國立交通大學

電子工程學系電子研究所

士論文

40 奈米 1.0Mb 6T 管線化靜態隨機存取記憶體與 三步階升壓型字元線和位元線降壓和適應性電 壓偵測

40nm 1.0Mb 6T Pipeline SRAM with Three Step-Up Word-Line, Bit-Line Under-Drive and Adaptive Voltage Detector

1896

研究生:廖偉男

指導教授:莊景德

中華民國一〇一年九月

40 奈米 1.0Mb 6T 管線化靜態隨機存取記憶體與三步

階升壓型字元線和位元線降壓和適應性電壓偵測 40nm 1.0Mb 6T Pipeline SRAM with Three Step-Up Word-Line,

Bit-Line Under-Drive and Adaptive Voltage Detector

研究生:廖偉男 Student:Wei-Nan Liao

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

國 立 交 通 大 學 電子工程學系 電子研究所碩士班 碩 士 論 文

A Thesis Submitted to Department of Electronics Engineering and Institute of Electronics College of Electrical and Computer Engineering National Chiao Tung University In Partial Fulfillment of the Requirements for the Degree of Master of Science in Electronics Engineering

> September 2012 Hsinchu, Taiwan, Republic of China

中華民國 一 〇 一 年 九 月

40 奈米 1.0Mb 6T 管線化靜態隨機存取記憶體與三步

皆升壓型字元線和位元線降壓和適應性電壓偵測

學生:廖偉男

指導教授:莊景德教授

國立交通大學電子工程學系電子研究所

摘要

近幾年來,記憶體在許多電子產品中被廣泛運用,因為記憶體的高操作速度與高效 能。另外,因為靜態隨機存取記憶體也比其他種類的記憶體具有更高的操作速度,所以 靜態隨機存取記憶體在高性能微處理器的快取記憶體和嵌入式系統中更是被廣泛應用。 過去 20 年間,6T 靜態隨機存取記憶體因為有較高的操作速度與較緊密的面積,因此在 設計上仍然以 6T 靜態隨機存取記憶體為設計主流。但是隨著製程演進至深次微米等級 之後,製程變異會是影響 6T 靜態隨機存取記憶體存活的關鍵因素。在先進製程下,這 些製程變異會讓 6T 靜態隨機存取記憶體的讀或寫的能力受到嚴重的退化。除了讀寫能 力受到影響之外,特別是在低壓操作時,6T 靜態隨機存取記憶體幾乎是無法正常的運 作。

為了設計出能在先進製程下正常運作的 6T 靜態隨機存取記憶體,我們提出三步階 升壓型字元線技術、適應性數據感知寫入輔助技術、位元線降壓技術以及適應性電壓偵 測技術來提高讀寫能力與降低開極氧化層被擊穿的機會。此外,為了提高操作速度我們 也運用管線化技巧。在本論文中,我們將這些技術、2 階級管線化技術與單電源電壓設 計在一顆1.0Mb 高性能 6T 靜態隨機存取記憶體,並且透過下線將該晶片實現在 40 奈米 低功耗互補金屬氧化物半導體技術上。該晶片可以工作在寬電壓範圍從 1.2V 至 0.7V, 具有工作平率 900MHz@1.1V 和 25°C。

I

40nm 1.0Mb 6T Pipeline SRAM with Three Step-Up Word-Line

and Bit-Line Under-Drive and Adaptive Voltage Detector

Student : Wei-Nan Liao

Advisor : Ching-Te Chuang

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

ABSTRACT

In recent years, memories have been widely used for the most electronic products due to their high operation speed and high performance. Besides, Due to SRAMs have higher operating speed than other memory family, SRAMs have been widely used for the high-performance microprocessor cache and embedded system. During the past 20 years, standard 6T SRAM cell becomes the mainstream of SRAMs design due to its highest speed and compact area. However, with the scaling into the deep sub-micron of process, the process variation affects the subsistence of the 6T SRAM cell. In advance technology node, the read and write ability suffer a serious degradation by theses process variation. Especially, at low operation voltage, 6T SRAM cell almost couldn't have normal operation.

In order to design the 6T SRAM that it can normal work in the advanced process, we proposed the Three Step-Up Word-Line technique, Adaptive-Data-Aware Write-Assist technique, Bit-Line Under-Drive Read-Assist technique, and Adaptive Voltage Detector technique to enhance the read/write ability and performance, and reduce the gate oxide to be punctured. Besides, in order to enhance operating speed, we also applied the pipeline technique to enhance the operating speed. In the thesis, we design a 1.0Mb high-performance 6T SRAM with these techniques with two stage pipeline technique with a single supply voltage, and implement by way of tape out in the 40nm Low- Power complementary metal-oxide semiconductor technology. The chip has wide voltage range from 1.5V to 0.6V, with operating frequency of 900MHz@1.1V and $25^{\circ}C$.

誌 謝

本論文能順利完成,首先誠摯的感謝指導教授莊景德教授。在這兩年多的 研究生涯裡,莊老師除了在專業領域上以他豐富的知識給予耐心的指導之外,更 重要的是,讓我們很快就搭上最先端的研究,而省去獨自慢慢摸索的無奈歷程。 莊老師為人和藹可親以及對於研究上的嚴謹態度,讓我學得更多專業知識以外的 事物。

感謝博士班的連南鈞(Patrick)學長,在研究過程中給了許多的幫助。Patrick 學長研究與實務經驗充足,也在許多方面給了我指導與建議,讓我能夠快速的在 研究上步入軌道。另外感謝張琦昕學長,在畢業後仍不離不棄的給予我鼓勵並且 在我壓力大的時候帶我去發洩壓力,讓我可以順利的完成論文。至於實驗室的包 家豪、蔡明甫、朱俐瑋同學以及學長吳尚霖、學弟張智皓、黃騰頡、鍾兆貴還有 學妹林毓柔、楊邵喻、王唐瑄,也感謝你們能夠和我一起討論、運動、烤肉,也 因為有你們實驗室才能如此歡樂並且快樂的做研究。

感謝 FARADAY 的學長們,使我可以提前了解業界的考量,並且以業界的標 準來設計晶片。也特別感謝各位學長們不吝嗇給予建議與協助,使晶片可以如期 下線。

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

廖偉男

于新竹交大

2012.9.20

Content

CHAPTE	R 1 INTRODUCTION	1
1.1	BACKGROUND	1
1.2	MOTIVATION AND GOALS	2
1.3	THESIS ORGANIZATION	2
CHAPTE	R 2 OVERVIEW OF THE DESIGN OF 6T SRAM	4
2.1	Memory Family	4
2.2	6T SRAM	5
2.2.1	Structure of 6T SRAM	5
2.2.2	Read Operation and Read Disturb of 6T SRAM	6
2.2.3	Hold Static Noise Margin and Read Static Noise Margin	8
2.2.4	Write Operation and Half-Selected Read Disturb of 6T SRAM	10
2.2.5	Write Static Noise Margin and Write Margin and AC Write Margin	12
2.2.6	The Size and Layout of 6T SRAM	14
2.3	SRAM ARRAY ARCHITECTURE	15
2.3.1	Memory Array	15
2.3.2	Differential Sensing and Large Signal Sensing Scheme	17
2.3.3	Non-Pipeline SRAM Design	21
2.3.4	Pipeline SRAM Design	22
2.4	GLOBAL VARIATION AND LOCAL VARIATION ISSUE	26
2.5	THE DESIGN METHODOLOGY OF 6T SRAM	28
2.5.1	Dual Supplies	29
2.5.2	Dynamic Bit-Line Level	31
2.5.3	Dynamic Word-Line Level	32
2.5.4	Negative Bit-Line Level	34
CHAPTE	R 3 DESIGN OF 1.0MB 6T PIPELINE SRAM WITH THREE STEP-UP	
WORD-L	INE AND BIT-LINE UNDER DRIVE AND ADAPTIVE VOLTAGE DETECTOR	
SKILL	37	
3.1	INTRODUCTION	37
3.2	PROPOSED BIT-LINE UNDER-DRIVE (BLUD) TECHNIQUE	40
3.3	PROPOSED THREE STEP-UP WORD-LINE (TSUWL) TECHNIOUE	47
3.4	PROPOSED ADAPTIVE VOLTAGE DETECTOR (AVD) TECHNIQUE	
3.5	MACRO IMPLEMENTATION AND SIMULATION RESULT.	
3.6	TEST FLOW	69

3.7	IMPLEMENTATION AND MEASUREMENT RESULT OF TEST CHIP7	<i>'</i> 0
СНАРТ	ER 4 CONCLUSIONS	/5
REFER	ENCE OF CHAPTER 27	/6
REFER	ENCE OF CHAPTER 3	30



List of Figures

Fig. 2-1	Traditional 6T SRAM cell	5
Fig. 2-2	Voltage Transfer Curve (VTC) of CMOS inverter [2-1]	6
Fig. 2-3	Read operation	7
Fig. 2-4	Stability ratio [2-2]	8
Fig. 2-5	Butterfly curve of HSNM	9
Fig. 2-6	The HSNM and RSNM butterfly curve	10
Fig. 2-7	Write operation	11
Fig. 2-8	Half-Selected Read Disturb Voltage of 6T SRAM cell [2-2]	12
Fig. 2-9	Half-Selected Disturb of the 6T SRAM array (For Write operation) [2-2]	12
Fig. 2-10	Butterfly curve of WSNM	13
Fig. 2-11	The definition of the Write Margin (WM) [2-3]	14
Fig. 2-12	Transistor size ratio in 6T SRAM[2-2]	15
Fig. 2-13	The layout of the 6T SRAM in advanced process technology	15
Fig. 2-14	Array architecture of an 2 ^N x2 ^M memory array [2-1]	16
Fig. 2-15	SRAM critical path [2-5]	17
Fig. 2-16	Differential sense amplifier [2-1]	18
Fig. 2-17	Large signal sensing scheme of IBM Cell processor [2-12, 2-13]	20
Fig. 2-18	Large signal sensing scheme [2-11]	20
Fig. 2-19	Non-Pipeline SRAM operation diagram	22
Fig. 2-20	The design of 11FO4 cycle time between cycle boundary [2-6]	23
Fig. 2-21	Local store macros in Streaming Processor Element (SPE) [2-6]	24
Fig. 2-22	Pipeline SRAM operation diagram	25
Fig. 2-23	Global variation and Local variation of threshold voltage [2-2]	26
Fig. 2-24	The effect of local variation, (a) Write mode worse case, and (b) Read mode worse c 28	ase
Fig. 2-25	Dual voltage domain of IBM Cell processor [2-12, 2-13]	30
Fig. 2-26	Dual voltage domain of 6T SRAM floor-plan [2-10]	30
Fig. 2 - 27	V _{MIN} and stability range of dual supply [2-22]	31
Fig. 2 - 28	Bit-Line charge-recycling technique [2-23]	32
Fig. 2 - 29	RSNM improvement with lower V_{WL} (b)WSNM decade with lower V_{WL} [2-24]	33
Fig. 2-30	RSNM is improved by suppressing WL level [2-16]	33
Fig. 2-31	Boosting WL technique of the 6T SRAM [2-15]	34
Fig. 2-32	Multi-Step WL technique [2-17, 2-18]	34
Fig. 2-33	Negative ground voltage of the 6T SRAM [2-14]	35
Fig. 2-34	Negative BL technique of [2-21]	36

