DEVELOPMENT OF AN AREA-EFFICIENT AND LOW-POWER FIVE-TRANSISTOR SRAM FOR LOW-POWER SOC

by

Hooman Jarollahi
B.A.Sc., Engineering Science
Simon Fraser University, 2008

THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF APPLIED SCIENCE

In the School
of
Engineering Science

© Hooman Jarollahi 2010

SIMON FRASER UNIVERSITY

Summer 2010

All rights reserved. However, in accordance with the *Copyright Act of Canada*, this work may be reproduced, without authorization, under the conditions for *Fair Dealing*. Therefore, limited reproduction of this work for the purposes of private study, research, criticism, review and news reporting is likely to be in accordance with the law, particularly if cited appropriately.

APPROVAL

Name: **Hooman Jarollahi**

Degree: **Master of Applied Science (M.A.Sc.)**

Title of Thesis: **DEVELOPMENT OF AN AREA-EFFICIENT AND LOW-POWER FIVE-TRANSISTOR SRAM FOR LOW-POWER SOC**

Examining Committee:

Chair: **Dr. Shahram Payandeh, P.Eng.**
Professor, School of Engineering Science

Dr. Richard F. Hobson, P.Eng.
Senior Supervisor
Professor, School of Engineering Science

Dr. Marek Syrzycki
Supervisor
Professor, School of Engineering Science

Dr. Glenn H. Chapman, P.Eng.
Internal Examiner
Professor, School of Engineering Science

Date Defended/Approved: June 3, 2010

SIMON FRASER UNIVERSITY
LIBRARY

Declaration of Partial Copyright Licence

The author, whose copyright is declared on the title page of this work, has granted to Simon Fraser University the right to lend this thesis, project or extended essay to users of the Simon Fraser University Library, and to make partial or single copies only for such users or in response to a request from the library of any other university, or other educational institution, on its own behalf or for one of its users.

The author has further granted permission to Simon Fraser University to keep or make a digital copy for use in its circulating collection (currently available to the public at the "Institutional Repository" link of the SFU Library website <www.lib.sfu.ca> at: <http://ir.lib.sfu.ca/handle/1892/112>) and, without changing the content, to translate the thesis/project or extended essays, if technically possible, to any medium or format for the purpose of preservation of the digital work.

The author has further agreed that permission for multiple copying of this work for scholarly purposes may be granted by either the author or the Dean of Graduate Studies.

It is understood that copying or publication of this work for financial gain shall not be allowed without the author's written permission.

Permission for public performance, or limited permission for private scholarly use, of any multimedia materials forming part of this work, may have been granted by the author. This information may be found on the separately catalogued multimedia material and in the signed Partial Copyright Licence.

While licensing SFU to permit the above uses, the author retains copyright in the thesis, project or extended essays, including the right to change the work for subsequent purposes, including editing and publishing the work in whole or in part, and licensing other parties, as the author may desire.

The original Partial Copyright Licence attesting to these terms, and signed by this author, may be found in the original bound copy of this work, retained in the Simon Fraser University Archive.

Simon Fraser University Library
Burnaby, BC, Canada

Last revision: Spring 09

ABSTRACT

The purpose of this thesis is to introduce a new low-power, reliable and high-performance five-transistor (5T) SRAM in 65nm CMOS technology, which can be used for cache memory in processors and low-power portable devices. An area reduction of ~13% compared to a conventional 6T cell is possible. A biasing ground line is charged by channel leakage current from memory cells in standby, and is used to pre-charge a single bit-line and bias the negative supply voltage of each memory cell to suppress standby leakage power. A major standby power reduction is gained compared to conventional 5T and 6T designs, and up to ~30% compared to previous low-power 6T designs. Read, write, and standby performance and reliability issues are discussed and compared with conventional and low-power 6T SRAM designs.

Keywords: Static Random Access Memory; Five-transistor SRAM; Low-power SRAM; Area-efficient SRAM; Leakage-power reduction

ACKNOWLEDGEMENTS

I would like to thank my supervisor, Professor Richard Hobson, who believed in me in the first place, taught me how to conduct research and apply the knowledge behind it to this thesis. His encouragement, guidance and support from the initial to the final level enabled me to complete this thesis.

In addition, I thank my parents who have always supported me unconditionally especially during the completion of this thesis.

Lastly, I offer my regards and gratitude to anyone who supported me in any respect in the completion of this thesis.

Hooman Jarollahi

TABLE OF CONTENTS

Approval ii
Abstract iii
Acknowledgements iv
Table of Contents v
List of Figures vii
List of Tables xi
Glossary xii

1: Introduction 1

1.1 SRAM block organization and structure 4
1.2 CMOS SRAM and technology scaling 6
1.3 Threshold voltage variations and causes 8
1.4 Leakage current sources 10
1.4.1 Reverse-biased junction leakage current (I_j) 12
1.4.2 Gate induced drain leakage (I_{GIDL}) 12
1.4.3 Gate direct-tunnelling leakage (I_g) 12
1.4.4 Sub-threshold (weak inversion) leakage (I_{sth}) 12
1.5 SRAM operation failure 14
1.6 Stability analysis 14
1.6.1 Static Noise Margin (SNM) 15
1.6.2 Read Noise Margin (RNM) 16
1.6.3 Butterfly diagrams 17
1.6.4 Static noise margin measurement using SPICE 19
1.7 Read noise margin and supply voltages 23
1.8 Technology corners and static noise margin in SRAM 24

2: 6T SRAM operation 26

2.1 Standby mode 27
2.2 Read operation 29
2.3 Effect of PMOS mismatch on RNM using non-rotated original butterfly diagrams 35
2.4 Effect of the cell differential voltage on SNM and RNM of a conventional 6T SRAM cell 37
2.5 Write operation 40
2.6 Threshold voltage variation of the PMOS device on write delay 43
2.7 Effect of V_{DDM} variations on write and write recovery delay 44
2.8 Effect of cell voltage, cell ratio, and threshold voltage on SNM 45
2.9 Differential biasing scheme to mitigate the effect of V_{th} variations 45

3: Proposed 5T SRAM (5TSDG) Design 48
3.1 Standby mode 50
3.2 Read operation 64
3.2.1 V_{SSM} stability in dynamic mode of operation 74
3.2.2 Read power consumption 79
3.3 Write operation 81
3.3.1 Write Margin (WM) 88
3.3.2 Write power consumption 89

4: Layout Design 90
4.1 Layout extraction and Post-Layout Simulations (PLS) 95

5: Discussions, Conclusions and future work 96

Appendices 98
Appendix A 98
Appendix B 99

Reference 100

LIST OF FIGURES

Fig. 1.1 Personal computer (PC) memory hierarchy. 2

Fig. 1.2 Prediction of microprocessor technology density and area trends of 6T SRAM and 4T logic gates vs. year and technology node (Courtesy of [3]). 2

Fig. 1.3 A conventional 6T SRAM cell. 5

Fig. 1.4 SRAM block structure. 6

Fig. 1.5 Leakage current sources in an NMOS transistor. 11

Fig. 1.6 Cross-coupled inverters with the presence of DC noise sources v_n. 16

Fig. 1.7 Schematic of a 6T SRAM cell in read mode with two adverse polarity noise sources (v_n). 17

Fig. 1.8 SNM (Left) and RNM (Right) measurement technique using VTC curves of cross-coupled inverters *f* and *g* in standby and when accessed respectively. 18

Fig. 1.9 Butterfly diagrams of a conventional 6T SRAM cell showing the effect of the magnitude of the noise source, v_n (0-400mV), on SNM and meta-stable point. 19

Fig. 1.10 SNM estimation of conventional 6T SRAM, based on “maximum squares” in a 45° rotated coordinate system. 21

Fig. 1.11 Circuit implementation of SNM measurement using rotated butterfly diagrams. 23

Fig. 2.1 A conventional 6T SRAM cell. 27

Fig. 2.2 The butterfly diagrams of a conventional 6T cell in rotated coordinate system and the difference between two VTC curves (Bottom) showing the effect of mismatch in the V_{th} of the PMOS device (load) on SNM. 28

Fig. 2.3 Effect of V_{th} mismatch of the PMOS device on SNM in a conventional 6T cell (TT coner). 28

Fig. 2.4 A sense amplifier circuit used in read operation for a conventional 6T SRAM cell. 31

Fig. 2.5 Read operation in a conventional 6T SRAM cell. 31

Fig. 2.6 Butterfly diagrams of a conventional 6T cell in rotated coordinate system and the difference between the two VTC curves (top) showing the effect of mismatch in the threshold voltage of PMOS device on RNM. 34

Fig. 2.7 Effect of PMOS mismatch on RNM. 34

Fig. 2.8 Original style butterfly diagrams showing the effect of P-type transistor V_{th} mismatch on RNM. 35

Fig. 2.9 SNM and RNM comparison and the effect of PMOS V_{th} mismatch. 37

Fig. 2.10 Family of butterfly diagrams (top) and the VTC difference function curves of a conventional 6T cell illustrating the effect of cell voltage variations on SNM. 38

Fig. 2.11 Family of butterfly diagrams and the VTC difference function curves (Bottom) of a conventional 6T cell illustrating the effect of cell voltage variations on RNM. 39

Fig. 2.12 Effect of cell voltage on SNM and RNM in a conventional 6T cell (V_{SSM}=0V). 39

Fig. 2.13 Simplified write circuits for 6T SRAMs (Bit-line pre-charge pull-up devices not shown). 41

Fig. 2.14 Write operation in a conventional 6T SRAM cell. 43

Fig. 2.15 PMOS V_{th} mismatch effect on write delay in the conventional 6T cell. 44

Fig. 2.16 Effect of V_{DDM} variation on write delay in the conventional 6T cell. 44

Fig. 2.17 Comparisons of write recovery period for single vs. differential V_{DDM} methods at V_{DD}=0.75V, V_{DDM1}=0.60, V_{DDM2}=0.85 (SS corner, 120℃). 47

Fig. 3.1 5TSDG structure with sub-column V_{SS} control. 50

Fig. 3.2 A 5T SRAM cell with two adverse polarity noise sources v_n. 53

Fig. 3.3 Relationship between Vmin=V_{DDM}-V_{SSM} and SNM of 5TSDG in various corners. 54

Fig. 3.4 RNM variations vs. bit-line biasing (=V_{SSM}) - Supply voltage (Worst Case, FS Corner). 54

Fig. 3.5 5TSDG cell: Storing a ‘0’ (Left) and a ‘1’ (Right). 55

Fig. 3.6 Normalized standby power consumption of various V_{th} combinations in a 5T cell. 57

Fig. 3.7 Effect of Temperature on V_{SSM} with and without voltage regulation for different corners. 59

Fig. 3.8 5TSDG vs. a low-power 6T cell comparison when storing a ‘0’. 60

Fig. 3.9 5TSDG vs. low-power 6T SRAM designs: (64Kb SRAM array and peripheral circuits, FF corner on the left, 120℃, 50% ‘0’s 50% ‘1’s stored in the array). 61

Fig. 3.10. Butterfly diagrams for different corners for 5T SRAM in standby (SF, SS, TT, FF, FS corners from bottom up). 62

Fig. 3.11 Sub-column structure of 5TSDG. 63

Fig. 3.12 Sense amplifier used in 5TSDG read operation. 65

Fig. 3.13 Read operation in a 5TSDG cell, typical corner (TT), 120℃, Q0-Q0z-BL0-Gbit0 and Q1-Q1z-BL1-Gbit1 are related to a cell storing a ‘0’ and ‘1’ respectively and sharing the same word-line (WL). 66

Fig. 3.14 Comparison between 5TSDG when accessed (Left) with the cell in standby with the presence of the noise sources, v_n. 67

Fig. 3.15 Effect of V_{th} variations on RNM varying each single transistor in a 5T cell (Worst case: Relative to FS corner, 120℃). 67

Fig. 3.16. RNM variations vs. bit-line biasing voltage (=V_{SSM}) for three threshold voltages of the access transistor (N_3) in FS corner. 68

Fig. 3.17 RNM variations vs. bit-line biasing (=V_{SSM}) for various corners in 5TSDG. 69

Fig. 3.18 Effect of the threshold of P_2 transistor on RNM. All transistors are HVT unless otherwise specified. 70

Fig. 3.19 Effect of changing the threshold voltage of the load transistors (P_1 and P_2) on optimum RNM (All transistors are HVT unless otherwise specified. 71

Fig. 3.20 Read '0' and read '1' operations in 5TSDG in different corners (superimposed). BL0-Gbit0, and BL1-Gbit1 are the bit-lines and sense amplifier outputs for cells storing a '0' and '1' respectively. 72

Fig. 3.21 Standard (Left) and rotated (Right) butterfly diagrams of 5TSDG when accessed (Read mode) in different corners. 73

Fig. 3.22 Butterfly diagrams of 5TSDG and low-power 6T cells in read and standby modes: a: RNM_{5T} , b: RNM_{6T}, c: $SNM_{5T} \approx SNM_{6T}$. 74

Fig. 3.23 Read delay comparison in 5TSDG for different interesting threshold voltages in the cell (All transistors are assumed to be HVT unless otherwise specified). 74

Fig. 3.24. The response of V_{SSM} when forced to '0' volts and left floating in standby for 64Kb, 128Kb, 256Kb, and 512Kb from left to right for 5T SRAM array (FF corner, 120℃). 77

Fig. 3.25 Worst case effect of read operation on V_{SSM} (64Kb array 16 bits/word (all '0's), FF, 120℃). 78

Fig. 3.26 V_{SSM} Saturation during several read operations for 64Kb, 1Mb and 2Mb 5T SRAM arrays (64 bits/word, FF, 120℃). 78

Fig. 3.27 Comparison of normalized read and standby power vs. V_{DDM} for 5TSDG, 64x16 bit block, reading 16 '0's continuously from 16-bit words (FF corner, 120℃). 80

Fig. 3.28 Case study results of the worst-case normalized power consumption for standby mode and read operation of 5TSDG in comparison with low-power 6T design (FF, 120℃), 1.00 ≈ 25.8 mW. 81

Fig. 3.29 The write circuitry for 5TSDG (Not including the ground control circuit needed in write operation). 84

Fig. 3.30 W1 operation of 5TSDG in slow corner (SS) (120℃). 84

Fig. 3.31 W1 operation and effect of various V_{th} variations applied on each transistor at a time (-6σ to 4σ) relative to SS corner in a writing and

a disturbed cell (Q, Qz and Qd, Qdz), 120℃, Note that Qdz is not disturbed. 85

Fig. 3.32 Write '1' (left) and Write '0' (right) delay comparison in different corners between a 5TSDG and a low-power 6T cell (120℃). 85

