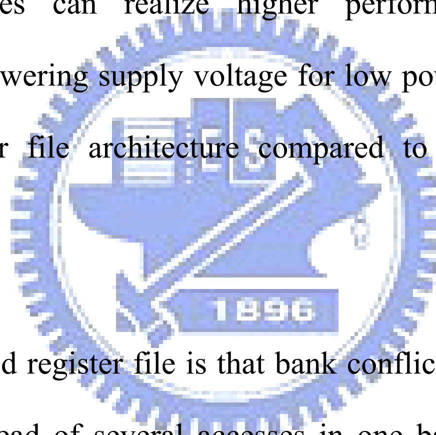


Fig. 5.4: Mux-based selected logic is the robust design in sub-threshold region.[5.7]

In [5.7], authors pointed out that static CMOS circuit is the better solution in sub-threshold circuit design. Using large mux-type circuits to replace the pass-gate on bit-line lets the leakage have the smallest influence on bit-line. As a result, the supply voltage in the circuit can scale down to 0.18V.

5.2.2 Banked register file architecture

Multi-banked register files achieve multi-port access by using less port bit-cells instead of multi-port bit-cell. The drawbacks of multi-port bit-cell are shown in above section. In advanced process technology, reduce wiring quantity in bit-cell, multi-bank register files can realize higher performance and lower power consumption, besides, lowering supply voltage for low power issue is more available in multi-banked register file architecture compared to conventional register file architecture.



The problem of banked register file is that bank conflict happens. One bank might only do one access instead of several accesses in one bank. System is not able to access several data in the same bank at once. In [5.16-17], authors point out several methods to address this issue. In [5.16], the scheme handles bank conflicts by scheduling groups of instructions without conflicts. This design increase significant logic into the critical wakeup-select loop. In other words, single-ported banks is evaluated, however, complex issue logic and functional unit datapaths to allow instructions where both operands originate from the same bank to be issued across two successive bank read cycles [5.16]. This key idea of design in [5.16] is to reduce complexity of register file but improve the whole system clock speed. It is dependent on chip application. The concept of bank conflict issue addressed by issue logic also

describes in [5.18-19]. Methods, such as renaming, reservation station implement, out-of-order scheme and combine access, to decrease access conflicts by system for a deeply pipelined dynamically scheduled processor present in [5.20-23].

5.3 Architecture of proposed low power multi-port register file

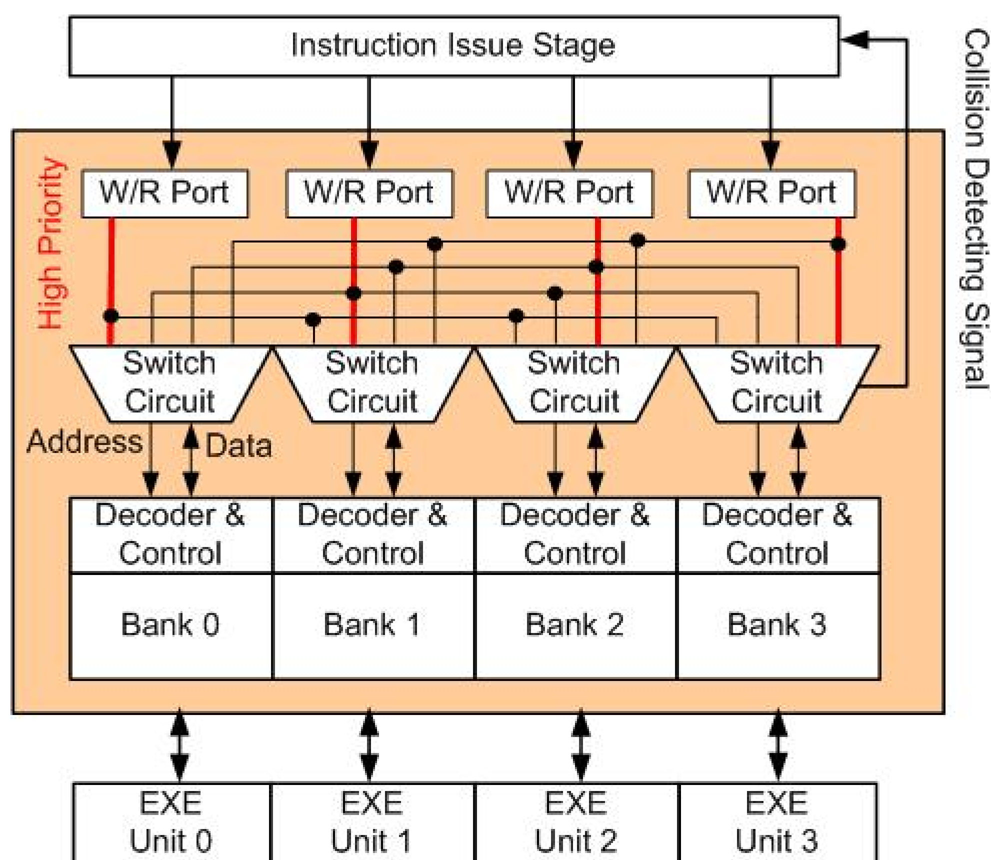


Fig. 5.5: Proposed register file in VLIW DSP

The proposed architecture of register file has 4 Read ports and 4 Write ports, and there are 4 banks, each bank capacity has 4KB with bit-interleaving design, as shown in Fig. 5.5. Similar architecture of VLIW DSP is shown in Fig. 5.6 [2.34]. Instruction issue stage can use simple control signal and address to determine each execution unit

to access the particular storage bank simultaneously. The main function of the switch circuit is to authorize which execution unit can access the corresponding bank. If address collision happens, switch circuit still correctly selects an access with high priority and issue the collision signal back to instruction issue stage. In this architecture, each execution unit has a higher priority to access the corresponding bank. For example, execute unit 0 has a higher priority to access the bank 0. Consequently, each bank can simultaneously perform write and read operation for the same or the different execution units dependent on application. In other words, the register file can support four different applications performing the access or support one program for the simultaneous multi-access like VLIW DSP.

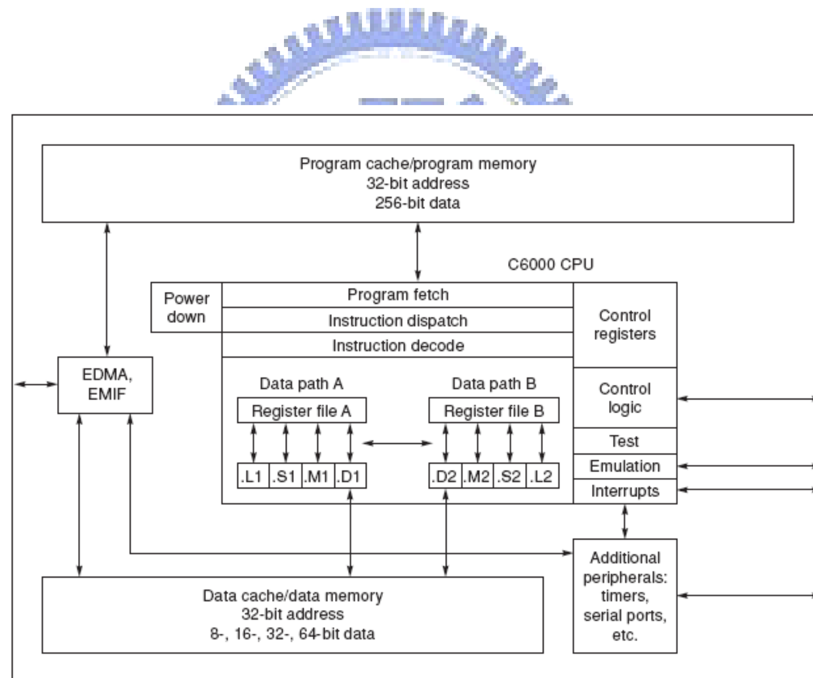


Fig. 5.6: Architecture of the TMS320C64x family of DSPs. The C6x is an eight-issue traditional VLIW processor. (Courtesy Texas Instruments.) [2.34]

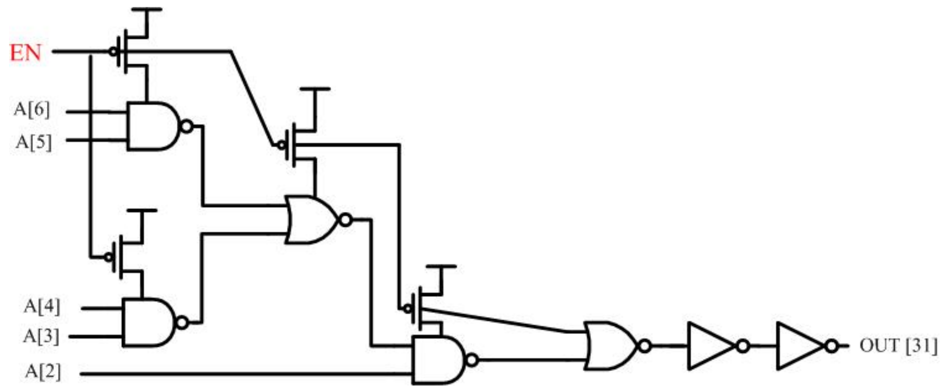


Fig. 5.7: Sub-decoder with power gating

Switch circuit addresses the collision and ensures exact access with high priority in this multibank structure. As a result, it is not necessary to add more ports on bit-cell which increases area and power consumption of register file. In addition, local decoder is turned off while the switch circuit detects no access in this bank, as shown in Fig. 5.7. While "EN" is logic-1, these logic gates in sub-decoder are turned off in this cycle. Decoder is designed in static CMOS logic instead of dynamic circuit or passgate logic. There are several reasons: First, static CMOS logic is the most robust design in sub-threshold region. Second, transmission gate and dynamic gate do not work in sub-threshold region due to the large leakage current.

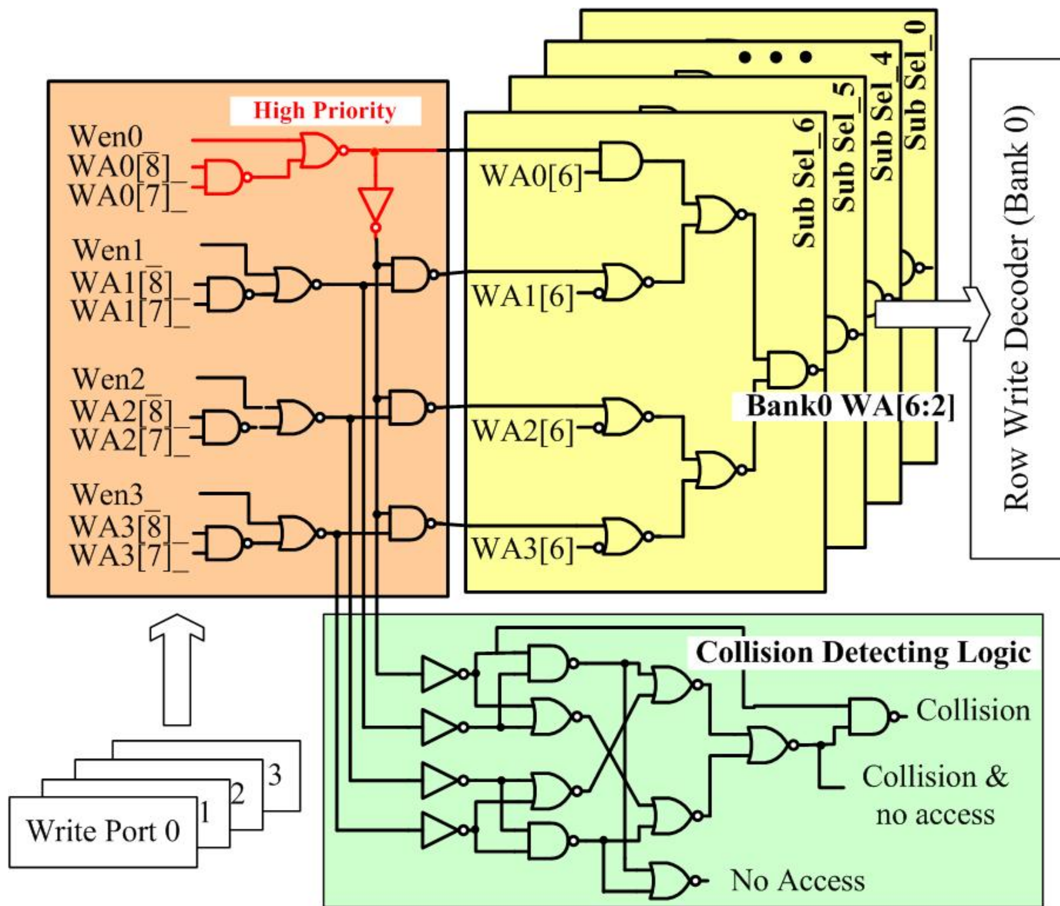


Fig. 5.8: Switch circuit with collision detecting logic.