Fig. 3-1	Standard 6T SRAM cell schematic in Read mode	.40
Fig. 3-2	Standard 6T SRAM cell butterfly curves under best and worst case	.41
Fig. 3-3	Bit-Line level in dual supply SRAM [3-22]	.42
Fig. 3-4	VBLH Bit-Line regulation system and yield improvement [3-21]	.42
Fig. 3-5	Bit-Line Under-Drive (BLUD) circuit	.43
Fig. 3-6	Large signal sensing circuit with Cross couple pair circuit	.44
Fig. 3-7	Timing diagram for BLUD during read cycle	.45
Fig. 3-8	The BLUD technique improves Read Margin with $3 ext{-}\sigma$ variation	.46
Fig. 3-9	The BLUD technique improves LBL falling time with $3-\sigma$ variation (read 0)	.47
Fig. 3-10	RSNM increase with suppress word-line supply	.48
Fig. 3-11	WSNM decrease with suppress word-line supply	.48
Fig. 3-12	(a) Word-Line Under-Drive (WLUD) circuit (b) Previous Read Assist circuit (PRA) (c)
Multi-Step V	Vord-Line Control (MWC) circuit (d) Step-Up Word-Line (SUWL) circuit	.50
Fig. 3-13	Three Step-Up Word-Line (TSUWL) circuit	.51
Fig. 3-14	Timing diagram for TSUWL during read cycle	. 52
Fig. 3-15	Spice simulation results for TSUWL with different delay time	.53
Fig. 3-16	Spice simulation results for read speed comparison of propose and precious	.54
Fig. 3-17	Spice simulation results for WL rising time comparison of propose and precious	. 55
Fig. 3-18	Spice simulation results for butterfly curve improvement with $3-\sigma$ of variation	
comparison	of TSUWL and BLUD	. 55
Fig. 3-19	Spice simulation results for Read Margin (RM) improvement with $3-\sigma$ variation	
comparison	of TSUWL and BLUD	.56
Fig. 3-20	Spice simulation results for Read Margin (RM) improvement with $3 \cdot \sigma$ variation	
comparison	of TSUWL and BLUD	. 56
Fig. 3-21	Adaptive Voltage Detector (AVD) circuit	. 57
Fig. 3-22	Timing diagram for AVD during read/write cycle	. 58
Fig. 3-23	Adaptive-Data-Aware Write-Assist (ADAWA) circuit of 6T SRAM	. 59
Fig. 3-24	Proposed ADAWA_WEB tracking control circuit	. 60
Fig. 3-25	Timing diagram for ADAWA during write cycle	. 60
Fig. 3-26	The ADAWA technique improves AC Write Margin (ACWM), V_{min} with 3- σ variation	.61
Fig. 3-27	The ADAWA technique improves WSNM with $3-\sigma$ variation	.61
Fig. 3-28	The ADAWA technique improves Write time with $3 \cdot \sigma$ variation	. 62
Fig. 3-29	Spice simulation results for Write Margin (WM) improvement with $3-\sigma$ variation	
comparison	of TSUWL and ADAA	. 62
Fig. 3-30	Critical path of 1.0Mb two stages pipeline 6T SRAM macro	.63
Fig. 3-31	Local Evaluation Circuit (LEV)	.65
Fig. 3-32	Read path (Word-Line to Output latch)	. 66
Fig. 3-33	Layout view of test chip	. 67

Chapter 1

Introduction

1.1 Background

During the past 20 years, Moore's Law told us that the density of the chip capacity is doubled per 18 month. Today, the CMOS technology still follows this rule. In addition to performance, the chip cost and complexity are enhanced with the advanced technology. However, with the scaling into the deep sub-micron of process, the size of device and V_{th} are reduced. But the process variation will become serious issue, because the sigma of local V_{th} variation is larger than that of the global V_{th} variation in advance technology. Therefore, we must consider the global variation and the local variation in previous simulation. However, the performance of manufactured transistor may be different to previous simulated value and lead to system functional error. This could result the degradation to the yield during chip manufacture.

In accordance with ITRS's predictions, memory area will occupy nine-tenths area of the chip. Static Random Access Memory (SRAM) is an important role, because it would dominate the area, performance and power of the SOC chip. Besides, we know that high performance multi-core processors and clouding computing usually need high speed and large capacity SRAM to do data processing. In order to implement these electronic application products, the most important issue is how to design a high performance SRAM.

1.2 Motivation and Goals

Nowadays, In order to implement high performance electronic application products, the Static Random Access Memory (SRAM) is an important role. However, with the reducing supply voltage and scaling process, the transistor characteristic variability would affect the subsistence of the standard SRAM in advance technology node. The degradation of the read and write static noise margin (SNM) is the most crucial issue. From past decades, much circuit technique solutions have been proposed in order to reduce the variation and shifting issue. However, static noise margin of SRAM is contradictory condition between read and write mode. Therefore, we try to propose different circuit technique to separately solve the read and write issue. In thesis, we must focus on the read and write assist circuit technique in order to enhance read and write ability of SRAM. In addition to these read/write assist circuit technique, we want to have widely operation voltage range. Even at low supply voltage, we wish 6T SRAM could also have good manufacturability and yield with these read and write assist circuit technique.

1.3 Thesis Organization

In the following of the thesis, Chapter 2 discuss the basic operation concept of traditional 6T SRAM and its design issue. Besides, we would compare the difference between non-pipeline SRAM design and pipeline SRAM design in Chapter 2. In addition to these concepts, the reliability issue and some design methodology would be also mentioned in this Chapter. Chapter 3 demonstrates "40nm 1.0Mb high performance 6T Pipeline SRAM with Three Step-Up Word-Line (TSUWL) and Bit-Line Under-Drive (BLUD) and Adaptive Voltage Detector (AVD)" design. In this Chapter, Variation Tolerant TSUWL is proposed to improve the read and write

stability of 6T SRAM. Variation Tolerant Bit-Line Under-Drive scheme for SRAM stability enhancement for low voltage operation. Variation Tolerant boost control scheme using Adaptive Voltage Detector circuit to mitigate gate dielectric over stress. The design issue and test flow and chip measurement result would be also discussed in Chapter 3. In the end, Chapter 4 makes a conclusion to this thesis.



Chapter 2

Overview of the design of 6T SRAM

2.1 Memory Family

Memory always occupies over 90% area of the current System on Chip (SOC). As a result, memory always dominates the overall performance of one system. In order to store data, we always used the Random Access Memory (RAM) in the integrated system (IC). Besides, memory family could basically be divided into two categories: volatile memory and nonvolatile memory. RAMs is always associated with volatile memory which the storage data would loss if the power turn off. In contrast, nonvolatile memory would keep the storage data if the power off.

RAMs have been widely used for the most embedded system due to their higher access speed than other memory family. Besides, volatile memory could basically be divided into two categories: Dynamic RAM (DRAM) and Static RAM (SRAM). DRAM has more compact density than SRAM, because DRAM can be built by one transistor and one capacitance. For the past decade, conventional 6T SRAM is always the mainstream to the cache memories in high performance system due to it has the highest operation speed which could reach several hundred Mega Hertz or even Giga Hertz than DRAM. Nowadays, DRAM is currently the major storage device of most SOC due to it has more compact density.

However, with the process technology node goes the deep sub-micron, the

design of 6T SRAM will be faced with several challenges. We must consider these issues which are the process variation and the leakage due to the Read/Write ability suffers a serious degradation in these issues. In order to reduce the process variation and the leakage, we must focus on how to design an efficient circuit technique and understand the basic operation of 6T SRAM.

2.2 6T SRAM



2.2.1 Structure of 6T SRAM

Fig. 2-1 Traditional 6T SRAM cell

In Fig. 2-1 shows the widely common used traditional 6T SRAM cell. For this cell, it includes three control signals and six transistors. Three control signals contain Word-Line (WL) and one pair of Bit-Lines (BL and BLB). Six transistors of the cell contain two pass-gate n-type transistors (M3 and M6), and two pull-up p-type transistors and two pull-down n-type transistors, so the cell is called 6"T" (Transistor)

cell. Two inverters (M1-M2 and M4-M5) are to combine to form one cross couple latch in this cell. This cell could use the cross couple latch to lock value at logic "1" or "0" due to the voltage transfer curve (CVT) of the cross couple latch has only two stable points. In fact, there is one meta-stable point when slope is positive one, but it's not easy to exist (show in Fig.2-2). This Bit-Line pair is connected to the source node of pass-gate n-type transistors. Besides, two pass-gate n-type transistors could be seen as port to access the storage data in this cell. And "Q" and "QB" are storage nodes. Word-Line signal is used to enable this cell, and then the data could be passed in or passed out from this Bit-Lin pair.



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

2.2.2Read Operation and Read Disturb of 6T SRAM

Before the read operation, all of Bit-Line pairs of 6T SRAM cell are pre-charged to high voltage (VDD) at standby mode. At read mode, assume the cell storage data which "Q" storage node is "0" (GND) and "QB" storage node is "1" (VDD) (Fig. 2-3). When once the signal of Word-Line goes high, two pass-gate n-type transistors (M3 and M6) are turned on for accessing storage data. The storage node "0" side will discharge the Bit-Line voltage to ground level (through M3 and M2). However, on the other side of the storage node "0", the storage node "1" will uphold the high level due to the storage node "1" and BLB are the same high level. According to this operation flow, the storage data can be easily passed to Bit-Line. And then, in order to get exact storage data on output pin, we must use additional peripheral read circuit to get the storage data from Bit-Line.



At read mode, this cell had a thorny problem that would to hurt the original storage data. Fig. 2-3 shows the pass-gate n-type transistor (M3) and the pull-down n-type transistor (M2) form a voltage divider. In this case, we assume node "Q" is "0" (GND). When the signal of Word-Line goes high level, node "Q" would be rose to a voltage rather than ground voltage. This situation was called Read-Disturb voltage. Besides, because of the Read-Disturb voltage, the read stability suffers a serious

degradation. When the Read-Disturb voltage goes over the trip voltage of the opposite inverter, the storage node "1" would be flipped to "0".

Fig. 2-4 shows the stability ratios. At 90 nm technology node, the cell switch-point and read down-level began to overlap. That would affect the design of a high yield SRAM in advance technology node.



2.2.3Hold Static Noise Margin and Read Static Noise Margin

In order to evaluate the read ability of 6T SRAM, the Static Noise Margin (SNM) is an important indicator. First of all, we can use the Voltage Transfer Curve (VTC) to get the butterfly curve through switch the axis of any one of the Voltage Transfer Curve (VTC). And then, we can get the Static Noise Margin (SNM) by the butterfly curve. Fig. 2-5 shows the butterfly curve of the Hold Static Noise Margin (HSNM). We can get the HSNM curve at standby mode due to the standby mode operation of 6T SRAM is exactly a cross coupled pair latch. Besides, we can see it has two "wings" and the largest tolerable square of these two wings chooses the smaller one to use definition the Hold Static Noise Margin (HSNM).



Fig. 2-6 shows the difference of curves between the Hold Static Noise Margin (HSNM) and the Read Static Noise Margin (RSNM). For read operation, the signal of Word-Line goes high and two pass-gate n-type transistors turn on for passing storage node data. In addition, the butterfly curve of standby mode is larger than the read mode, and the largest square in either wing that is the HSNM (standby mode) larger than the RSNM (read mode). Therefore, the minimum RSNM also could directly defined as the voltage difference between the trip voltage of inverter and the Read-Disturb voltage. If any other of RSNM wing becomes to "0" or under the zero, the destructive read operation will occur the read fail.