Fig. 3.33 Effect of Vth mismatch on write '1' delay in worst case corner (SS), V_{g1}=0.07, V_{g2}=0.47 86

Fig. 3.34 SNM in W1 disturbed cells 87

Fig. 3.35 The SNM of W1 disturbed cells in different corners and different M_{eq} strengths. 88

Fig. 3.36 Case study results of the worst-case write power consumption in comparison with read and standby power for 5TSDG vs. low-power 6T design (FF, 120℃), 1.00 ≈ 33.8 mW 89

Fig. 4.1 5TSDG (a) and conventional 6T (b) cell layouts. 92

Fig. 4.2 5TSDG (a) and conventional 6T (b) pair of cells sharing same word-line (WL), Total lengths and widths not to precise scale. 93

Fig. 4.3 A 4x4 (a), and a 64x16 cell array of 5TSDG. Total lengths and widths not to precise scale. 94

LIST OF TABLES

TABLE 2.1 Maximum vertical distance measurements of RNM using non-rotated butterfly diagrams. 36

TABLE 3.1. Different types of SRAM cells used in this chapter, $V_{DD}=V_{DDM}=1.3V$, ... 50

TABLE 3.2. Transistor biasing voltages in standby for a cell storing a ‘0’ (Q=’0’). 55

TABLE 3.3 Transistor biasing voltages in standby for a cell storing a ‘1’ (Q=’1’). 56

TABLE 3.4. Worst-case standby leakage current and RNM comparison in different SRAM types (not including peripheral circuits). 60

TABLE 3.5 5TSDG: Trip point voltage vs. Q_{max} and Q_{min}, 120℃). 68

TABLE 3.6. RNM and W1 comparison for different V_{th} for N_3, at worst case RNM (FS corner, 120℃). 83

GLOSSARY

BL	Bit-line
BSIM	Berkeley Short-channel IGFET Model
CMOS	Complementary Metal Oxide Semiconductor (PMOS and NMOS)
CR (β)	Cell Ratio
DRAM	Dynamic Random Access Memory
DRV	Data Retention Voltage
DIBL	Drain Induced Barrier Lowering
GIDL	Gate Induced Drain Leakage
Gwr	Global Write
Gbit	Global Bit-line (Global Read)
HVT	High Threshold Voltage
I_{sth}	Sub-threshold Current
I_g	Gate Leakage Current
LER	Line Edge Roughness
LG	Gate Length
LVT	Low Threshold Voltage
NBTI	Negative Bias Temperature Instability
PBTI	Positive Bias Temperature Instability
PC	Personal Computer

RD Random Dopant

RNM Read Noise Margin

SNM Static Noise Margin

SPICE Simulation Program with Integrated Circuit Emphasis

SRAM Static Random Access Memory

SVT Standard Threshold Voltage

V_{DDM} Positive Supply Voltage of Memory

V_{SSM} Negative Supply Voltage of Memory

VTC Voltage Transfer Curve

V_t Thermal Voltage

V_{th} Threshold Voltage

WL Word-line

WM Write Margin

W1M Write ‘1’ Margin

W0M Write ‘0’ Margin

5T Five-Transistor

6T Six-Transistor

β_p Load Ratio

σ_N Number of standard deviations of the threshold voltage of N-type MOS

1: INTRODUCTION

With the recent aggressive CMOS technology scaling, standby leakage power is increased nearly five times each technology generation while active power remains constant [7]. In addition, process variations and hence performance fluctuations are widely noticed in 65nm and beyond CMOS technologies [5]. Fig. 1.1 illustrates the memory hierarchy of a personal computer (PC). A significant part of today's microprocessor chips consists of SRAM cells and it is predicted to reach 90% of the chip area in some applications by 2013 [3][9]. Although not the most power and area efficient type, *Static Random Access Memories* (SRAMs) are attractive due to their higher reliability and performance compared to other types of memories. Thus, they are used extensively in processors as cache memories in different levels, and consequently account for a large amount of power consumption, and the chip area in most System on Chip (SoC) applications. Fig. 1.2 demonstrates a prediction of the transistor density in high volume microprocessors consisting of 6T SRAM cells and 4T logic gates and area trends with respect to year and technology nodes [3]. Therefore, SRAM density growth is much more rapid than the logic elements and hence, it shows the significance of development of new power and area reduction techniques for this type of memory. Leakage power reduction, especially in more advanced technologies, is particularly important in portable and battery-powered electronics such as cellular phones, PDAs,

wireless, and low-power biomedical devices since standby leakage power ultimately determines the battery life.

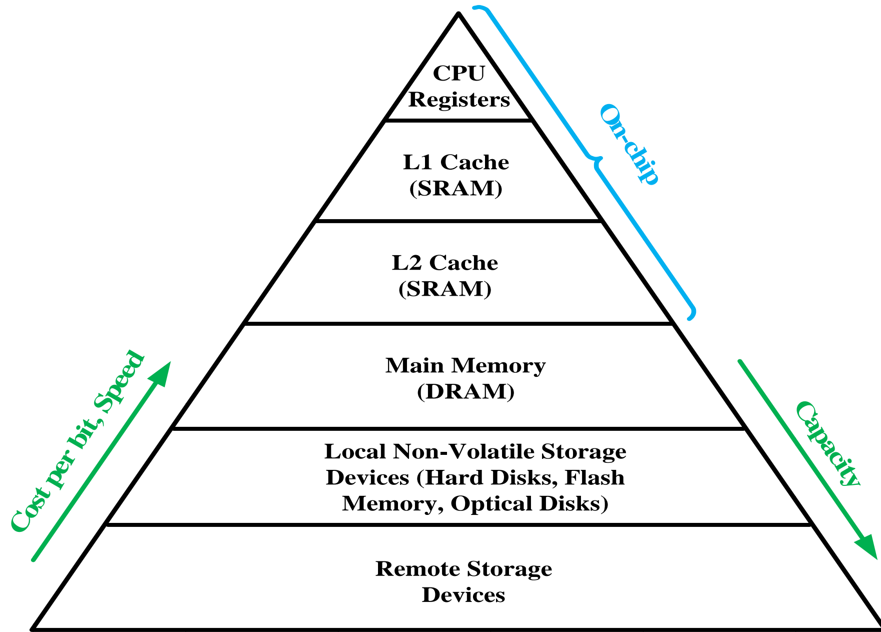
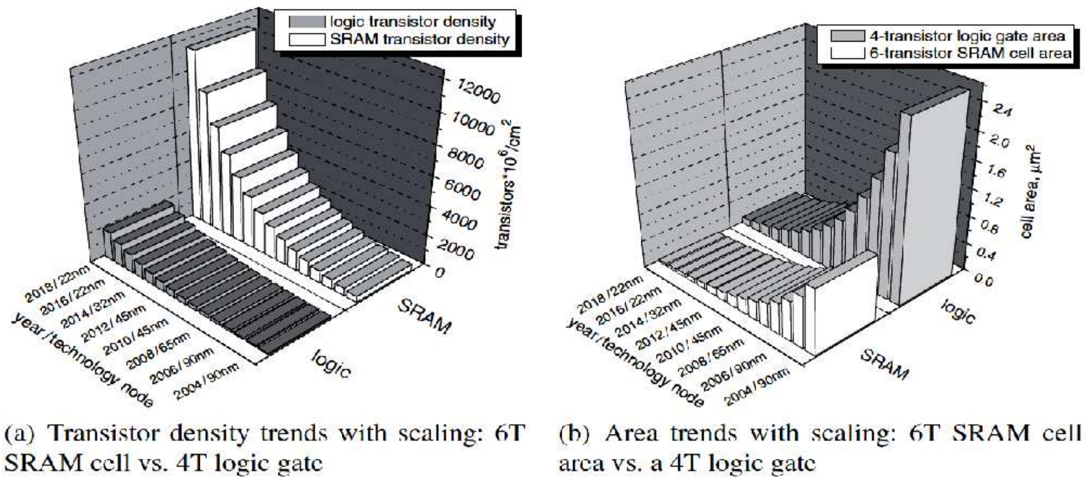


Fig. 1.1 Personal computer (PC) memory hierarchy.



(a) Transistor density trends with scaling: 6T SRAM cell vs. 4T logic gate

(b) Area trends with scaling: 6T SRAM cell area vs. a 4T logic gate

Fig. 1.2 Prediction of microprocessor technology density and area trends of 6T SRAM and 4T logic gates vs. year and technology node (Courtesy of [3]).

Unlike other types of storage devices, such as DRAMs, where data needs to be refreshed due to the destructive nature of their operations, SRAMs have higher performance in the sense that the stored value needs to be written to the cell only once. Therefore, SRAMs are storage elements, which are suitable to high performance applications although they consume more power. 6T SRAM cells are typically used in high performance cache memories in microprocessors due to their symmetry in cell layout, high stability and reliability. *Five-transistor* (5T) SRAMs are attractive due to their advantage in area and power efficiency compared to 6T SRAMs due to one fewer transistor and a single bit-line [1][2][14][15]. Research in the past on this type of memory has been mostly focused on improving performance and stability while maintaining the promised area saving in a particular technology node. On the other hand, with continuous scaling down of CMOS transistors, new techniques have been developed in 6T SRAMs such as Dynamic Standby Mode [9][17], *Data Retention Voltage* (DRV) method [7], and well biasing, some of which are summarized in [7] and [9]. Therefore, in order to suppress leakage power consumption and combat performance fluctuations due to process variations, the previous research in 5T SRAMs such as [14] and [15], can no longer compete with current 6T SRAMs and that is why 6T SRAMs are still predominantly used in current systems.

In this thesis, an improved low-power design of a 5T SRAM cell featuring differential ground lines (5TSDG), and a novel biasing technique are introduced that will guarantee operation under all process variations and temperatures while taking benefit of area reduction. Also, this design has improved performance

compared to previous research results in [2][14][15]. Another 5T SRAM type known as the "Portless" memory, is introduced in [23] which appears to need larger PMOS and access transistors than 5TSDG.

Throughout this thesis, all simulation results have been obtained using ST 65nm CMOS models with BSIM4 and HSPICE. In the following sections of this chapter, some related background and previous research work is presented. In Chapter 2, operation of a 6T SRAM and simulation results addressing common issues are discussed. In Chapter 3, the proposed 5T SRAM (5TSDG) is introduced, and the results are compared with previous research in this area. In Chapter 4, layout design of 5TSDG structure is presented in comparison with the conventional 6T cell, and the area saving opportunity is illustrated. Finally, In Chapter 5, a summary of this research, potential future work, and conclusions are presented.

1.1 SRAM block organization and structure

A typical SRAM cell is, like other types of storage devices, a device that is capable of holding a ‘0’ or a ‘1’, and is also capable of being written to and read from without causing disturbance in neighbouring cells or the destruction of its own stored data. Essentially, a 6T SRAM cell consists of a pair of cross-coupled inverters connected to two wires known, as true and complementary bit-lines (BL, BLZ), via two access transistors enabling the cell to communicate with outside when word-line (WL) is enabled as shown in Fig. 1.3. The operation of a 6T SRAM cell is discussed in detail in Chapter 2.

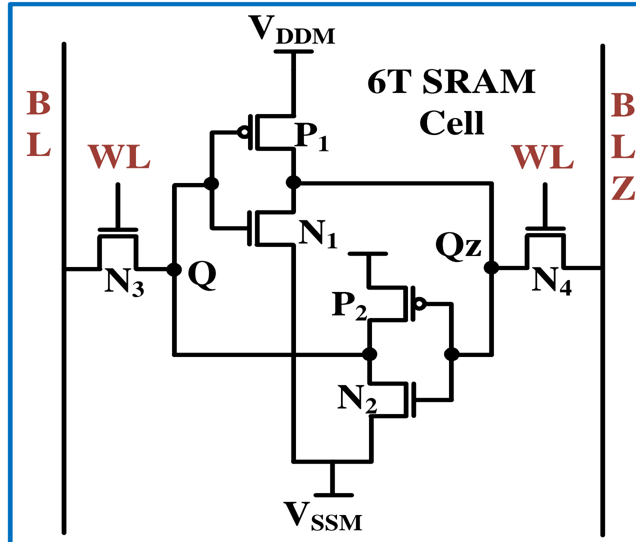


Fig. 1.3 A conventional 6T SRAM cell.

A simplified and popular 6T SRAM block structure is demonstrated in Fig. 1.4. A memory array is usually divided into blocks and sub-blocks, each including columns. Each column can be divided into two or more sub-columns. Row and column decoders can be staged with pre-decoders to optimize efficiency. In Fig. 1.4, an $M \times N$ sub-block is shown with M rows and N columns. In this example, there are no sub-columns and thereby, the sense amplifier and write circuits are required one per column. In a case where more sub-columns are required (for power and performance requirements), a sub-column and its adjacent one can share a sense amplifier. The row and column decoders decode M rows and N columns given the row and column addresses. For example, if a cell in row i and column j is to be accessed (C_{ij}), the row decoder will provide the word-line WL_i, and the column decoder will provide the column address to the write and sense amplifier circuits as well as the input output bus (Global I/O). The column decoder allows sharing a sense amplifier between 2, 4 or more sub-columns.

The control circuitry generates the timing signals required to control the memory array.