Fig. 5.8 shows the circuit diagram of write switch circuit with collision detecting logic. Switch circuit in each bank receives write data and its write address from 4 write ports. By using write enable (Wen) signal and priority logic, switch circuit selects the exact address or data for corresponding bank decoder and write driver. Collision detecting logic can represent several statements in this bank. "No Access" represents there is no access request in this bank, and control circuit will turn off the decoder and relative circuits to reduce active power consumption. "Collision & no access" means that address conflict happens and there is no access required in high priority port. Control circuit turns off the decoder and transfers the collision signal back to instruction issue stage. On the contrary, if there is an access with higher priority when conflict happens, "Collision" is pulled up and the bank still performs

the access operation and transfers a collision signal back to instruction issue stage. Besides, there is another switch circuit to deal “bank conflict” of read in the same bank. Read from the same address from different ports in the same clock cycle still achieves by adding other logic in the switch circuit.

5.4 Design concept

In this section, several design techniques which make the register file work in sub-threshold voltage are represented. These techniques include bit-cell design, architecture, write scheme, read scheme and output latch. Also, analysis of these proposed circuit techniques also shown in each sub-section.

5.4.1 Dual Vt 8T bit- cell with high-Vt passgate

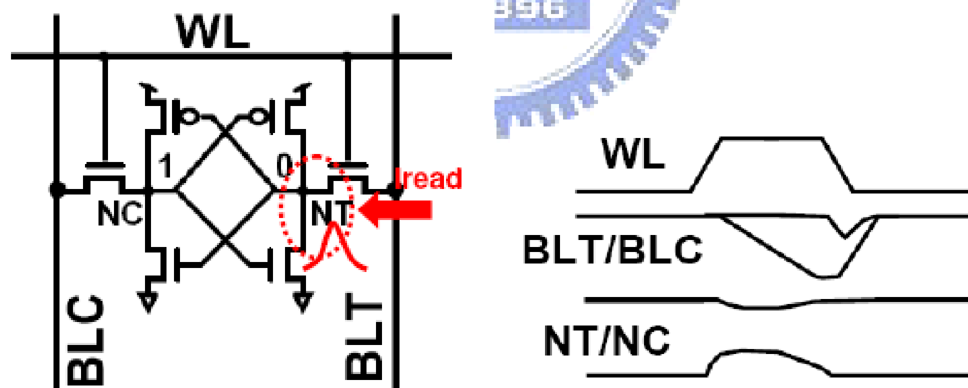
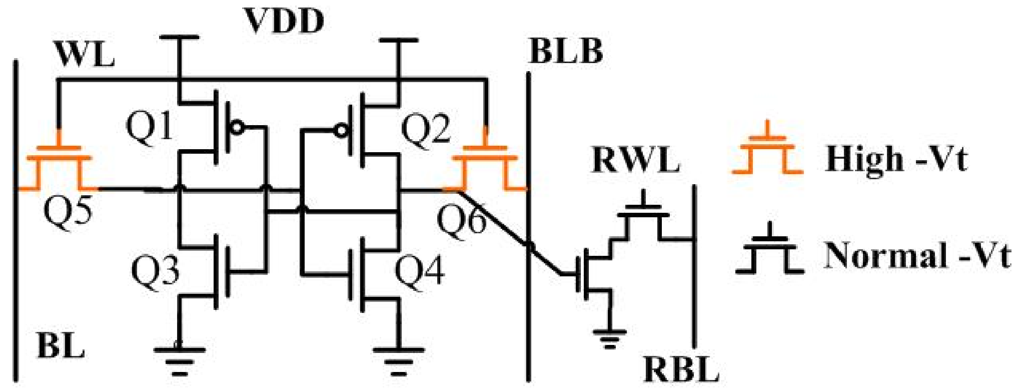
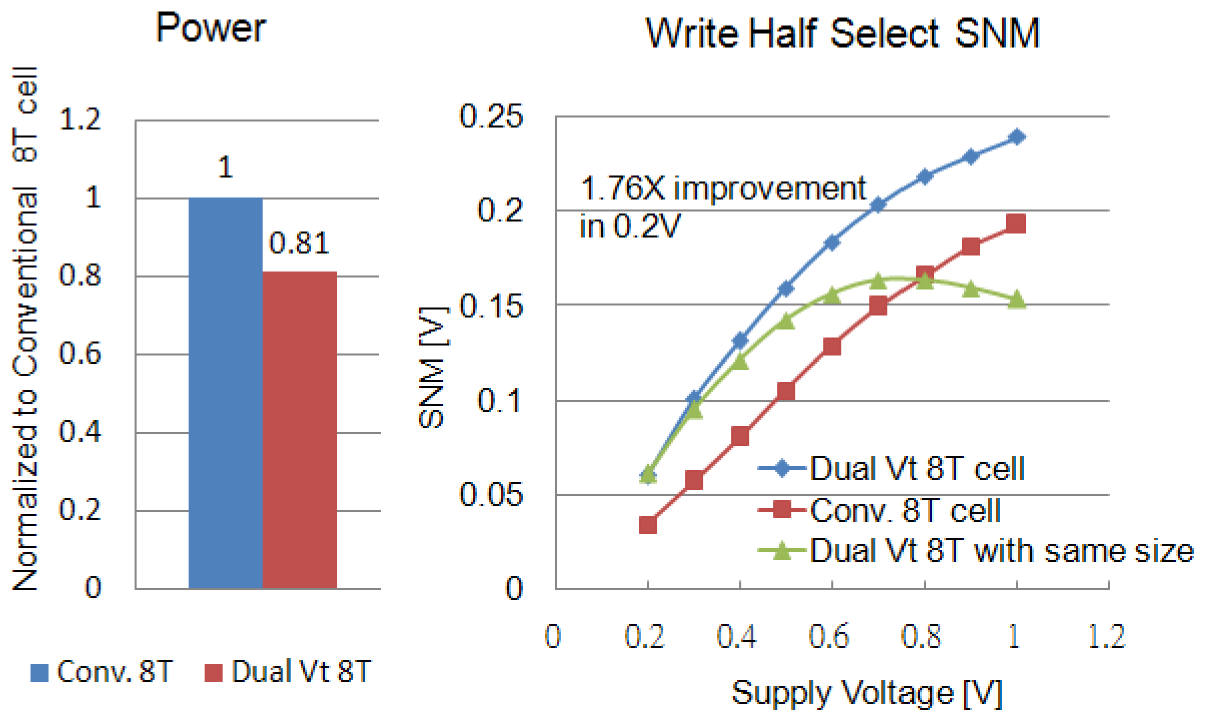


Fig.5.9: Read disturb in traditional 6T bit-cell.



(a)



(b)

Fig. 5.10: Properties of dual Vt 8T bit-cell (a) Power consumption in Write half select mode; (b) Write half select noise margin of dual Vt 8T bit-cell compared with conv. 8T bit-cell.

Conventional 6T bit-cell only performs at high operating voltage. In read mode, 6T cell discharge the bit-line charge from VDD to GND through the passgate, as shown in Fig. 5.9. If the voltage in NT voltage is higher than trip-point voltage of inverter,

the stored data is flipped out.

Fig. 5.10 (a) shows the circuit diagram of dual Vt 8T bit-cell which is adopted in proposed register file. The extra read port make read noise margin much better than conventional 6T bit-cell [5.12]. However, in bit-interleaving architecture, conventional 8T bit-cell has the same noise margin compared to conventional 6T bit-cell in write half select mode [5.3]. This issue let the conventional 8T bit-cell work fail in low voltage or below. Dual Vt 8T bit-cell with high-Vt passgate can increase noise margin which is very important for sub-threshold region operation and decrease the power in half-write select mode. By using high Vt transistor in passgate, write half-select margin improves 1.3X in 1V and 1.8X in 0.2V. Fig. 5.11 shows the Monte Carlo simulation result of write half-select noise margin, and the result is expected.

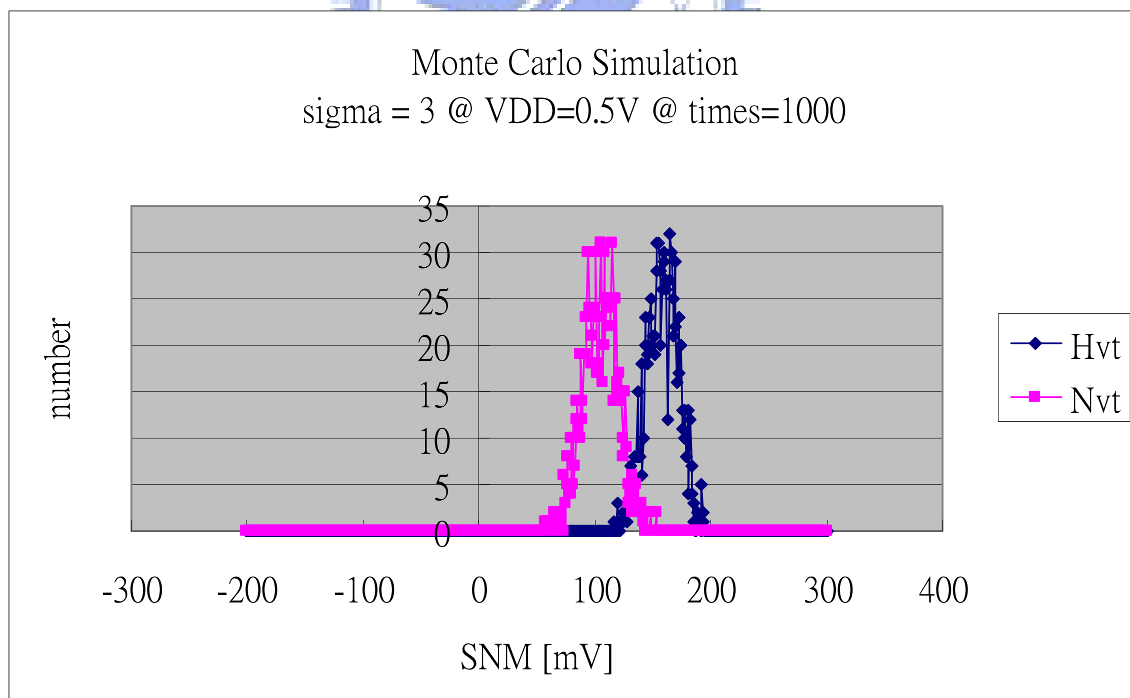


Fig. 5.11: Monte Carlo simulation, sigma=3, VDD=0.5V, times=1000

If this system only operates at low voltage, such as 0.4V to 0.6V, bit-cell designed

with the same size in all gate also performs suitable noise margin, as shown in Fig. 5.8 (b), since the size issue has a little influence on transistor current driving. However, this sub-threshold bit-cell area is larger than conventional 8T bit-cell which only works in high voltage region. The reason is that capability of NMOS and PMOS driving current in sub-threshold region is different from high voltage region and the variations in Vt caused by random dopant fluctuations (RDF) [5.35]. As a result, this dual Vt 8T bit-cell is almost 2X larger than that of a standard 8T bit-cell, [4.1], to work successfully in sub-threshold region across various corner. Similarly, the sub-threshold 6T bit-cell of [5.35] is also ~2X larger than standard 6T bit-cell.