2.2.4Write Operation and Half-Selected Read Disturb of 6T SRAM

Before the signal of Word-Line goes high, the write data must be ready on the Bit-Line pair. Fig. 2-7 shows a write mode, we assume the storage node "Q" is "0" and the storage node "QB" is "1" and we want to write "0" data to the storage node "QB". In this case, the Bit-Line of the storage node "QB" side should be prepared to "0". And then, when the signal of Word-Line goes high, the storage node "QB" data will discharge to ground level by pull-up p-type transistor (M4) and pass-gate n-type transistor (M6). This write operation is successfully, but it still has chance to happen write fail. If the storage node "QB" data is not lower to trigger the tip voltage of opposite inverter, then this write operation is fail.



However, the most common seem problem to 6T SRAM Array is the Half-Selected Read Disturb issue (Fig. 2-8). Fig. 2-9 shows a write mode example. When the signal of WL1 goes high, the Column 1 (COL1) is selected for write operation and the Column 0 (COL0) is not selected for read operation. Under this situation, the Column 0 (COL0) cell can occur the Half-Selected Read Disturb. But the Half-Selected Read Disturb issue is unwanted. Because of the other standby cells would affect the storage data by this issue.



Fig. 2-8 Half-Selected Read Disturb Voltage of 6T SRAM cell [2-2]



can occur

Fig. 2-9 Half-Selected Disturb of the 6T SRAM array (For Write operation) [2-2]

2.2.5Write Static Noise Margin and Write Margin and AC Write Margin

Fig. 2-10 shows the butterfly curve of write operation. In a successfully write operation, the butterfly curve must be open with only one interest point. By definition, this butterfly curve like be combined by RSNM and HSNM. Besides, we also can find a largest tolerable square like RSNM or HSNM on this curve. But there is also write

fail problem for write operation. If the WSNM becomes to 0 or under the zero or more than one interest point, the write operation will fail.

To evaluate the write performance, the Write Margin (WM) [2-3, 2-4] is also an important indicator. Before the write operation, both of BL and BLB have to set at high level (logic one). During the write operation, we sweep down the Bit-Line voltage of the storage node "1" side of the cell from high level to ground level. Afterwards, when the storage node "1" flip, the Bit-Line voltage at this moment is defined as Write Margin (WM) (Fig. 2-11).

During write operation at the same Word-Line pulse width, we change the Bit-Line voltage of the storage node "1" side of the cell from high level to ground level. At some Bit-Line voltage the cell storage node "1" will be suddenly flip, and the Bit-Line voltage at this moment is defined as ACWM.



Fig. 2-10 Butterfly curve of WSNM



2.2.6The Size and Layout of 6T SRAM

In order to keep the read stability, the V_{READ} (Read-Disturb voltage) must be small. So, the pass-gate n-type transistor should be weaker than pull-down n-type transistor (Fig. 2-12). To maintain the write ability, the pass-gate n-type transistor should be stronger than pull-up p-type transistor (Fig. 2-12). In addition, for keep the stability at the standby mode, the pull-down n-type transistor cannot be too stronger compares to pull-up p-type transistor (Fig. 2-12). As a result, the size of each transistor of the 6T SRAM cell is specific designed to ensure maximize the read and hold stability and the write ability.

Starting around 90nm node [2-2], the Thin-Cell layout of the 6T SRAM cell becomes the mainstream due to the Thin-Cell of layout style could reduce BL loading to improve performance and noise immunity. Fig. 2-13 shows the layout of 6T SRAM which uses a single direction poly-silicon to improve manufacturability and yield.

$$I_{PUP} \sim \beta_{1} = \frac{\mu_{p}(W/L)_{PUP}}{\mu_{pD}(W/L)_{PD}} \Rightarrow V_{TRIP}$$

$$I_{AX} \leftrightarrow (1) \longrightarrow (1)$$

Fig. 2-12 Transistor size ratio in 6T SRAM[2-2]



Fig. 2-13 The layout of the 6T SRAM in advanced process technology

2.3 SRAM Array Architecture

2.3.1 Memory Array

Inner For current System on Chip (SOC), memory always be build in the integrated system (IC) as the storage media. These cells are usually formed into an array to enhance the area efficiency and to easy access. In traditional SRAM architecture design, all of the SRAM cells are put together with the peripheral circuit such as Row/Column decoder and Sense Amplifier (SA) are placed next to the SRAM to control Read/Write operation (Fig. 2-14). In this architecture, in order to control pass-gate of all the row direction cells, the signal of Word-Line (WL) is usually a row direction signal. And the Bit-Line (BL) is a signal of column direction that can pass in or out the data from the SRAM cell. So, if we want to select one cell for read or write operation, both the Word-Line (WL) and Bit-Line (BL) must be activated. When the interest cell is selected, the read or write operation is depend on the signal of write enable. However, if the number of the row or column cell increase, the total capacitance and resistance will increase on Word-Line (WL) and Bit-Line (BL), and thus increasing the transient response. In order to reduce the transient response issue, we can use the Hierarchical Word-Line technique and the Hierarchical Bit-Line technique. These two techniques not only reduce the Word-Line loading and the Bit-Line loading but also reduce the charge injection into SRAM cell and the transient response, and thus improving the performance, power and noise margin [2-2].



Fig. 2-14 Array architecture of an 2^Nx2^M memory array [2-1]



2.3.2Differential Sensing and Large Signal Sensing Scheme

In order to get exact storage data on output pin, we must use additional peripheral read circuit to get the storage data from Bit-Lin. The sensing scheme could basically be divided into two categories: differential sensing scheme and large signal sensing scheme. The differential sensing is also celled the small signal sensing in conventional sensing scheme. In order to get the logic "0" or "1" signal from the amplified signal of differential sense amplifier, the basic idea of the differential sensing scheme is to sense the voltage difference between BL and BLB with amplify. To use a cross-couple latch (Q1, Q2, Q3 and Q4) and two access transistors (Q5 and Q6) are to combine to form a conventional differential sensing scheme (Fig. 2-16). It is similar to 6T SRAM but the sizing and the design are different. During read operation, after the Word-Line signal is enabled and the storage node data passed to the Bit-Line, one of the Bit-Line will begin to go low. If the voltage difference between BL and BLB has enough voltage to enable the differential sense amplifier,

the sense amplifier enable (SAE) signal would go high and activate the sense amplifier. Then, the Bit-Line pair will be fully separated and we can get a fully logic "0" or "1". The differential sensing scheme usually co-operate with long Bit-Line structure which means there are many cells along the Bit-Line (usually more than hundred cells). In the long Bit-Line structure, due to the Bit-Line loading would become very heavy, the read time would suffer a serious degradation from the storage node data pass to Bit-Line. Therefore, in order to improve the read time, the long Bit-Line structure must be co-operated with the differential sensing scheme.



Fig. 2-16 Differential sense amplifier [2-1]

With the technology node goes the deep sub-micron of process, the leakage issue becomes a critical issue due to the charge into the cell will become very serious in the long Bit-Line structure. Even the Word-Line was turned off, the standby cell could be flip by the large leakage current that would result retention fail. And the differential sensing scheme also may fail to sense Bit-Line signal by the leakage current. Base on this issue, the Bit-Line length must be decreased and co-operate with short Bit-Line structure. In the short Bit-Line structure, the most common used length is about 8, 16, or 32 cells on Bit-Line. So, the total leakage current could be reduced and the read time could be also reduced. Therefore, the short Bit-Line structure must be co-operated with the large signal sensing scheme. The large signal sensing scheme has been used to sense the data on local Bit-Line [2-10, 2-11, 2-12 and 2-13]. At this scheme, the most common used one transistor or one inverter to detect the signal on the Bit-Lines. When the Bit-Line voltage goes lower than the sensing transistor or sensing inverter, the data of the Bit-Line will be passed out such as to Global Bit-Line and next stage circuit. Besides, the short Bit-Line structure can reduce the Bit-Line loading, and the storage node data pass to the Bit-Line will be faster. Due to the large signal sensing scheme is a single ended sensing, so the leakage is half to the differential sensing where both BL and BLB must be connected on the differential sensing amplifier. And the large signal scheme is easier to implement than the differential sensing scheme which is not easier to optimize the gain. For short Bit-Line structure with the large signal scheme, the area overhead was a disadvantage. Because of the short Bit-Line structure needs many sensing transistors or sensing inverters in each column, the area overhead is larger than the long Bit-Line structure. Fig. 2-17 and Fig. 2-18 show the typical the large signal sensing scheme.



Fig. 2-18 Large signal sensing scheme [2-11]

2.3.3Non-Pipeline SRAM Design

Fig. 2-19 shows a Non-Pipeline SRAM operation diagram. When the Clock rising edge coming, the input address will be launched and decoded at the same time. Next the Finite State Machine (FSM) will control the WLE signal to enable Word-Line signal to perform read or write operation. Due to the WLE signal enable, the pre-charge circuit is turned off. In Non-Pipeline SRAM design, the WLE signal can be seen as internal Clock. For read operation, Local Bit-Line (LBL) and Global Bit-Line (GBL) will get the read data from the storage node data of the 6T SRAM cell. Then, we can utilize the replica control circuit to perform dummy read or write operation for making the WLE signal goes low to turn off Word-Line and enable Global Bit-Line (GBL)/Output latch. When the WLE signal goes low, the pre-charged circuit will be turned on again. This is a completely procedure of read operation in Non-Pipeline SRAM.

mm

m



2.3.4Pipeline SRAM Design

Ultra high performance system utilizes the Pipeline SRAM to enhance the performance. Fig. 2-20 [2-6] shows the best of the operating time in a cycle is 11 Fan-Out-Of-4 (FO4). Fan-Out-Of-4 (FO4) means an inverter can drive four identical copies. The time of only 5 to 8 FO4 is used to operate for function and distribution due to the output delay of L2 and the setup time of the L1 should be removed. Hence, in Pipeline SRAM design, how to balance the operating time in every cycle is an important issue.



Fig. 2-20 The design of 11FO4 cycle time between cycle boundary [2-6]

Fig. 2-21 [2-6] shows the macro of IBM fully Pipelined Embedded SRAM in the Streaming Processor of the cell processor. Starting operating SRAM is from the 3^{rd} cycle to 5^{th} cycle. At the 3^{rd} cycle, one of Local Word-Line (LWL) signals will be decoded and latched in Word-Line driver. At the 4^{th} cycle, the Local Word-Line (LWL) will be enabled to perform read or write operation. By the way, the write operation is finished in this cycle. In the read operation, the Bit-Line (BL) data utilizes the sense amplifier to sense and keep the data until the Read Latch (RL) captures it in the beginning of the 5^{th} cycle. The 5^{th} cycle is used to pass the read data from the Read Latch (RL) to next stage.



Fig. 2-21 Local store macros in Streaming Processor Element (SPE) [2-6]

Fig. 2-22 shows the timing diagram of the two stage pipeline design. In order to implement the two stage pipeline design, it has to need three components which are input latch, middle latch and output latch. Input latch is composed of the L1 latch and the L2 latch, it is used to capture input the address data and launch the pre-decode data to middle latch. Middle latch is L1 latch and always placed next to Word-Line (WL) driver, it is used to latch the pre-decode data and enable Word-Line (WL) driver. Output latch is composed of the L1 latch and the L2 latch, it is used to latch the pre-decode data and enable Word-Line (WL) driver. Output latch is composed of the L1 latch and the L2 latch, it is used to capture the signal of Global Bit-Line (GBL) and launch to output node. The L1 latch is utilized capture pre-data at negative edge of Clock and launch data to next stage at positive edge of Clock. Next, the L2 latch captures pre-data at positive edge of Clock and launch data to next stage at negative edge of Clock. The L1 latch and the L2 latch are to combine to form a Flip-Flop.



