國立交通大學

電子工程學系 電子研究所碩士班

碩士論文

超低功率抗雜訊 8T 靜態隨機存取記憶體的設計與實



Design of Ultra-Low-Power Disturb-Free Cross-Point 8T SRAMs

研究生:夏茂墀

指導教授: 莊景德 教授

中華民國九十九年十一月

國立交通大學

電子工程學系 電子研究所碩士班

碩士論文

超低功率靜態隨機存取記憶體的設計與實現 ES

Design of Ultra-Low-Power Disturb-Free Cross-Point 8T

SRAM

研究生: 夏茂墀

指導教授: 莊景德 教授

中華民國九十九年十一月

超低功率靜態隨機存取記憶體的設計與實現

Design of Ultra-Low-Power Disturb-Free Cross-Point 8T SRAMs

研究生: 夏茂墀 Student: Mao-Chih Hsia

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



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 EngineeringSeptember 2010 Hsinchu, Taiwan, Republic of China

中華民國九十九年十一月

超低功率靜態隨機存取記憶體的設計與實現

學生: 夏茂墀

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

低功率靜態隨機存取記憶體(SRAM)設計在手持式裝置高使用率的加持下已 經漸漸成為主流,因為手提式製品都需要低功率晶片以延長工作時間。根據 ITRS2005的預測,嵌入式靜態隨機存取記憶體佔整顆晶片面積將可達到90%, 意思也就是說減少 SRAM 的功率損耗是降低晶片耗能最直接的辦法。

摘要

本篇論文提出一種降低單端讀取 8T-SRAM 陣列功率損耗的架構,藉由取得 陣列 keeper 與記憶體單元漏電流的平衡,可以使記憶體陣列的 virtual cell supply 降低。這樣將可以同時達到降低記憶體陣列的能量號損與改善記憶體單元的寫 入能力。HSNM 由此論文建構的一套選擇 keeper size 檢測法,將被保證是安全 無虞的。我們更提出了一個新的架構改善讀取動作的能量損耗。本篇論文的 ULP-SRAM 與 ALP-SRAM 在跟對照組 Novel 8T 比較下,皆可達到低功率的目 的且最低操作電壓可達 0.45V。

i

Design of Ultra-Low-Power Disturb-Free Cross-Point 8T SRAMs

Student : Mao-Chih Hsia

Advisor : Prof. Ching-Te Chuang

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

ABSTRACT

Low power design in Static Random Access Memory (SRAM) has become one of the mainstreams as a respond to the increasing usage of handheld device in that portable device requires a less power consumption chip to extend its working time. Nearly 90% area of a chip will be occupied by embedded SRAM in 10 years to come in accordance with ITRS2005 predictions which means that diminish power on SRAM will directly lead to chip power reduction.

This thesis presents a power control technique to minimize the array power consumption of single-ended Read/Write 8T SRAM. By obtaining current balance between keepers of virtual cell supply and cell leakage, it allows a fall of virtual cell supply which is power supply of cell array. That leads to reduction of array power and improvement on write ability. Half Select Noise Margin (HSNM) is guaranteed safe by constructing an algorithm to size virtual cell supply keepers. To further improve read power, another structure is proposed to charge the virtual cell supply of selected column solely instead of all columns of selected rows. Both ALP and ULP SRAM achieves the power saving purpose compared to novel 8T SRAM and its VCC_{MIN} is 0.45V.

誌 謝

本篇論文可以順利完成,要感謝的人很多,首先要感謝的就是我的指導教授莊景德教授,在兩年的研究所生涯裡面不僅提供了我氣氛良好與資源充足的研究環境,並且在專業領域上給我悉心的指導,讓我能夠大幅成長,能夠充分發揮。 最重要的是在老師身上,我看到了對工作的熱誠與責任,待人的謙和與真誠,比 起專業知識,這將更使我終生獲益。

再來要感謝的就是 LPMD Lab 的楊皓義學長,從碩一上開始便帶領我學習 SRAM 專業知識,陪我熬夜 Debug,讓我在短時間內就可以上手。 另外感謝 LAB 同學 林宜緯、謝建宇、蔡銘謙、林勇維 等多位同學在這段期間的互相討論與砥 礪,也感謝 MSCS Lab 王紹丞學弟在我做論文研究期間給予的協助。感謝智原科 技的學長姐們(Jack、Patrick、Sevic、Eric、Brenda、信吉、Willis)在科專計 畫中給予的指導與協助,讓我可以提早了解業界的考量並從業界的觀點來設計晶 片。感謝所有參與我研究所生活的同學與學弟妹,陪我打球、打 BG、看電影、 烤肉、健身,讓我這兩年非常多彩多姿。

最後感謝我的父母從小到大給予的照顧與支持,讓我可以走出一條精彩的道路。

Contents

Chapter	1 Introduction1
1.1	Background1
1.2	Motivation1
1.3	Thesis Organization2
Chapter	2 Review of Low-Power SRAM in recent years3
2.1	Introduction
2.2	Power consumption
2.2.	1 The Dynamic Power Dissipation
2.2.2	2 The Static Power Dissipation
2.3	The definitions of reliabilities on SRAM
2.3.	1 Hold-mode Static Noise Margin (HSNM)6
2.3.2	2 Read-mode Static Noise Margin (RSNM)7
2.3.3	3 Write Margin (WM) & Write Ability
2.4	Overview of recent low power SRAM design10
2.4.	1 Error-Tolerant SRAM Design for Ultra-Low Power Standby
	Operation
2.4.2	2 Write power saving by reducing swing of bit-line11
2.4.3	3 AC Current Reduction13
2.4.4	4 Leakage Current Reduction14
2.5	Nanoscale CMOS Design Challenge and Techniques16
2.6	Summary19
Chapter	3 Single cell stage of Ultra Low Power SRAM Design20
3.1	Background20
3.1.	1 Prior Art20
3.1.2	2 Conventional 8T SRAM24
3.1.3	3 Introduction to disturb-free cross-point double-layer pass-gate 8T
	SRAM cell24
3.2	Introduction of ULP design27
3.3	Power Reduce Concept
3.4	Stage1_Single cell analysis
Chapter	4 Stage2_16 x 4 array base simulation and Stage3_Pre-simulation
	Result
4.1	Stage2_16 x 4 array base simulation
4.2	Stage3_Pre-simulation44

Chapter5 Power comparison & Absolute Low Power mode & Test Flow49			
5.1	Absolute Low Power Mode	49	
5.2	Power Comparison and Simulation Result	51	
5.3	Test Flow	58	
5.4	Design implement	61	
Chapter6 Conclusions			
Reference	e	64	
Vita		68	

List of Figures

Fig 2-1	Circuit diagram of dynamic nower 4
Fig. 2 T	The diagram of lookage neth
Fig. 2-2	
F1g. 2-3	HSNM detect circuit/
Fig. 2-4	The hold mode butterfly curve7
Fig. 2-5	The Read mode SNM
Fig. 2-6	The write mode of 6T cell
Fig. 2-7	The butterfly curve of write mode9
Fig. 2-8	Write ability diagram
Fig. 2-9	SNM degrade to zero under DRV10
Fig. 2-10	Measured SRAM leakage current
Fig. 2-11	Write amplifier
Fig. 2-12	2 (a)Write voltage generator (b)shared additional NMOS (c) Write cycle
	waveform
Fig. 2-13	VTH independent BL voltage provider12
Fig. 2-14	(a) Concept of HBLSA SRAM
Fig. 2-14	(b) Architecture of HBLSA SRAM
Fig. 2-15	Row Decoder of three bit input14
Fig. 2-16	A two-stage decoder architecture14
Fig. 2-17	Dual VTH CMOS circuit technique14
Fig. 2-18	(a) Dynamic Leakage Cut-off scheme (b) Its operation15
Fig. 2-19	Global and local variation of VT16
Fig. 2-20	(a) Number of dopant atoms in the channel as function of effective channel
	length (b) "sigma" of VT variation as function of technology node16
Fig. 2-21	Scaling and cell stability margin of 6T SRAM17
Fig. 2-22	Thin cell layout17
Fig. 2-23	Process and temperature variation tolerant Read-Assist Circuits (RAC)18
Fig. 3-1 A	A traditional 6-T SRAM cell structure20