5.4.2 Negative voltage write scheme

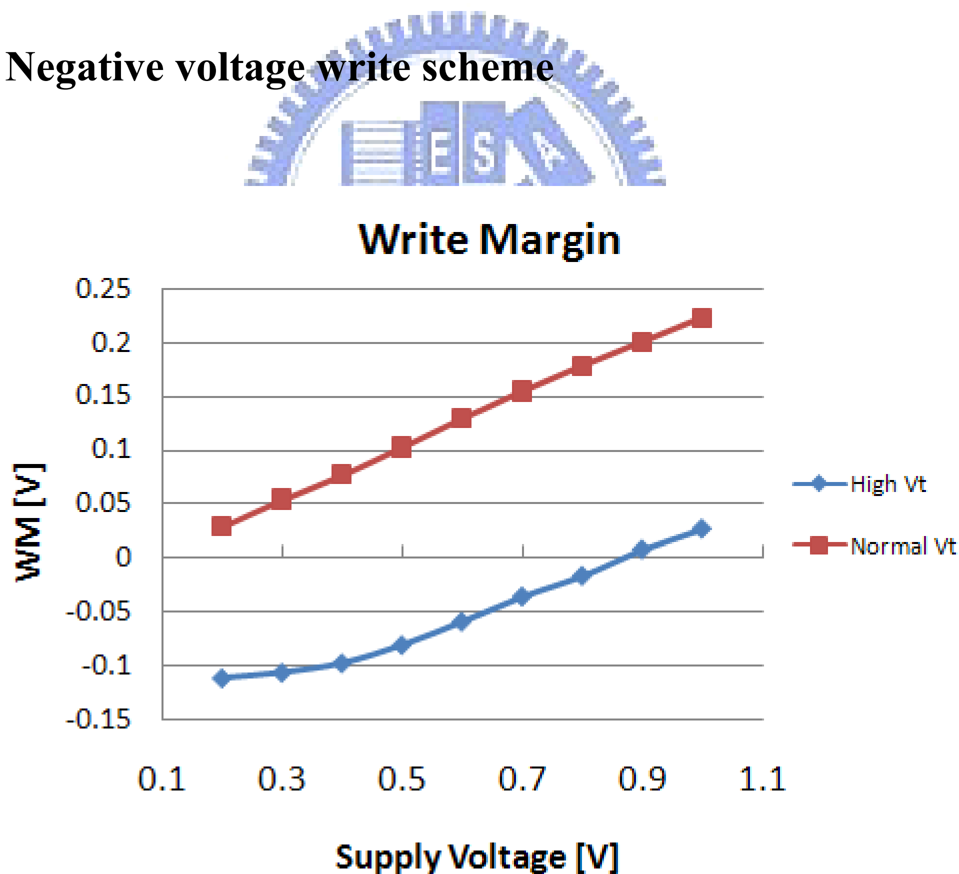
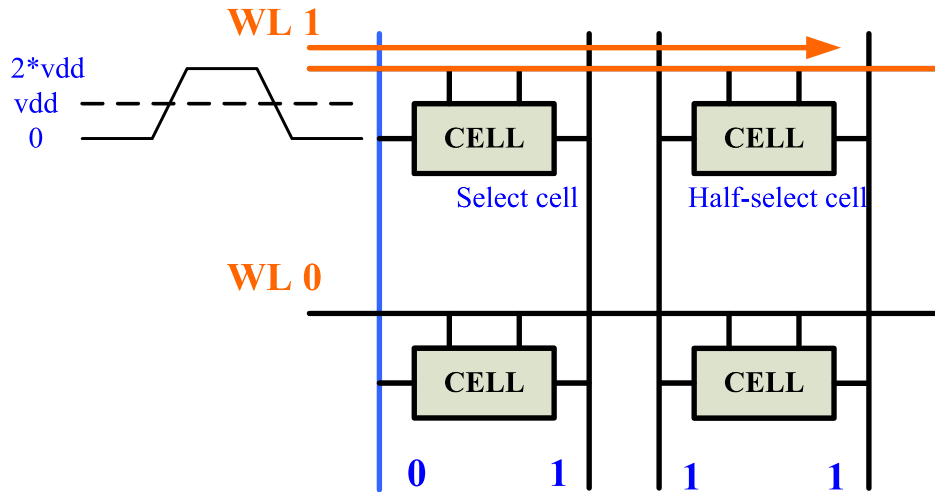
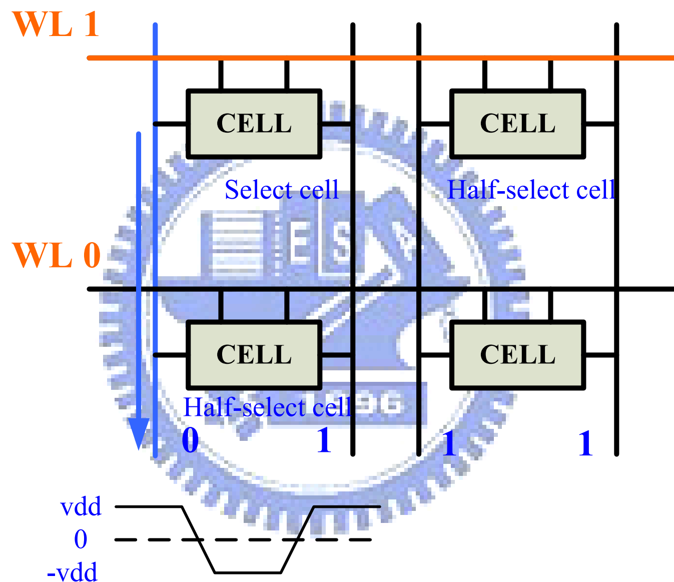


Fig. 5.12: Write margin of 8T cell with Hvt passgate and normal Vt passgate.



(a)



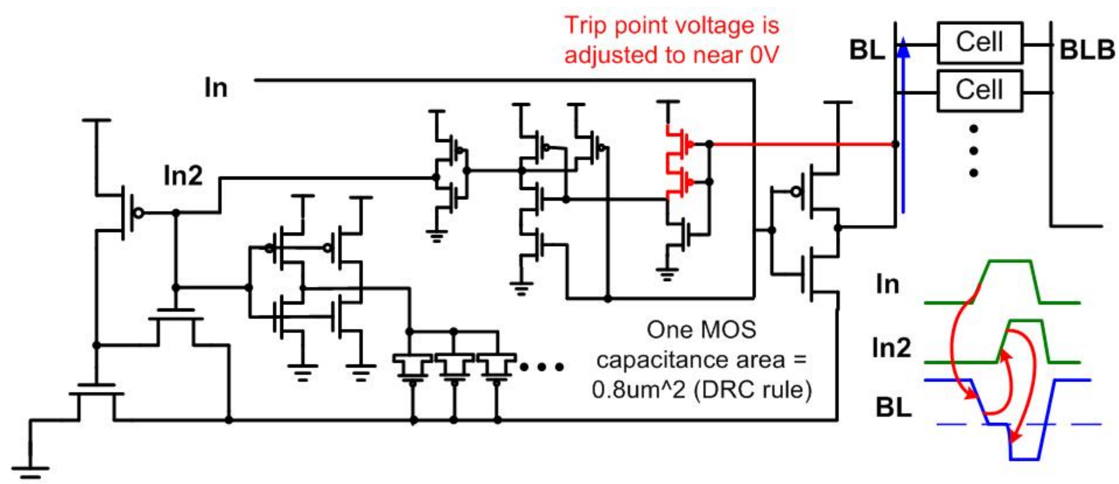
(b)

Fig. 5.13: Write scheme comparison. (a) Boosting WL voltage (b) Negative bit-line voltage

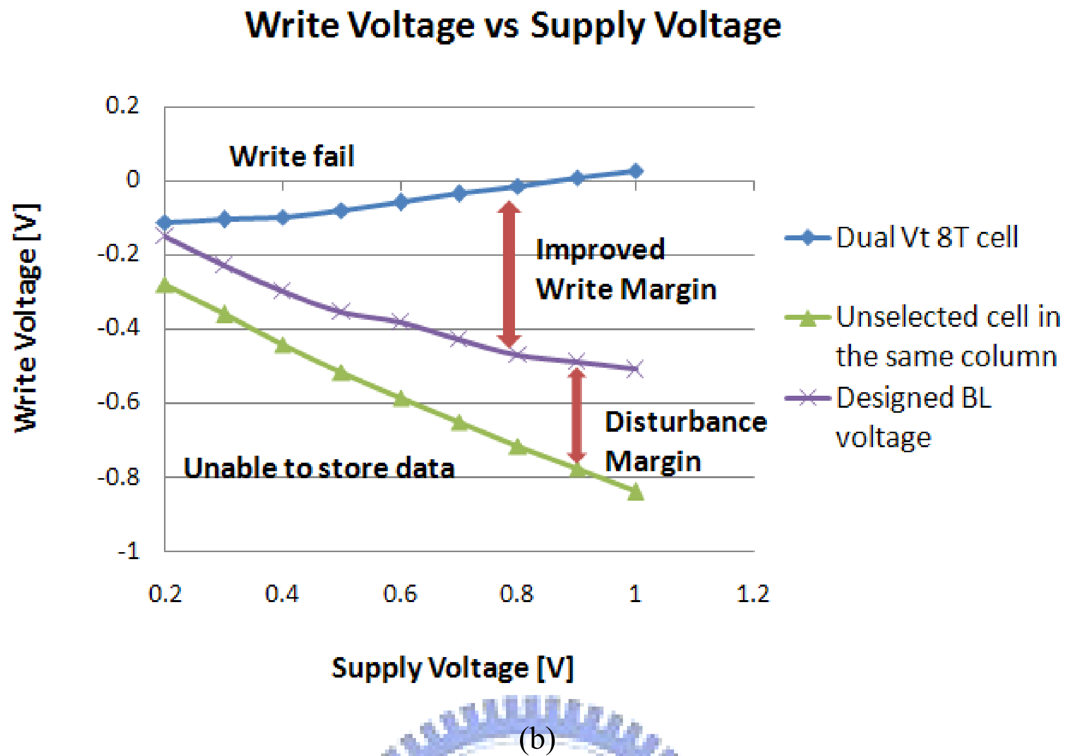
Dual Vt 8T bit-cell improves the noise margin but decrease write margin, as shown in Fig.5.12. Here, the definition of write margin is that the maximum bit-line voltage can destroy the data in selected bit-cell. In conventional write scheme, by discharging bit-line voltage near zero and turning on the passgate of bit-cell, the bit-cell can be written into new data. In DC analysis, 8T bit-cell with high-Vt passgate needs

negative bit-line voltage to perform a successful write operation. Boosting word-line voltage or using negative voltage on bit-line can address the write issue, as shown in Fig.5.13. However, the write scheme in Fig. 5.13 (a) makes half-selected cell have worse noise margin due to the efficiency of high-Vt passgate is eliminate by boosting word-line voltage in selected row. Therefore, negative voltage write strategy in this register file replaces conventional write scheme.