Because of this architecture is edge control, so there is no Finite State Machine (FSM) and replica circuit to control any internal signals in this architecture. Besides, all of the operating such as decoding and enable Word-Line (WL) to data output latch are only operated at the positive edge Clock, and the negative edge Clock is only used to pre-charge Bit-Line (BL) and Global Bit-Line (GBL). The fully operation in pipeline: at the 1st cycle for decoding address signal, at the 2nd cycle for enable Word-Line (WL)

and sensing data to Global Bit-Line (GBL) latch, at the 3rd cycle for available output data. By the way, when the operating of middle latch is ongoing, the new address will be continuing to decode in the input latch.

2.4 Global Variation and Local Variation Issue

When the real chip was implemented, there must exist variation factor due to foundry makes the real physical device is different. At the CMOS technology process, because of these variations can appear on the threshold voltage (V_{th}), the current drive ability and the leakage of the transistor may be decrease and larger. These problems would affect the functionality and the power consumption. The most common reasons are the lithography variation at each process, the concentration fluctuation of the doping and the line edge roughness, etc. The threshold voltage (V_{th}) variation could basically be divided into two categories: Global variation and Local variation [2-7]. Global variation is also celled intra-die variation that means it between die to die. Local variation is also called intra-die variation that means the device variation at the same die. Fig. 2-23 shows the case for threshold voltage variation (δV_t), the variation can be expressed as (1) [2-7]





Local device mismatch

Fig. 2-23 Global variation and Local variation of threshold voltage [2-2]
For Global variation, when every time we make the real chip at different corners, the gate length, gate width, gate oxide thickness and channel doping concentration would be different [2-8, 2-9]. The different process condition may affect different die. If we compare with PSNS and PFNF, the characteristic would vary inter dies. In addition, assume the two same devices at different location in the same die. We can find the threshold voltage (V_{th}) of device not the same with the same size of device. In this case, it called Local variation. But the doping concentration is random and follows the statistical Gaussian-Distribution. For this reason, Local variation is unpredictable and hard to be controlled.

However, with the process technology node goes the deep sub-micron, the gate length is very short and the doping concentration is much less. At the advance technology design phase, the variation would become a critical problem. Both Global variation and Local variation issues also affect the characteristic of 6T SRAM. For the read operation, if Global variation at PSNF and Local variation of the each transistor of 6T SRAM cell shows in Fig. 2-24(b), the Read Static Noise Margin (RSNM) would be reduce even equal to zero or less than zero. Because of the high V_t M4 and the low V_t M5 are to combine to form a higher Read-Disturb voltage with the high V_t M1 and the low V_t M2 are to combine to form a smaller trip voltage of inverter such as a worst case for read mode. Next, for the write operation, if Global variation at PFNS and Local variation of the each transistor of 6T SRAM cell shows in Fig. 2-24(a), the Write Static Noise Margin (WSNM) would be reduce even equal to zero or less than zero. Because of the high V_t M3 are to combine to form a smaller trip voltage of inverter such as higher write trip voltage with the high V_t M1 and the low V_t M2 are to combine to form a smaller trip voltage to zero or less than zero. Because of the high V_t M5 and the low V_t M3 are to combine to form a higher write trip voltage with the high V_t M1 and the low V_t M2 are to combine to form a smaller trip voltage to zero or less than zero. Because of the high V_t M5 and the low V_t M3 are to combine to form a higher write trip voltage with the high V_t M1 and the low V_t M2 are to combine to form a smaller trip voltage for write mode.



Fig. 2-24 The effect of local variation, (a) Write mode worse case, and (b) Read mode worse case

2.5 The Design Methodology of 6T SRAM

There several critical issue of 6T SRAM such as Read-Disturb voltage, Half-selected Disturb problem, Static Noise Margin (SNM), leakage and variation, etc have been discussed in the Chapter 2. Unfortunately, with the technology node goes the deep sub-micron of process, these issues will become more serious than before the discussing. In recent years, in order to enhance survival of 6T SRAM at advance technology node, there has been a dramatic increase research concerned with how to improve the read/write ability and Static Noise Margin (SNM). In fact, the most common methodology to improve the read/write ability is four: 1) Dual Supplies 2) Dynamic Bit-Line level 3) Dynamic Word-Line level 4) Negative Bit-Line level. We would discuss these methods below.

2.5.1 Dual Supplies

In order to improve the performance of 6T SRAM such as the read and write ability, the memory array and local control circuits use the difference supply voltage. For increasing the read ability or the Static Noise Margin, we could increase the cell supply voltage or a negative supply voltage of cell. For increasing the write ability, we could reduce the supply of the cell. However, in order to achieve these targets, the easiest way is to use the second power source, so that the supply voltage of cell and logic circuit are separated. In order to have the best performance of the circuit, the most common used higher supply voltage is the critical path of the circuit. It usually sets a supply voltage 150mV – 200mV higher than the logic supply voltage to ensure cell stability and improve performance. And the critical path of circuit can also be on higher supply such as the Word-Line driver, decoder and write driver. However, the implementation of dual supplies is too expensive, so we need to so consider impact to overall system cost overhead when designing with dual supply grids. Fig. 2-25 and Fig. 2-26 show the dual supplies. These two examples used higher supply voltage in memory cell, the Word-Line driver, second level decode and write driver. Fig. 2-27 shows V_{MIN} and stability range of dual supply.



Fig. 2-26 Dual voltage domain of 6T SRAM floor-plan [2-10]



2.5.2 Dynamic Bit-Line Level

Another method to improve the Read V_{min} and the Read Static Noise Margin (RSNM) is to utilize decrease a voltage about 30% of the supply voltage on Bit-Line. For read operation, before the Word-Line signal is activated, the Bit-Line voltage must be decreased about 30% of the supply voltage. However, it can reduce the Bit-Line loading to improve the Read Static Noise Margin (RSNM), read speed and no degradation of write ability, Write Margin (WM), and Write V_{min} . Assume at the short Bit-Line structure with large signal sensing scheme with Dynamic Bit-Line Level, the sensing transistor or the inverter would be enabled too early. When we want to sense logic "0" or "1", the sensing inverter may be enabled too early to sense the wrong data and read fail. So, we need to use a timing control circuit. But the timing is not easier to be controlled and to be a crucial issue to the Dynamic Bit-Line Level. Fig. 2-28 shows Bit-Line charge-recycling technique [2-23].



Fig. 2-28 Bit-Line charge-recycling technique [2-23]

2.5.3 Dynamic Word-Line Level

In order to improve cell stability, we can use another method which is the Dynamic Word-Line Level. The basic concept of the method is to use a higher Word-Line voltage to enhance the driving effect of the pass-gate n-type transistor, and then the write ability and read speed would be improved. But higher Word-Line voltage would increase the Read-Disturb voltage of the 6T SRAM cell which would decrease the Read Static Noise Margin (RSNM). Even the Half-selected Read Disturb issue would be more serious. In order to reduce the Read-Disturb voltage (such as Word-Line Under-Drive). Rather than adopt the higher Word-Line voltage, the driving effect of the pass-gate n-type transistor would be decrease. With the lower Word-Line Voltage, the Read Static Noise Margin (RSNM) of all cells could be improved (Fig. 2-29(b)), and the Half-selected Read Disturb issue could be reduced. But the Write

Static Noise Margin (WSNM) would be decreased (Fig. 2-28). Hence, we can utilize these two dual supply skills to improve the read speed, the Read Static Noise Margin (RSNM) and the write ability. Fig. 2-30, Fig. 2-31, and Fig. 2-32 show the Multi-Step Word-Line technique [2-17, 2-18].





Fig. 2-30 RSNM is improved by suppressing WL level [2-16]



Fig. 2-31 Boosting WL technique of the 6T SRAM [2-15]



2.5.4 Negative Bit-Line Level

Next, in order to improve the write ability, the Negative Bit-Line (NBL) is another method [2-19, 2-20, 2-21]. The basic concept is to use a boosting capacitor to couple a negative bias on Bit-Line into the storage node of cell to improve the write performance. If the Bit-Line voltage was lower than zero, the cross-couple pair latch could be easier enabled to flip the storage node data. So, this skill usually sets a negative voltage about -200mV at the Bit-Line to enhance the write ability and the Write Margin (WM). Although, from the Negative Bit-Line (NBL) skill, we can get many advantages that such as the Write Margin, the read speed and the write ability. If the implementation of the boosting capacitor uses the Metal-Insulator-Metal (MIM) structure, the cost will be increased. Rather than adopt the MOS capacitor, it is easier to achieve, and not increase the cost. But, the MOS capacitor will increase the area overhead and the capacitance is unstable. So, we need to so consider impact to overall system area overhead when designing with this kind of skills. By the way, the Negative Bit-Line (NBL) scheme must have a very precisely timing control to maximize the improvement of write. Basically, the negative voltage should be generated before Word-Line turned on. But the timing is quite hard to be controlled and to be a crucial issue to the Negative Bit-Line (NBL). Fig. 2-32 and Fig. 2-33 show the Negative Bit-Line (NBL) scheme by using the boosting capacitor.



Fig. 2-33 Negative ground voltage of the 6T SRAM [2-14]

TUT

m



Fig. 2-34 Negative BL technique of [2-21]

Chapter 3

Design of 1.0Mb 6T Pipeline SRAM with Three Step-Up Word-Line and Bit-Line Under Drive and Adaptive Voltage Detector skill

3.1 Introduction

In accordance with ITRS's predictions, memory area will occupy nine-tenths area of the chip [3-1]. Static Random Access Memory (SRAM) is an important role, because it would dominate the area, performance, speed, die yield and power of the SOC chip. Besides, we know that high performance multi-core processors and clouding computing usually need high speed and large capacity SRAM to do data processing. However, with the reducing supply voltage and scaling process, the transistor characteristic variability would affect the subsistence of the standard SRAM in advance technology node. When the real chip was implemented, there must exist variation factor due to foundry makes the real physical device is different. The most common reasons are the systematic Global variation and Local random variations due to microscopic effects such as Random Doping Fluctuation (RDF) and Line Edge Roughness (LER). So, with the technology node goes the deep sub-micron of process, these issues will become more serious than before the discussing.

When the SRAM cell is scaled, the cell stability, Static Noise Margin (SNM), and V_{min} are limited by leakage, variation, and supply voltage. In order to facilitate read and minimize Read-Disturb (V_{READ}) to make sure the cell won't flip during read operation, the designing of 6T SRAM must follow strong pull-down n-type transistor and weak pass-gate n-type transistor. And in order to improve write ability (write margin) during write operation, the designing of 6T SRAM must follow strong pass-gate n-type transistor and weak pull-up n-type transistor. But the read/write operation of 6T SRAM cell is conflicting. So, the optimization of 6T SRAM must consider the read/write requirements. Besides, in order to overcome these problems, several approaches have been proposed to reduce the leakage (sub-threshold leakage and gate leakage), variations (V_{th} variation, Global variation, and Local variation), disturbs (Read-Disturb and Half-selected Disturb), etc. Such as using high-k metal technique, suppressing LER, optimizing channel profile, new device structure, read/write access circuits can reduce these issues [3-2~3-5].