Fig. 3-2 Column Based Dynamic VCC scheme	21
Fig. 3-3 Methods for resolving Read/Write conflict and reducing write failure	22
Fig. 3-4 Floating Power-Line Write scheme	23
Fig. 3-5 Differential Data-Aware VSS scheme	23
Fig. 3-6 Conventional 8T SRAM	24
Fig. 3-7 Proposed novel 8T SRAM cell	24
Fig. 3-8 Novel 8T cell in standby mode	26
Fig. 3-9 Novel 8T cell in read mode	26
Fig. 3-10 Novel 8T cell in write 0 mode	26
Fig. 3-11 Novel 8T cell in write 1 mode	26
Fig. 3-12 Novel 8T cell schematic	28
Fig. 3-13 Desired virtual cell supply	30
Fig. 3-14 Ultra Low Power structure	31
Fig. 3-15 HSNM gets worse if VDDQ and VDDQB aren't identical	32
Fig. 3-16 8T-cell layout black frame rounded are two individual cell supplies	33
Fig. 3-17 ULP in current aspect	33
Fig. 3-18 VT shift of two extreme conditions	35
Fig. 3-19 Single cell simulation on write ability improvements	37
Fig. 3-20 Three criterions of VTL and flip point. "10%VDD+asym" is adopt	oted
criterion	37
Fig. 3-21 The determined VTL is about 100mV higher than VTRIP of inverter	r in
cell	38
Fig. 4-1 Waveform of read write operation	39
Fig. 4-2 Keeper sizing by V _{TL}	40
Fig. 4-316x4 array simulation	41
Fig. 4-4 The lowest virtual cell supply of write operation	41
Fig. 4-5 Monte Carlo result Shows the lowest VDDQ when write	42
Fig. 4-6 Write ability improvement. MAX VRBL for PFNS rise up by 97mV	43
Fig. 4-7 The VRBL increase amount in 5 corners and 4 temperatures	44
Fig. 4-8 Architecture of ULP	45
Fig. 4-9 VDDQ post-simulation waveform	46
Fig. 4-10 Current replica circuit	46
Fig. 5-1 Absolute Low Power Structure	50
Fig. 5-2 Local Evaluation Circuit	50
Fig. 5-3 Array Power Comparison (ALP/ULP vs COM) at 0.6V	54
Fig. 5-4 Split ALP array power into three parts	
- spin - inter and power into an or parts	54
Fig. 5-5 ALP power recalculated without counting INV and NOR	54 55

Fig. 5-7 Array Power Comparison (ALP/ULP vs COM) at 1.0V	56
Fig. 5-8 Total Power Comparison (ALP/ULP vs COM) at 1.0V	56
Fig. 5-9 Virtual cell supply waveform of ALP	57
Fig. 5-10 Two flow of power measuring	58
Fig. 5-11 Test flow of ULP SRAM	59
Fig. 5-12 Test flow of ALP SRAM	60
Fig. 5-13 Floor Plan and layout view of ULP and ALP	62
Fig. 5-14 Power save ratio of ULP SRAM	52
Fig. 5-15 Power amount ratio of ULP SRAM	52
Fig. 5-16 Power save ratio of ALP SRAM	53
Fig. 5-17 Power amount ratio of ALP SRAM	53

List of Tables

4-1 Control signal content
4-2 Truth table of 8T cell and virtual cell supply
5-1 Post Simulation Result



Chapter 1 Introduction

1.1 Background

Based on Moore's law, the density of the chip density is doubled every 18 months. Chip design is getting more complex and large. Low power design has become one of important topics on booming usage of portable devices such as PDA, cell phone, NB... and so on. In deep submicron process, leakage power dissipation usually takes the lion's share of total power consumption and that limits working time of devices whose power are supplied by battery.

By ITRS Roadmap predicting, memory area will occupy 90% of a system on a chip in 10 years to come. The prediction implies that the Static Random Access Memory (SRAM) will dominate both the system power and chip area. In recent years, the bioelectronics, portable consumer IC and embedded application, such as implanted medical instruments and wireless body sensing network, become a kind of important application in SOC system design. Due to limited battery lifetime, the power consumption of SRAM has a great influence on the kinds of SOC application. Diminish power on SRAM will direct lead to power reduction of a chip.

1.2 Motivation

Nowadays, various kinds of electronic devices have become necessities of everyday life. In order to meet different demands of each product, more and more companies design custom SRAM by themselves for particular application instead of making use of general SRAM compiler. The traditional 6T SRAM cannot satisfy the noise toleration in low voltage and nanoscale CMOS process. The core SRAM cell of

this thesis is featured for Read/Write disturb free and that allows a robust operation in low voltage supply. By the help of proposed structure, we can lower array power consumption in a wide range of operating voltage. Moreover, traditional register file is not capable of working with wide-operating -voltage range. Thus a custom SRAM is needed for ultra low voltage devices.

1.3 Thesis Organization

The rest of this thesis is organized as following.

Chapter 2

- 1. Low power SRAM design in recent years
- 2. Overview of power dissipation and definition of cell reliability, SNM,...
- Challenges of SRAM in nanoscale process and low supply voltage 3.

Chapter3

- Introduction to traditional 6T, traditional 8T. 1.
- Basic operation and feature of novel 8T SRAM which is core cell of this 2. Design concept of ULP and ALP. thesis.
- 3.
- 4. Stage1_Single cell analysis

Chapter4

- 1. Stage2_16 x 4 array base simulation
- 2. Stage3_Pre-simulation result

Chapter5

- 1. Absolute Low Power Mode
- 2. **Power Comparison and Simulation Result**
- Test Flow 3.
- 4. Design implement

Conclusions and test flow are drawn on chapter6.

Chapter 2 Review of Low-power SRAM Design in recent years

2.1 Introduction

This chapter offers an introduction to power dissipation and reliability of CMOS. Power dissipation is composed of dynamic power ($P_{DYNAMIC}$) and static power (P_{STATIC}) and is presented in section2.2.

$$P_{\text{TOTAL}} = P_{\text{DYNAMIC}} + P_{\text{STATIC}}$$
(2.1)

Reliability issue is made up of Hold mode Static Noise Margin (HSNM), Read mode Static Noise Margin (RSNM), Write Margin (WM) and write ability. Section 2.4 is an overview of low power SRAM design in recent years. Considerable techniques have been paid to reduce power dissipation in bit-line (LBL) and WL whose capacitance is quite large.

1896

2.2 Power Consumption

2.2.1 The Dynamic Power Dissipation

Generally speaking, dynamic power is made up of switching power, internal power and short-circuit power.

$P_{\text{DYNAMIC}} = P_{\text{SWITCH}} + P_{\text{INTERNAL}} + P_{\text{SHORT-CIRCUIT}}(2.2)$

Switch power is due to data transition (high to low or low to high) in logic gates. Output load capacitance is charged or discharged according to data in (V_{IN}) of logic gate shown in Fig.2-1 and the charge transferred in circuit is referred to as switch power dissipation.

The internal power is caused by parasitic capacitance in circuits such as lump and

couple capacitance. Those capacitances are also charged or discharged in operations and result in power dissipation. When signal is changing to complementary values, there exists a period that both NMOS and PMOS are partially turn on. A short current path is appeared and leads to power consumption during this period. Usually, short circuit power can be expressed as [34]:

 $P_{\text{SHORT-CIRCUIT}} = I_{\text{MEAN}} * V_{\text{DD}}(2.3)$

where I_{MEAN} (mean value of short circuit current) can be expressed as:

$$I_{MEAN} = \frac{1}{12} * \frac{\beta}{V_{DD}} * (V_{DD} - 2V_T)^3 * \frac{\tau}{T} \quad (2.4)$$

Where β is the gain factor(uA/V²) of MOS transistor, V_T is threshold voltage, and τ is the signal rise/fall time. P_{SWITCH} and P_{INTERNAL} are often modeled as:

$$P = \alpha * C * V_{DD}^2 * f$$
 (2.5)

Where f is frequency of switching and C is the capacitance involved. Since dynamic power is data dependent, the switching probability α is introduced. From equation 2.4 and 2.5 we can observe that lowering VDD is a way to reduce P_{DYNAMIC}.



Fig. 2.1: Circuit diagram of dynamic power

2.2.2 The Static Power Dissipation

In ideal conditions, there is no DC current path for logic devices in steady state. However, take non-ideal situation into considerations. Off-state leakage takes a considerable portion of power dissipation which is called static power in deep submicron process. In general, leakage current is composed of four components which are the Gate Induced Drain Leakage (I_{GIDL}), the PN reverse bias junction leakage ($I_{REVERSE}$), the tunneling current of gate oxide (I_G) and the sub-threshold conduction current (I_{SUB}). Leakage power has become not negligible as process scale down.

Fig. 2-2 shows the four sources of leakage. I_{GIDL} results from electric field between overlap area of gate and drain. Both thinner oxide and high supply voltage will increase I_{GIDL} . Though GIDL current becomes less significant for digital logics whose supply voltage is under 1.1V; it is still important issue for DRAM whose data retention time is massively degraded by GIDL current.