Fig. 5.14 (a) shows the write scheme and the proposed negative voltage generator. In addition, the negative voltage generator with MOS capacitance is on-chip design for reducing influence of process variation and the chip cost. However, capacity of MOS capacitance degrades in sub-threshold region especially in SS corner, as shown in Fig. 5.15 (a). That means more area required in MOS capacitance for successful operation while the supply voltage scaled down. In order to decrease the area overhead, a negative voltage generator with local BL sensing logic is proposed. Fig. 5.14 (b) shows the improved write margin and the appropriate disturbance margin which won't flip out the data of unselected cell in the same column.



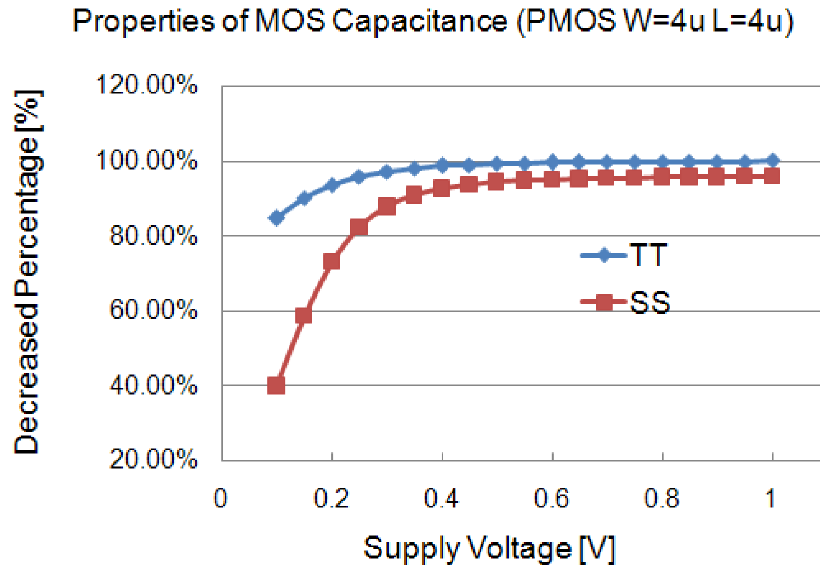
(a)



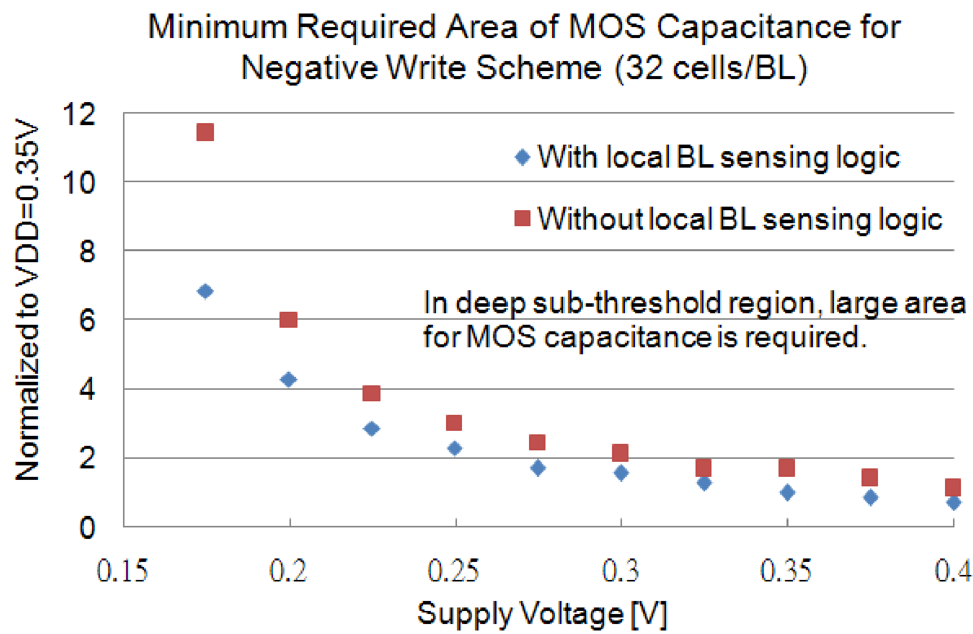
(b)

Fig. 5.14: Negative voltage generator with local BL sensing logic.(a) Circuit diagram; (b) Bit-line write voltage vs. VDD.

When a write operation begins, negative voltage generator pulls bit-line voltage down. The device generates the negative voltage for bit-line after the sensing device detects voltage zero on bit-line. This similar work also reported in [5.4-6]. However, timing of generating negative voltage in [5.4-6] is not a suitable design for sub-threshold operating. Required time of discharging bit-line fluctuates a lot in deep sub-threshold region due to the significant influence caused by PVT variation and bit-line leakage with data dependence. By using the local BL sensing logic, required area of negative voltage generator for successful operation in sub-threshold region can be reduced significantly, as shown in Fig. 5.10 (b).



(a)



(b)

Fig. 5.15: (a) Capacity of MOS capacitance degraded in sub-threshold region (b) Write scheme with local BL sensing logic reduces required area of MOS capacitance in deep sub-threshold region with PVT variation.

5.4.3 Improved read buffer footer, controllable pre-charge scheme and read replica circuit

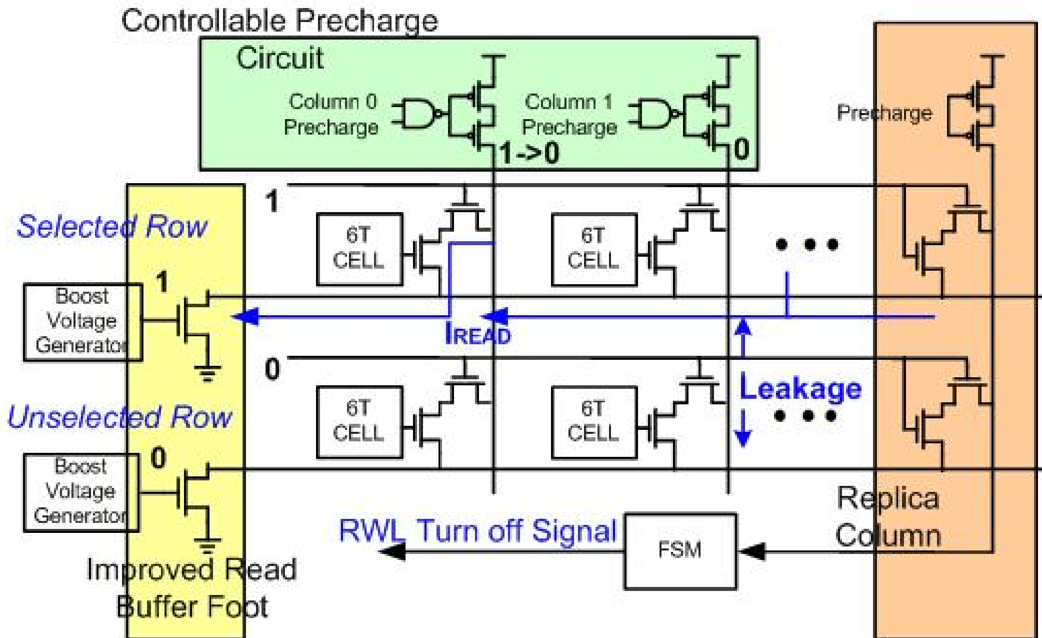


Fig. 5.16: Read architecture and I_{READ} tracing replica circuit

In contrast with super-threshold region, leakage is an important issue in sub-threshold region. The ratio of I_{ON} and I_{OFF} declines from 10e+5 to less than 100 [5.7]. As process scaled down, the large gate and junction leakage degrade circuit design. Unfortunately, this impact also makes 8T bit-cell work fail while performing read operation under ultra low voltage. The charge of read bit-line is discharged by the read port of other unselect 8T cells. Many researches provide new 10T or 11T bit-cell to address leakage issue [5.8-10]. However, the new bit-cell increases area, power consumption and cost of chip. One of methods to address read fail of conventional 8T bit-cell can be found in [5.11]. Adding a read buffer footer for eliminating leakage path makes 8T bit-cell perform read operation successfully. In original design, the read buffer footer is driven by boosting voltage instead of increase transistor size due to the cost consideration and sizing transistor has little help for current driving in sub-threshold region. Besides, capability of driving current of read

buffer footer determines the read access time.

However, this scheme limits the performance especially in high voltage region. Although the proposed wide operating voltage range register file is mainly applied to low power product which usually operating under low voltage to sub-threshold voltage region, it is still necessary to improve the performance at high voltage operating since system might increases supply voltage for requiring higher throughput in a short time. Fig. 5.16 shows the read architecture of the proposed register file. Only using NMOS in read buffer footer is enough due to the stack effect almost eliminate the leakage. Read buffer footers without PMOS can provide shorter read access time with some pattern due to eliminating leakage toward the I_{READ}, as shown in Fig. 5.17. Longer read access time increases the leakage power consumption. Relative simulation in Fig. 5.18 shows the comparison in varies voltage and bit-cell number/bit-line. A large improvement provided by improved read buffer footer presents except 0.3V-0.25V because capacity of MOS capacitance in deep sub-threshold region degrades, and voltage is not the key influence on current driving at high voltage region.

The read replica circuit traces the selected row's I_{READ} and leakage to generate the most appropriate width of RWL pulse since I_{READ} determines read access time. The quantity of I_{READ} and leakage is dependent on stored data, variation and operation in the corresponding row. Appropriate RWL pulse width each time is important for sub-threshold operating since the fluctuation of RWL pulse width from simulation is more than 30% while reading different column but same row.