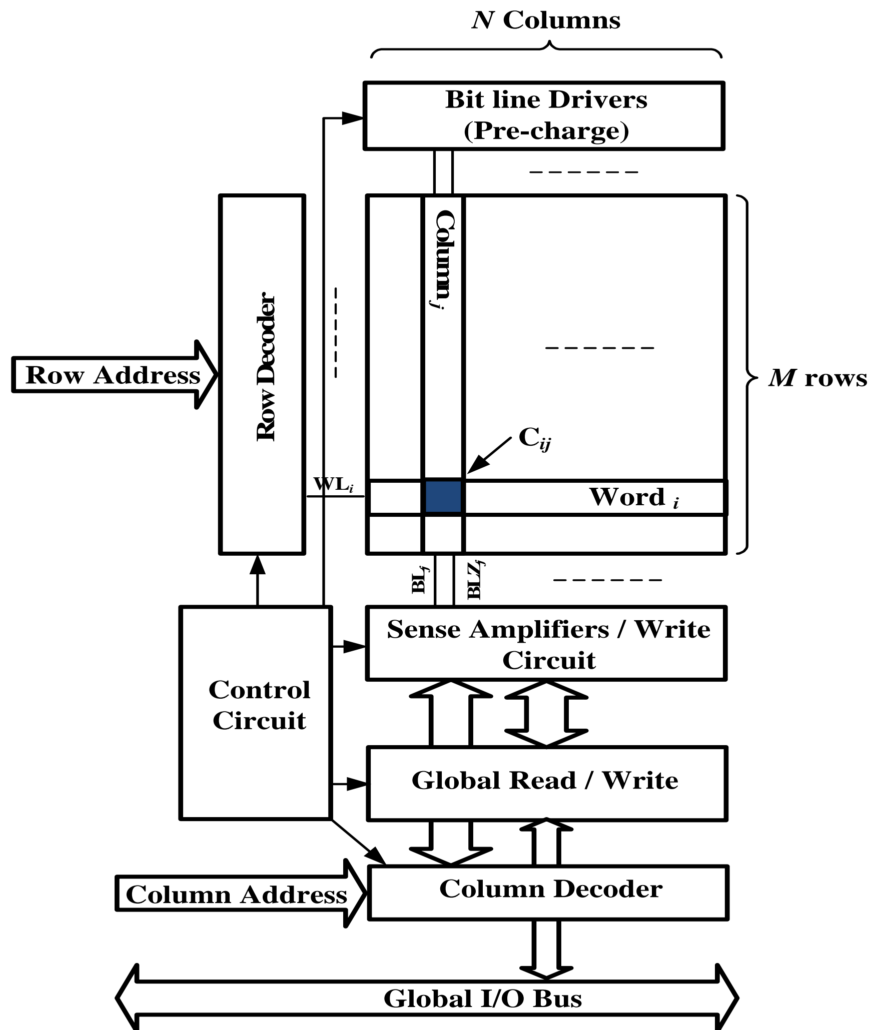


Fig. 1.4 SRAM block structure.

1.2 CMOS SRAM and technology scaling

A microprocessor chip mainly consists of cache memory and computing core(s). As shown in Fig. 1.1, SRAM cells are used as on-chip cache in different levels and are designed based on performance and power requirements. The demand for high-density SRAM structures continues to grow so rapidly that millions of SRAM cells, tightly packed beside each other, occupy most of the SoC

area. Hence, SRAMs will ultimately account for a major part of the chip cost and power consumption. Chip area efficiency, minimized to its design limits, low-power consumption, high performance, reliability in data storage, and a yield in ultra-large scales are the main objectives in any SoC. Therefore, it is essential to find a balance between cost, memory density, power consumption and performance depending on the design requirements. With increasing demand on these types of memories and growth of CMOS scaling, designers encounter complications and effects, such as leakage power and parameter mismatch, which were more or less negligible in previous generations. Technology scaling has the following major aspects:

- Scaling down of the transistor critical dimensions: Results in reduction of the total chip area and power consumption, and improves the performance due to reduction parasitics, and hence signal swing delay.
- Scaling the operation voltage: Results in lower static and dynamic power consumption, but may affect performance.

On the other hand, scaling transistor dimensions increases the design's susceptibility to random fluctuations in the transistor parameters, and therefore results in more challenges in manufacturing repeatability and reliability. In technology scaling, gate oxide thickness needs to be scaled to maintain effective gate control on the channel region. However, smaller gate oxide thickness will exponentially add to the leakage power [20]. Therefore, more accurate simulation models, such as BSIM v.4, are required for design purposes. Major types of process variations appear in the transistor's threshold voltage, V_{th}, and mobility,

which will cause a mismatch in drive current. SRAMs are low-current, sensitive, and bi-stable devices in which performance and reliability can be affected by parameter mismatch.

The significance of the above issues is better recognized in high density and large-scale memory structures containing millions of storage cells.

1.3 Threshold voltage variations and causes

Some of the major factors that cause threshold voltage mismatch and operation failure in neighbouring transistors specifically in an array of SRAM cells are summarized in [3][5][6][11] and are pointed out in the following:

- **Discrete Dopant (RD) effect:** Random fluctuations in the number of location of the dopant atoms in the channel,
- **Variation of poly-silicon critical dimensions:** gate length (Lg) variations,
- **Line Edge Roughness (LER):** Caused by 1) limitation of the resolution of lithography and 2) the grainy nature of photo resist and poly gate. The effect of line width roughness is random and cannot be corrected by optical proximity correction,
- **Short channel effects** (V_{th} roll-off, Drain Induced Barrier Lowering (DIBL, see section 1.4),
- **Gate dielectric material and thickness:** Reduction in thickness results in the reduction of the threshold voltage. Increase of thickness

will add to the probability of mismatch due to more surface potential perturbations.

The doping level of the channel needs to be increased to maintain a reasonable short-channel threshold voltage roll-off. As a result, the mobility of electrons and holes is decreased and the junction leakage, either from band-to-band tunnelling or from Gate-Induced Drain Leakage (GIDL), will be increased [11].

A comparison of threshold voltage variations in different technology nodes and transistor structures such as Poly gate with Oxide, Metal gate with high K dielectric, Omega FinFET, and Nanowire FinFet is reported in [11]. The comparison shows that in each transistor type, smaller dimensions will result in an increase in the V_{th} fluctuations and Nanowire FinFet has the smallest number of V_{th} variations among other types of transistors.

The standard deviation of the total threshold voltage variation, $\sigma_{Vth-total}$, can be modelled as the sum of variances of three major variation sources: Lg, LER and RD standard deviations, σ_{Vt-Lg}, σ_{Vt-LER} and σ_{Vt-RD} respectively [11], as expressed in equation 1.1.

$$\sigma_{Vth-total}{}^{2} = \sigma_{Vth-Lg}{}^{2} + \sigma_{Vth-LER}{}^{2} + \sigma_{Vth-RD}{}^{2} \qquad (1.1)$$

These V_{th} variation components for a nominal gate length of 35nm were studied by [10]. For a gate length of larger than 35nm, the effect of fluctuations due to random doping dominates the other two sources [10].

A comparison between different gate oxide thicknesses and corresponding variations in V_{th} is reported in [11]. The three sources of V_{th} variations can be reduced with decreased gate dielectric thickness due to less surface potential perturbation under the enhanced gate controllability. The reduction in thickness can be achieved by using metal-gate and high-k dielectric for low standby power devices. However, this improvement will depend on timely implementation of reliable high-k gate dielectric, and well work-function modulated metal gate [11].

There are other limiting factors in scaling transistors than process variation and leakage that affect the stability, cell area and performance of the SRAM. For example, NBTI (Negative Bias Temperature Instability), is a limiting factor in the scaling of PMOS. At high negative bias and elevated temperature, the PMOS V_{th} slowly shifts to become more negative and hence reducing PMOS current drive and affecting cell stability, margin, and minimum operating voltage of SRAM. This long-term V_{th} drift and other gate oxide related V_{th} degradation mechanisms must be accounted for over the life span of usage.

In addition, PBTI (Positive Bias Temperature Instability), is a limiting factor of NMOS scaling. With high-K gate materials, similar to NBTI V_{th} shift in NMOS also can occur [10].

1.4 Leakage current sources

SRAM cells are storage devices arranged in a large array of cells in which they are mostly in standby mode, holding either a ‘0’ or a ‘1’, and few are actually

in an active read/write mode. Before analyzing leakage sources in an SRAM cell, leakage sources in a standard CMOS transistor are studied. Fig. 1.5 shows major sources of leakage current in typical CMOS transistors. In general, leakage is increased by temperature and higher differential voltage between two nodes. In SRAM cells, a mixture of active and leakage current sources constitute the total power consumption of the memory array. The magnitude of the leakage current is relatively increased as technology is scaled, and therefore, leakage current can no longer be neglected as it used to be in older generations of CMOS devices. On the other hand, technology scaling has helped reduce the cell differential V_{DD}-V_{SS}, one of the effective methods known to suppress leakage power. Unfortunately, this causes delay in digital logic design by reducing the gate drive [20]. A thorough analysis of leakage current sources is reported in [20] and is summarized in the following sections.

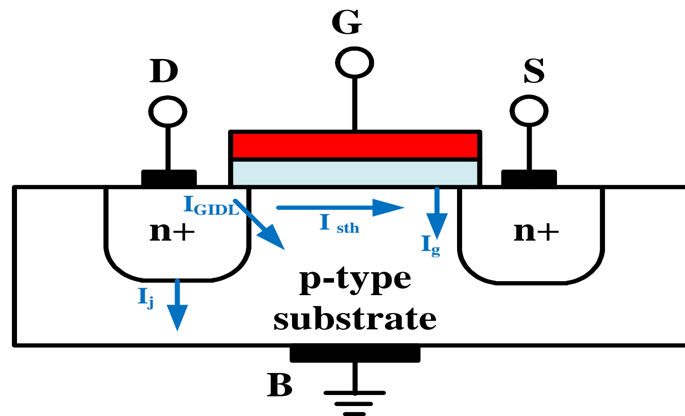


Fig. 1.5 Leakage current sources in an NMOS transistor.

Major sources of leakage in MOS devices are as follows:

1.4.1 Reverse-biased junction leakage current (I_j)

Also known as junction leakage, flows from the drain or source of a MOS device to the substrate. In an NMOS, the substrate is often tied to ground to reduce threshold voltage and hence optimize performance. Therefore, with increase of drain voltage, the reverse biased PN junction leakage is increased. This type of leakage depends on the drain area and doping concentration. In addition, it has a high temperature dependency and is generally negligible compared to other types of leakage.

1.4.2 Gate induced drain leakage (I_{GIDL})

This type of leakage occurs due to the field effect in the drain junction and is highly dependent on the drain to substrate voltage V_{DB}, drain to gate voltage V_{DG}, and oxide thickness Tox. Higher V_{DB} and smaller Tox will increase GIDL with a fixed V_{DG}.

1.4.3 Gate direct-tunnelling leakage (I_g)

This type of leakage flows from gate to substrate through the direct tunnelling of electrons through a leaky insulating gate oxide. PMOS transistors have one order of magnitude smaller I_g compared to NMOS when using silicon dioxide in the gate. This type of leakage has an exponential relation with gate oxide thickness. Thicker oxide thickness will result in lower leakage current.

1.4.4 Sub-threshold (weak inversion) leakage (I_{sth})

This type of leakage flows from drain to source of the transistor operating in weak inversion region. In addition, this is the dominant leakage source in

current CMOS technologies in comparison with other types of leaking currents. In more advanced technologies, I_{sth} is noticed more widely since as the technology advances and transistor dimensions scale down, the threshold voltage of the transistor is also scaled. At lower threshold voltages I_{sth} is increased. Equation 1.2 shows how I_{sth} is calculated and how it is related to V_{th} [20].

$$I_{sth} = \frac{W}{L}\mu v_t C_s e^{\frac{V_{GS} - V_{th} + \eta V_{DS}}{n v_t}} (1 - e^{\frac{-V_{DS}}{v_t}}) \quad (1.2)$$

where, W and L denote the width and the length of the transistor, μ denotes the carrier mobility, v_t is thermal voltage which is a function of temperature, C_s is the summation of depletion region capacitance and the interface trap capacitance per unit area of the gate, η is the DIBL coefficient, n is the slope shape factor and is a function of oxide thickness and C_s, V_{th} is the threshold voltage of the MOS device. Therefore, reducing the oxide thickness which results in the reduction of V_{th}, will result in the sub-threshold current increasing exponentially.

In summary, the total leakage current of a CMOS transistor when it is OFF (V_{GS} = 0) will lack the gate leakage factor and can be expressed as

$$I_{lkg} = I_{sth} + I_j + I_{GIDL} \quad (1.3)$$

The total leakage current I_{lkg} does not include gate tunnelling leakage current, I_g, since that current only occurs when the MOS device is in the “ON” mode of operation whereas I_{lkg} is only indicating the total leakage sources when the gate to source voltage (V_{GS}) is zero.

1.5 SRAM operation failure

SRAM operation failure (soft errors) can be categorized to the following types:

- **Storage Hold (Standby) failure:** It can occur when the stored data state is, *flipped*, during the memory standby state due to the existence of noise in the environment. Some of the sources of disturbances are caused by α particles, cross talk, and thermal noise [16].
- **Read upset:** Also known as destructive read, occurs when the stored data in a cell is corrupted when accessed, due to the bit-line current flow into the memory.
- **Write operation failure:** Unsuccessful attempt to store new data into the cell.
- **Write disturb:** Also known as destructive write, occurs in an attempt to write into a cell causing a neighbouring cell to change its state.

1.6 Stability analysis

Cell stability is an important factor in SRAM design since it determines the sensitivity of the cell to process variations and operating conditions. Another important aspect in memory design is cell area since more than 75% of the total chip area is determined by the SRAM area. These two factors are correlated in the sense that increasing stability requires larger cells. The basic cross-coupled inverter pairs may seem deceptively simple in appearance; however, limited

success has been achieved to analytically model stability. Therefore, simulation methods are used in this thesis to measure and analyze stability margins.

A popular methodology to measure the stability and reliability of an SRAM cell in different operation modes is by measuring noise margins. “In general, Noise Margin (NM) is the maximum spurious signal that can be accepted by the device when used in a system while still maintaining the correct operation [3].” The noise margin of a cell in standby is called Static Noise Margin (SNM). In read mode, when the cell is accessed, the noise margin is known as Read Noise Margin (RNM). Methods to measure and calculate these stability parameters are explained in detail in the following sections.

1.6.1 Static Noise Margin (SNM)

Static noise margin (SNM) indicates the maximum tolerance of an SRAM cell to perturbations and is measured using the voltage transfer curves (VTC) of the inverter pairs used in the SRAM cell, and drawing a curve known as *Butterfly diagram* discussed in section 1.6.3.

Fig. 1.6 shows the schematic of the cross-coupled inverters *f* and *g* used in the SRAM cell in standby (access transistors OFF) and two DC static noise sources. SNM is defined to be the maximum value of the noise sources (v_n) that can be tolerated by the cell before initiating the state toggle. The presence of the two DC noise sources with opposite polarities applied to the input of each inverter will create a pessimistic disturbance condition for a cell in standby.

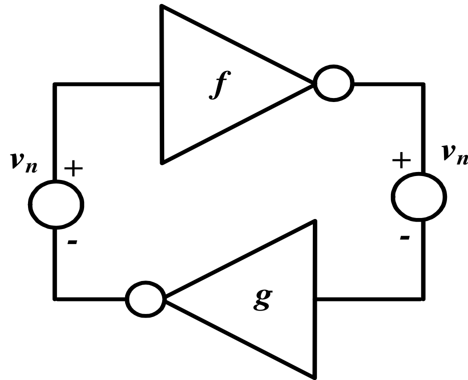


Fig. 1.6 Cross-coupled inverters with the presence of DC noise sources v_n.