In this work, we proposed a Variation-Tolerant Three Step-Up Word-Line (TSUWL) technique to improve the Read Static Noise Margin (RSNM), Write Static Noise Margin (WSNM), and read/write speed. We use the Step-Up Word-Line (previous design is called Word-Line Under-Drive (WLUD) [3-6 and 3-7]) scheme (SUWL) and Boosting Word-Line scheme [2-17 and 2-18] to form a Variation-Tolerant Three Step-Up Word-Line (TSUWL) technique. When the Word-Line signal is activated, the first step Word-Line voltage is lower about 90% of supply voltage that can reduce the Read-Disturb and Half-Selected Disturb issues. And the second step Word-Line voltage is pre-charge to original supply voltage that can evolute and write ability. And the third step Word-Line voltage

is higher about increasing 200mV that can improve the read speed and write ability. During we use the Boosting Word-Line scheme, we propose an Adaptive Voltage Detector (AVD) reduce technique to avoid mitigate gate dielectric over-stress. Thin gate dielectric in deeply scaled technology such as EOT \leq 1.8nm at 90nm, EOT \approx 0.9nm at 32nm, and EOT \approx 0.65-0.75nm at 22nm. So, the Gate Dielectric Reliability (GDR) must be considered. The basic concept of the Adaptive Voltage Detector is when voltage higher than designed voltage, the Booster will be not enabled; when voltage lower than designed voltage, the Booster will be enabled. We also propose another circuit technique "Bit-Line Under-Drive Read-Assist (BLUD)" to enhance read ability, Read Static Noise Margin (RSNM), and read speed. For read/write operation, before the Word-Line signal is activated, the Bit-Line voltage must be decreased about 30% of the supply voltage. However, it can reduce the Bit-Line loading to improve the Read Static Noise Margin (RSNM), read speed and no degradation of write ability, Write Margin (WM), and Write V_{min} (previous design are called Bit-Line Charge-Recycling (BLCR) [3-23] and Adaptive BL Bleeder [3-24]).Implemented the TSUWL and BLUD and AVD schemes in a 40nm 1.0Mb 6T SRAM with two stages Pipeline require 36.36% area overhead. This macro can operate across wide voltage range from 1.5V down to 0.6V, with operating frequency of 900MHz@1.1V and 25°C. The remainder of this work is organized as follows. Section 3.2 presents Bit-Line Under-Drive (BLUD) Read-Assist scheme. Section 3.3 presents Three Step-Up Word-Line (TSUWL) Read/Write-Assist scheme. Section 3.4 presents Adaptive Voltage Detector (AVD) scheme. Section 3.5 presents the macro implementation and Post- Simulation result. Section 3.6 briefly introduces the test flow. Section 3.7 presents the measurement result of Test Chip.

3.2 Proposed Bit-Line Under-Drive (BLUD) Technique

With the scaling into the deep sub-micron of process, there many problems affect the subsistence of the 6T SRAM cell. However, at read operation, the 6T SRAM cell had a thorny problem which is Read-Disturb that would to hurt the original storage data (Fig. 3-1). The pass-gate n-type transistor (M3) and the pull-down n-type transistor (M2) form a voltage divider. In this case, we assume node "Q" is "0" (GND). When the signal of Word-Line goes high level, node "Q" would be rose to a voltage rather than ground voltage. And the rising voltage is called Read-Disturb that could decrease the Read Static Noise Margin (RSNM), even fail in read operation.



Fig. 3-1 Standard 6T SRAM cell schematic in Read mode

In read mode without Local V_{th} variation, the best case is at PFNS Global variation and high temperature (125°C); the worst case is at PSNF Global variation and low temperature (-40°C). Compare best case with worst case on butterfly curve, the wing of the worst case is smaller than the best case. During at PSNF corner, all of

n-type transistors become stronger. So, the Read-Disturb voltage would increase and may to flip the trip voltage of opposite inverter. Fig. 3-2 shows the best case had larger square than the worst case. However, the most common seem problem to 6T SRAM Array is the Half-Selected Read Disturb issue. At the worst case, the Half-Selected Disturb will cause retention failure. Therefore, the first consideration of our methodology is to improve stability in read operation by improving the RSNM.



Fig. 3-2 Standard 6T SRAM cell butterfly curves under best and worst case
In this work, we proposed a Variation-Tolerant Bit-Line Under-Drive (BLUD)
technique to enhance SRAM stability for low-voltage operation, and improve Read
V_{min}. If the Bit-Line Under-Drive (BLUD) technique is achieved in dual supply
SRAM, the high cell supply (VCS) for cell stability and performance critical
immediate neighboring circuits, and the low supply (VDD) for peripheral circuits to
reduce power, and Word-Line connected to VCS for write ability tracks cell VCS. But
the BLUD in dual supply SRAM had some disadvantages such as dual supply
expensive, not suitable for cost-effective SRAM compiler applications (Fig. 3-3)
[3-22].



Fig. 3-4 VBLH Bit-Line regulation system and yield improvement [3-21]

Fig. 3-4 shows the op-amp based push-pull Bit-Line voltage regulator. It can set Bit-Line level between 68%-78% of VDD to reduce the Read-Disturb voltage and Half-Selected Disturb. This technique had some advantages such as precise pining of Bit-Line level, PVT compensated design. But it still had some disadvantages such as Op-Amp (analog circuits, high supply voltage, area overhead, routing) not suitable for SRAM compiler applications (distributed SRAM macros with various sizes and configurations).

In this work, we used large signal sensing scheme (Fig. 3-6) for SRAMs with Bit-Line Under-Drive (BLUD) (Fig. 3-5) with following desired features: 1) Maintain the larger sense margin and better scalability of large signal single-ended sensing to enable further scaling, 2) Eliminate the leakage (hence possible sensing error) in large signal sensing with BLUD, 3) Simple implementation, 4) Minimum transistor count, 5) High-speed sensing, 6) Can be implemented in either single-supply or dual-supply SRAMs, 7) Easy implementation for SRAM compiler applications, 8) Variation tolerant.



Fig. 3-5 Bit-Line Under-Drive (BLUD) circuit



Fig. 3-6 Large signal sensing circuit with Cross couple pair circuit

OSB<0> for Bit-Line discharge 40% of VDD, OSB<1> for Bit-Line discharge 30% of VDD, OSB<2> for Bit-Line discharge 20% of VDD, OSB<3> for no enable BLUD scheme; OSP<0> for setting BLUD level at 60% of VDD, OSP<1> for setting BLUD level at 70% of VDD, OSP<2> for setting BLUD level at 80% of VDD. SELE signal is selected bank signal; LXP is selected local word-line driver signal. Cross-coupled PMOS to develop full-rail signal, Complementary pre-charge PMOS/NMOS pair to neutralize coupling noise to LBL. Fig. 3-7 shows timing diagram for BLUD during read cycle. Assume OSB<1> and OSP<1> are high (Bit-Line discharge 30% of VDD and setting BLUD level at 70% of VDD). For read operation, if SELE goes high in the negative edge of CLK, PWR (Bit-Line power source) stars to discharge Bit-Line to desired level by M5, M6 (Fig. 3-5) and transmission gate (M2&M3 and M4&M5 (Fig.3-6)). Final the BLUD level set by M7&M8 voltage divider (Fig. 3-5). When positive edge of CLK comes and LXP goes high, (Local) Word-Line and cross couple pair will be activated to access the storage node data. For access "0", the P1 signal can activate with Word-Line signal at the same time. But for access "1", before the P1 signal is activated, the Bit-Line voltage must be at 90% of VDD. Due to we use PMOS to sense data with BLUD, if P1 is activated with Word-Line signal at the same time and Bit-Line level at 70% of VDD for access "1", the read operation may fail. During Bit-Line level at 70% of VDD, the sensing PMOS is weakly activated that may to sense wrong data to Global Bit-Line.



Fig. 3-7 Timing diagram for BLUD during read cycle

Read Margin (RM) of the 6T SRAM cells can be defined as: $RM=V_{trip} - V_{disturb}$; Where V_{trip} represents the trip voltage of the inverter which is composed of a pull-up p-type transistor and a pull-down n-type transistor. Fig. 3-8 shows the RM could be further improved if we adopt the BLUD technique. At supply 0.8V (PSNF @125°C) in read operation, the RM with BLUD technique has 31mV improvement. As supply voltage goes high, the improvement of RM is small than low supply voltage.



Fig. 3-8 The BLUD technique improves Read Margin with 3-σ variation

Fig. 3-9 shows the LBL falling time could be further reduced if we adopt the BLUD technique. At supply 0.8V (PSNF @125 $^{\circ}$ C) in read "0" operation. The LBL falling time with BLUD technique has 104ps improvement. As supply voltage goes high, the improvement of LBL falling time is small than low supply voltage.



Fig. 3-9 The BLUD technique improves LBL falling time with 3-σ variation (read 0)

3.3 Proposed Three Step-Up Word-Line (TSUWL)

Technique

In recent years, there has been a dramatic increase research concerned with read/write access circuits. All of read/write access circuits were utilized to improve the RSNM, WSNM, read/write ability, and V_{min} which such as Suppress Word-Line supply, and Multi-Step Word-Line, Boosting Word-Line, Negative Bit-Line, etc. For Suppress Word-Line supply, it could reduce Read-Disturb and Half-Selected Disturb, but the read speed and write ability would be degraded due to the pass-gate n-type transistor would become weaker. Then, if we consider the Global variation at PSNF with high temperature (125°C) in Suppress Word-Line supply, the WSNM and read speed would suffer more serious degradation [3-6, 3-7, 3-8, 3-9, 3-10, 3-11]. Fig. 3-10 shows the RSNM increase with Suppress Word-Line supply.



Fig. 3-11 WSNM decrease with suppress word-line supply

For Boosting Word-Line, it was utilized to enhance write performance because the higher Word-Line voltage could increase current driving of pass-gate n-type transistor. But this skill had some disadvantages that Half-Selected Disturb would become more serious in row direction during read/write operation. The cost would be increased because it needed a large capacitor. For Negative Bit Line (NBL) skill, it could improve Write Margin (WM) and write ability by using a large capacitor to couple a negative level to increase the current driving of pass-gate n-type transistor. But it had the same issue with Boosting Word-Line skill which the cost increased due to a large capacitor. Besides, both Boosting Word-Line and Negative Bit Line (NBL) needed a precise timing control circuit, and the timing control is not easy to design in advantage process [3-6, 3-7, 3-15, 3-16, 3-17, 3-18, 3-19].

In our previous work, we proposed Word-Line Under-Drive (WLUD) and Step-Up Word-Line (SUWL) circuit techniques [3-6, 3-7, and 3-25] (Fig. 3-4 (a) and (d)) to enhance read stability. This resistance-free circuit technique can be easily implemented in SRAM compiler with minimum device/area overhead. Both WLUD and SUWL circuit techniques could suppress Word-Line level to reduce Read-Disturb and "Half-Select" Disturb issues. During read/write operation, Word-Line voltage was higher than V_{th} of NMOS (M4) to enable WLUD and SUWL. These two techniques had the faster speed than Previous Read Assist circuit (PRA) (Fig. 3-4 (b)) [3-20] and Multi-Step Word-Line Control (MWC) circuit (Fig. 3-4 (c)) [3-12, 3-13]. For PRA, the power source node of WL driver used resistance and NMOS (M1) to form a divided to decrease Word-Line level and rising time. Besides, PRA couldn't cover all of corner variation due to it only used NMOS (M1) to track WL driver pull-up PMOS (M2). For MWC, it only used weakly PMOS (M4) to decrease WL slew-rate. But both WLUD and SUWL circuit techniques could track all of corner variation due to these two skills use a NMOS (M4) and a PMOS (M3) (Fig. 3-12 (a) and (d)). For example, we compare PSNF corner and PFNS corner. In order to reduce Read-Disturb at the worst case which is PSNF corner, we could use M3 to track M1 of WL driver to get a lower Word-Line level that could decrease on-current of pass-gate n-type transistor of 6T SRAM cell to reduce Read-Disturb. At the PFNS corner, the pass-gate n-type transistor of 6T SRAM cell became weaker to degrade the read/write ability, so we could use M4 to track pass-gate n-type transistor of 6T SRAM cell to get a higher Word-Line level than PSNF corner.