Comparing to other three component of leakage, $I_{REVERSE}$ is generally negligible. Junction leakage arises from reverse bias of PN diode junction between drain/source and substrate as illustrated in Fig. 2-2. In nanometer-scale CMOS technologies, gate oxide thickness is getting thinner. The tunneling of electrons from bulk to gate or from gate to bulk forms the leakage current (I_G) and this current usually dominates the total leakage power. Sub-threshold current (I_{SUB}) is dominated by diffusion current in off-state rather than drift current in on-state. I_{SUB} is a function of V_{DS} and supply voltage. Larger V_{DS} will lead to a larger I_{SUB} between drain and source. The following is an expression for I_{SUB} :

$$I_{SUB} = \frac{W}{L} \mu v_{th}^2 C_{sth} e^{\frac{V_{GS} - V_T + \eta V_{DS}}{nV_{th}}} (1 - e \frac{-V_{DS}}{V_{th}}), n = 1 + \frac{C_{sth}}{C_{ox}}$$
(2.6)



Fig. 2.2: The diagram of leakage path

2.3 The definitions of reliabilities on SRAM

Reliability has become a major concern for SRAM design in deeply scaled technologies. Both global/ local mismatch and lowered VDD degrade read /write stability. A safety guaranteed operation of SRAM can be detected through these definitions on stability. Traditional 6T SRAM will be considered as an example in this section.

2.3.1 Hold-mode Static Noise Margin (HSNM)

In hold cycles, word line will be turned off (WL=0). Both BL and BLB are pre-charged high. The cross couple inverter in 6T cell can store a complementary data just like a basic latch. The way to measure HSNM is to connect two DC noise source with 6T cell as in Fig. 2-3. The tolerable noise without flipping storage node is defined as signal noise margin (SNM). Once VN in Fig. 2-3 increased higher than margin, data would be flipped. Fig. 2-4 is the butterfly curve for presenting HSNM graphically. The revolved inverter transfer curve is overlapped with another one. The

corresponding HSNM in this butterfly curve is the maximum square side embedded. Noise in V_L or V_R must be kept below HSNM to ensure the safety of hold mode cell.



Fig. 2.3: HSNM detect circuit



Fig. 2.4: The hold mode butterfly curve

2.3.2 Read-mode Static Noise Margin (RSNM)

Suppose $V_R = 0$ and the cell is performing a read operation. Word line is turned on (WL=VDD). Because of the voltage dividing effect between M4/M5 and M2/M6, V_R is experiencing a noise higher than ground. This will change voltage transfer curve (VTC) shown in Fig. 2-5. Obviously, the square in read mode is smaller than that in hold mode. This implies a worse SNM in read mode. However, the core cell in this

thesis is read disturb free. So there isn't degradation in SNM when read operation is issued.



Fig. 2.6: The write mode of 6T cell

In write cycles, WL is turned on with BL or BLB equal to VDD as in Fig. 2-6. The VTC curve of high bit-line side is the same as in read mode. Another BL offers a strong 0 which results in a different VTC shown in Fig. 2-7. The common definition for write margin is the maximum side of square embedded in these two curves.

Beside write margin, there is another definition called write ability. Write ability of

a cell offers an indication of how easy or hard it is to write a cell. The way we test write ability is to rise BL up little by little in write operation until write fail. Max BL Voltage (V_{BL}) defines the voltage to flip storage node. The higher the V_{BL} is, the easier it is to write the cell. Fig. 2-8 is the waveform of V_R and V_L with BL voltage as x-axis. Node flipped when BL declined from 1V to 500mV. That means data will flip when one of the BLs fall to 500mV. 500mV is the V_{BL} mentioned above.



Fig. 2.7: The butterfly curve of write mode



Fig. 2.8: Write ability diagram

2.4 Overview of recent low power SRAM design2.4.1 Error-Tolerant SRAM Design for Ultra-Low PowerStandby Operation

Dual supply is one method to reduce standby leakage power dissipation. A low standby SRAM V_{DD} is used while retaining memory data in sleep mode. The data retention voltage (DRV) denotes the voltage needed to hold data in SRAM cells. By lowering DRV, the leakage in standby mode decreases. However, DRV has a lot to do with HSNM. Fig. 2-9 shows that when VDD is equal to DRV, the HSNM of SRAM cell will degrade to 0. The techniques proposed in this paper reduce DRV to 255mV and lead to 98% leakage power saved [7]. Fig. 2-10 shows results of leakage reduction. The overheads of this design are: 1.) Balanced P/N ratio helps to gain a larger HSNM under low DRV, yet it also reduces performance in read/write mode. 2.) Length and width of MOS are enlarged by 50% under fixed P/N ratio to mitigate process variation. This has caused 50% area overhead. 3.) Error-correction power overhead is inevitable since Error-Correct-Code (ECC) is used to save low reliability under low DRV.



Fig. 2.9: SNM degrade to zero under DRV



Fig. 2.10: Measured SRAM leakage current.

2.4.2 Write power saving by reducing swing of bit-line

Charging and discharging bit-line (BL) always dominate power in write operation because capacitances in BL are large. Reduce either capacitance or voltage swing of BL is method to reduce write power. Fig. 2-11 shows one technique to reduce voltage swing of BL to half VDD and 75% write power is saved consequently [36]. In this scheme, it is difficult to save more power due to write error problem. Moreover, if read write cycle comes alternatively, there will be extra power wasted owing to mismatch of BL levels in both operaiotns.

Fig. 2-12 proposed another scheme called Sense-Amplifying Cell (SAC) that is able to further save write power by 90% [9]. Bit line swing of this design is only $1/6V_{DD}$. An additional NMOS is connected to memory cell which is off-state in write cycles. This NMOS is sharable along ROW direction. A V_{TH} independent BL voltage provider is shown in Fig. 2-13. When VTH is fluctuated by ±0.15V, ΔV_{BL} fluctuation cab be kept as low as ±30mV.



Fig. 2.12: (a)Write voltage generator (b)shared additional NMOS (c) Write cycle waveform



Fig. 2.13: V_{TH} independent BL voltage provider

Hierarchical bit line and local sense amplifier SRAM (HBLSA SRAM) is another way to reduce write power. This technique reduce write power on bit line by applying full swing to low capacitive bit-line (sub-BL) and low swing to high capacitive bit-line (BL) shown in Fig.2-14 [4]. The hierarchical bit line and local sense amplifier help to reduce the effective bit line capacitance.



2.4.3 AC Current Reduction

Continuously charge and discharge high capacitive node is power consuming. Decoder must work in read or write cycles, therefore reduce active power in decoder is a good method to minimize total power dissipation. Fig. 2-15 shows a row decoder of three bit input. Address line capacitance is large owing to interconnected data bus between NAND gate and INV gate. Since active power is a function of C and f, larger C would cause larger AC power in decoders. Fig. 2-16 is two-stage decoder whose number of transistors, fanin and loading on the data bus are reduced. As a result, both speed and power are optimized.



Fig. 2.15: Row Decoder of three bit input.

2.4.4 Leakage Current Reduction

Low threshold voltage (V_T) is an effective way to reduce dynamic power. Yet it increase leakage power on the other way. Two techniques are proposed to reduce leakage power without degrading chip performance. Fig. 2-17 is Dual-V_{TH} circuit technique.



Fig. 2.17: Dual V_{TH} CMOS circuit technique

SWP and SWN are high V_{TH} devices and are turned off in standby cycles. Leakages are greatly decreased under this structure. SWN and SWP are both gate driven by 1.5V while V_{DD} =1.0V for peripheral circuits. This ensures them considerable conductivity to gain benefit in performance. A technique similar to variable V_{TH} CMOS is shown in Fig. 2-18 called dynamic leakage cut-off scheme (DLC). The non-selected SRAM cells are biased with $2V_{DD}$ for N-WELL and $-V_{DD}$ for P-WELL. In the mean while, selected cells N/P-WELL are biased to V_{DD} and V_{SS} . Through this technique, the V_{TH} of selected cells are low in order to operate high speed. The V_{TH} of unselected cells are high to reduce leakage current.



Fig. 2.18: (a) Dynamic Leakage Cut-off scheme (b) Its operation

2.5 Nanoscale CMOS Design Challenge and Techniques

SRAM is vulnerable to negative impact of CMOS technology scaling such as signal loss due to leakage, V_T scatter owing to process variation and random dopant fluctuation (RDF) and NBTI (Negative Bias Temperature Instability) caused by gate bias and temperature [11]. V_T shift severely limits the scaling of SRAM cell size in that V_T mismatch in cell will exasperate the worse SNM. Global V_T variation affects the correctness of an operation and local V_T shift has influence on SNM of cell (Fig. 2-19). The sigma (σ) of V_T variation also increased by a factor of 4 in 30 nm technology Fig. 2-20.



Fig. 2.19: Global and local variation of V_T



Fig. 2.20: (a) Number of dopant atoms in the channel as function of effective channel length (b) "sigma" of VT variation as function of technology node

The conflicting requirement of read /write is inevitable in SRAM cell. Small β ratio is required to mitigate read disturb which is formed by voltage divider of access pass-transistor NMOS and the pull-down NMOS, However we need a strong access pass-transistor to conduct a robust write operation. Once the read disturb is a larger than inverter trip point, the cell flips. Fig. 2-21 shows that due to V_T variation, cell switch point and read disturb level may overlap in process under 90nm. This means stability of SRAM gets worse in deep submicron technology. There are bunch of methods to solve the reliability problem in nanoscale SRAM:



Fig. 2.21: Scaling and cell stability margin of 6T SRAM

Large signal sense: For higher stability, large signal read-out scheme has been adapt in recent SRAM design unlike small signal sensing in traditional design.