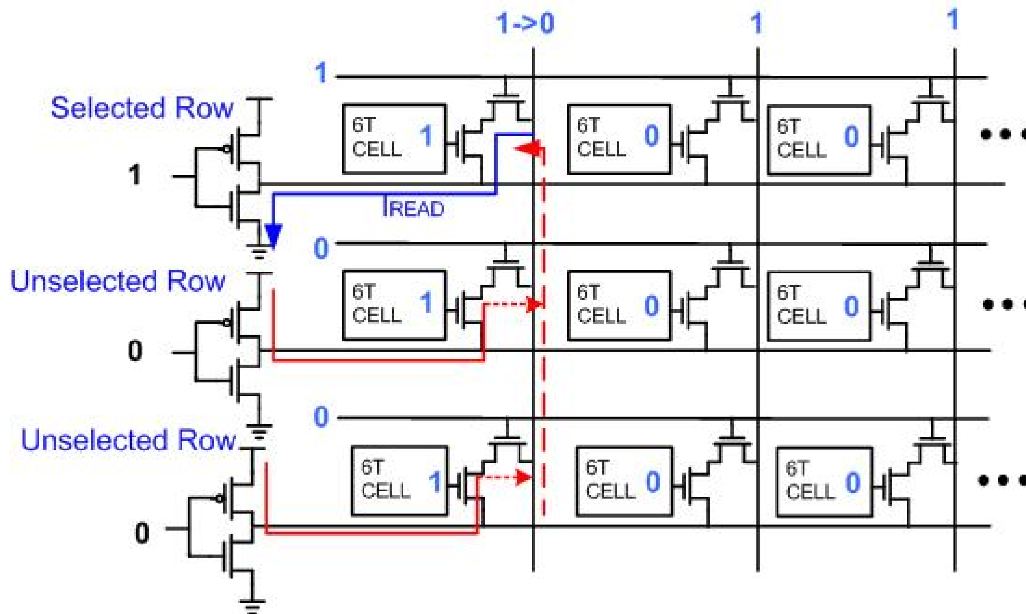
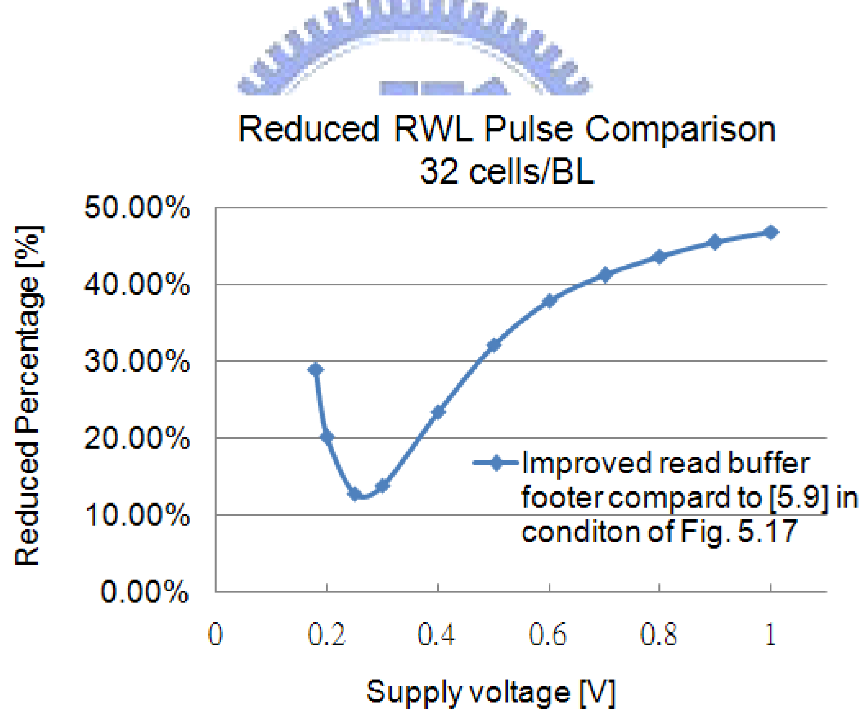
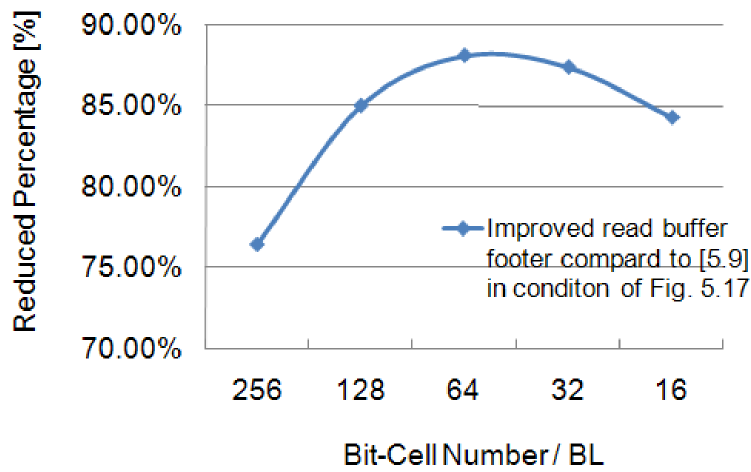


Fig. 5.17: PMOS of read buffer footer in [5.9] generates leakage to increase access time in worst pattern.



(a)



(b)

Fig. 5.18: Read buffer footer with PMOS increases array read access time in Fig. 5.17. (a) 0.18V – 1V (b) 16-256 cells/BL

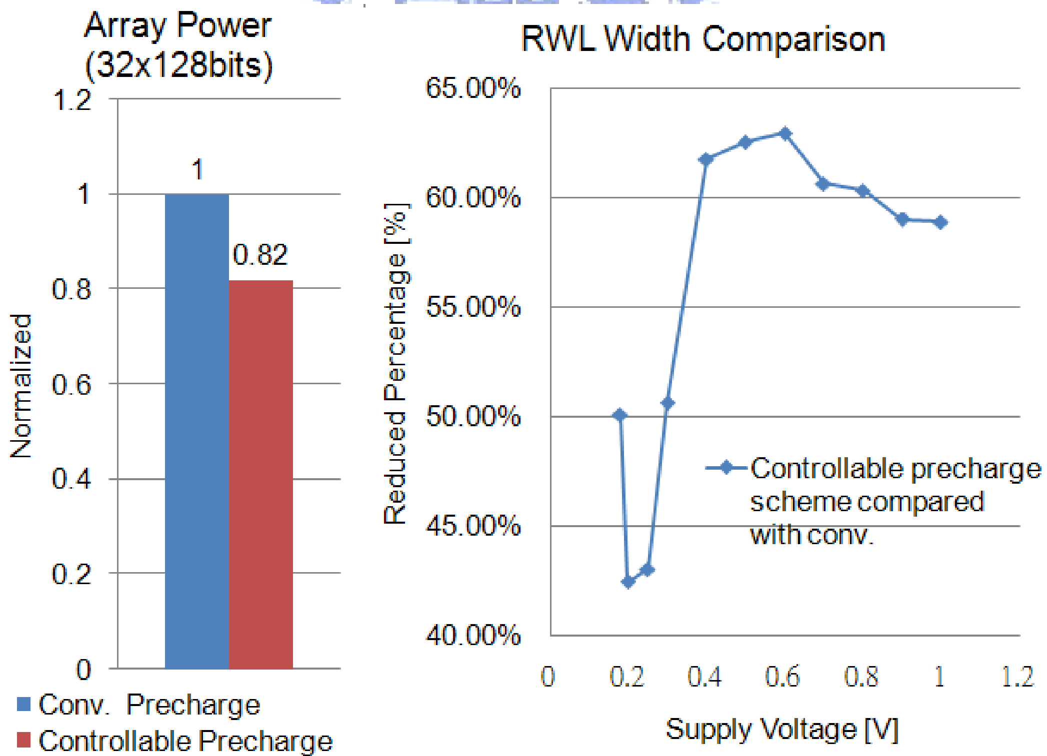
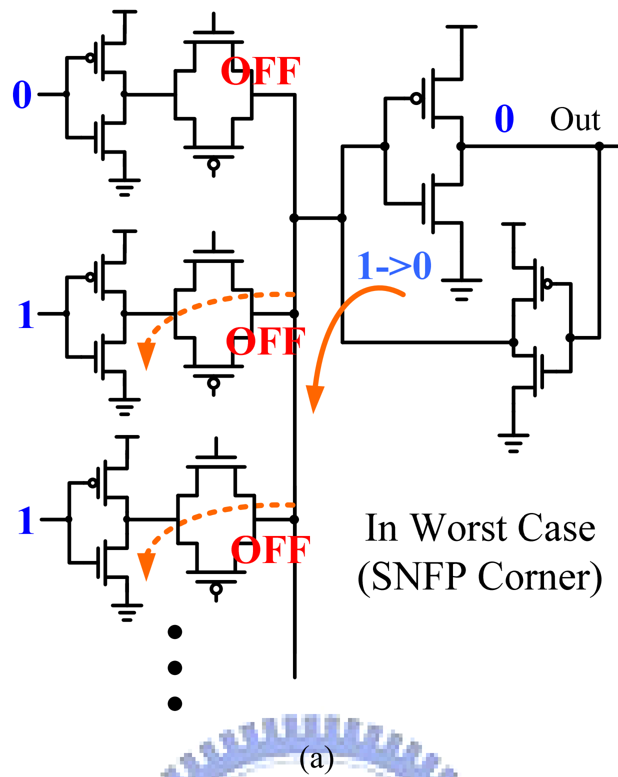


Fig. 5.19: Controllable pre-charge scheme reduces both power and access time in this read architecture.

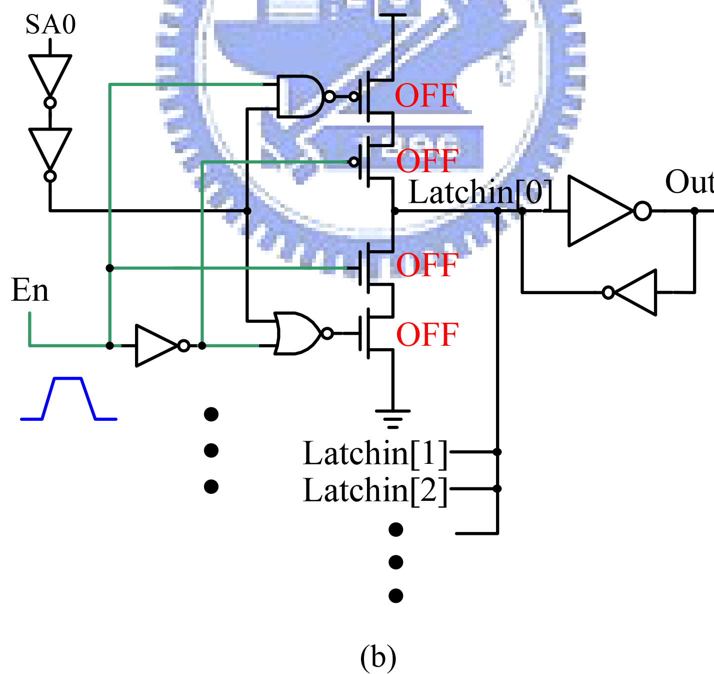
Besides, a controllable pre-charge circuit is important to this read architecture. In bit-interleaving architecture, not all columns need to perform read operation. Therefore, controllable scheme not only reduces lots of power consumption in array, but also decreases IREAD which reduces overhead of current driving of read buffer footer especially in high voltage operating. With 4-bit bit-interleaving architecture, the IREAD can be reduced to 1/4 in the worst pattern. Access time in array reduces to 42-63% and array power reduces to 82% in the proposed scheme, as shown in Fig.5.19. Proposed improved read buffer footer and controllable pre-charge scheme decrease the read access time and power, in addition, read replica circuit also generates the shortest and the most appropriate RWL pulse width by tracing IREAD for successful read operation each time.

5.4.4 Improved output latch

In bit-interleaving architecture, Fig. 5.20 (a) is a common design to store the selected data. First, transmission gate is turned on to transfer the updated data to latch. Then, after successfully updating the data, transmission gate is turned off. However, the updated data in latch might lose in sub-threshold operating voltage after a long time due to these leakage paths. By using stack effect, as shown in Fig.5.20 (b), improved output latch can stored data stably for a long time since leakage path are eliminated.



(a)



(b)

Fig. 5.20: Improved output latch can ensure data stability.