Fig. 3-12 (a) Word-Line Under-Drive (WLUD) circuit (b) Previous Read Assist circuit (PRA) (c) Multi-Step Word-Line Control (MWC) circuit (d) Step-Up Word-Line (SUWL) circuit

However, during read/write operation, WLUD circuit technique always suppressed in our first previous work, so we improve WLUD circuit technique to SUWL circuit technique with a tracking circuit in our second previous work. SUWL utilized two steps Word-Line level to fix the disadvantage of WLUD because at second step, Word-Line level was pre-charge to full voltage to improve access speed and write performance. In this work, we proposed a Three Step-Up Word-Line (TSUWL) circuit technique (Fig. 3-13) which combines SUWL and Boost Word-Line (BWL).



Fig. 3-13 Three Step-Up Word-Line (TSUWL) circuit

Fig. 3-14 shows the circuit diagram of the proposed TSUWL. This technique consists of three phases. For the first phase, When Word-Line level higher than V_{th} of M4, Word-Line voltage is decreased a voltage about 10% of VDD by a discharge path

which is a PMOS (M3) and a NMOS (M4). Before the Word-Line level higher than V_{th} of NMOS (M4), this skill can get better Word-Line rising speed than PRA and WMC. For the second phase, we pre-charge Word-Line level to full voltage due to a delay chain circuit which could control the operating time of PMOS (M3). Compare with WLUD, the second phase of TSUWL can avoid RSNM and WSNM to suffer a serious degradation. For the third phase, we utilize Boosting Word-Line (BWL) to improve read/write V_{min} at low voltage read/write operation. This skill could enhance on-current of pass-gate n-type transistor of 6T SRAM cell, even increasing read/write speed time at low voltage read/write operation.



Fig. 3-14 Timing diagram for TSUWL during read cycle

Fig. 3-15 shows the SPICE simulation result for TSUWL with different delay time without BWL. The operating time of PMOS (M3) is decided by delay chain. The maximum delay time of delay chain could have more time to suppress Word-Line level and then to reduce driving capability of pass-gate n-type transistor of 6T SRAM cell to improve cell stability (RSNM) [3-2]. By the way, the TSUWL circuit technique is very suitable for SRAM compiler with delay chain control circuit because we choose the different delay time for different Local Bit-Line length. For long Bit-Line length, we could use maximum delay time to postpone suppress Word-Line level.



THE REPORT OF A TABLE

Fig. 3-15 Spice simulation results for TSUWL with different delay time

Fig. 3-16 shows to compare the read speed with previous (Fig. 3-12 (a) (b) (c) (d)) and propose (Fig. 3-13) work in various process corners and temperature (Initial rising edge = 100p, PTNT, Temperature = 25° C). The read speed is measured from 50% of VDD of Word-Line voltage (rise) to 50% of VDD of Local Bit-Line (fall). We assume these technique use the same area of Word-Line driver to drive the same Word-Line length (32 bits) and Local But-Line length (16 bits). Our proposed technique has faster read speed, because the TSUWL technique use BWL to gain faster read speed. Besides, Word-Line level and driving capability almost doesn't decrease.



Fig. 3-16 Spice simulation results for read speed comparison of propose and precious

Fig. 3-17 shows to compare the read speed with previous (Fig. 3-12 (b) (c)) and propose (Fig. 3-13) work in various process corners and temperature. The TSUWL technique has significantly faster Word-Line rising time than previous work. Due to previous work utilize resistance and NMOS (M1) to form a divided to decrease Word-Line level and rising time (Fig. 3-12 (b)) or weakly PMOS (M4) to decrease WL slew-rate (Fig. 3-12 (c)), so Word-Line rising time is longer than the TSUWL technique. Fig. 3-18 shows to compare the butterfly curve with TSUWL and BLUD. Fig. 3-19 shows to compare the Read Margin (RM) with TSUWL and BLUD. If TSUWL and BLUD are to combine to reduce Read-Disturb, we could get the best improvement. At low supply voltage (PSNF @125°C) in read operation, the RM with TSUWL and BLUD technique has 37mV improvement.



Fig. 3-18 Spice simulation results for butterfly curve improvement with 3-σ of variation comparison of TSUWL and BLUD



Fig. 3-20 Spice simulation results for Read Margin (RM) improvement with 3- σ variation comparison of TSUWL and BLUD

3.4 Proposed Adaptive Voltage Detector (AVD) Technique

With the scaling into the deep sub-micron of process, tine gate dielectric will become flimsy. Such as EOT ≤ 1.8 nm at 90nm, EOT ≈ 0.9 nm at 32nm, and EOT \approx 0.65-0.75nm at 22nm. So, the Gate Dielectric Reliability (GDR) must be considered. In this work, we propose an Adaptive Voltage Detector (AVD) circuit technique. The basic concept of Adaptive Voltage Detector is the voltage higher than designed voltage, the Booster will be not enabled; when voltage lower than designed voltage, the Booster will be enabled. In our desired features, this scheme is all digital-based circuit on comparison of a voltage generated by diode-connected resistor-loaded transistor circuit and the trip voltage of a reference inverter (Fig. 3-21). Then, it is a binary decision for boosting action as follow: 1) VDD higher than designed voltage \rightarrow Booster "Off", 2) VDD lower than designed voltage \rightarrow Booster "On". Besides, it also has an inherently variation tolerant as follow: 1) Booster off at lower VDD for FF corner, 2) Booster off at lower VDD for FF corner.



Fig. 3-21 Adaptive Voltage Detector (AVD) circuit

This circuit operating flow: chip-selected bar signal is used generate ST, CLKB and CLK signal to enable the detector circuit, then CLKB and CLK signal generate a pulse width signal to enable M2 and M5. Next, compare VD0 and V_{inv} , if VD0 lower than V_{inv} , the booster is enabled. Besides, in order to enhance judgment at PFNF, PTNT and PSNS corner variation, we could use these option pins which are OSPD<0>, OSPD<1> and OSPD<2> to solve corner variation. Fig. 3-22 shows timing diagram for AVD during read/write cycle.



Fig. 3-22 Timing diagram for AVD during read/write cycle

3.5 Macro Implementation and Simulation Result

In this work, we reserve Adaptive-Data-Aware Write-Assist (ADAWA) circuit technique (Fig. 3-23) to improve write ability and WM. M1 and M2 is a pair of power switch which is controlled by ADAWA_WEB (ADAWA_WEB means write enable signal and high-active during write cycle). M1 and M2 are turned off if ADAWA would enable in write cycle. M3 and M4 is also another pair of power switch which is

controlled by Bit-Line, so only the power switch of write "0" side will be turned off during write operation. We only turn off one power switch (M3 or M4), so the PMOS pull-up ability and latch feedback effect of selected column won't be weakened.



Fig. 3-23 Adaptive-Data-Aware Write-Assist (ADAWA) circuit of 6T SRAM

Fig. 3-24 shows the Proposed ADAWA_WEB tracking control circuit. M2 is used to track pass-gate n-type transistor of 6T SRAM cell. M3, M4, M5 and M6 are used to track pull-up p-type transistor of 6T SRAM cell. VCS_LOAD is a dummy cell. Fig. 3-25 shows the timing diagram for ADAWA during write cycle. During write operation, AW0 signal goes low due to SELE and WE signal goes high, so the VCS_LOAD voltage can discharge by M2 with M3, M4, M5 and M6. When AW1 signal goes low, ADAWA_WEB signal goes low to turn on M1 and M2 power switch that means write operation finished. This tracking control circuit can cover five corner variations. At the write worst case which is PFNS corner, the VCS of write column cell must drop more voltage to enhance write ability. So we can choose OSD option pins to make discharge of VCS_LOAD voltage become slow. For PSNF corner, the VCS of write column cell do not need to drop more voltage to enhance write ability. So we can choose OSD option pins to make discharge of VCS_LOAD voltage become fast to protect HSNM of the column half-selected cell.



Fig. 3-24 Proposed ADAWA_WEB tracking control circuit



Fig. 3-25 Timing diagram for ADAWA during write cycle

We measure AC Write Margin (ACWM) with the Word-Line pulse width at 1ns. Fig. 3-26 shows the ADAWA technique improves AC Write Margin (ACWM), V_{min} with 3- σ variation. At the worst case which is PFNS and high temperature (125°C), ADAWA could aggressively improve the ACWM about 48mV at supply 1.1V. Fig. 3-27 shows the ADAWA technique improves WSNM with 3- σ variation large than original. Fig. 3-28 shows the ADAWA technique improves Write time with 3- σ variation. At supply 0.8V, the write time could reduce about 19% with the ADAWA technique. But, at high supply voltage, the Write time is almost the same as the original. Fig. 3-29 shows to compare the Write Margin (WM) with TSUWL and ADAWA.



Fig. 3-26 The ADAWA technique improves AC Write Margin (ACWM), V_{min} with



Fig. 3-27 The ADAWA technique improves WSNM with $3-\sigma$ variation



Fig. 3-28 The ADAWA technique improves Write time with 3- σ variation



Fig. 3-29 Spice simulation results for Write Margin (WM) improvement with 3-σ variation comparison of TSUWL and ADAA

We have fabricated a 1.0Mb SRAM test macro with the large signal sensing for BLUD, TSUWL, AVD and ADAWA circuit technique using the 6T SRAM cell size of $0.303\mu m^2$ with a single power supply in the 40-nm advanced Low-Standby-Power bulk CMOS technology. In this chip, there are 8192 Word-Lines and 1024 Columns with inter-leaving 16. This macro consists of 64 I/O bits and a column multiplex of 16 per one data-I/O. Local Word-Line length is 32 bits and local bit line partition is 16
bits. The feature of fabricated SRAM macros is summarized in Table I. Fig. 3-30 shows the critical path of 1.0Mb two stages pipeline 6T SRAM macro.



Fig. 3-30 Critical path of 1.0Mb two stages pipeline 6T SRAM macro

TABLE I

Technology	40-nm LP bulk CMOS			
Number of cells / LBL	16			
Cell size	0.303µm ²			
SRAM size	1Mbit (8192 x 128)			
Supply voltage	0.6 - 1.5 V			
Access time 1.1ns @1.1V 25°C PTNT				

Feature of the fabrication SRAM macro

Fig. 3-31 shows the Local Read/Write circuit, also called Local-Evaluation circuit (LEV). During read operation, when CLK rising edge goes high, Word-Line signal, PC and PCHP goes high with PRESA and PCHN signal goes low to make Local Bit-Line floating. P1 signal must delay few times to go low for protecting data "1". Then, the storage node data of 6T SRAM cell is passed into Local Bit-Line and to use sensing PMOS (M24) to pass data into GBL. M17 and M18 (cross couple pair) are used to separate Bit-Line pair. During write operation, when CLK rising edge goes high, Word-Line signal, PC and PCHP goes high with PRESA, YMUX and PCHN signal goes low to make write data signal pass into Local Bit-Line. M1 to M8 are our write assist circuit (ADAWA). M9 to M14 are our BLUD and pre-charge circuit. Fig. 3-32 shows a read path (Word-Line to Output latch). In order to improve read speed, Bit-Line latch is use to latch data at the positive cycle of latch signal and we utilize the negative cycle of latch signal to sense GBL data to output latch (L1).