SNM is reported to be reduced four times when transistors are scaled down from 250nm CMOS technology down to 50nm technology [3].

1.6.2 Read Noise Margin (RNM)

Similar to SNM, RNM is measured by drawing and mirroring the VTC curves of the cross-coupled inverters. The difference is that since during read operation the access transistors conduct and are connected to pre-charged bit-lines, in measuring RNM, the gates of the access transistors are clamped to V_{DD} and the bit-lines are clamped to the biasing voltage (in this case also V_{DD}) as shown in Fig. 1.7. Therefore, RNM indicates the tolerance of the cell to environmental perturbation or process variations when the cell is accessed.

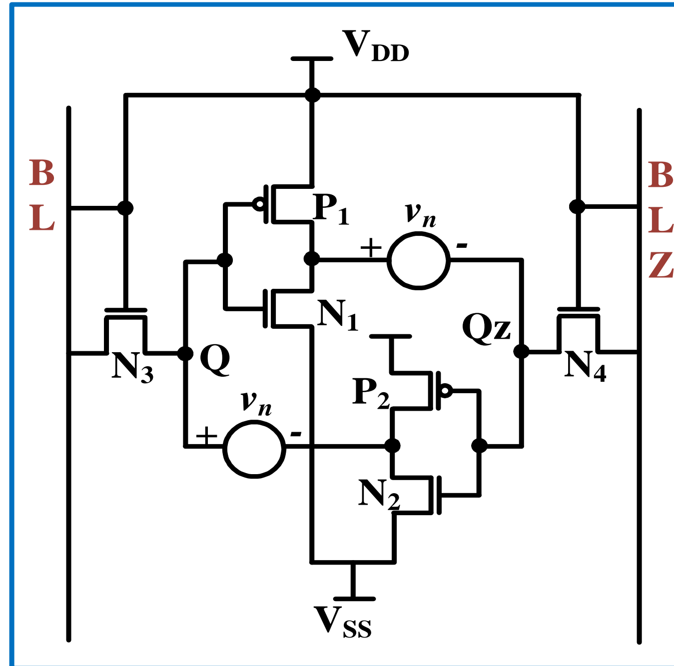


Fig. 1.7 Schematic of a 6T SRAM cell in read mode with two adverse polarity noise sources (v_n).

Fig. 1.7 is the schematic of a conventional 6T SRAM cell with added adverse static noise sources to the input of the inverters. The gates and the bit-line sides of the access transistors, N_3 and N_4, are connected to V_{DD} to construct a worst-case read condition. For measuring the static noise margin, the two access transistors are removed from the circuit while keeping the noise sources similar to the read case.

1.6.3 Butterfly diagrams

A butterfly diagram consists of the VTC curves of the two cross-coupled inverters in an SRAM cell. Fig. 1.8 depicts the graphical representation of the butterfly diagrams used to determine SNM and RNM of an SRAM cell with inverters *f* and *g*. Graphically, the diagonal length of the maximum square that can fit in each side of the butterfly diagram is defined to be the noise margin. This

curve has two steady state points at each end indicating the state of the cell storing either a ‘1’ or a ‘0’. The midpoint of the butterfly diagram, where the two VTC curves cross each other is known as the meta-stable point. If the amount of disturbance reaches this point, small perturbations will cause the cell to either restore the disturbed cell to its original stored value through the feedback path or initiate the state toggle or the data corruption depending on the direction of the perturbation. The meta-stable point is a good estimation of the trip point, though trip point is usually measured using a different circuit structure. In summary, if a perturbation occurs in the cell, the feedback inverter will restore the value to the steady state points as long as the perturbation is not larger than the trip point of the cell. If larger, the cell is flipped to read the opposite steady state point on the other side of the butterfly diagram.

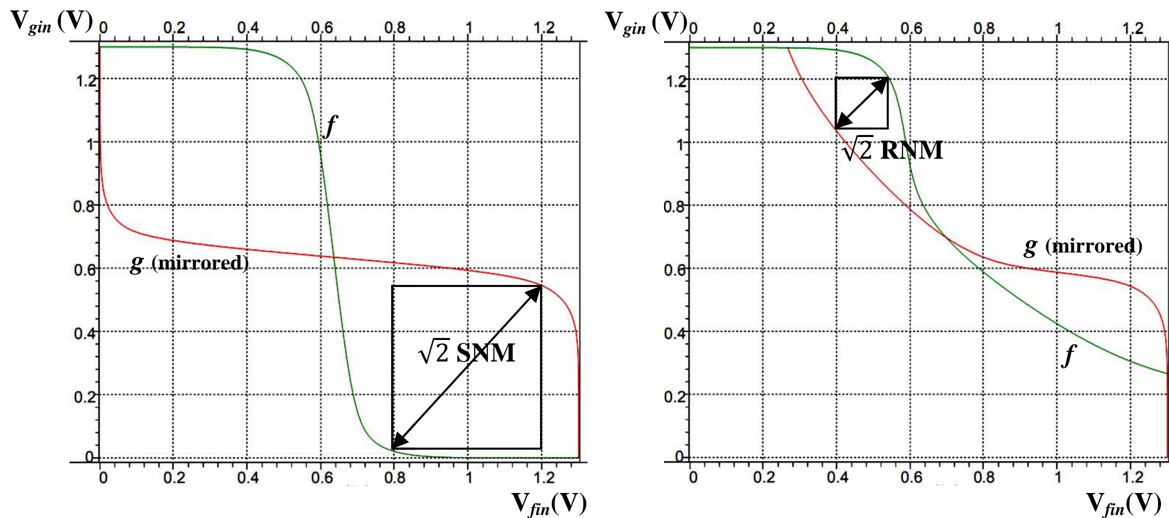


Fig. 1.8 SNM (Left) and RNM (Right) measurement technique using VTC curves of cross-coupled inverters *f* and *g* in standby and when accessed respectively.

The meta-stable point is shifted as the noise source increases in magnitude. With higher amount of the noise disturbance, SNM is decreased, meaning that the probability of data corruption in standby is increased.

This decay in the stability margin during read can affect the reliability of the data storage significantly meaning that once a bit is stored in a cell, it is more vulnerable to change its state after small perturbations in the system environment. Using carefully sized transistors can limit this decay in RNM.

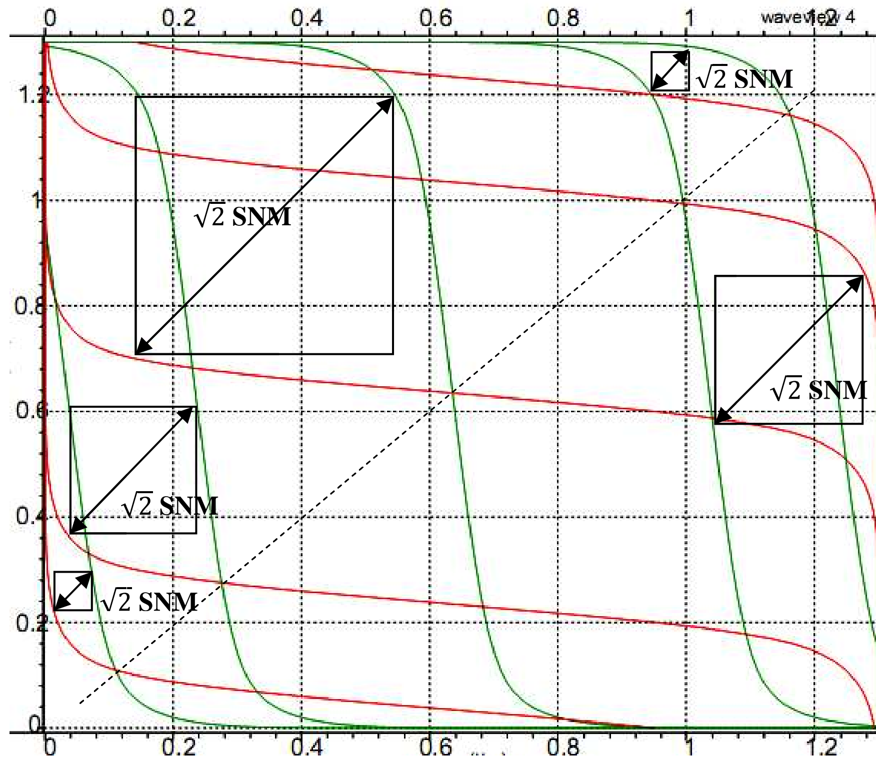


Fig. 1.9 Butterfly diagrams of a conventional 6T SRAM cell showing the effect of the magnitude of the noise source, v_n (0-400mV), on SNM and meta-stable point.

1.6.4 Static noise margin measurement using SPICE

A typical method to measure SNM or RNM is by first drawing and mirroring the VTC curves. The length of the diagonal of the maximum square fitting in each side of the butterfly diagram is graphically determined. Then the

minimum value of these two is selected to represent SNM for worst-case scenario. However, this method requires numerical work post-circuit analysis. In addition, many designers rely on visual graphical observations.

A method proposed by [16] is an alternative and “easy to use” technique in estimating SNM using Spice circuit simulations. The idea is to rotate the butterfly diagram by 45 degrees so that the diagonal of the squares appears to be vertical in the new coordinate system to facilitate measurements. Given that the original coordinate system in which VTC curves are drawn in a typical way has *x* and *y* coordinates, the new coordinate system will have *u* and *v* coordinates.

As shown in Fig. 1.10, *f* and *g* are the voltage characteristic curves of the cross-coupled inverters in the SRAM cell. *g*' is the mirrored version of *g*. Curve D is the difference between the two curves in the new 45 degree rotated coordinate system. The peaks of the difference curve are where in the original curve, the SNM values are maximized. By comparing the absolute values of peaks and choosing the minimum, RNM or SNM is calculated depending on the purpose.

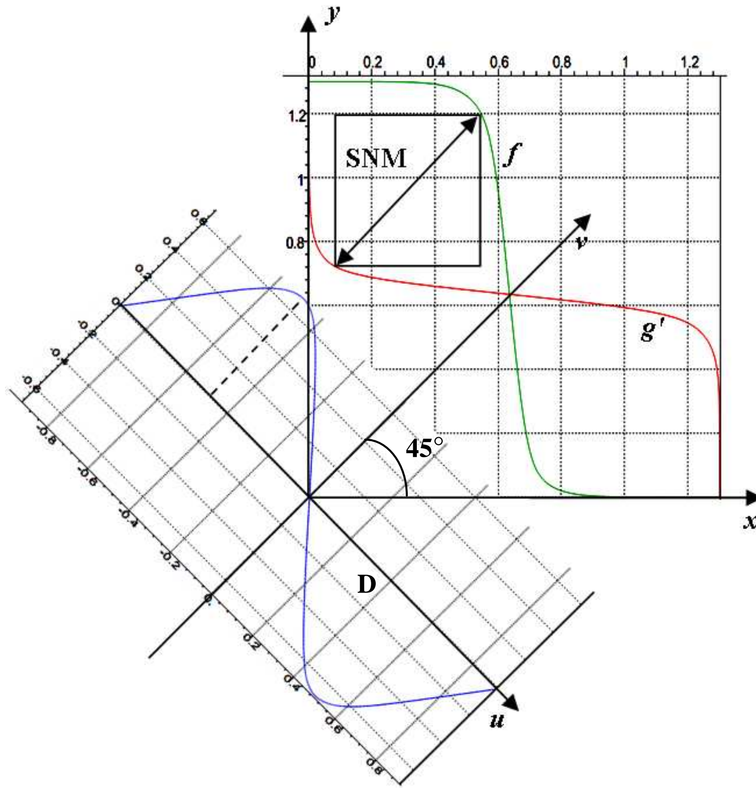


Fig. 1.10 SNM estimation of conventional 6T SRAM, based on “maximum squares” in a 45° rotated coordinate system.

In order to transform (*x,y*) coordinate to (*u,v*), the 45 degree rotated system is defined as:

$$x = \frac{1}{\sqrt{2}}u + \frac{1}{\sqrt{2}}v \qquad (1.4)$$

$$y = -\frac{1}{\sqrt{2}}u + \frac{1}{\sqrt{2}}v \qquad (1.5)$$

There are two inverter characteristic curves:

$$y_1 = f(x) \text{ and } y_2 = g(x) \qquad (1.6)$$

Substituting equation 1.4 and 1.5 in 1.6 we get:

$$-\frac{1}{\sqrt{2}}u + \frac{1}{\sqrt{2}}v_1 = f(\frac{1}{\sqrt{2}}u + \frac{1}{\sqrt{2}}v_1)$$

$$v_1 = u + \sqrt{2} f(\frac{1}{\sqrt{2}} u + \frac{1}{\sqrt{2}} v_1)$$

Similarly, $g'(x)$ can be transformed into the (u, v) coordinate system by first applying the inverse function operation on $F_2(x)$ to calculate $F'_2(x)$ and second by applying the rotation operation.

$$y_2 = g(x)$$

$$x = g'(y_2)$$

$$v_2 = -u + \sqrt{2}\, g(-\frac{1}{\sqrt{2}} u + \frac{1}{\sqrt{2}} v_2)$$

The circuit implementation of this method, which can be simulated using SPICE models, is discussed in further detail in [16] and is used in this thesis to measure and compare stability margins for both 6T and 5TSDG designs. Comparison results are illustrated in detail in Chapter 3. The schematic of the stability measurement circuit is depicted in Fig. 1.11.