Fig. 2.22: Thin cell layout

Thin cell layout: Fig. 2-21 shows uni-directional poly thin cell layout which is widely used in sub-90nm technology to reduce bit line loading for performance and noise immunity. The thin cell layout can improve the yield.

RAC & WAC: To solve the problem in read write operation, read assist circuit (RAC) and write assist circuit (WAC) are proposed. The WL turn on voltage level can be optimized by tracing the local variation of pass-transistor MOS. In fast NMOS corner, WL will turn on below VDD to reduce read disturb. The overhead of this technique is that there always exists a DC current path between P0 and RAT shown in Fig. 2-23.



Fig. 2.23: Process and temperature variation tolerant Read-Assist Circuits (RAC)

Dual supply & Adaptive Read/Write supply

In dual supply design, peripheral critical path circuit supply and cell are set at a level higher than logic block. Supply can also be adaptively switched according to operation issued. High cell supply voltage is used in read cycles for better SNM while low cell supply is offered to improve write margin.

WL pulsing

Proper control of word-line (WL) pulse width is another way to maintain balance between SNM and write ability. WL must be long enough to generate ΔV_{BL} for read-out sensing or write margin; while short enough to mitigate read disturb.

2.6 Summary

This chapter introduces definitions of SNM in each operation by which we are used in our design. Scaling in process has imposed many design challenge such as V_T scattering, NBTI/PBTI, GIDL, leakage and so forth. The scaling of 6T cell has almost come to the end; however many alternative cell are also proposed and are introduced in next chapter. Many skills are used to reduce power consumption. Active and standby power can be cut down in different ways. Capacitance reduction by using divided word-line or single-bit line cross point SRAM cell. Reduce signal swing of high capacitive wires and data bus. AC current can be decreased by multi-stage decoding method. Leakage power is suppressed by DLC, Dual- V_{TH} scheme or Auto-Backgate-Controlled multiple- V_{TH} .

1896

Chapter 3 Single cell stage of Ultra Low Power SRAM Design

3.1 Background

3.1.1 Prior Art

Fig. 3-1 is the schematic of a 6T SRAM cell. Due to process and supply scaling along with conflicting requirement between Read and Write operation, the degradation of RSNM has become the bottleneck of 6T SRAM in sub-100nm technologies.



Fig. 3.1 : A traditional 6-T SRAM cell

Both BL and BLB are precharged to high voltage during read cycles and discharge through (PG2 & PD2) or (PG1 & PD1). The access transistor and pull-down NMOS form a voltage divider that results in Read-disturb. Once the Read-disturb voltage is larger than the V_{TRIP} of the other half-cell inverter, the cell will flip.

In order to reduce Read-disturb, it is desirable to have small β ratio(PG/PD). On the other hand, access transistor must be large enough to facilitate write margin, Write ability. In the following, many skills are introduced to solve these problems. Besides, the drawbacks are also revealed. In this section, the proposed novel 8T SRAM cell and the Ultra-Low-Power structure are sown to overcome these drawbacks.



Fig. 3.2: Column Based Dynamic V_{CC} scheme

Fig. 3-2 shows a column based dynamic V_{CC} scheme [11]. The cell supply is dynamically switched in accordance with operation issued such as: Read (High V_{CC}), Write (Low V_{CC}), Standby (Low V_{CC}). The drawbacks of this structure are:

- (1) 2 external supplies or on-chip regulators are needed.
- (2) Routing 2 supply lines.
- (3) It switches supply for whole cells, and as such, there is large amount of charge to 1896 move; resulting in slow settling time.
- (4) It impedes the pull-up of the "Low" cell storage node, and weakens the latch feedback effect as discussed previously.
- (5) The timing of supply switching is derived from control logic and delay chains, thus subject to PVT variation and V_T scatter.

Three methods to solve read/Write conflict to reduce write failure are illustrated in Fig. 3-3. The WL can be raised up in write cycles to strengthen AXL/AXR and thus improve WM. Nevertheless, high WL will lead to more serious Half-select disturb in cells along the same WL. Therefore this approach is not applicable.

Negative Bit-Line (NBL) technique is a common way to improve write margin. By increasing the V_{GS} across the access transistor, AXL/AXR is equipped with higher

conductivity. This method requires either a negative supply or on-chip negative generation circuit. The DC negative power supply may induce leakage and reliability



Fig. 3.3: Methods for resolving Read/Write conflict and reducing write failure. problem. On-chip circuit needs a large boosting capacitor which indicates a large area overhead.

Fig. 3-4 is a "Floating Power-Line Write" scheme [6]. The VDD of selected column is switched off by a PMOS power switcher. In write cycles, write current and leakage would lower the floating VDD together to improve write margin and write performance. The drawbacks of this scheme are:

- It switches supply for whole cells, and as such, there is large amount of charge to move; resulting in slow settling time.
- (2) It impedes the pull-up of the "Low" cell storage node, and weakens the latch feedback effect as discussed previously.
- (3) The timing of supply switching is derived from control logic and delay chains, thus subject to PVT variation and V_T scatter.
- (4) The virtual supply level may drop too low, thus degrading SNM and stability of half-select cells along selected column and caused failure.



Fig. 3-5 is a column based data aware V_{SS} structure. The sources of pull-down NMOS (N4 and N5) are separated and connect to V_{SSM1} and V_{SSM2} respectively. In write "1" operation, V_{SSM1} is rise up to reduce pull down strength of N4. In write "0" cycle, V_{SSM2} is raised to increase the V_{TRIP} of the right inverter so as to help pull up of the right cell node. The drawbacks of this scheme are:

(1) V_{SSM1} and V_{SSM2} are supplied through BL muxes, hence subject to BL mux timing, (2) The timing of supply switching is derived from control logic and delay chains, thus subject to PVT variation and V_T scatter.

3.1.2 Conventional 8T SRAM



Fig. 3.6: Conventional 8T SRAM

Since the cell storage node is decoupled from read path, conventional 8T SRAM doesn't suffer from read disturb anymore. However, the half-selected cells on the same wordline still experience storage node disturb similar to "Read-Disturb" in write operation especially worse for dual supply SRAM when V_{MAX} applied to WWL.

Introduction to disturb-free 3.1.3 cross-point double-layer pass-gate 8T SRAM cell



Fig. 3.7: Proposed novel 8T SRAM cell

Fig. 3-7 is the scheme of core cell adapted in Ultra Low Power (ULP) SRAM and Absolute Low Power (ALP) SRAM. This cell features the following:

- Read Operation
 - Read (Selected cell) disturb free
 - Read Half-Selected (along WL) disturb free
- Write Operation
 - Write (Selected cell) disturb free if VVSS = 0
 - Write Half-Selected (along BL) disturb free or containable (< 6T)
 - Super V_T : VVSS = 1 (*Also, P. 31, sizing needed)
 - Sub V_T : VVSS = 0

	Standby	Write 0	Write 1	Read				
WL	0	1	1	1				
WWL	0	0	1	0				
WWLB	0	1	0	0				
VVSS	0	0	1	0				
RBL	1	0	0	floating				

Table 3.1: Truth table of novel 8T SRAM cell

Fig.3-8~Fig. 3-10 show three operation mode of novel 8T SRAM cell. Both WWL pair and RWL is turned off when standby mode. In read cycles, RBL is pre-charged to VDD by LEV. RWL is turned on so that if Q=0 is being read out, RBL will discharge through N5 and N6 in the selected cell. Since storage node is decoupled from read current path, there is no read disturb in this cell. In write cycles, RBL is pre-discharged to ground. Either WWL or WWLB will be turned on in accordance with data to be written in the cell.


3.2 Introduction of ULP design

In terms of power reduction, there are 3 operation mode can be thought over which are write, read, and standby. Considerable attentions have been paid to reduce either read power or write power. Floating bit-line (BL) is proposed to reduce Read/Write power. Lessen the voltage swing or capacitance in BL is another way to achieve the same purpose and so on.

This Ultra Low Power (ULP) SRAM design combines several techniques to minimize the power consumption in peripheral circuits. Such as floating BL, hierarchical BL, two-stage decoding and so forth. Along with peripheral, we proposed a novel way to reduce array power consumption in Read/Write/Standby mode simultaneously and mostly in standby mode which contributes to 42% array power saved in consequence. To reduce array power, we propose a way to make virtual cell supply drop to its min value lower than VDD. However, it may still be pulled down to some degree by write current. In order to stabilize the virtual cell supply and match up with interleaving four structure, we connect four columns of virtual cell supplys together to relief its sensitivity from write current and establish a parameter V_{TL} (the lowest voltage virtual cell supply i.e. VDDQ (VDDQB) can be) to ensure the safety of half-selected cells. Along with saving power, this ULP design is capable of improving write ability without negative write assist circuit compared to novel 8T version. Thus, it takes less area in ULP design because negative write assist circuits requires PMOS CAP and some peripheral circuits which is responsible for a large LEV area. Please refer to for the read/write operation of 8T core cell (Fig 3-12). The rest of this paper is organized as follows. Section2 introduce the basic concept of this design and a brief narration to how we construct this chip.