Fig. 3-31 Local Evaluation Circuit (LEV)



Fig. 3-32 Read path (Word-Line to Output latch)

The layout plot of the proposed 1Mb two stages pipeline 6T SRAM macro is shown in Fig. 3-33. This macro area size of the 1.0Mb 6T SRAM is 2966.80um x 1412.84um. The area penalty of the read and write assist circuit is about 36.36%. From the comparison of Table II, we can see that the increased area overhead of the TSUWL, AVD, BLUD, and ADAWA is 26.24%, 0.01%, 3.54%, and 6.57%.



Fig. 3-33 Layout view of test chip

TABLE I

Area comparison

TABLE II Area comparison						
	Original	W/ TSUWL	W/ AVD	W/ BLUD	W/ ADAWA	
Area	2.732m ²	3.450 m ²	2.733m ²	2.829m ²	2.912m ²	
%	1	1+26.24%	1+0.01%	1+3.54%	1+6.57%	

Fig. 3-34 shows the write/read "0" simulated waveforms of a 1.0Mb two stages pipeline 6T SRAM in read/write cycle in the typical case (PTNT corner, 1.1V, $(@25^{\circ}C)$). Fig. 3-35 shows the write/read "1" simulated waveforms of a 1.0Mb two stages pipeline 6T SRAM in read/write cycle in the typical case (PTNT corner, 1.1V, $(@25^{\circ}C)$). Output data (DO) becomes available after next 2 cycles and it's only available within 1 cycle.



Fig. 3-35 Simulation waveform in Write/Read "1" operation

3.6 Test Flow

This section introduces the test flow of implemented chip. Fig. 3-36 and Table III show the test flow of 1.0Mb two stages pipeline 6T SRAM. This flow provides a solution when we encounter any error. According to the test flow, we can get all the information we want on this chip.



3.7 Implementation and Measurement Result of Test Chip

A 1.0Mb test chip is fabricated using UMC 40nm advanced Low-Standby-Power (LP) bulk CMOS technology. Fig. 3-37 shows the die photo.



Fig. 3-38 shows measured error free full functionality die yield (without redundancy) versus VDD (=VCC) for FF (58 dies), TT (65 dies), and SS (53 dies) corners (without read/write assist technique). At 0.7V, we still have die yield of about 70% (FF) and 30% (TT). The V_{MIN} of this SRAM is limited by Write operation.



Fig. 3-38 Measured error free full functionality die yield (without redundancy) versus VDD (=VCC) for FF (58 dies), TT (65 dies), and SS (53 dies) corners (without read/write assist technique)

Fig. 3-39 shows with TSUWL and ADAWA technique. At 0.7V, we still have die yield of about 40% (FF) and 15% (TT). Fig. 3-40 shows with TSUWL and BLUD technique. At 0.7V, we still have die yield of about 60% (FF) and 20% (TT).



Fig. 3-39 Measured error free full functionality die yield (without redundancy) versus VDD (=VCC) for FF (58 dies), TT (65 dies), and SS (53 dies) corners (with



Fig. 3-40 Measured error free full functionality die yield (without redundancy) versus VDD (=VCC) for FF (58 dies), TT (65 dies), and SS (53 dies) corners (with TSUWL and BLUD technique)



Fig. 3-41 Measured Bit Failure Rate (BFR) at TT corner (Write-Assist: TSUWL and ADAWA; Read-Assist: TSUWL and BLUD)



Fig. 3-42 Measured Bit Failure Rate (BFR) at FF corner (Write-Assist: TSUWL and ADAWA; Read-Assist: TSUWL and BLUD)



Fig. 3-43 Measured Bit Failure Rate (BFR) at SS corner (Write-Assist: TSUWL and ADAWA; Read-Assist: TSUWL and BLUD)

Fig. 3-41 shows Bit Failure Rate (BFR) at TT corner (Write-Assist: TSUWL and ADAWA; Read-Assist: TSUWL and BLUD). When read/write operation with write-assist technique or read-assist technique, the BFR is better than no technique. Fig. 3-42 shows Bit Failure Rate (BFR) at FF corner. Fig. 3-43 shows Bit Failure Rate (BFR) at SS corner. Fig.3-44 shows failure bit count improvement with TSUWL and Boosting WL technique at three corners which are TT, FF, and SS corner. Fig.3-45 shows failure bit count improvement with TSUWL and BLUD technique at three corners which are TT, FF, and SS corner. Fig. 3-45 shows failure bit count improvement with TSUWL and BLUD technique at three corners which are TT, FF, and SS corner.



Fig. 3-44 Measured Failure Bit Count Improvement with TSUWL and Boosting WL



Fig. 3-45 Measured Failure Bit Count Improvement with TSUWL and BLUD technique

Chapter 4 Conclusions

In this thesis, we presented a high-performance 1.0Mb 6T SRAM using 40nm Low Power (LP) 1P9M CMOS technology. Banking architecture, hierarchical WL, and hierarchical BL were used to improve the access performance. Large signal sensing for BLUD and TSUWL were utilized to mitigate Read-Disturb and Half-Selected Disturb while maintaining adequate sensing margin. AVD was used to mitigate gate dielectric over-stress with booster while maintaining adequate gate dielectric reliability. ADAWA was used to improve write ability while maintaining adequate WM and WSNM. The SRAM operated from 1.5V down to 0.6V. The operating frequency is <u>900MHz@1.1V</u>.

277 TITLE

Reference of Chapter 2

- [2-1] Adel S. Sedra, Kenneth C. Smith, "Microelectronic Circuits" 5rd ed. Oxford University Press, 2003.
- [2-2] Ching-Te Chuang, S. Mukhopadhyay, Jae-Joon Kim, Keunwoo Kim, and R. Rao, "High-Performance SRAM in Nanoscale CMOS: Design Challenges and Techniques," *IEEE International Workshop on Memory Technology, Design and Testing*, 2007, pp.4-12.
- [2-3] T. Fischer, E. Amirante, P. Huber, T. Nirschl, A. Olbrich, M. Ostermayr, and D. Schmitt-Landsiedel, "Analysis of read current and write trip voltage variability from a 1-mb sram test structure," *IEEE Trans on Semiconductor Manufacturing*, 2008, vol. 21, no.4, pp. 534–541.
- [2-4] J. Wang, S. Nalam, and B.H. Calhoun, "Analyzing static and dynamic write margin for nanometer SRAMs," *IEEE International Symposium on Circuits and Systems*, 2008, pp.129-134.
- [2-6] Sridhar Ramalingam, Elakkumanan Praveen, Natarajan Sreedhar,
 "Tutorial 6: Design Challenges and Solutions for Nanoscale Memories",
 IEEE International Symposium on Circuits and Systems, 2007, nil28 nil29
- [2-6] S. H Dhong, O. Takahashi, M. White, T. Asano, T. Nakazato, J. Silberman, A. Kawasumi, and H. Yoshihara, "A 4.8GHz fully pipelined embedded SRAM in the streaming processor of a CELL processor," *IEEE International Solid-State Circuits Conference*, 2005, pp.486-612.
- [2-7] A. Bhavanagarwala, X. Tang, and J. Meindl, "The impact of intrinsic

device fluctuations on CMOS SRAM cell stability," *IEEE Journal of Solid State Circuits*, Apr. 2001, vol. 36, no. 4, pp. 658-665.

- [2-8] M. Yamaoka, and H. Onodera, "A Detailed Vth-Variation Analysis for Sub-100-nm Embedded SRAM Design," *IEEE International SOC Conference*, 2006, pp.315-318.
- [2-9] 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*, April 2007, vol.42, no.4, pp.820-829.
- [2-10] J. Davis, D. Plass, P. Bunce, Y. Chan, A. Pelella, R. Joshi, A. Chen, W. Huott, T. Knips, P. Patel, K. Lo, and E. Fluhr, "A 5.6GHz 64kB Dual-Read Data Cache for the POWER6TM Processor," *IEEE International Solid-State Circuits Conference*, 2006, pp.2564-2571.
- [2-11] D. W. Plass, and Y. H. Chan, "IBM POWER6 SRAM arrays," IBM Journal of Research and Development, Nov. 2007, vol.51, no.6, pp.747-756.
- [2-12] J. Pille, C. Adams, T. Christensen, S. Cottier, S. Ehrenreich, T. Kono, D. Nelson, O. Takahashi, S. Tokito, O. Torreiter, O. Wagner, and D. Wendel, "Implementation of the CELL Broadband Engine in a 65nm SOI Technology Featuring Dual-Supply SRAM Arrays Supporting 6GHz at 1.3V," *IEEE International Solid-State Circuits Conference*, 2007, pp.322-606.
- [2-13] J. Pille, C. Adams, T. Christensen, S. Cottier, S. Ehrenreich, T. Kono, D. Nelson, O. Takahashi, S. Tokito, O. Torreiter, O. Wagner, and D. Wendel,

"Implementation of the Cell Broadband Engine[™] in 65 nm SOI Technology Featuring Dual Power Supply SRAM Arrays Supporting 6 GHz at 1.3 V," *IEEE Journal of Solid-State Circuits*, Jan. 2008, vol.43, no.1, pp.163-171.

- [2-14] M. Yabuuchi, K. Nii, Y. Tsukamoto, S. Ohbayashi, and Y. Nakase, H. Shinohara, "A 45nm 0.6V cross-point 8T SRAM with negative biased read/write assist," *Symposium on VLSI Circuits*, 2009, pp.158-159.
- [2-15] A. Raychowdhury, B. Geuskens, J. Kulkarni, J. Tschanz, K. Bowman, T. Karnik, Shih-Lien Lu, V. De, and M.M. Khellah, "PVT-and-aging adaptive wordline boosting for 8T SRAM power reduction," *IEEE International Solid-State Circuits Conference*, 2010, pp.352-353.
- [2-16] K. Takeda, T. Saito, S. Asayama, Y. Aimoto, H. Kobatake, S. Ito, T. Takahashi, K. Takeuchi, M. Nomura, and Y. Hayashi, "Multi-step word-line control technology in hierarchical cell architecture for scaled-down high-density SRAMs," *IEEE Symposium on VLSI Circuits*, 2010, pp.101-102.
- [2-17] K. Takeda, T. Saito, S. Asayama, Y. Aimoto, H. Kobatake, S. Ito, T. Takahashi, K. Takeuchi, M. Nomura, and Y. Hayashi, "Multi-Step Word-Line Control Technology in Hierarchical Cell Architecture for Scaled-Down High-Density SRAMs," *IEEE Journal of Solid-State Circuits*, April 2011, vol.46, no.4, pp.806-814.
- [2-18] H. Yamauchi, "A Discussion on SRAM Circuit Design Trend in Deeper Nanometer-Scale Technologies," *IEEE Transactions on Very Large Scale Integration Systems*, May 2010, vol.18, no.5, pp.763-774.
- [2-19] K. Nii, M. Yabuuchi, Y. Tsukamoto, S. Ohbayashi, Y. Oda, K. Usui, T. Kawamura, N. Tsuboi, T. Iwasaki, K. Hashimoto, H. Makino, and H.

Shinohara, "A 45-nm single-port and dual-port SRAM family with robust read/write stabilizing circuitry under DVFS environment," *IEEE Symposium on VLSI Circuits*, 2008, pp.212-213.