5.5 Design implement

Fig. 5.21 shows the floorplan and layout of proposed register file. Considering to

shuttle of UMC 90nm process in 2009, it is necessary to decrease area to ensure this design can tapeout. Therefore, the capacity of proposed register file in layout is 4KB. In other words, the data-length is decreased from 32-bit to 8-bit, and the other design and feature is remained, such as bit-interleaving - 4, 4 banks, 4W/4R …etc. The area composition is following: input/output circuit in 9%, switch circuit in 24%, decoder and driver in 27%, array in 15%, and MOS capacitance in 2.9%...etc.

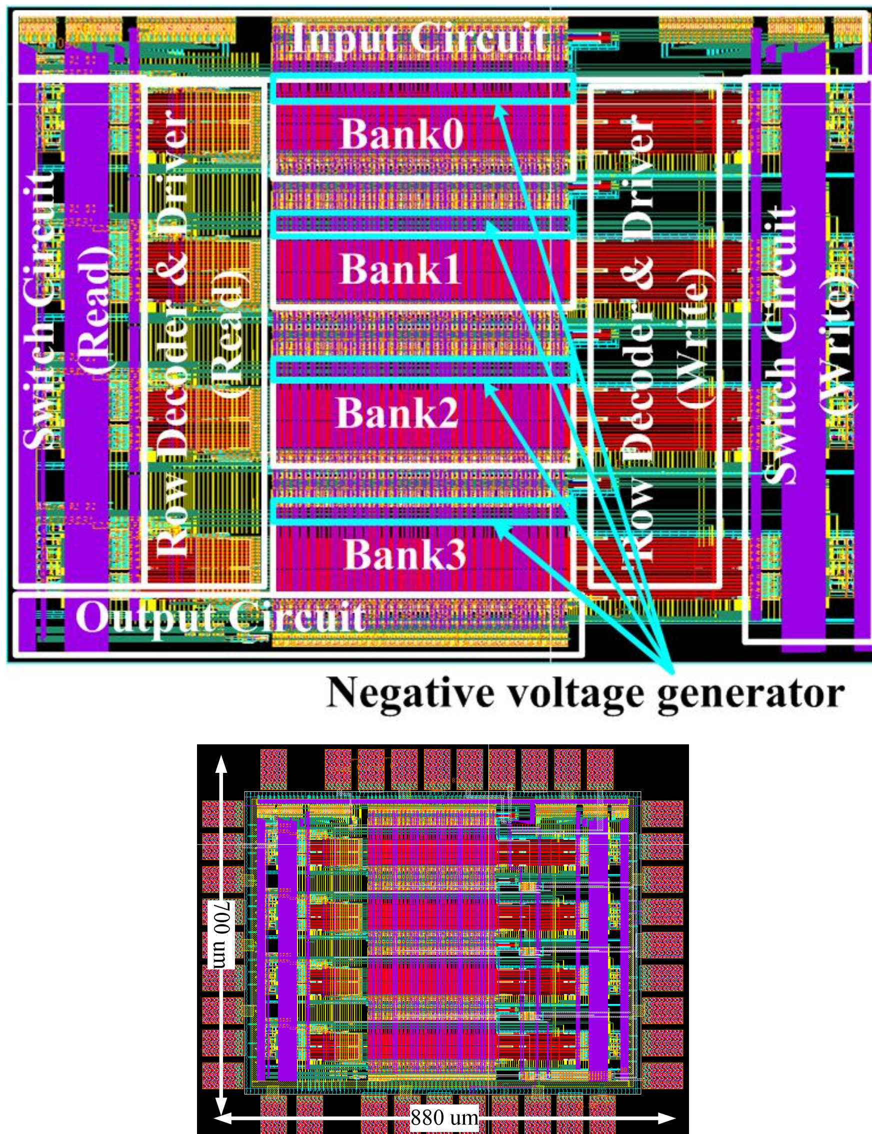


Fig. 5.21: Layout photograph of the proposed sub-threshold multi-port register File.

5.6 Simulation results

The proposed 4W/4R 16KB low power multi-port register file with wide operating voltage range is implemented in UMC 90um CMOS technology. It can operate at 433MHz at 1V, 48MHz at 0.5V and 485 KHz at 0.25V, respectively. While the register file works under 433MHz for 4 simultaneous accesses, it consumes 4.97mW and 2.53mW during write and read operations respectively. When it works under 485 KHz for 4 simultaneous accesses, it consumes 22.3uW and 22.9uW during write and read operations respectively. In most of the time, operating voltage of low power application is under 0.5V or below. High voltage operating for performance is only in a short period of time. This proposed register file can achieved the requirement. Besides, here is the composition of power consumption: 48% in array, 26% in decoder and driver, 16% in switch circuit and 10% in other circuits.

Table 5.1 shows the layout micrograph of proposed register file. Furthermore, by increasing more area of MOS capacitance of boosting and negative voltage generator in original design, operating voltage of this proposed register file can scale down to 0.18V. In 0.18V, proposed register file work successfully across varies process and temperature variation. Even though increasing more area on MOS capacitance, it works still fail at 0.17V or below due to the large gate leakage and sub-threshold leakage in worst corner. Fig. 5.22 shows access time in FF 75°C corner is 254X times in SS -15°C corner while operating in deep sub-threshold region.

Table 5.1: The simulation results.

Configuration	4W/4R 4x128x32bits		
Technology	UMC 90nm CMOS		
Operating Voltage	1V	0.5V	0.25V
Max. Frequency	433MHz	48MHz	485KHz
Max. Read Power	2.53mW	443uW	22.3uW
Max. Write Power	4.97mW	823uW	22.9uW

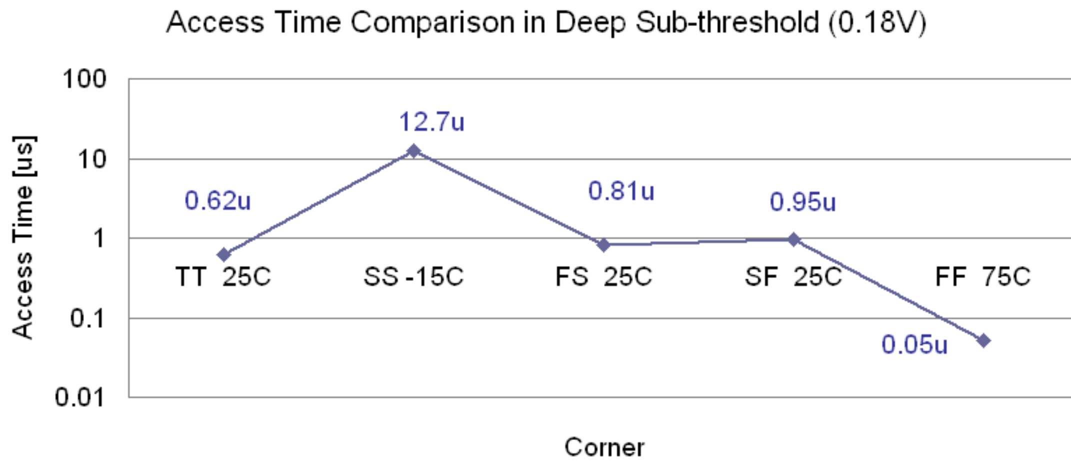


Fig. 5.22: Access time significantly varies across process and temperature variation.

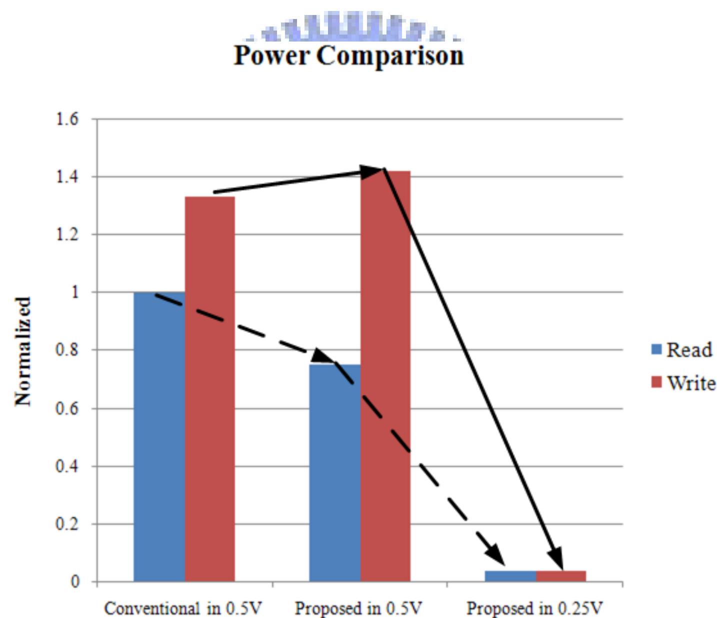


Fig. 5.23: The power comparison between this work and conventional design.

In Fig. 5.23, power comparison between this work and conventional design is shown. Conventional 8T bit-cell design without the proposed read scheme works fail due to large leakage current below 0.5V. By using the proposed scheme such as the controllable precharge scheme and improved read stack, the read power consumption is reduced to 75%. The reason of write power increase is that the large capacitance of

negative voltage generator. However, theses proposed scheme make the register file still operates successfully in 0.25V or below, and it is the main target. In 0.25V, the power consumption is less than 0.05% of conventional design in 0.5V. Since this proposed register file is for low power / low voltage application instead of high performance, it can be applied to sub-threshold application.

5.7 Summary

A low power multibank architecture for simultaneous access with collision detecting technique is presented. For the case the performance is non-critical, the supply voltage can operate at sub-threshold region (<0.5V). A new dual Vt 8T bit-cell, negative voltage write scheme with local BL sensing logic, and read scheme with read footer improvement, controllable pre-charge scheme and read replica circuit are proposed. A 4W/4R 16KB register file under wide operating voltage range between 1V to 0.25V has been designed and implemented in UMC 90nm CMOS technology. The results shows that register file are operated properly at ultra low voltage. The power consumption and operating frequency are 823uW, 48MHz at 0.5V and 22.9uW, 485 KHz at 0.25V, respectively. The proposed register file will be useful for the future micro-power applications.

Chapter 6 Conclusions

By improving algorithm, architecture, circuit design and process technology, chip performance increases steady. According to ITRS roadmap, memory will occupy most part of chip area in ten years. One of embedded memory is 6T SRAM which is common to be applied to high performance chip. Although future process technology provides chip higher clock speed, lower cost and area, serious PVT variation, production reliability issue, etc. become the significant impact on traditional 6T SRAM.

High-k metal-gate process which is a probable substitute for conventional poly-gate device supplies higher performance for circuit design. However, NBTI/PBTI effect increases Vt value of transistor within usage time and also degrades SRAM performance, like cycle time increasing. Detailed analysis is presented in this thesis. A proposed scheme in this thesis can significantly reduce degradation of SRAM performance by using power switch and change SRAM architecture.

One of methods for improving traditional 6T SRAM is to design a new bit-cell. A new 8T bit-cell which eliminates disturb issue of 6T bit-cell is presented in this thesis. Read noise margin of new 8T bit-cell has 1.75X improvement compared to tradition 6T bit-cell. Besides, this 8T bit-cell can remain original peripheral circuit of traditional 6T SRAM. Proposed interface circuit is a bridge between traditional SRAM peripheral circuit and new 8T array. Furthermore, interface circuit doesn't decline performance.