Fig.3.12: Novel 8T cell schematic Features: 1. Read disturb free 2. Read Half-selected (along RWL) disturb free 3. Write (selected cell) disturb free 4. Write Half-selected cell (along RBL) disturb free

Exhibit the power management structure. The simulation is separated into three columns which are single cell stage, 16 x 4 array stage and pre-simulation stage. A brief introduction to each stage is shown below:

• Stage1 (Single cell analysis stage):

In this stage, we are desired to define the maximum difference of virtual cell supply (i.e. between VDDQ and VDDQB) because of bad HSNM when two supplies of a cell are not identical and name it by Stable Margin.

- 1. The way we define and find V_{TL} .
- 2. Prove that write ability has an improvement with unfastened virtual cell supply.
- 3. Take apart two supplies of one cell in layout view.
- 4. Verification: comparisons between three definitions of V_{TL}
- 5. Verification: V_{TRIP} V.S V_{TL}
- 6. Find worst threshold voltage (V_T) shift to two conditions that used in stage
 - 2. "Prone to flip" and "Difficult to write".

• Stage2 (16 x 4 array base stage):

Effect of V_{TL} on ensuring the security of half-selected cell must be strictly verified in this stage. Lump and couple capacitance plays an important role in this design because the size of cap has great influence on how deep virtual cell supply would fall and how long it takes to recover from read/write operation. Therefore caps are extracted from layout and join simulation.

- 1. Methods to size keeper of virtual cell supply.
- 2. Verification: 16x4 array simulation with V_T shifted.
- Run Monte Carlo in two extreme cases: "prone to flip" and "difficult to write".
- Comparisons on Write Ability improvement between ULP and Novel 8T design.

896

5. Power management structure and its control manners.

• Stage3 (Pre-simulation stage):

The ULP SRAM design is an extension from Novel 8T, hence most peripheral circuits remains the same; nevertheless, new control signals are required in ULP SRAM and the target of VDD is 0.6V which is 100mV lower. Some technique and change must be added to guarantee a successful completion such as LRC [37].

There are 5 key points to focus in stage1 and stage2. 1) Algorithm and consideration for keeper size of array power. 2) Monte Carlo simulation result. 3) Analysis on HSNM (Half selected cell noise margin). 4) Improvements on Write Ability. 5) Virtual cell supply monitor in 16x4 array stage. Stage3 shows the post simulation result inclusive of detailed power comparisons. Conclusions are drawn in chapter6.

3.3 Power Reduce Concept

Fig3-13 shows the leading idea. Virtual cell supply (VDDQ) is the power supply for 8T cell whose strength is decided by programmable keepers which are enclosed by dotted line and is connected to VDDQB (REG=0)when standby and write operation (Fig3-14). We intend to lower VDDQ i.e. VDDQB in standby mode in order to decrease cell power consumption while peripheral circuits are connected to full VDD. Virtual cell supply will be charged up only when read. During read cycles, RBL will discharge through M7 and M8 in 8T cell if data 0 is being read out. Gate of M8 has to be VDD to fully open M8 that is why we must charge VDDQB up. Table2 shows the truth table of virtual cell supply and novel 8T design. Virtual cell supply equals to VDD only in read cycles. Two targets must be reached by carefully sizing keepers. 1) Maintain a current balance between leakage and keepers so that virtual cell supply stays in a value lower than VDD. 2) Virtual cell supply must be kept above V_{TL} when pulled down by performing write operation to keep half select cells away from flip.



Fig.3.13: Desired virtual cell supply (i.e. VDDQ or VDDQB) waveform. V_{TL} is defined as 10%VDD HSNM with asymmetric threshold-voltage (V_T) shift in 8T cell. " prone to flip"



Fig. 3.14: Ultra Low Power structure: 16x4 array as one unit .Cell virtual cell supply supplied by programmable keepers. Gates of keepers are controlled by I/O pins.

Fig3-14 shows how we realize the power management method. 4 columns of 16-cell-length local bit-line are connected together with virtual cell supply supplied by programmable keepers. As long as the keeper size is decided in stage 2, we set the size in the 8th of programmable keeper so that there is room for tuning if the default size is not suitable for implemented chip. For example, if the default size of keeper is W/L=800n/100n then the combination of keeper is $100n/100n \cdot 200n/100n \cdot 400n/100n$ 800n/100n. 100n/100n Thus keeper strength ranges from to (100+200+400+800)n/100n. The length of keepers is 100n in order to minimize process variation in that the precision of size is very important in our design. The rest MOS length is 55n.

The followings are operation introductions. In read cycle, first set REG=1 to turn MPG off. Then REB=0 to turn on MPB to charge VDDQB to full VDD so as to speed up read operation while VDDQ stays in standby value lower than VDD . However, unequal value of VDDQB and VDDQ results in a worse HSNM (Fig3-15). For safety of half selected cells, REG can be set to 0, external controlled, to charge both VDDQB and VDDQ to full VDD when read in price of larger read power. In a word, it's a tradeoff between HSNM and read power.



Fig. 3.15: HSNM gets worse (Black square to dotted one) if VDDQ and VDDQB aren't identical



Fig. 3.16: 8T-cell layout. Black frame rounded are two individual cell supplies, VDDQB and VDDQ.



Fig. 3.17: ULP in current aspect

Fig3-16 is the way to take apart two supplies of 8T cell. VDDQ and VDDQB are two isolated power lines by length of 16 cells height. Surroundings of each power line must keep as symmetric as possible to maintain same loadings of power lines.

Fig3-17 tells how this idea works in current aspect. Regardless of data saved in storage node, there always exits a leakage path from virtual cell supply to ground. As a consequence, I-keeper must be larger than M x 2 x 4 leakages so that virtual cell supply can pull up back to standby value when write operation is over. M is the number of cells in a local bit-line. 2 power paths, i.e. node Q and node QB of a cell are connected in 16 x 4 structure. 4 columns are linked up (Fig3-14) together. In write operation, I-keeper will be smaller than "I-write + leakages" that virtual cell supply would be pull down a little. When standby, I-keeper and leakages are in a balanced situation. Therefore decreased virtual cell supply shall reduce array leakage power consumption.

As a solution of Soft Error Rate (SER), interleaving4 structure is adopted in this design. This ULP chip has 128 columns. 1word=32bit.To match up with, we connect virtual cell supply i.e. VDDQ (VDDQB) of four 16x1 arrays to each other to form a base unit. By joining four columns, virtual cell supply gets more stable.



Fig. 3.18: VT shift of two extreme conditions

3.4 Stage1_Single cell analysis

Since pseudo powers i.e. VDDQ and VDDQB exists difference in either read or write, V_{TL} plays an important role in making certain the safety of half select cell. V_{TL} is determined through one 8T cell analysis with threshold voltage shifted as Fig3.18 (prone to flip) which is easier for data1 to flip in node QB.

In our design, pseudo power is movable based on operation issued, read or write. VDDQ and VDDQB would be different in read cycles, if we choose to charge single side, and write cycles. A difference between two supplies, i.e. VDDQ and VDDQB, of one cell will lead to a bad HSNM shown in Fig3.15. Thus, we are desired to know the max difference and name it by Stable Margin. That is, the max difference between VDDQ and VDDQB is "VDD - V_{TL} =Stable Margin". Remember that VDDQ and VDDQB are connected together during write cycle. Thus, merely a slight difference, even none, exists due to RC effect. To find V_{TL} , a single cell is simulated with one power fixed, say, VDDQ and drops another. Considering the variation between HSPICE and real chip, the threshold voltage (V_1) of cell under simulation is shifted to condition "prone to flip" (Fig3.18) in order to get a highly safe stable margin. V_{TL} is the lowest value count downward from VDD, i.e. the bottom line of pseudo power. V_{TL} is detected whenever 10% VDD asymmetric HSNM is measured. The word "asymmetric" means that V_T shifts to "prone to flip".

To exam the effectiveness of our criterion, Fig3.20 presents comparisons between three criterions which are 1000N without flip HSNM, 10%VDD HSNM and 10% VDD asymmetric HSNM. The "1000ns without flip HSNM" mean that data will hold for more than 1000ns whenever VDDQ is unequal to VDDQB. Usually VDDQ and VDDQB will recover to same value in the end of one cycle. Obviously criterion "10% VDD asymmetric HSNM" is the safest one for it always has the farthest value from the flip point. A more concrete way to check is to compare the V_{TL} to the trip point of inverter (V_{TRIP}) within a cell because node QB will equal to VDDQB and if QB is beneath transition point of inverter then Q will flip. That is so called hold fail. V_{TL} of 5 corners and the highest V_{TRIP} of inverter are shown on the Fig3.21. We must take the highest one as our criterion V_{TL} for sizing keeper because the higher pseudo power is the safer HSNM guaranteed.