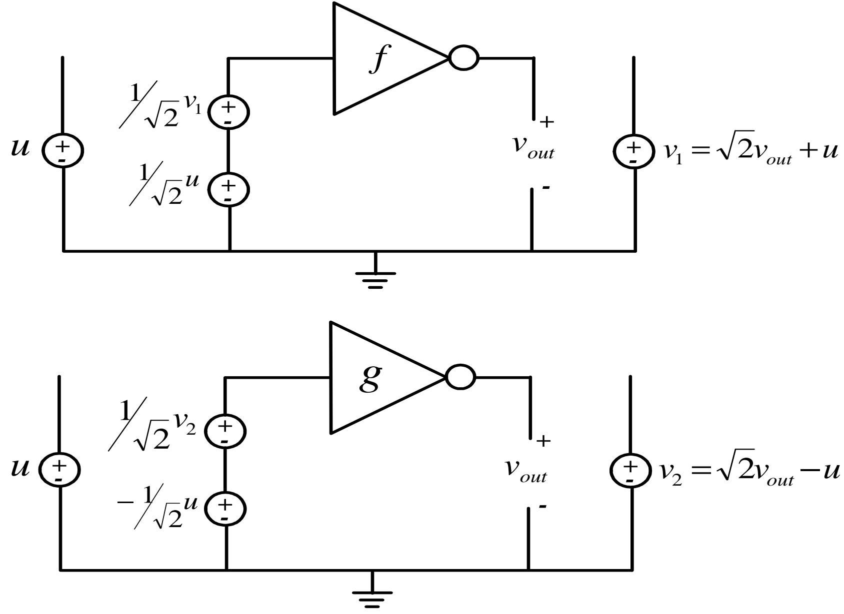


Fig. 1.11 Circuit implementation of SNM measurement using rotated butterfly diagrams.

The absolute values of the maximum and minimum are the values of the diagonals of the maximum squares. Finally, multiplying the smaller of the two by $1/\sqrt{2}$ will result in the SNM of the SRAM cell as shown in equation 1.7.

$$NM = \frac{1}{\sqrt{2}}.\min \begin{cases} \max\left(|v_1(u) - v_2(u)|,\ -\frac{1}{\sqrt{2}} < u < 0\right. \\ \max\left(|v_1(u) - v_2(u)|,\ 0 < u < +\frac{1}{\sqrt{2}}\right. \end{cases} \tag{1.7}$$

where v_1 and v_2 are the characteristic curves of the inverters and *u* is the sweeping DC voltage used to obtain v_1 and v_2.

1.7 Read noise margin and supply voltages

One of the important design factors in SRAM design is sensitivity to voltage scaling. As discussed earlier, more advanced technologies will enable standard digital IC designers to lower the supply voltage to reduce power consumption, and limit the voltage swing period, but of course with performance

penalties. An analysis on the effect of supply voltages on SNM and RNM in a conventional 6T SRAM is reported in [8]. The authors show the relationship between SNM of a 6T SRAM cell when accessed (RNM), cell supply voltage and bit-line pre-charge voltage. Higher cell voltages will cause more stability in the cell since the current drive of the storage inverters is increased due to higher differentials making it harder for the noise to take over the cell voltage and flip the cell value. On the other hand, higher bit-line pre-charge voltages will reduce the noise margin since a higher read upset voltage is induced during read due to higher current flow from the bit-line through the cell. On the other hand, lowering bit-line voltage below the trip point will also disturb the read operation by either flipping the cell (writing a '0' on the cell node storing a '1') or cause a false trigger on the sense amplifier. Therefore, the amount of charge stored on the bit-lines, which is directly related to the pre-charge voltage and bit-line wire length, must be limited by the designer. When the pre-charge voltage of the bit-line reaches a certain biasing level (near the cell supply voltage), [8] shows that RNM is saturated.

1.8 Technology corners and static noise margin in SRAM

In a conventional 6T SRAM design, yield is directly related to SNM distribution. If μ is the mean and σ is the variance of the SNM distribution in a memory block, it is reported that, $\mu-6\sigma$ of SNM is required to exceed $0.04\times V_{DD}$ to reach a 90% yield on 1MB SRAM. Transistor mismatch and parameter variations result in increased variations in threshold voltages. This will lead to a reduction in $\mu-6\sigma$ and increases the number of unstable SRAM cells which affects SRAM

yield. Typically, this translates into a requirement that minimum SNM, be greater than 20% of a typical SNM [3].

In [3], the authors report that in all corners, the symmetry point is mostly at 0% deviation from μ. In other words, the most stable case is when there is no mismatch in V_{th}. The effect of V_{th} variation is increased in devices with lower threshold voltages. In the following a summary of the mismatch effect on SNM is given for each transistor (see Fig. 1.7 for transistor names and schematic):

Driver transistors (N_1, N_2): Seem to have the largest effect on the butterfly diagram and hence, the SNM due to their larger W/L ratio compared to other transistors in the SRAM cell.

Access Transistors (N_3, N_4): As V_{th} decreases in the access transistor, a strong negative impact is observed on the RNM. In an experiment by [3], the measurements were taken in a worst-case read-accessed SRAM cell. Thus, the access transistors were effectively connected in parallel with the load transistors. Thus, reducing the V_{th} of the access transistor compromises the low level stored in the cell, which in turn reduces the RNM.

Load transistors (P_1 and P_2): Seem to have the least impact on the SNM due to its weaker current drive and typically smaller W/L ratio.

In Chapter 2, the design and operation of a conventional SRAM is presented in further detail.

2: 6T SRAM OPERATION

In this chapter, the design and operation of a conventional *six-transistor* (6T) SRAM is presented in various modes of operation. A conventional 6T cell consists of two cross-coupled inverters (P_1-N_1, P_2-N_2) connected to two bit-lines (BL, BLZ) through access transistors (N_3, N_4) as shown in Fig. 2.1. The word-line (WL) is used to access the cell during read or write. Therefore, the operations of a 6T SRAM cell can be categorized into standby mode, read, and write operations which will be explained in this chapter. In more advanced and low-power 6T designs, the positive and negative supply voltage, V_{DDM} and V_{SSM}, can be controlled by additional circuitry to limit the cell differential especially in standby mode to save power consumption. This procedure is also known as sleep mode method (see [20]) both in memory and digital logic design when the device is in standby. The transistors constituting the 6T cell are ratioed and depending on performance, area and power requirements, length and width parameters can be adjusted. Use of high threshold MOS devices is also another technique to reduce power with performance penalties. There are two important ratios in designing a conventional 6T cell denoted by cell ratio, β, and load ratio, β_p, in this thesis. In Fig. 2.1, P_1 and P_2 are also known as load transistors and N_1 and N_2 are known as driver, and N_3 and N_4 are called access transistors. The cell and load ratios are defined by

$$\beta = \frac{W_{N_2}/L_{N_2}}{W_{N_3}/L_{N_3}} = \frac{W_{N_1}/L_{N_1}}{W_{N_4}/L_{N_4}} \text{ and } \beta_p = \frac{W_{N_3}/L_{N_3}}{W_{P_2}/L_{P_2}} = \frac{W_{N_4}/L_{N_4}}{W_{P_1}/L_{P_1}}$$

where W and L denote the width and length of the MOS devices in Fig. 2.1.

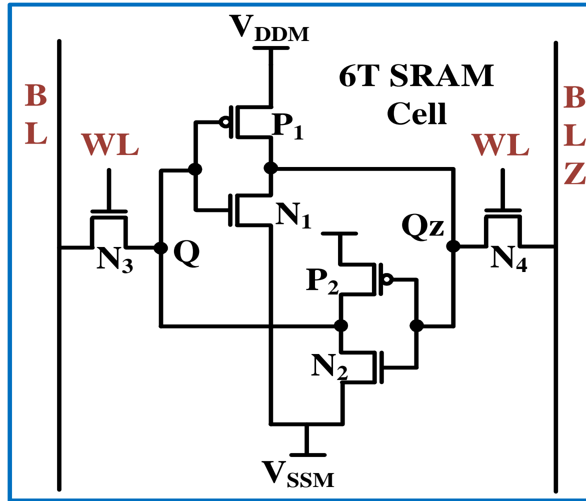


Fig. 2.1 A conventional 6T SRAM cell.

2.1 Standby mode

The two cross-coupled inverters will construct a true and a complementary storage node, Q and Qz, to store either a ‘0’ or a ‘1’ into the cell. These inverters play the role of a closed loop feedback system that can cancel the effect of perturbations to a degree, depending on the stability of cell, restoring the disturbed data to the original value.

As shown in Fig. 2.2, it can be observed that the SNM value will reach its highest when there is no mismatch in the PMOS threshold voltage i.e. $\sigma_{Vthp} = 0$. Also the higher the V_{th} variations on the PMOS in a 6T cell, the lower the value for SNM. As shown in Fig. 2.3, there is a symmetry across standard deviation

factor $\sigma_{Vthp} = 0$. The larger the σ_{Vthp} value, the threshold voltage for the mismatch PMOS will increase making it harder to turn on and work well as a pull up transistor and thereby reducing the SNM.

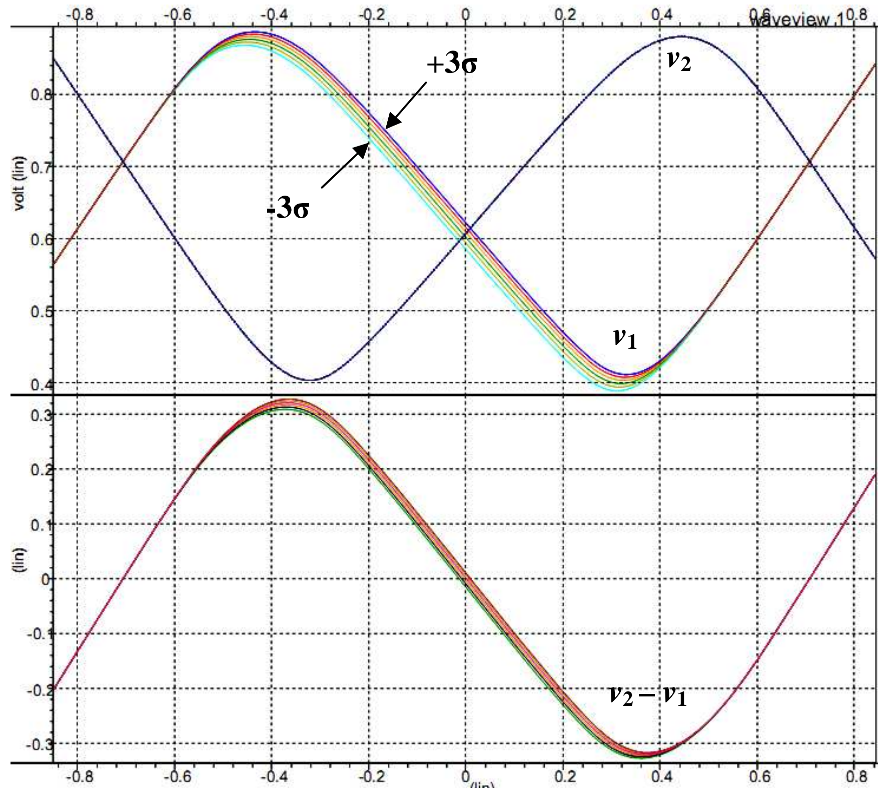


Fig. 2.2 The butterfly diagrams of a conventional 6T cell in rotated coordinate system and the difference between two VTC curves (Bottom) showing the effect of mismatch in the V_{th} of the PMOS device (load) on SNM.

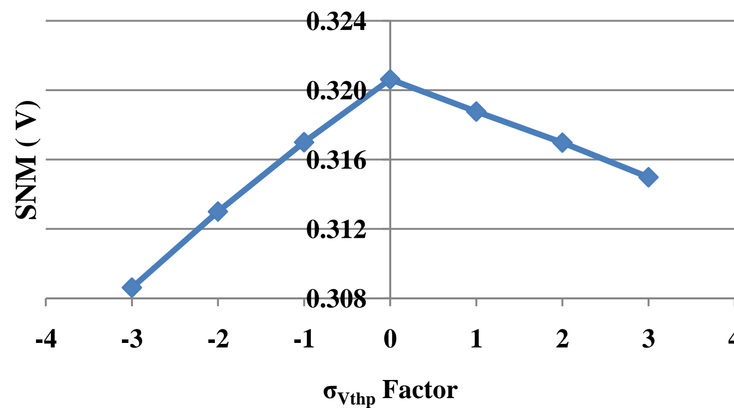


Fig. 2.3 Effect of V_{th} mismatch of the PMOS device on SNM in a conventional 6T cell (TT coner).

2.2 Read operation

In read operation, the bit-lines are first pre-charged high to V_{DD} and left floating. Since the bit-lines are shared with other cells in the same sub-column (see Fig. 1.4), they are capacitive elements consisting of mostly wire and junction capacitors. The schematic of a simple sense amplifier used in this thesis for the read operation of the 6T SRAM cell is depicted in Fig. 2.4. In a case when a ‘0’ is stored in the cell ($Q=V_{SSM}$, $QZ=V_{DDM}$), when the word-line, WL, is raised, the true bit-line (BL) is pulled down to V_{SSM} via N_2 while the complementary bit-line (BLZ) remains floating at the pre-charged value (V_{DD}) as shown in Fig. 2.5 (see also [3]). The sense amplifier can be designed to function with small differentials between BL and BLZ. Therefore, in order to increase performance and reduce the dynamic power consumption, the bit-line swing in read does not have to be from pre-charge voltage all the way down to V_{SSM} (see Fig. 2.5). The two PMOS devices connected to BL and BLZ (M_6 and M_7) are to isolate the sense amplifier from the bit-lines when the sub-column is not selected for read. With the presence of two cross-coupled inverter pairs, M_1-M_2 and M_3-M_4, this sense amplifier operates in a similar way as a memory cell does without the access transistors. M_5 is connected to the source nodes of the inverter pairs to disconnect them from V_{SS} in standby for pre-charge purposes. Small voltage differentials between BL and BLZ when column is selected i.e. select signal is active (*sel*), and the sense enable pulse (*se*) will cause the inverter pairs to turn on and drive the sense amplifier outputs *Gbit* and *Gbitz* to full rail voltages. M_8

and M_9 are pre-charge devices to pre-charge and let Gbit and Gbitz float prior to read, and are enabled by *prez* pulses. Generally, design parameters characterizing a sense amplifier are defined by

- The amplitude of the differential input signal
- The minimum detectable signal
- The gain (G = $\frac{V_{out}}{V_{in}}$)
- The tolerance against environmental disturbance and mismatches
- Sense delay which can be measured in different ways such as from when the word-line reaches 50% V_{DD} to when sense output (global bit-line) reaches 50% V_{DDM}.

The gain of the sense amplifier is a function of the bit-line pre-charge voltage. The sense delay is partly related to the SRAM cell ratio (β), and can be adjusted depending on the power and delay requirements. Larger β values will reduce read delays up to a saturation point, and will also add to the power consumption and the chip area.