In traditional register file design, designer increases the number of SRAM ports by adding more access transistors. However, stability, cell area, access time and power consumption in this kind design become more serous in future process technology. On the other hands, traditional dual-port, three port SRAM design can’t work under low voltage. A sub-threshold multi-port register file for VLIW DSP is proposed. By using multi-bank, dual-Vt 8T SRAM bit-cell, negative voltage write scheme and improved read footer buffer, proposed register file can work from 1V to 0.25V across various temperature and process corner. The proposed register file is useful for the low-power applications.

Bibliography

Reference of Chapter 2

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

[2.2] K. 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, pp. 305-327, February 2003.

[2.3] K. Roy and S. C. Prasad, Low-Power CMOS VLSI Circuit Design. New York: Wiley, 2000, ch. 5, pp. 214-222.

[2.4] K. M. Cao, W. C. Lee, W. Liu, X. Jin, P. Su, S. K. Fung, J. X. An, B. Yu, C. Hu, "BSIM4 Gate Leakage Model Including Source-Drain Partition," in IEDM Technical Digest, December 2000, pp. 815-818.

[2.5] S. Mukhopadhyay, C. Neau, R. T. Cakici, A. Agarwal, C. H. Kim, and K. Roy, "Gate Leakage Reduction for Scaled Devices Using Transistor Stacking," IEEE Trans. VLSI System, vol. 11, no. 4, pp. 716-730, August 2003.

[2.6] N. Yang, W. K. Henson, and J. Wortman, "A Comparative Study of Gate Direct Tunneling and Drain Leakage Currents in N-MOSFETS with Sub-2100-nm Gate Oxides," IEEE Trans. Electron Devices, vol. 47, pp. 1636-1644, August 2000.

[2.7] K. Nii, Y. Tsukamoto, T. Yoshizawa, S. Imaoka, Y. Yamagami, T. Suzuki, A. Shibayama, H. Makino, and S. Iwade, "A 90-nm Low-Power 32-kB Embedded SRAM With Gate Leakage Suppression Circuit for Mobile Applications," IEEE J. Solid-State Circuits, vol. 39, no. 4, pp. 684-693, April 2004.

[2.8]Semiconductor Industry Association, International Technology Roadmap for Semi-conductors, 2003 ed., http://public.itrs.net.

[2.9] H. J. M. Veendrick, "Short-Circuit Dissipation of Static CMOS Circuitry and Its Impact on the Design of Bu_er Circuits," IEEE J. Solid-State Circuits, vol. sc-19, no. 4, pp. 468-473, August 1984.

[2.10] T. Sakurai, "Perspectives on Power-Aware Electronics," in ISSCC Dig. Tech. Papers, February 2003, pp. 26-29.

[2.11] W. Hwang, (2008), "Embedded Memory Design", Lecture/Class, National Chiao Tung University.

[2.12] E. Seevinck, F. List, and J. Lohstroh, "Static Noise Margin Analysis of MOS

SRAM Cells," IEEE J. Solid-State Circuits, vol. SC-22, no. 5, pp. 748-754, October 1987.

[2.13] B. H. Calhoun and A. P. Chandrakasan, "Static Noise Margin Variation for Sub-threshold SRAM in 65-nm CMOS," IEEE J. Solid-State Circuits, vol. 41, no. 7, pp.1673-1679, July 2006.

[2.14] A. Rychowdhury, S. Mukhopadhyay, and K. Roy, "A Feasibility Study of Sub-threshold SRAM Across Technology Generations," in IEEE Proc. ICCD, October 2005, pp. 417-412.

[2.15] R. Heald and P. Wang, "Variability in Sub-100nm SRAM Designs," in IEEE/ACM Proc. ICCAD, November 2004, pp. 347-352.

[2.16] E. Grossar, M. Stucchi, K. Maex, and W. Dehaene, "Read Stability and Write-Ability Analysis of SRAM Cells for Nanometer Technologies," IEEE J. Solid-State Circuits, vol. 41, no. 11, pp. 2577-2588, November 2006.

[2.17] K. Takeda, H. Ikeda, Y. Hagihara, M. Nomura and H. Kobatake, “Redefinition of Write Margin for Next-Generation SRAM and Write-Margin Monitoring Circuits”, IEEE International Conference on Solid-State Circuits, Session 34, No.5, pp. 630-632, Feb. 2006

[2.18] Ching-Te Chuang, Saibal Mukhopadhyay, Jae-Joon Kim, Keunwoo Kim, and Rahul Rao, “High-Performance SRAM in Nanoscale CMOS: Design Challenges and Techniques,” Invited Plenary Paper, Proc. IEEE International Workshop on Memory Technology, Design, and Testing, Taipei, Taiwan, Dec. 3-5, 2007, pp. 4-11.

[2.19] S. Ishikura, M. Kurumada, T. Terano, Y. Yamagami, N. Kotani, K. Satomi, K. Nii, M. Yabuuchi, Y. Tsukamoto, S. Ohbayashi, T. Oashi, H. Makino, H. Shinohara and H. Akamatsu, “A 45 nm 2-port 8T-SRAM Using Hierarchical Replica Bitline Technique With Immunity From Simultaneous R/W Access Issues,” IEEE Journal of Solid-State Circuits, Vol. 43, Issue 4, pp. 938 – 945, April 2008.

[2.20] L. Chang et al., “A 5.3GHz 8T-SRAM with Operation Down to 0.41V in 65nm CMOS,” Digest of Tech. Papers, Symp. VLSI Circuits, 2007, pp. 252-253.

[2.21] R. Joshi et al., “6.6+ GHz Low Vmin, read and half select disturb-free 1.2 Mb SRAM,” Digest of Tech. Papers, Symp. VLSI Circuits, 2007, pp. 250-251.

[2.22] Y. Morita et al., “An Area-Conscious Low-Voltage-Oriented 8T-SRAM Design under DVS Environment,” Digest of Tech. Papers, Symp. VLSI Circuits, 2007, pp. 256-257.

[2.23] Ik Joon Chang, Jae-Joon Kim, Sang Phill Park, and Kaushik Roy, “A 32kb 10T Subthreshold SRAM Array with Bit-Interleaving and Differential Read Scheme in 90nm CMOS,” Digest of Tech. Papers, ISSCC, 2008, pp. 388-389.

[2.24] Tae-Hyoung Kim, J. Liu, J. Keane and C.H. Kim, "A 0.2 V, 480 kb Subthreshold SRAM With 1 k Cells Per Bitline for Ultra-Low-Voltage Computing," IEEE Journal of Solid-State Circuits, Vol. 43, Issue 2, pp. 518 – 529, Feb. 2008

[2.25] Sheng Lin, Y. B. Kim and Fabrizio Lombardi, "A low leakage 9t sram cell for ultra-low power operation," Proceedings of the 18th ACM Great Lakes symposium on VLSI, pp. 123-126, 2008

[2.26] K. Nii, M. Yabuuchi, Y. Tsukamoto, S. Ohbayashi, S. Imaoka, H. Makino, Y. Yamagami, S. Ishikura, T. Terano, T. Oashi, K. Hashimoto, A. Sebe, S. Okazaki, K. Satomi, H. Akamatsu and H. Shinohara, "A 45-nm Bulk CMOS Embedded SRAM With Improved Immunity Against Process and Temperature Variations," IEEE Journal of Solid-State Circuits Vol. 43, Issue 1, pp. 180 - 191 Jan. 2008

[2.27] M. Yamaoka, K. Osada and K Ishibashi, "0.4-V logic-library-friendly SRAM array using rectangular-diffusion cell and delta-boosted-array voltage scheme," Volume 39, Issue 6, pp. 934 – 940, June 2004.

[2.28] S. Ohbayashi, M. Yabuuchi, K. Nii, Y. Tsukamoto, S. Imaoka, Y. Oda, T. Yoshihara, M. Igarashi, M. Takeuchi, H. Kawashima, Y. Yamaguchi, K. Tsukamoto, M. Inuishi, H. Makino, K. Ishibashi and H. Shinohara, "A 65-nm SoC Embedded 6T-SRAM Designed for Manufacturability With Read and Write Operation Stabilizing Circuits," IEEE Journal of Solid-State Circuits, Vol. 42, Issue 4, pp.820 – 829, April 2007

[2.29] Y. Wang, H. Ahn, U. Bhattacharya, T. Coan, F. Hamzaoglu, W. Hafez, C.-H. Jan, R. Kolar, S. Kulkarni, J. Lin, Y. Ng, I. Post, L. Wel, Y. Zhang, K. Zhang and M. Bohr, "A 1.1GHz 12ýA/Mb-Leakage SRAM Design in 65nm Ultra-Low-Power CMOS with Integrated Leakage Reduction for Mobile Applications," ISSCC 2007 pp. 324 – 606, Feb. 2007

[2.30] Byung-Do Yang and Lee-Sup Kim "A low-power SRAM using hierarchical bit line and local sense amplifiers," IEEE Journal of Solid-State Circuits, Vol. 40, Issue 6, pp. 1366 – 1376, June 2005

[2.31] T. Suzuki, Y. Yamagami, I. Hatanaka, A. Shibayama, H. Akamatsu, H. Yamauchi, "A sub-0.5-V operating embedded SRAM featuring a multi-bit-error-immune hidden-ECC scheme," IEEE Journal of Solid-State Circuits, Vol. 41, pp. 152 – 160, Jan. 2006

[2.32] K. Osada, Jin-Uk Shin, M. Khan, Yu-De Liou, K. Wang, K. Shoji, K. Kuroda, S.Ikeda and K. Ishibashi, "Universal-Vdd 0.65-2.0V 32 kB cache using voltage-adapted timing-generation scheme and a lithographical-symmetric cell." IEEE International Solid-State Circuits Conference, pp.168 - 169, 443, Feb. 2001

[2.33] Meng-Fan Chang, Shu-Meng Yang, Kuang-Ting Chen, Hung-Jen Liao and R.Lee, “Improving the speed and power of compilable SRAM using dual-mode self-timed technique,” IEEE International Workshop on Memory Technology, Design and Testing, pp.57 – 60, Dec. 2007

[2.34] N. Shibata, H. Kiya, S. Kurita, H. Okamoto, M. Tan'no and T. Douseki, “A 0.5-V 25-MHz 1-mW 256-kb MTCMOS/SOI SRAM for solar-power-operated portable personal digital equipment - sure write operation by using step-down negatively overdriven bitline scheme,” EEE Journal of Solid-State Circuits, Vol. 41, Issue 3, pp. 728 – 742, March 2006

[2.35] J.P. Kulkarni, , K. Kim and K. Roy, “A 160 mV Robust Schmitt Trigger Based Subthreshold SRAM,” IEEE Journal of Solid-State Circuits, Vol. 42, Issue 10, Oct. 2007, pp. 2303 - 2313

Reference of Chapter 3

[3.1] http://www.eas.asu.edu/~ptm/

[3.2] G. Chen, K.Y. Chuah, M.F. Li, D.S.H. Chan, C.H. Ang, J.Z. Zheng, Y. Jin, D.L. Kwong, "Dynamic NBTI of PMOS transistors and its impact on device lifetime," IEEE International Reliability Physics Symposium Proceedings, pp.196 – 202, April 2003

[3.3] K. Kang et al., "Impact of Negative-Bias Temperature Instability in Nanoscale SRAM Array: Modeling and Analysis," IEEE T- CAD, pp. 1770 – 1781, Oct. 2007[3.4] R. Fernandez, B. Kaczer, A. Nackaerts, S. Demuynck, R. Rodriguez, M. Nafria, G. Groeseneken, "AC NBTI studied in the 1 Hz Â¿ 2 GHz range on dedicated on-chip CMOS circuits," International Electron Devices Meeting, pp. 1-4, 11-13 Dec. 2006

[3.5] S. Zafar, Y. H. Kim, V. Narayanan, C. Cabral, V. Paruchuri, B. Doris, J. Stathis, A. Callegari, M. Chudzik, "A Comparative Study of NBTI and PBTI (Charge Trapping) in SiO2/HfO2 Stacks with FUSI,TiN,Re Gates," Symp. VLSI Tech., pp. 23-25, 2006.