The result meets our expectation because the highest V_{TL} is always above the highest V_{TRIP} . Since V_{TL} stays about 120mV above V_{TRIP} , safety of half-selected cell is ensured.

Before entering stage2, we conduct a brief exam on write ability improvement of movable pseudo power. The worst corner for write operation is often PFNS at low temperature. Novel 8T suffers a write fail in PFNS at -20°C writing QB. Once VDDQB floats, write will succeed shown in Fig3.19. Fig3.16 is the way to take apart two supplies of 8T cell. VDDQ and VDDQB are two isolated power lines by length of M cells height.

Surroundings of each power line must keep as symmetric as possible to maintain same loadings of power lines. In stage 1, we obtain V_{TL} through a rigorous process. This criterion is used in stage2 when sizing keepers.



Fig. 3.19: Single cell simulation to verify that movable virtual cell supply helps to improve write ability.





Fig. 3.20: Three criterions of V_{TL} and flip point. "10%VDD+asym" is adopted criterion.



VTL verify by TRIP POINT

Fig. 3.21: The determined V_{TL} is about 100mV higher than V_{TRIP} of inverter in cell.



Chapter 4 Stage2_16 x 4 array base simulation and Stage3_Pre-simulation Result

4.1 Stage2_16 x 4 array base simulation

In this stage, we are eager to see what value virtual cell supply stays when standby and how deep it falls as write operation is issued. A large capacitance helps to stabilize the virtual cell supply and that's one of the reasons we link four columns side by side. The target VDD of ULP design is 0.6V so we starts sizing by VDD=0.6V with V_{TL} =0.46V which is obtained in stage1. By simulation, we found that the lowest virtual cell supply take place in write QB to 0 and the reason is revealed in Fig3-16. VVSS is just by the side of VDDQB in layout. During write QB, VVSS is 1 and is 0 when write QB operation is over as in Fig4-1. Therefore the virtual cell supply is coupled down by VVSS.



Fig. 4.1: Waveform of read write operation.



HOLD_Q=0_PFNF_125 Fig. 4.2: Keeper sizing by V_{TL}

To size the keeper, first we set L=100n to alleviate process variation then sweeps width from 100n to 2000n. Fig4-2 shows hold and lowest virtual cell supply under each keeper size. We wish have a lowest value higher than V_{TL} (0.46V), hence W=600n is chosen in this step and exams its reliability later. Fig4-4 is the lowest virtual cell supply of each corner mixed with temperature and data saved in other (M-1) cells in write QB operation. Note that the lowest occurs in PFNF & 125°C & all QB=1 in (M-1) cells. In the following steps we will focus our exams in this situation. The size of MPB has nothing to do with HSNM. It affects the speed of read only, therefore the size of MPB is chosen to make slew rate of virtual cell supply the same as control signal. Size of MPG matters when write. If MPG is not large enough, VDDQ and VDDQB exists difference when write cycles. Thus size for MPG must be large enough to make VDDQ equals VDDQB in write operation.

The lowest virtual cell supply is getting away from V_{TL} as VDD goes higher under the keeper size chosen which means safer (Fig4-3). To be more precisely, keeper size





Fig. 4.4: The lowest virtual cell supply of write operation. $QB=0_-40^{\circ}C$ means that data 0 is saved in all node QB except for the one to be written. Lowest value occurs in write QB=1 PFNF_125°C

Smaller keeper through programmable mechanism. Monte Carlo simulation is conducted to check two features: 1) Virtual cell supply is always above V_{TL} 2) A successful write operation. We use Fig3 16 x 4 array base to do it with cell#1 V_T shift to "difficult to write" and cell#M VT shift to "prone to flip".

Cell #1 is placed nearest to keeper and that offers cell#1 a strong power line, thus data1 isn't easy to pull down. Even though the position and V_T shift makes cell#1 the most difficult cell to write, Fig4-5 shows that Monte Carlo has write successfully by MC=17000 at PFNS corner. Another Monte Carlo simulation is conducted to check cell#M. Cell#M places farthest from keepers along with V_T shift to "prone to flip" makes it easiest cell to flip. Fig4-5 reveal that all virtual cell supply is above 0.46V, therefore no cell flipped in this condition at PFNF corner. The Monte Carlo simulation is issued with V_T shifted to the worst condition already. Hence Monte Carlo count is 17000 which are relatively few, yet it covers the worst condition.



Fig. 4.5: Monte Carlo result Shows the lowest VDDQ when write.

Write ability of a cell offers an indication of how easy or hard it is to write a cell. Capacitance has great influence on how much write ability improves in ULP design because it determines how easy virtual cell supply moves and how deep it falls. In stage 1 we have already confirm the validity of write ability improvements by single cell analysis. A 16 x 4 array base simulation with lump and couple capacitance added is used to measure write ability. The way we test write ability is to rise RBL up little by little in write operation until write fail. Max RBL Voltage (V_{RBL}) defines the voltage to flip storage node. The higher the V_{RBL} is, the easier it is to write the cell. The DC sweep measurement result is shown in Fig4-6 (VDD=0.6V). Fig.4-7 is the VRBL increase amount in each corner and temperature. Obviously, write ability is improved in all condition. Compared to Novel 8T version without negative write assist circuit, V_{RBL} increases from 25mV to 116mV at 125°C and from write fail to 55mV at -40°C in worst write corner PFNS. The better write ability is achieved with this structure and it takes less area and lower power consumption comparing to negative write assist circuit.



Fig. 4.6: Write ability improvement. MAX V_{RBL} for PFNS rise up by 97mV



Fig. 4.7: The V_{RBL} increase amount in 5 corners and 4 temperatures.

4.2 Stage3_Pre-simulation 1896

Schematic simulation was carried out to verify the functionality of ULP SRAM. A critical path of 1/4 block is implemented in pre-simulation as in Fig4-8. The ULP SRAM consists of 16 blocks with 128 x 32 bits. Each block has 32 hierarchical bit lines. The hierarchical bit line consists of 8 local bit lines with 16 memory cells. Each local bit line is equipped with one Local Evaluation circuit (LEV). Each WL driver drives 64 cells in row direction. Dummy local bit line pulse width is controlled by pin [C0, C1] rather than trace circuit because this paper targets low power issue, thus a safety guaranteed pulse width is priority. The peripheral circuit is fabricated based on novel 8T design with few block altered. Take Leakage Current-Replica (LCR) for instance. Compare to last novel 8T design. Leakage gradually becomes not negligible in deep submicron process. A floating node in MUX suffers from four paths of NMOS

leakage that may drop floating 1 to 0 results in an error in read operation. To deal with those leakages, traditional solution is to add a feedback PMOS. (Fig4-10 left side) However, it may lead to pulling down fail, especially in fast-P corner. As a better solution, leakage current-replica circuit was brought in (Fig4-10 right side) and successfully solved this problem. The output of LRC is connected to the floating node of MUX and offers exactly the minor current needed to fight the leakage.

Fig4-9 shows the simulation waveforms of ULP SRAM. The simulations are performed at 2.15ns with VDD=0.6 V at PTNT corner. When standby, virtual cell supply is about 100mV beneath the VDD and is pulled down by I_{WRITE} to 489mV in write cycles. During read operations, virtual cell supply is charged up by MPB and fall back to its standby position when read over. The ULP SRAM significantly reduces the array power consumption by lowering virtual cell supply in standby and write cycles. Table4.2 shows that virtual cell supply equals to VDD only when read cycles. If we



Fig. 4.8: Architecture of ULP



Fig. 4.9: VDDQ post-simulation waveform, VDD=0.6V



Fig. 4.10: Current replica circuit provide enough leakage current to maintain the safety of the floating 1 in MUX

choose to charge VDDQ only when read, signal REG comes before REB for about 3 inverter delay to isolate VDDQ and VDDQB first then charge VDDQ solely.

In order to make sure ULP will work successfully. Several control pins are added to adjust pulse width and I-V supply condition (Table4.1).

1. Sig-nor: In case that the longest pulse width is not wide enough. We can lengthen the pulse width through this pin until a success read or write operation.

- 2. Sig-and: A pin used to determine to charge VDDQ solely or both VDDQ and VDDQB in read operation. A trade-off between lower power and better HSNM.
- **3. [C0, C1]:** A control of discharge speed of dummy local bit-line. High speed discharge will make a short pulse width of control signal. Through fine adjust of pulse width, read and write will be performed successfully.
- 4. [K0, K1, K2, K3]: Gate control signal of programmable keepers. There are total 16 combinations of keeper size. The chosen size in stage2 is positioned in the 8_{TH} one. As a result, size can be tuned through various VDD. This chip is a confirmation of ULP structure, thus the programmable keepers are not equipped with corner tracking technique.