- [2-20] Y. Fujimura, O. Hirabayashi, T. Sasaki, A. Suzuki, A. Kawasumi, Y. Takeyama, K. Kushida, G. Fukano, A. Katayama, Y. Niki, and T. Yabe, "A configurable SRAM with constant-negative-level write buffer for low-voltage operation with 0.149µm2 cell in 32nm high-k metal-gate CMOS," *IEEE International Solid-State Circuits Conference*, 2010, pp.348-349.
- [2-21] H. Pilo, I. Arsovski, K. Batson, G. Braceras, J. Gabric, R. Houle, S. Lamphier, F. Pavlik, A. Seferagic, Liang-Yu Chen, Shang-Bin Ko, and C. Radens, "A 64Mb SRAM in 32nm High-k metal-gate SOI technology with 0.7V operation enabled by stability, write-ability and read-ability enhancements", *IEEE International Solid-State Circuits Conference*, 2011, pp.254-256.
- [2-22] M. Khellah, Nam Sung Kim, J. Howard, G. Ruhl, Yibin Ye, J. Tschanz, D. Somasekhar, N. Borkar, F. Hamzaoglu, G. Pandya, A. Farhang, K. Zhang, V. De, "A 4.2GHz 0.3mm2 256kb Dual-V/sub cc/ SRAM Building Block in 65nm CMOS," *IEEE International Solid-State Circuits Conference*, 2006, pp.2572-2581.
- [2-23] Kim Keejong, H. Mahmoodi, K. Roy, "A Low-Power SRAM Using Bit-Line Charge-Recycling," *IEEE Journal of Solid-State Circuits*, Feb. 2008, vol.43, no.2, pp.446-459.
- [2-24] M. Alam and S. Mahapatra, "A comprehensive model of PMOS NBTI degradation," Microelectron. Reliab., vol. 45, no. 1, pp. 71–81, 2005.

Reference of Chapter 3

- [3-1] International Technology Roadmap for Semiconductors, ITRS, http://public.itrs.net
- [3-2] Ching-Te Chuang, S. Mukhopadhyay, Jae-Joon Kim, Keunwoo Kim, and R. Rao, "High-Performance SRAM in Nanoscale CMOS: Design Challenges and Techniques," *IEEE International Workshop on Memory Technology, Design and Testing*, 2007, pp.4-12.
- [3-3] E. Josse, S. Parihar, O. Callen, P. Ferreira, C. Monget, A. Farcy, M. Zaleski, D. Villanueva, R. Ranica, M. Bidaud, D. Barge, C. Laviron, N. Auriac, C. Le Cam, S. Harrison, S.Warrick, F. Leverd, P. Gouraud, S. Zoll, F. Guyader, E. Perrin, E. Baylac, J. Belledent, B. Icard, B. Minghetti, S. Manakli, L. Pain, V. Huard, G. Ribes, K. Rochereau, S. Bordez, C. Blanc, A. Margain, D. Delille, R. Pantel, K. Barla, N. Cave, and M. Haond, "A cost-effective low-power platform for the 45-nm technology node," *International Electron Devices Meeting*, 2006, pp. 1–4.
- [3-4] H. Fukutome, Y. Momiyama, T. Kubo, E. Yoshida, H. Morioka, M. Tajima, and T. Aoyama, "Suppression of Poly-Gate-Induced Fluctuations in Carrier Profiles of Sub-50nm MOSFETs", *International Electron Devices Meeting*, 2006, pp. 1–4.
- [3-5] T. Hayashi, M. Mizutani, M. Inoue, J. Yugami, J. Tsuchimoto, M. Anma,
 S. Komori, K. Tsukamoto, Y. Tsukamoto, K. Nii, Y. Nishida, H. Sayama,
 T. Yamashita, H. Oda, T. Eimori, and Y. Ohji, "Vth-tunable CMIS platform with high-k gate dielectrics and variability effect for 45nm node," *International Electron Devices Meeting*, 2005, pp. 906–909.

- [3-6] Yi-Wei Lin, "A 55nm 6T SRAM with Variation-Tolerant Word-Line Under-Drive and Data-Aware Write-Assist," *master thesis of Department of Electronics Engineering of National Chiao Tung University*, 2010.
- [3-22] Chi-Shin Chang, "40nm 1.0Mb 6T Pipeline SRAM with Step-Up Word-Line and Adaptive-Data-Aware Write-Assist Design," *master thesis of Department of Electronics Engineering of National Chiao Tung University*, 2011.
- [3-7] S. Ohbayashi, M. Yabuuchi, K. Nii, Y. Tsukamoto, S. Imaoka, Y. Oda, M. Igarashi, M. Takeuchi, H. Kawashima, H. Makino, Y. Yamaguchi, K. Tsukamoto, M. Inuishi, K. Ishibashi, and H. Shinohara, "A 65 nm SoC Embedded 6T-SRAM Design for Manufacturing with Read and Write Cell Stabilizing Circuits", *Symposium on VLSI Circuits*, 2006, pp.17-18.
- [3-8] 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*, April 2007, vol.42, no.4, pp.820-829.
- [3-9] M. Yabuuchi, K. Nii, Y. Tsukamoto, S. Ohbayashi, S. Imaoka, H. Makino, Y. Yamagami, S. Ishikura, T. Terano, T. Oashi, K. Hashimoto, A. Sebe, G. Okazaki, K. Satomi, H. Akamatsu, H. Shinohara, "A 45nm Low-Standby-Power Embedded SRAM with Improved Immunity Against Process and Temperature Variations," IEEE International Solid-State Circuits Conference, 2007, pp.326-606.
- [3-10] 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*, Jan. 2008, vol.43, no.1, pp.180-191.

- [3-11] K. Takeda, T. Saito, S. Asayama, Y. Aimoto, H. Kobatake, S. Ito, T. Takahashi, K. Takeuchi, M. Nomura, and Y. Hayashi, "Multi-step word-line control technology in hierarchical cell architecture for scaled-down high-density SRAMs," *IEEE Symposium on VLSI Circuits*, 2010, pp.101-102.
- [3-12] K. Takeda, T. Saito, S. Asayama, Y. Aimoto, H. Kobatake, S. Ito, T. Takahashi, K. Takeuchi, M. Nomura, and Y. Hayashi, "Multi-Step Word-Line Control Technology in Hierarchical Cell Architecture for Scaled-Down High-Density SRAMs," *IEEE Journal of Solid-State Circuits*, April 2011, vol.46, no.4, pp.806-814.
- [3-13] A. Kawasumi, T. Yabe, Y. Takeyama, O. Hirabayashi, K. Kushida, A. Tohata, T. Sasaki, A. Katayama, G. Fukano, Y. Fujimura, and N. Otsuka, "A Single-Power-Supply 0.7V 1GHz 45nm SRAM with An Asymmetrical Unit-ÄŸ-ratio Memory Cell," *IEEE International Solid-State Circuits Conference*, 2008, pp.382-622.
- [3-14] K. Zhang, U. Bhattacharya, Z. Chen, F. Hamzaoglu, D. Murray, N. Vallepalli, Y. Wang, B. Zheng, and M. Bohr, "A 3-GHz 70Mb SRAM in 65nm CMOS Technology with Integrated Column-Based Dynamic Power Supply," *IEEE International Solid-State Circuits Conference*, 2005, pp. 474-475.
- [3-15] M. Yamaoka, N. Maeda, Y. Shinozaki, Y. Shimazaki, K. Nii, S. Shimada,K. Yanagisawa, and T. Kawahara, "Low-Power Embedded SRAM

Modules with Expanded Margins for Writing," *IEEE International Solid-State Circuits Conference*, 2005, pp. 480-481.

- [3-16] T. Suzuki, H. Yamauchi, Y. Yamagami, K. Satomi, H. Akamatsu, "A Stable SRAM Cell Design Against Simultaneously R/W Disturbed Accesses," *Symposium on VLSI Circuits*, 2006, pp. 11-12.
- [3-17] Meng-Fan Chang, Jui-Jen Wu, Kuang-Ting Chen, Hiroyuki Yamauchi, "A Differential Data Aware Power-supplied (D2AP) 8T SRAM Cell with Expanded Write/Read Stabilities for Lower VDDmin Applications," *Symposium on VLSI Circuits*, 2009, pp. 156-157.
- [3-18] Meng-Fan Chang, Jui-Jen Wu, Kuang-Ting Chen, Yung-Chi Chen, Yen-Hui Chen, R. Lee, Hung-Jen Liao, H. Yamauchi, "A Differential Data Aware Power-supplied (D2AP) 8T SRAM Cell with Expanded Write/Read Stabilities for Lower VDDmin Applications," *IEEE Journal* of Solid-State Circuits, June 2010, vol.45, no.6, pp.1234-1245.
- [3-19] Y. Fujimura, O. Hirabayashi, T. Sasaki, A. Suzuki, A. Kawasumi, Y. Takeyama, K. Kushida, G. Fukano, A. Katayama, Y. Niki, and T. Yabe, "A configurable SRAM with constant-negative-level write buffer for low-voltage operation with 0.149µm2 cell in 32nm high-k metal-gate CMOS," *IEEE International Solid-State Circuits Conference*, 2010, pp.348-349.
- [3-20] H. Pilo, I. Arsovski, K. Batson, G. Braceras, J. Gabric, R. Houle, S. Lamphier, F. Pavlik, A. Seferagic, Liang-Yu Chen, Shang-Bin Ko, and C. Radens, "A 64Mb SRAM in 32nm High-k metal-gate SOI technology with 0.7V operation enabled by stability, write-ability and read-ability enhancements", *IEEE International Solid-State Circuits Conference*, 2011, pp.254-256.

- [3-21] J. Pille, D. Wendel, O. Wagner, R. Sautter, W. Penth, T. Froehnel, S. Buetter, O. Torreiter, M. Eckert, J. Paredes, D. Hrusecky, D. Ray, M. Canada, "A 32kB 2R/1W L1 data cache in 45nm SOI technology for the POWER7TM processor," *IEEE International Solid-State Circuits Conference*, 2010, pp.344-345.
- [3-23] Kim Keejong, H. Mahmoodi, K. Roy, "A Low-Power SRAM Using Bit-Line Charge-Recycling," *IEEE Journal of Solid-State Circuits*, Feb. 2008, vol.43, no.2, pp.446-459.
- [3-24] Hao-I Yang, Yi-Wei Lin, Mao-Chih Hsia, Geng-Cing Lin, Chi-Shin Chang, Yin-Nien Chen, Ching-Te Chuang, Wei Hwang, Shyh-Jye Jou, Nan-Chun Lien, Hung-Yu Li, Kuen-Di Lee, Wei-Chiang Shih, Ya-Ping Wu, Wen-Ta Lee, and Chih-Chiang Hsu, "High-Performance 0.6V VMIN 55nm 1.0Mb 6T SRAM with Adaptive BL Bleeder," Proc. 2012 IEEE International Symposium on Circuits and Systems (ISCAS), Seoul, Korea, May 20-23, 2012, pp. 1831-1834.
- [3-25] Yi-Wei Lin, Hao-I Yang, Geng-Cing Lin, Chi-Shin Chang, Ching-Te Chuang, Wei Hwang, Chia-Cheng Chen, Willis Shih, Huan-Shun Huang, "A 55nm 0.55V 6T SRAM with Variation-Tolerant Dual-Tracking Word-Line Under-Drive and Data-Aware Write-Assist," Proc. 2012 IEEE International Symposium on Low Power Electronics and Design (ISLPED), Redondo Beach, CA, USA, July 30 August 1, 2012, pp. 79-84.
- [3-26] Yi-Wei Lin, Ming-Chien Tsai, Hao-I Yang, Geng-Cing Lin, Shao-Cheng Wang, Ching-Te Chuang, Shyh-Jye Jou, Wei Hwang, Nan-Chun Lien, Kuen-Di Lee and Wei-Chiang Shih, "An All-Digital Read Stability and Write Margin Characterization Scheme for CMOS 6T SRAM Array,"

2012 IEEE International Symposium on VLSI Design, Automation and Test (VLSI-DAT), Hsinchu, Taiwan, April 23-25, 2012