[3.6] R. Vattikonda, Wenping Wang and Y. Cao, "Modeling and minimization of PMOS NBTI effect for robust nanometer design," IEEE Design Automation Conference, 24-28 July 2006, pp. 1047 – 1052

[3.7] J.C. Lin, A.S. Oates and C.H. Yu, "Time Dependent Vccmin Degradation of SRAM Fabricated with High-k Gate Dielectrics," *Reliability physics symposium*, 15-19 April 2007, pp. 439 – 444

[3.8] T. Suzuki, Y. Yamagami, I. Hatanaka, A. Shibayama, H. Akamatsu, H. Yamauchi, "A sub-0.5-V operating embedded SRAM featuring a multi-bit-error-immune hidden-ECC scheme," *IEEE Journal of Solid-State Circuits*, Vol. 41, Jan. 2006, pp. 152 – 160.

[3.9] M. Sharifkhani and M. Sachdev, "Segmented Virtual Ground Architecture for Low-Power Embedded SRAM," IEEE Transactions on Very Large Scale Integration Systems, Vol. 15, Issue 2, Feb. 2007, pp. 196 - 205

[3.10] L. Chang, D.M. Fried, J. Hergenrother, J.W. Sleight, R.H. Dennard, R.K. Montoye, L. Sekaric, S.J. McNab, A.W. Topol, C.D. Adams, K.W. Guarini, W.Haensch, "Stable SRAM cell design for the 32 nm node and beyond," VLSI Technology Symposium, 14-16 June 2005, pp.128 – 129

Reference of Chapter 4

[4.1] Y. MORITA, H. FUJIWARA, H. NOGUCHI1 Y. IGUCHI, K. NII, H. KAWAGUCHI and M. YOSHIMOTO, "Area Optimization in 6T and 8T SRAM Cells Considering *V*th Variation in Future Processes," IEICE Transactions on Electronics, pp. 1949-1956, October 2007

[4.2] US Patent 7,385,840 B2, Issued June 10, 2008

[4.3] D.P. Wang and Wei Hwang, "A 45nm Dual-Port SRAM with Write and Read Capability Enhancement at Low Voltage," IEEE System-on-Chip Conference, pp. 211 -214, Sep. 2007.

[4.4] S. Mukhopadhyay, R. Rao, J.J. Kim, C.T. Chuang, "Capacitive coupling based transient negative bit-line voltage (Tran-NBL) scheme for improving write-ability of SRAM design in nanometer technologies," IEEE International Symposium on Circuits and Systems, pp. 384 – 387, 18-21 May 2008

[4.5] N. Shibata, H. Kiya, S. Kurita, H. Okamoto, M.Tan'no, and T. Douseki, "A 0.5-v 25mhZ 1-Mw 256-Kb MTCMOS/SOI SRAM for Solar-Power-Operated Portable Personal Digital Equipment --- Sure Write Operation by Using Step-Down Negatively Overdriven Bitline Scheme," IEEE JSSC, vol. 41, pp. 728, 2006.

Reference of Chapter 5

[5.1] H. Jessica Tseng, and Krste Asanovic, “A Speculative Control Scheme for an Energy-Efficient Banked Register file ,“ IEEE Transactions on Computers, Vol. 54, No. 6, pp. 741 – 751, June, 2005.

[5.2] S. Eric Fetzer, David Dahle, Casey Little, and Kevin 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, January 2006.

[5.3] S. Ishikura, M. Kurumada, T. Terano, Y. Yamagami, N. Kotani, K. Satomi, K. Nii, M. Yabuuchi, Y. Tsukamoto, S. Ohbayashi, T. Oashi, H. Makino, H. Shinohara and H. Akamatsu, “A 45 nm 2-port 8T-SRAM Using Hierarchical Replica Bitline Technique With Immunity From Simultaneous R/W Access Issues,” IEEE JSSC, Vol. 43, pp. 938, 2008.

[5.4] N. Shibata, H. Kiya, S. Kurita, H. Okamoto, M.Tan’no, and T. Douseki, “A 0.5-v 25mhZ 1-Mw 256-Kb MTCMOS/SOI SRAM for Solar-Power-Operated Portable Personal Digital Equipment --- Sure Write Operation by Using Step-Down Negatively Overdriven Bitline Scheme,” IEEE JSSC, vol. 41, pp. 728, 2006.

[5.5] D.P. Wang and Wei Hwang, "A 45nm Dual-Port SRAM with Write and Read Capability Enhancement at Low Voltage," IEEE System-on-Chip Conference, pp. 211 -214, Sep. 2007.

[5.6] S. Mukhopadhyay, R. Rao, J.J. Kim, C.T. Chuang, “Capacitive coupling based transient negative bit-line voltage (Tran-NBL) scheme for improving write-ability of SRAM design in nanometer technologies,” IEEE International Symposium on Circuits and Systems, pp. 384 – 387, 18-21 May 2008

[5.7] A. Wang and A. Chandrakasan, “A 180-mV subthreshold FFT processor using a minimum energy design methodology,**”** IEEE JSSC, Vol. 40, pp.310, 2005.

[5.8] Ik Joon Chang, Jae-Joon Kim, Sang Phill Park, and Kaushik Roy, “A 32kb 10T Subthreshold SRAM Array with Bit-Interleaving and Differential Read Scheme in 90nm CMOS,” Digest of Tech. Papers, ISSCC, 2008, pp. 388-389.

[5.9] N. Verma and A.P. Chandrakasan, “A 256 kb 65 nm 8T Subthreshold SRAM Employing Sense-Amplifier Redundancy,” IEEE Journal of Solid-State Circuits, Vol. 43, Issue 1, pp.141 – 149. Jan. 2008

[5.10] F. Moradi, D.T. Wisland, S. Aunet, H. Mahmoodi, Tuan Vu Cao, “65NM sub-threshold 11T-SRAM for ultra low voltage applications,” IEEE International SOC Conference , pp. 113 – 118, 17-20 Sept. 2008

[5.11] N. Verma and A.P. Chandrakasan, “A 256 kb 65 nm 8T Subthreshold SRAM Employing Sense Amplifier Redundancy,” IEEE JSSC, Vol. 43, pp. 141, 2008.

[5.12] L. Chang, D.M. Fried, J. Hergenrother, J.W. Sleight, R.H. Dennard, R.K. Montoye, L. Sekaric, S.J. McNab, A.W. Topol, C.D. Adams, K.W. Guarini, W.Haensch, “Stable SRAM cell design for the 32 nm node and beyond,” *Symposium on VLSI Technology*, 14-16 June 2005, pp.128 – 129

[5.13] L. Liu, R. Sridhar and S. Upadhyaya, “A 3-port Register File Design for Improved Fault Tolerance on Resistive Defects in Core-Cells,” Defect and IEEE International Symposium on Fault Tolerance in VLSI Systems, Oct. 2006, pp. 545 - 553

[5.14] J. Chen, L. T. Clark and T.-H. Chen, ”An Ultra-Low-Power Memory With a Subthreshold Power Supply Voltage,” IEEE Journal of Solid-State Circuits, Vol. 41, Issue 10, Oct. 2006, pp. 2344 - 2353

[5.15] T. H. Chen, J. Chen, L.T. Clark, J.E. Knudsen and G. Samson, “Ultra-Low Power Radiation Hardened by Design Memory Circuits,” IEEE Transactions on Nuclear Science, Vol 54, Issue 6, Part 1, Dec. 2007, pp. 2004 - 2011

[5.16] R. Balasubramonian, S. Dwarkadas, and D.H. Albonesi, “Reducing the Complexity of the Register File in Dynamic Superscalar Processors,” Proc. 34th Ann. IEEE/ACM Int’l Symp. Microarchitecture (MICRO-34), Dec. 2001.

[5.17]S. Wallace and N. Bagherzadeh, “A Scalable Register File Architecture for Dynamically Scheduled Processors,” Proc. Int’l Conf. Parallel Architectures and Compilation (PACT), Oct. 1996.

[5.18] J. E. Smith and G. S. Sohi, “The microarchitecture of superscalar processors,” IEEE of the Proceedings, Vol. 83, pp. 1609 – 1624, Dec. 1995.

[5.19] T. Hironaka, M. Maeda, K. Tanigawa, T. Sueyoshi, K. Aoyama, T. Koide, H.J. Mattausch and T. Saito, “Superscalar processor with multi-bank register file,” Innovative Architecture for Future Generation High-Performance Processors and Systems, pp.10, Jan. 2005.

[5.20] Jessica H. Tseng, and Krste Asanovic, “A Speculative Control Scheme for an Energy-Efficient Banked Register File ” IEEE Transactions on Computers, Vol. 54, No. 6, pp. 741 - 751,June 2005.

[5.21] Tadashi Saito, et al, “Design of superscalar processor with multi-bank register file,” ISCAS, Vol.4, 2005.

[5.22] T. Saito, M. Maeda, T. Hironaka, K. Tanigawa, T. Sueyoshi, K. Aoyama, T. Koide, H.J. Mattausch, “Design of superscalar processor with multi-bank register file” ISCAS, vol. 4 pp.3507 – 3510, 2005.

[5.23] T. Monreal, V. Vinals, J. Gonzalez, A. Gonzalez, M. Valero, “Late allocation and early release of physical registers,” IEEE Transactions on Computers, Vol. 53, no. 10, pp.1244-1259, Oct. 2004.

[5.34] John L. Hennessy and David A. Patterson, “Computer architecture: a quantitative approach,” forth edition, pp. D-9, 2007.

[5.35] B. Zhai, S. Hanson, D. Blaauw, D. Sylvester, “A Variation-Tolerant Sub-200 mV 6-T Subthreshold SRAM,” IEEE Journal of Solid-State Circuits, vol. 43, Issue 10, pp. 2338 - 2348, Oct. 2008

Vita

PERSONAL INFORMATION

Name: Shyh-Chyi Yang

Birth Date: Jan. 16, 1985

Birth Place: Keelung, Taiwan, R.O.C.

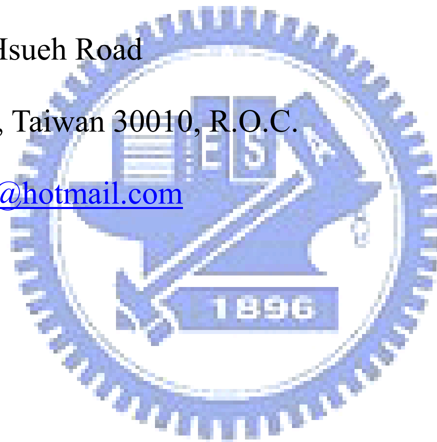
Address: Department of Electronics Engineering

National Chiao Tung University

1001 Ta-Hsueh Road

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

E-Mail Address: a90914@hotmail.com



EDUCATION

B.S. [2007] Department of Electronics Engineering, National Chung-Hsing University.

M.A. [2009] Institute of Electronics, National Chiao-Tung University.