SIGNAL	INTRODUCE				
K[0,1,2,3]	Control of PMOS with WP=[100n,200n,400n,800n]				
	=[0,1,1,1] → KPR=100n/100n				
	=[0,0,0,0] → KPR=1500n/100n				
C[0,1]	Control of WL pulse width				
	=[0,0]→Longest WL pulse				
	=[1,1] → Shortest WL pulse				
SIG_NOR	Control of WL pulse width from outside pin				
	=[0] \rightarrow work as INV, pulse controlled by C[0,1]				
	=[1]→WL pulse keeps high(active)				
SIG_AND(NBL)	Control of VDDQ and VDDQB <u>precharge</u> when read (only makes difference in ULP)				
	=[0] \rightarrow VDDQ and VDDQB charge up simultaneously when read (better HSNM , larger power consumption)				
	=[1] → Only VDDQB charge up when read (worse HSNM, smaller power consumption)				

Table.4.1: Control signal content

	Standby	Read	Write"1"	Write"0"
REB	1	0	1	1
RBL	1	Х	0	0
RWL	0	1	1	1
WWL	0	0	0	1
WWLB	0	0	1	0
VVSS	x	0	1	0

	Standby	Read	Write
VDDQ	<vdd< td=""><td><vdd< td=""><td><vdd< td=""></vdd<></td></vdd<></td></vdd<>	<vdd< td=""><td><vdd< td=""></vdd<></td></vdd<>	<vdd< td=""></vdd<>
VDDQB	<vdd< td=""><td>VDD</td><td><vdd< td=""></vdd<></td></vdd<>	VDD	<vdd< td=""></vdd<>

Table4.2: Truth table of 8T cell and virtual cell supply.



Chapter5 Power comparison & Absolute Low Power mode & Test Flow

5.1 Absolute Low Power Mode

In ULP SRAM structure of read operation, VDDQ and VDDQB of selected row will all charge up in order to speed up read. That is 128columns in our macro. In fact, only a quarter columns (32columns) of selected row need to be raised up due to interleaving-four scheme. An extended structure called absolute low power mode 1111 (ALP) is shown in Fig. 5-1. Signal-REB is a row base control signal and is 0 when selected. Column enable signal (COL [0:3]) is column base and is 0 when selected. REB and COL [0:3] together control the P2 and P1. P2 is turned off before P1 turned on to charge virtual cell supply to full VDD in read cycles. Due to restriction of control signals, a difference between ULP and ALP is that ALP can only charge both VDDQ and VDDQB in read cycles while ULP has the choice of charging VDDQB or both of them. Compared with ULP, one NOR gate and one INVERTER gate is added to each local column as area overhead. However, both ULP and ALP doesn't need negative write assist circuit which is quite are consuming. Thus 3/4 read power saving with only a NOR and INV is still a bargain. Fig. 5-2 is local evaluation circuit (LEV) of original design with negative write assist circuit. The part enclosed by dotted line is negative write assist circuit which is removed in ULP and ALP design. Since each 16x 1column has one LEV. ALP and ULP save a considerable area.



Fig. 5.2: Local Evaluation Circuit

5.2 Power Comparison and Simulation Result

Fig4-1 is the post-simulation waveform of virtual cell supply and control signals. The waveform and value in each operation are quiet fit into our anticipations. It stays in 517mV when standby and pulled down to 489mV as a half-select cell friendly write because V_{TL} is 460mV when VDD=600mV, thus HSNM is guaranteed save. In fact both VDDQ and VDDQB will be pulled down simultaneously, so cell won't suffer from different power value. During read cycles, RBL must discharge through M7 and M8 in 8T cell. Gate voltage of M8 has to be VDD to fully open M8 that is why we must charge VDDQB up. Fig. 5-3 is array power save ratio of ALP/ULP to comparison group (COM). Note that ALP array save ratio is supposed to beat ULP's in every corner. However, ALP only wins in PFNF and PFNS corner. After diving into the reason, we found that it is due to INVERTER and NOR-GATE added in ALP structure shown in Fig. 5-1. The power dissipated in these INV and NOR may cancel the power saved in read cycles as in Fig. 5-4. It is obvious that the read power saved is not enough to pay the power dissipated by NOR-GATE and INV. We normalize the power of the three portions by INV's current and recalculate the ALP array power save ratio by

New ALP Array Power = Original ALP Array Power
$$\times \frac{2.5x}{1x+2.5x+3.75x}$$

That is categorizing INV and NOR-GATE into peripheral logic gate and not count into array power dissipation. A new result is shown in Fig. 5-5 which fits our anticipation that ALP saves more array power than ULP. This formula is how we count the power save ratio:

power saved ratio =
$$\frac{(COM - ULP \text{ or } ALP)}{COM}$$
 (5.1)



Fig. 5.14: Power save ratio of ULP SRAM



Fig. 5.15: Power save amount of ULP SRAM



ALP_ARRAY POWER SAVE RATIO

Array_COM-Array_ALP



Fig. 5.17: Power save amount of ALP SRAM



Fig. 5.3: Array Power Comparison (ALP/ULP vs COM) at 0.6V





Fig. 5.4: Split ALP array power into three parts.



Fig. 5.5: ALP power recalculated without counting INV and NOR





Fig. 5.6: Total Power Comparison (ALP/ULP vs COM) at 0.6V



Fig. 5.7: Array Power Comparison (ALP/ULP vs COM) at 1.0V



Fig. 5.8: Total Power Comparison (ALP/ULP vs COM) at 1.0V

Regardless of comparison between ALP and ULP, both of them save a considerable portion of array power in Fig. 5-3~ Fig. 5-8. The result is quite inspiring that array power is reduced in every corner from 24% (PFNS corner) to 90% (PSNS) in V_{DD} =0.6V. Based on simulation result, this 128-Kb 8T SRAM is able to operate at 1.45GHz when VDD=1.0V, 246MHz when VDD=0.45V.

Table5.1 is the post-simulation result of access time and minimum virtual cell supply in write operation of ALP and ULP when VDD=0.6V.

VDD=0.6V	ULP		AI	LP
	Access Time	Min Pseudo Power	Access Time	Min Pseudo Power
PSNS	3.18Ns	531mV	3.15Ns	566mV
PTNT	1.79Ns	527mV	1.75Ns	551mV
PFNF	1.17Ns	467mV	1.12Ns	486mV
PFNS	1.98Ns	523mV	1.83Ns	547mV
PSNF	1.71Ns	496mV	1.69Ns	529mV

Table.5.1 Post Simulation Result

Fig. 5-9 is the virtual cell supply wave form of ALP SRAM. In read cycle, only the selected column power line (V_{DDQB}) is charged up to 598mV while VDD is 0.6V. Other three V_{DDQB} of the four remain the same value as in standby mode (483mV). Thus the read power wasted comparing to standby mode is decreased by factor 0.75.



Fig. 5.9: Virtual cell supply waveform of ALP

5.3 Test Flow

This section introduces the test flow of implemented chip. There are three conditions we want to test if the chip works successfully.

 Continuous write may drop virtual cell supply too low. Storage node that saves high voltage may flip.

 \rightarrow In fact there is enough time for virtual cell supply to recover to its standby voltage.

- 2. After a long period of sleep mode, virtual cell supply may go to unknown voltage.
- 3. Immediate write after continuous read. Since virtual cell supply is raised up in read cycles, write operation may suffer from too high virtual cell supply that make it uneasy to pull high storage node down.

In section 4.2 we have introduced the control pin from outside such as SIG-NOR, SIG-AND, K [0:3] and so on. Fig. 5-11 shows the test flow of ULP SRAM. This flow includes the three conditions we want to exam. Detailed elaboration on how this flow going is not drawn here. Fig. 5-12 is a similar test flow for ALP SRAM.



Fig. 5.10: Two flow of power measuring.