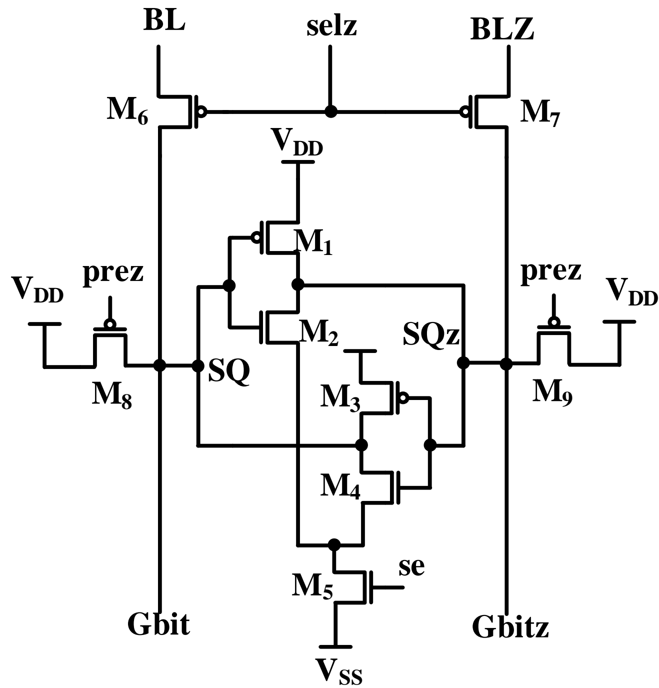


Fig. 2.4 A sense amplifier circuit used in read operation for a conventional 6T SRAM cell.

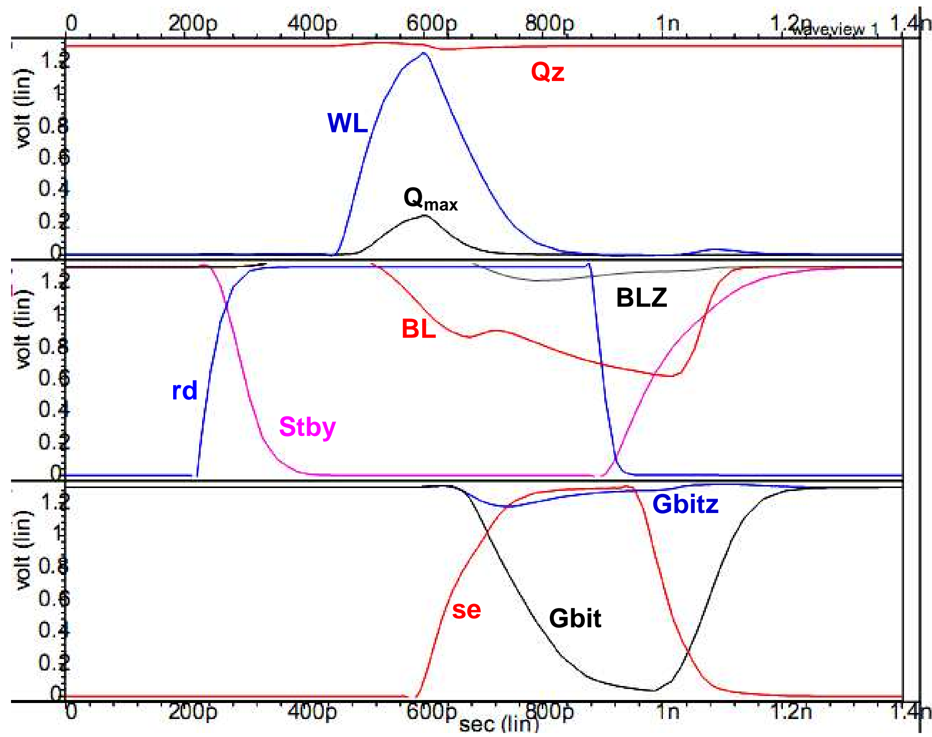


Fig. 2.5 Read operation in a conventional 6T SRAM cell.

Due to the capacitive nature of the bit-line, the current flow from either BL or BLZ through N_1 or N_2 during read when the word-line is raised, will cause a temporary rise on the '0' node of a 6T cell noted as Q_{max} shown in Fig. 2.5. Once the word-line is raised, the drain to source voltage (V_{DS}) of N_2 is 0V since a '0' is stored into the cell, and hence, N_2 is weakly ON. The cell node Q has a capacitance of C_Q which will be charged by the current flow from the bit-line, and once charged, the V_{DS} of N_2 will no longer be 0V, which then will turn N_2 on more strongly and thereby the charges on C_Q will be drained to V_{SS}. The amount of charges flown from BL to Q is dependent on the capacitance of the bit-line, C_{bit} and the cell ratio, β.

The length of the bit-lines, which is determined by the number of cells shared on the same wire, determines the bit-line capacitance, and consequently influences the value of Q_{max} along with β. The value of Q_{max} during read has a direct effect on read-stability since if Q_{max} is larger than the trip point of the inverter pairs, the read operation will turn into a write and hence the stored data is corrupted. Furthermore, read performance is also affected by longer bit-lines since it will take longer for the bit-lines to be pulled down by the driver transistor (N_1 or N_2) in the cell depending on the stored value. Therefore, in conventional 6T cells, cell ratio (β) is designed based on the length of the bit-lines to maintain stability, and improve performance. Smaller bit-line segments will result in less power consumption during read but require more peripheral circuits and hence more area is needed. Therefore, a designer should find a reasonable balance between required read performance, area and power. In addition, cell β plays an

important role in read-stability meaning that larger values of β will cause Q_{max} to be reduced since the driver transistor (N_1 or N_2), driving the '0' side of the cell, is turned on more strongly during read.

Mismatch in the threshold voltage of a cell can result in a cell to become more susceptible to data corruption. As shown in Fig. 2.6, it is observed that the RNM value will reach its highest when there is no mismatch in the PMOS threshold voltage factor i.e. $\sigma_{Vthp} = 0$. In addition, V_{thp} variations on the PMOS in a 6T cell will result in lower values for RNM. It can be easily observed that there is a symmetry across $\sigma_{Vthp} = 0$. Larger σ_{Vthp} values translate into increase of threshold voltage making the transistor harder to turn on and hence reducing the RNM. Another phenomenon to observe is that the static noise margin value is nearly twice as much the RNM margin when compared. This reduction is due to the fact that the cell is much more vulnerable to state loss when it is being accessed than when it is in the standby mode. When the cell is being accessed, the two access transistors are on and one of the bit-lines is pulled down by one of the NMOS devices of the cross-coupled inverter pairs (N_1 or N_2). If the current flowing in the channel of N_3 or N_4 is higher than N_1 or N_2 respectively, the read operation will cause a read upset turning a read into initiating a state toggle.

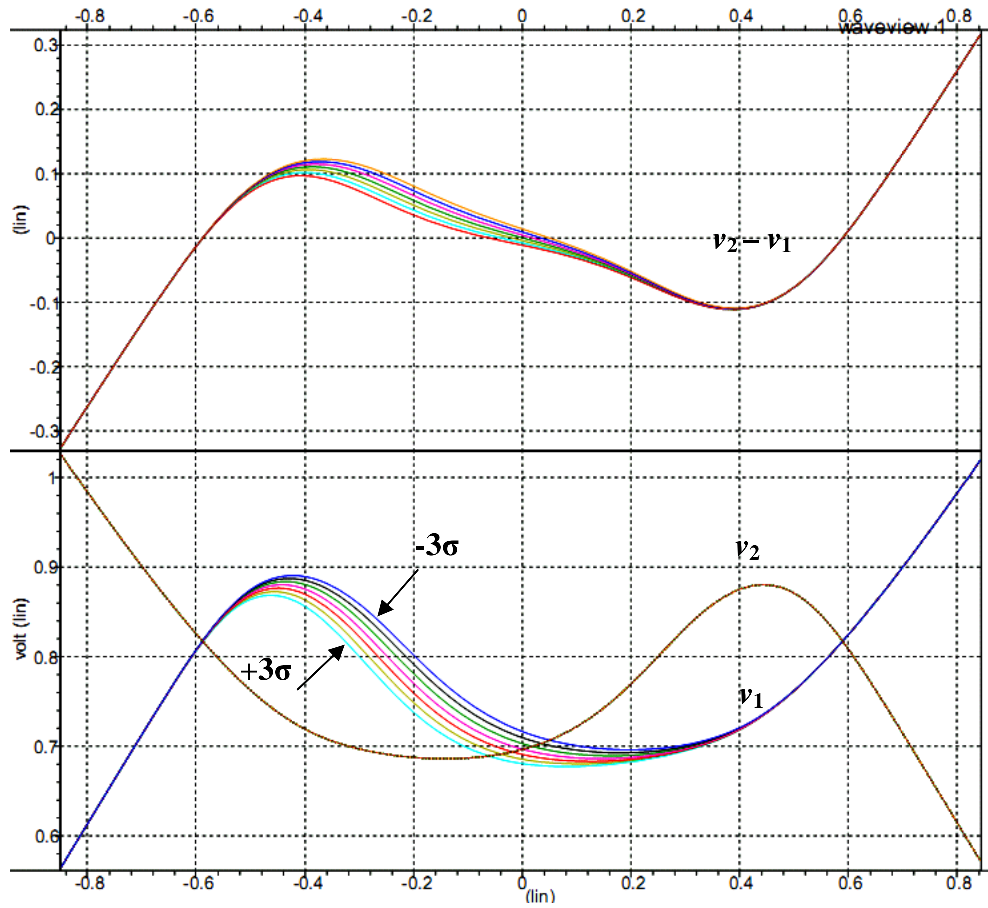


Fig. 2.6 Butterfly diagrams of a conventional 6T cell in rotated coordinate system and the difference between the two VTC curves (top) showing the effect of mismatch in the threshold voltage of PMOS device on RNM.

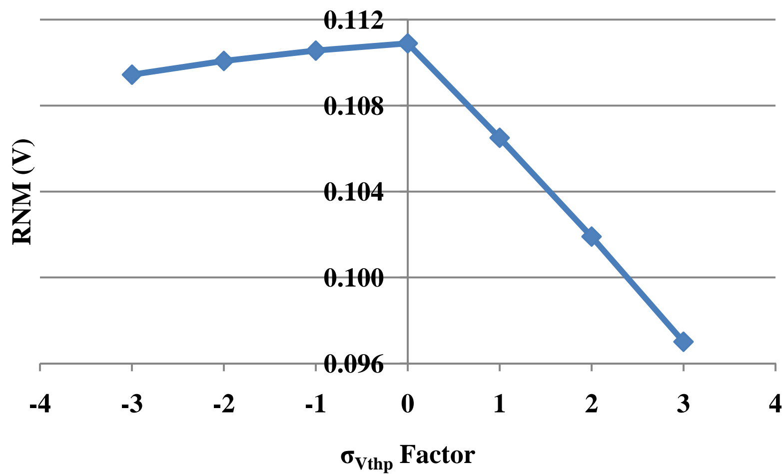


Fig. 2.7 Effect of PMOS mismatch on RNM.

2.3 Effect of PMOS mismatch on RNM using non-rotated original butterfly diagrams

In order to verify the simulation results from the rotated coordinate system method, another experiment was run using regular VTC curves. Fig. 2.8 depicts the original 6T SRAM butterfly diagram consisting of the two VTC curves of the inverters. Comparing these curves with the rotated ones in Fig. 2.6, it is observed that the results are consistent numerically. The central solid line belongs to $\sigma_{Vthp} = 0$. Graphically observing this figure, the larger the σ_{Vthp} value gets numerically, the diagonal length of the biggest square in each side will become smaller.

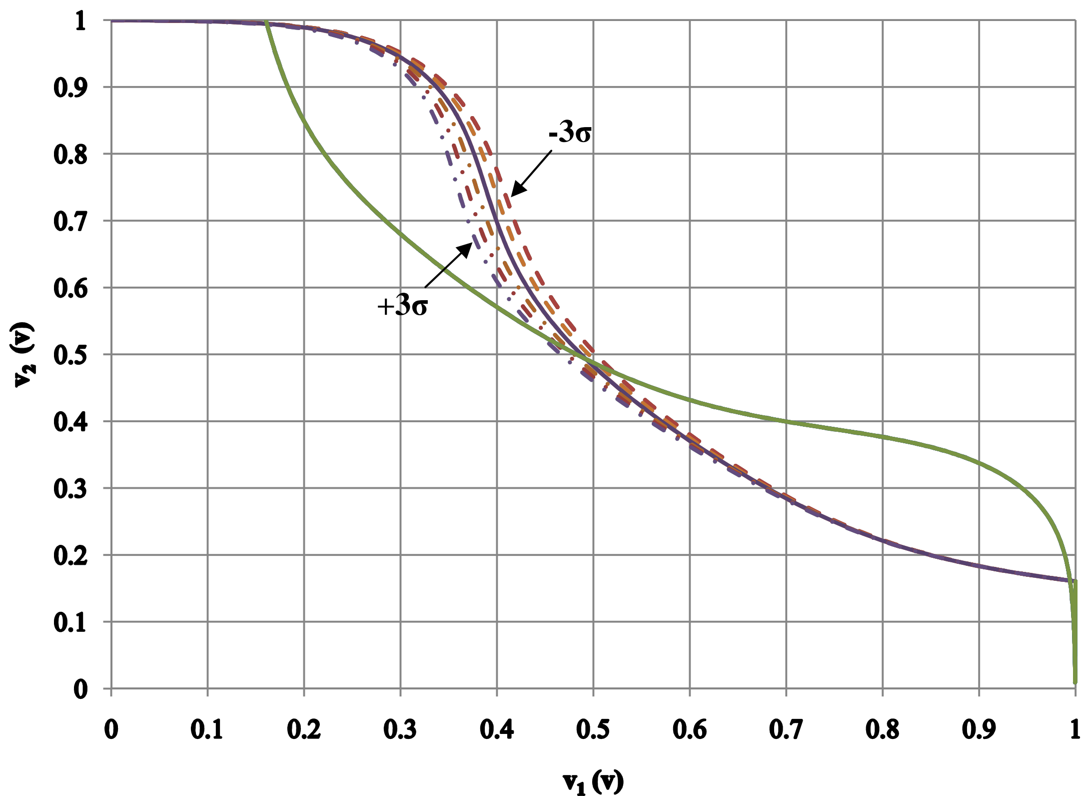


Fig. 2.8 Original style butterfly diagrams showing the effect of P-type transistor V_{th} mismatch on RNM.

It might be an interesting question to investigate how vertical measurements between the standard VTC curves translate into SNM. Fig A.1 depicts the non-rotated butterfly family of curves with maximum vertical distance measurements. The disadvantage of the vertical measurement is that the maximum distance never switches to the other side of the curve since it is not diagonal. These graphs are illustrated just to compare the results with the rotated versions explained earlier. The maximum vertical measurements are shown in TABLE 2.1.

TABLE 2.1 Maximum vertical distance measurements of RNM using non-rotated butterfly diagrams.

σ factor	**-3**	**-2**	**-1**	**0**	**1**	**2**	**3**
Vertical Distance(mV)	280	280	274	268	261	254	246

Finally, Fig. 2.9 depicts a comparison between the SNM and RNM measured from rotated curves proposed by [16]. SNM values are about two times larger than RNM. Therefore, a design must be tolerant to environmental disturbances to achieve reasonable RNM values, which will hence result in acceptable SNM values accordingly.

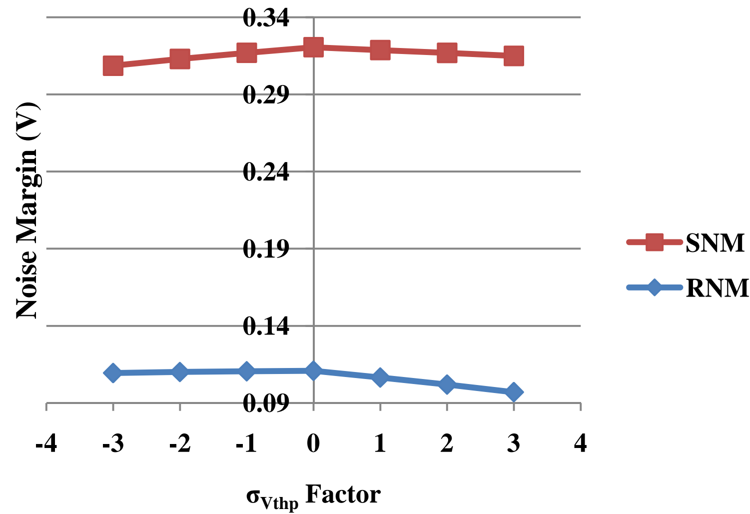


Fig. 2.9 SNM and RNM comparison and the effect of PMOS V_{th} mismatch.

2.4 Effect of the cell differential voltage on SNM and RNM of a conventional 6T SRAM cell

The voltage differential across the memory cell has a significant importance in the determination of its noise margins. Smaller differentials will result in the inverter transistors turning on weaker and therefore becoming less effective to in tolerating environmental disturbances. Fig. 2.10 and Fig. 2.11 depict simulations results for various cell differential voltages and the effects on SNM and RNM respectively. These results show that higher cell voltage will increase the noise margin by making the VTC curves capable of fitting larger squares. The difference function shown at the bottom of these two figures shows how the peak difference is shifted and lowered as the cell voltage is changed. It is important to note that since the structure of the conventional 6T SRAM is symmetrical in terms of the transistor's sizes and their threshold voltages, the butterfly diagrams in both standby and read modes are also symmetrical across the meta-stable point (see Fig. 2.10 and Fig. 2.11).

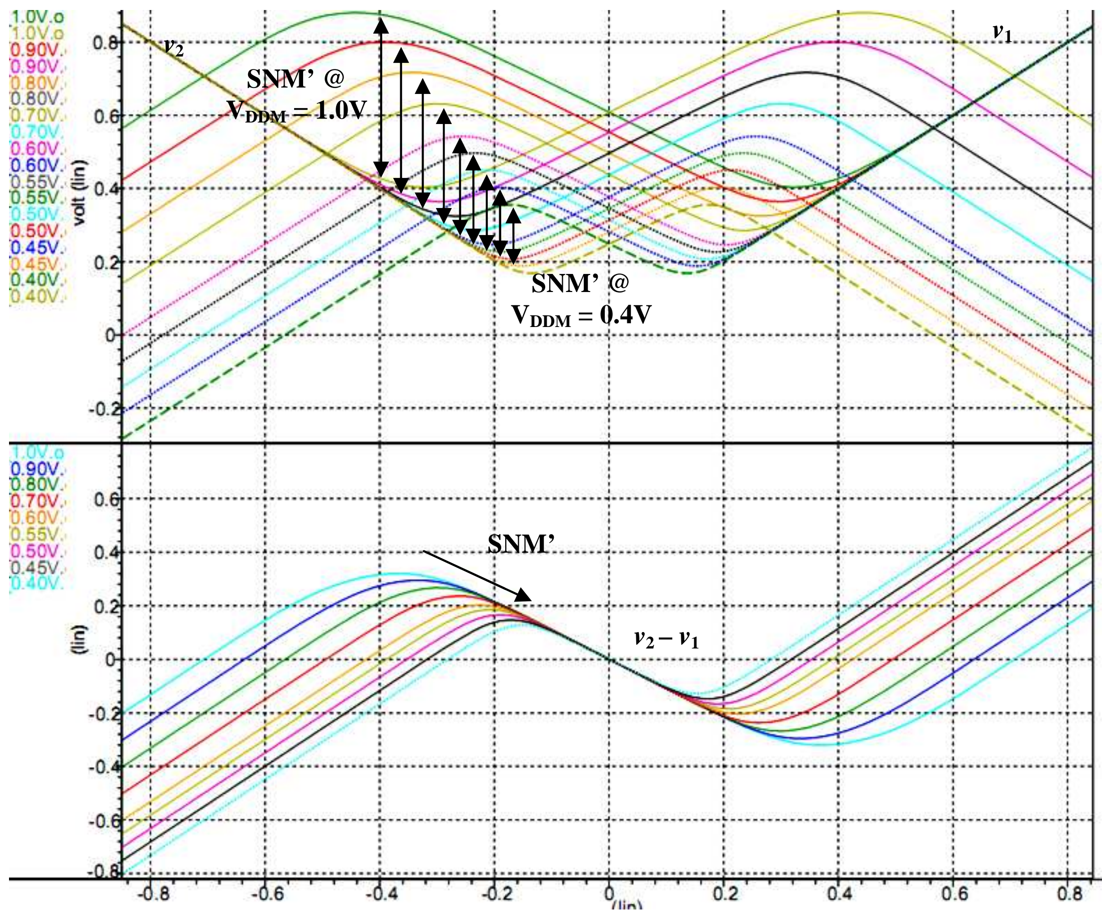


Fig. 2.10 Family of butterfly diagrams (top) and the VTC difference function curves of a conventional 6T cell illustrating the effect of cell voltage variations on SNM.

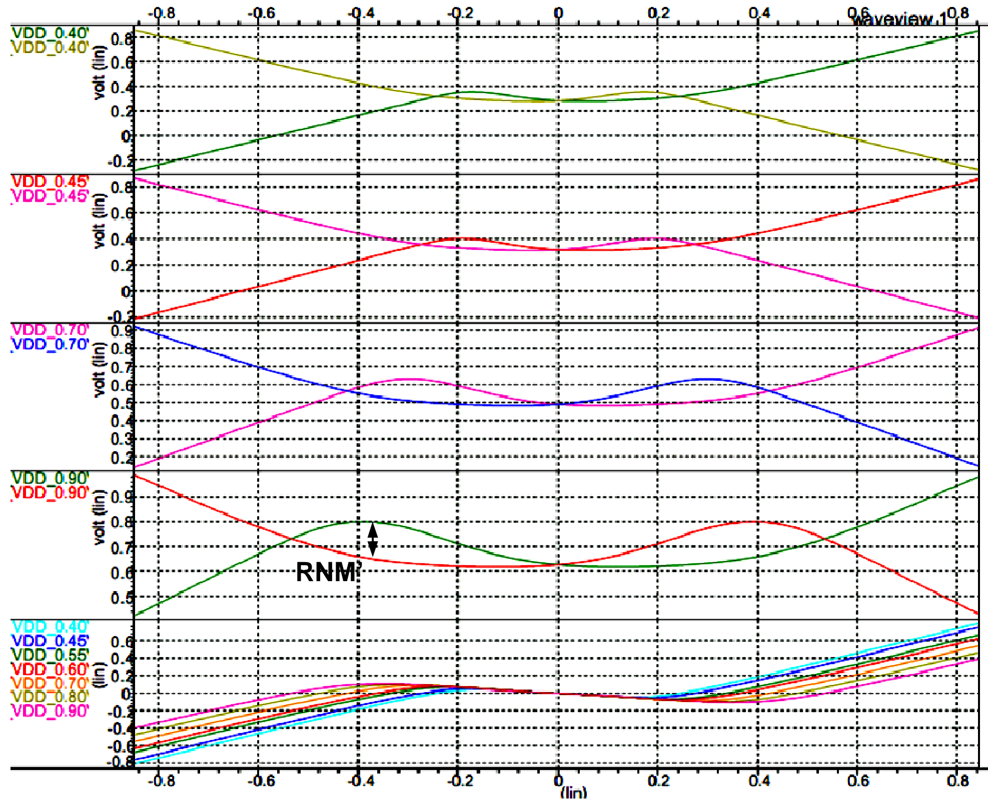


Fig. 2.11 Family of butterfly diagrams and the VTC difference function curves (Bottom) of a conventional 6T cell illustrating the effect of cell voltage variations on RNM.

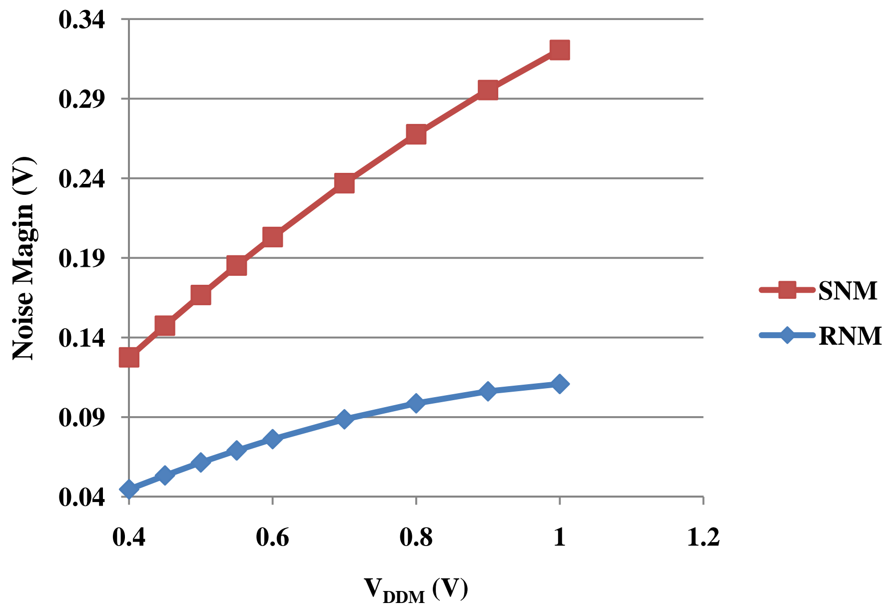


Fig. 2.12 Effect of cell voltage on SNM and RNM in a conventional 6T cell (V_{SSM}=0V).

2.5 Write operation

In write operation, whether writing a ‘0’ (W0) or writing a ‘1’ (W1), the bit-lines are pre-charged high similar to read. The write circuit drives the global write signals, and is connected to the bit-line through transmission gates (activated by TG_ON) or pass transistors (see Fig. 2.13). Therefore, in a write operation BL or BLZ is pulled down depending on whether data is a ‘0’ or a ‘1’ respectively. In case of W1, assuming that the cell is storing a ‘0’ (Q=0, Qz=1), when the word-line is raised, since BLZ is driven low by the write driver, Qz is pulled down and hence P_2 starts to turn on. When P_2 is turned on, it pulls Q high to V_{DDM} causing N_1 to turn on and pull Qz down to V_{SSM}. The performance of the write operation, also known as write-ability, is highly dependent on the cell ratio β. In this example, P_1 opposes the attempt to pull QZ down since it is turned on. On the other hand one can not write a ‘1’ by pulling BL high due to the read stability requirement as explained in the read section. Therefore, the load transistors (P_1 and P_2) must be weaker than the access transistors (N_3 and N_4) and are usually designed to have minimum size.

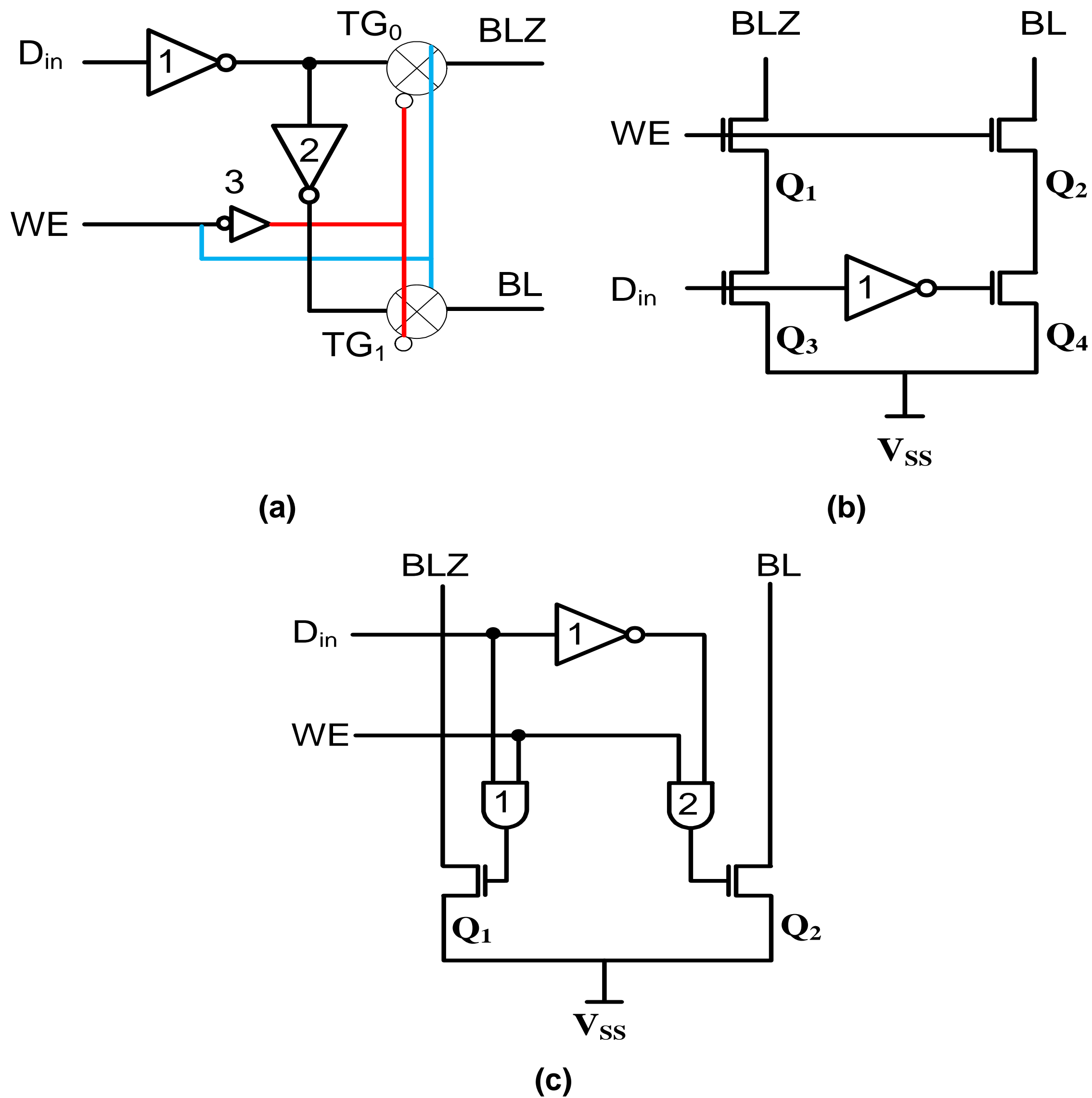


Fig. 2.13 Simplified write circuits for 6T SRAMs (Bit-line pre-charge pull-up devices not shown).