Fig. 5.11: Test flow of ULP SRAM



Fig. 5.12: Test flow of ALP SRAM

5.4 Design implement

Fig. 5-13 shows the floor plan and layout view of ULP SRAM. The difference between ULP and ALP is the power management structure, therefore the layout view of Alp is not shown here. Two 128 K bit chip is fabricated using UMC 55nm process. Below is the feature of these deigns:

- Instance : ULP(1024x128 bits) , ALP(1024x128 bits)
- Input pin : I[0:31]A[0:11],C[0:1],K[0:3],NBL,SIG_NOR,CK,CSB,WEB
- Output pin : DO[0:31]
- > Chip feature :
 - 1. Interleaving_4 : 1word=32bits
 - 2. 2 mechanisms of pulse width control
 - 3. Hierarchical bit-line : 16x1
 - 4. MUX current replica circuit : floating1
 - 5. WWL DRIVER : VVSS enclose WWL pulse
 - 6. Power control :
 - A. Programmable keeper

B. ALP,ULP(SRP,DRP)


Chapter6 Conclusions

In order to extend working time of battery-supplied devices, low power SRAM design is booming recently. Since Power = Current * VDD, lowering supply voltage is a direct way to reduce power consumption. However, low VDD induce problems in reliability. Along with some known skill to reduce peripheral power, this paper presents a method to minimize array power and improves write ability simultaneously. By constructing an algorithm to sizing the keepers, this structure enables array virtual cell supply movable according to operations and most time it stays below VDD. V_{TL} defines the lowest virtual cell supply voltage as write operation is carried out. Maintaining current balances between cell leakage and keeper helps to lower virtual cell supply when standby thus leakages in cells are reduced. It saves about 60% array power averagely and varies from corner to corner. Beside low power feature, write ability is also improved without help of negative write assist circuit which leads to 43% area saving in local evaluation circuit (LEV). Two 128Kb SRAM design are implemented in UMC 55nm CMOS technology.

To summarize:

- Features of Ultra Low Power Design :
- The virtual cell supply scheme improves write ability successfully in each corner
- > The power management structure reduces array standby power
- 43% LEV area is saved under this technique. Since every 16x1 array require one LEV, considerable area can be saved under this structure.
- RSNM and read speed doesn't decrease in that virtual cell supply is charged to VDD in read cycles
- > Only ¹/₄ read power is need in ALP SRAM compared to ULP SRAM

Reference

 [1] H. I. Yang, S. Y. Lai, W. Hwang, "LOW-POWER FLOATING BITLINE 8-T SRAM DESIGN WITH WRITE ASSISTANT CIRCUITS" IEEE SOC Conference, 17-20 Sept. 2008, pp.239 – 242.

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

[3] B. D. Yang, "A Low-Power SRAM Using Bit-Line Charge-Recycling for Read and Write Operations" IEEE Solid-State Circuit Journal, Oct. 2010, pp.2173-2183.

[4] B. D. Yang, L. S. Kim, "A Low-Power SRAM Using Hierarchical Bit Line and Local Sense Amplifiers" Solid-State Circuits Journal, June. 2005, pp.1366 - 1376.

[5] S. P. Cheng, S. Y. Huang, "A Low-Power SRAM Design Using Quiet-Bitline Architecture" IEEE International Workshop on Memory Technology, Design, and Testing, 5-5 Aug. 2005, pp.135 – 139.

[6] H. Qin, A. Kumar, K. Ramchandran, J. Rabaey, "Error-Tolerant SRAM Design for Ultra-Low Power Standby Operation" ISQED Conference Quality Electronic Design, 17-19 March. 2008, pp.30 – 34.

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

[8] Ramy E. Aly and Magdy A. Bayoumi, "Low-Power Cache Design Using 7TSRAM Cell" IEEE TCSII, April. 2007, pp.318 – 322.

[9] S. Hattori, T. Sakurai, "90% Write Power Saving SRAM Using Sense-Amplifying Memory Cell" Digest of Tech. Papers, Symp. VLSI Circuits, 2002, pp. 46 - 47.

[10] T. B. Hook, M. Breitwisch, J. Brown, P. Cottrell, D. Hoyniak, C. Lam, R. Mann, " Noise Margin and Leakage in Ultra-Low Leakage SRAM Cell Design" IEEE Electron Device, Aug. 2002, pp.1499 – 1501. [11] C. T. Chuang, S. Mukhopadhyay, J. J. Kim, K. Kim, R. Rao, "High-Performance SRAM in Nanoscale CMOS: Design Challenges and Techniques" IEEE International Workshop on Memory Technology, Design and Testing, 3-5 Dec. 2007, pp.4-12.

[12] M. Margala, "Low-Power SRAM Circuit Design" IEEE International Workshop on Memory Technology, Design, and Testing, 09 Aug. 1999, pp.115 – 122.

[13] S. K. Jain, P. Agarwal, "A Low Leakage and SNM Free SRAM Cell Design in Deep Sub micron CMOS Technology" IEEE VLSI Design Conference, 3-7 Jan. 2006, pp.4.

[14] R. Keerthi, C.-i H. Chen, "Stability and Static Noise Margin Analysis of Low-Power SRAM" IEEE IMTC Conference, 12-15 May. 2008, pp.1681 – 1684.
[15] H. Noguchi, S. Okumura, Y. Iguchi, H. Fujiwara, Y. Morita, K. Nii, H. Kawaguchi, M. Yoshimoto, "Which is the Best Dual-Port SRAM in 45-nm Process Technology?– 8T, 10T Single End, and 10T Differential –" IEEE ICICDT Conference, 2-4 June. 2008, pp.55 – 58.

[16] E. Morifuji, T. Yoshida, M. Kanda, S. Matsuda, S. Yamada, F. Matsuoka,
"Supply and Threshold-Voltage Trends for Scaled Logic and SRAM MOSFETs"
IEEE Transaction on Electron Device, June. 2006, pp.1427 – 1432.

[17] A. Kotabe, K. Osada, N. Kitai, M. Fujioka, S. Kamohara, M. Moniwa, S. Morita,
Y. Saitoh, "A Low-Power Four-Transistor SRAM Cell With a Stacked Vertical
Poly-Silicon PMOS and a Dual-Word-Voltage Scheme" IEEE Solid-State Circuit
Journal, April. 2005, pp.870 – 876.

[18] A. J. Bhavnagarwala, S. Kosonocky, C. Radens, Y. Chan, K. Stawiasz, U.
Srinivasan, S. P. Kowalczyk, M. M. Ziegler, "A Sub-600-mV, Fluctuation Tolerant
65-nm CMOS SRAM Array With Dynamic Cell Biasing" IEEE Solid-State Circuit
Journal, April. 2008, pp. 946 – 955.

[19] H. Yamauchi,"A Discussion on SRAM Circuit Design Trend in Deeper

Nanometer-Scale Technologies" IEEE Transactions on Very Large Scale

Integration Systems, Vol. 18, Issue 5, May. 2010, pp. 763 – 774.

[20] N. N. Mojumder, S. Mukhopadhyay, J. J. Kim, C. T. Chuang, K. Roy,
"Self-Repairing SRAM Using On-Chip Detection and Compensation" IEEE
Transactions on Very Large Scale Integration Systems, Vol. 18, Issue 1, May. 2010,
pp. 75 – 84.

[21] C. H. Kim, J. J. Kim, S. Mukhopadhyay, K. Roy, "A Forward Body-Biased Low-Leakage SRAM Cache:Device, Circuit and Architecture Considerations" IEEE Transactions on Very Large Scale Integration Systems, Vol. 13, Issue 3, March. 2005, pp. 349 – 357.

[22] M. H. Abu-Rahma, M. Anis, S. S. Yoon "Reducing SRAM Power Using Fine-Grained Wordline Pulsewidth Control" IEEE Transactions on Very Large Scale Integration Systems, Vol. 18, Issue 3, March. 2010, pp. 356 – 364.

[23] K. Nii, Y. Tsukamoto, T. Yoshizawa, S. Imaoka, Y. Yamagami, T. Suzuki, A.

Shibayama, H. Makino, S. Iwade "A 90-nm Low-Power 32-kB Embedded SRAM

With Gate Leakage Suppression Circuit for Mobile Applications" IEEE Solid-State Circuit Journal, April. 2004, pp. 684–693.

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

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

[26] 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 VLSIC Symposium, 18-20 June. 2008, pp. 212 – 213.

[27] M. Yabuuchi, K. Nii, Y. Tsukamoto, S. Ohbayashi, Y. Nakase, H. Shinohara., "A45nm 0.6V Cross-Point 8T SRAM with Negative Biased Read/Write Assist," Digest of Tech. Papers, Symp. VLSI Circuits, 2009, pp. 158-159.

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

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

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

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

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

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

[34] 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.

[35] International Technology Roadmap for Semiconductors, ITRS, 2010 [Online].Available: <u>http://public.itrs.net</u>

[36] K. W. Mai, T. Mori, B. S. Amrutur, R. Ho, B. Wilburn, M. A. Horowitz, I.

Fukushi, T. Izawa, S. itarai, "Low-Power SRAM Design Using Half-Swing

Pulse-Mode Techniques" IEEE J. Solid-State Circuits, Nov. 1998, pp. 1659-16714.

[37] Y. Lih, N. Tzartzanis, W. W. Walker, "A Leakage Current Replica Keeper for

Dynamic Circuit" IEEE Solid-State Circuit Journal, Jan. 2007, pp.48 – 55.

Vita

PERSONAL INFORMATION

Name: Mao-Chih Hsia

Birth Date: May. 23, 1983

Birth Place: Kaohsiung, 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: mao.1126@gmail.com

EDUCATION

- B.S. [2007] Department of Electronics Engineering, National Central University.
- M.A. [2009] Institute of Electronics, National Chiao-Tung University.