Three typical types of write drivers for a 6T SRAM column are shown in Fig. 2.13 (see also [3]). The purpose of a write driver is to discharge one of the bit-lines (BL or BLZ) depending on the data status when the write enable is active (WE). The circuit in (a) writes into a cell by pulling down BL or BLZ driven by inverters 1 and 2 and via transmission gates TG_0 and TG_1 depending on data value (D_{in}). Inverter 3 is used to drive WE and its complement to activate the

transmission gates. Write circuit in (b) uses Q_1-Q_3 and Q_2-Q_4 NMOS devices to discharge BL or BLZ. Inverter 1 is used to drive the complement value of data to drive BL's gate. If D_{in} is a '0', inverter 1 will turn Q_4 on and once WE is active, it will pull BL to V_{SS} resulting a write '0' operation. On the other hand, if D_{in} is a '1', it will turn on Q_3 and once WE is active, BLZ is pulled down via Q_1 and Q_3 resulting a '0' to be written onto QZ of the cell and activating write recovery process to store a '1' into the cell. In (c) the two AND gates 1 and 2 will prepare gate voltage for two pass transistors Q_1 and Q_2 depending on D_{in} and WE status. Inverter 1 drives AND gate 2 such that if D_{in} is '0', and WE is active, BL is driven low. The strength of the driver circuit is dependent on the capacitance of the bit-lines and power vs. performance requirements. Fig. 2.14 shows the write operation in a conventional 6T cell where, TG_ON is the signal that turns on the transmission gate in Fig. 2.13 (a), Gwr and Gwrz are the global write signals connected to the bit-lines via the transmission gates, and WE is the write enabler.

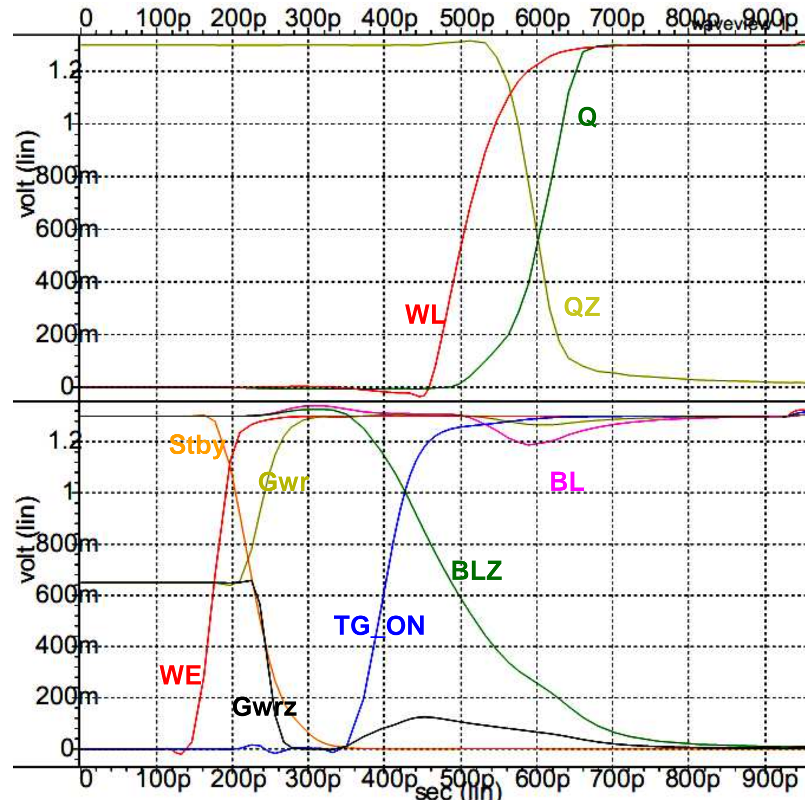


Fig. 2.14 Write operation in a conventional 6T SRAM cell.

2.6 Threshold voltage variation of the PMOS device on write delay

The effect of threshold voltage variation on PMOS devices is interesting since they determine the write recovery performance and are weakest devices in a 6T SRAM cell. In addition, the current drive in PMOS devices is lower than that of NMOS due to having lower carrier mobility in these devices. In an experiment asserting threshold voltage mismatch between PMOS devices in a conventional 6T cell, the write recovery delay was measured from WL=50% V_{DDM} voltage to QZ =80% V_{DDM} (see Fig. 2.15). Simulation results show that higher values in the standard deviation factor for the threshold voltage in the PMOS device (σ_{Vthp}), will make it harder for the PMOS (in this case P_2 raising QZ) to turn on during write

recovery process and hence adding to the delay. This effect is exponentially more severe in lower V_{DDM} voltages as shown in Fig. 2.15 and Fig. 2.16.

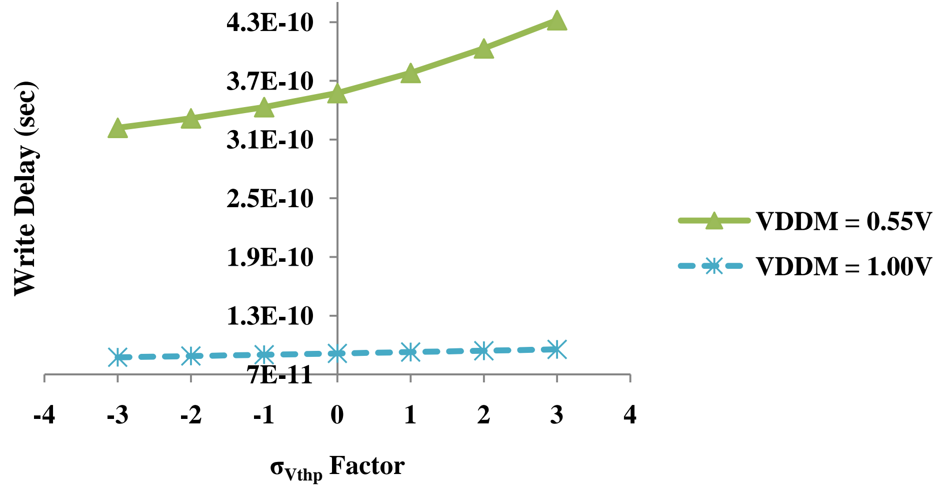


Fig. 2.15 PMOS V_{th} mismatch effect on write delay in the conventional 6T cell.

2.7 Effect of V_{DDM} variations on write and write recovery delay

In some low-power applications, it is required to reduce cell differential voltage for power saving purposes. This reduction of voltage has an inverse relationship with write delay shown in Fig. 2.16.

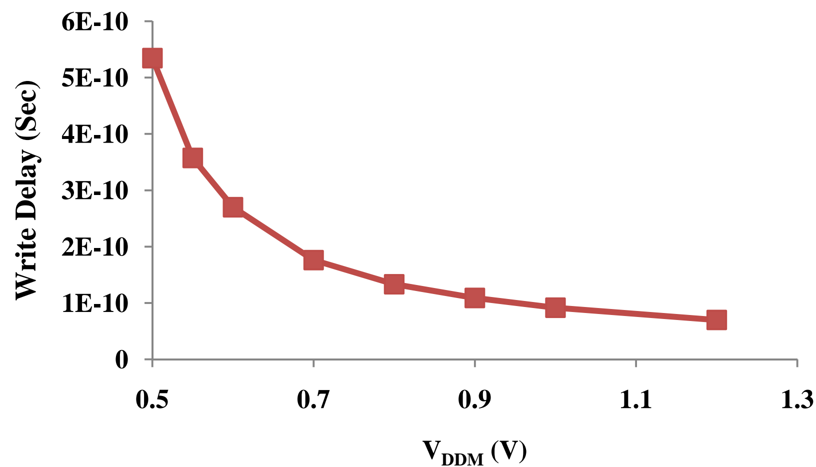


Fig. 2.16 Effect of V_{DDM} variation on write delay in the conventional 6T cell.

2.8 Effect of cell voltage, cell ratio, and threshold voltage on SNM

In an experiment by [10], SNM of a conventional 6T SRAM was measured in 0.13µm technology as a function of $V_{DD,}$ cell ratio β, and threshold voltage V_{th}. The cell was simulated up to 2.5 V V_{DD}. The authors report that for low V_{DD}, the SNM increases linearly with V_{DD} increase , since the pull-up and pull-down devices of the inverter pairs operate more strongly due to higher gate-source voltage, V_{GS}, and drain-source voltage V_{DS}, resulting in higher trip point thresholds, and hence securing the cell against perturbations. However, at higher V_{DD} the noise margin levels off and may even decrease when V_{DD} is approximately twice the threshold voltage. In addition, SNM is reduced at lower threshold voltages since it would be easier to turn on transistors by noise sources at lower V_{th} , and hence the stability of the cell is affected. On the other hand, higher cell ratio values (β) will result in more stability of the cell with the cost of area.

2.9 Differential biasing scheme to mitigate the effect of V_{th} variations

As discussed earlier, due to process variations in threshold voltage of modern transistors with scaled down sizes, SRAM circuits are affected in terms of performance and stability. In a conventional 6T cell, P_1 and P_2 transistors are usually selected to have the smallest sizes for read stability and write ability requirements. Threshold variations in pull up or pull down transistors can affect the write ability and reduce the performance or even causes the write operation

to fail. As an example, when writing a ‘1’ (W1) into the cell (Q=0 -> Q=1), the complementary bit-line (BLZ) is pulled down and the true bit-line (BL) is driven high through the write circuit prior to the rise of the word-lines. As a result of Qz being driven low by BLZ, P_2 is turned on and pulls up Q. The process of Qz being pulled up is often called *write recovery* meaning that the feedback loop recovers the opposite side of the node being pulled down to the opposite value. Therefore, write recovery process can become slower and may fail with the presence of threshold voltage variations especially in the P_1 and P_2. Differential biasing scheme proposed by [5] helps improve the write margin by introducing a differential voltage in which the pull up side of the cell (being turned on in write recovery) will have a higher source voltage than the other side. The differential biasing voltage scheme for the cell voltage was simulated in a conventional 6T SRAM for a write operation. As shown in Fig. 2.17, Q and Qz are true and complementary cell node voltages in a conventional 6T cell, and WL is the world-line. The results show improvements in write performance especially in the write recovery delay in SS corner. For some design cases, [5] reports to achieve 70% improvement in write performance by using this method.

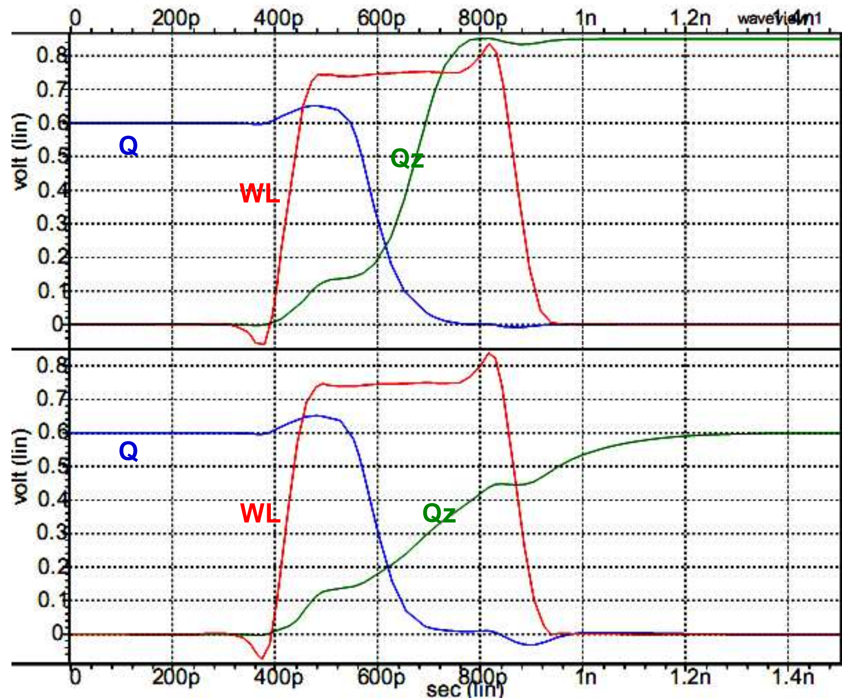


Fig. 2.17 Comparisons of write recovery period for single vs. differential V_{DDM} methods at V_{DD}=0.75V, V_{DDM1}=0.60, V_{DDM2}=0.85 (SS corner, 120℃).

A comparison between the current drive of the load transistors (P_1, P_2) in conventional single-V_{DDM} 6T cell and the differential biasing scheme is reported by [5]. The authors show that the differential biasing scheme increases the current drive of the load transistors by 3.4 times when V_{DD} = 1V and 13.5 times when operating at 0.7V. This comparison is another experiment that shows how the differential biasing scheme can help write-ability and write performance especially at lower cell voltages.

3: PROPOSED 5T SRAM (5TSDG) DESIGN

In this chapter, the proposed *five-transistor* (5T) SRAM architecture is presented, discussed, and compared with conventional 6T and 5T counterparts. A 5T SRAM cell consists of two cross-coupled inverters (P_1-N_1, P_2-N_2), an access transistor (N_3), and a single bit-line (BL) as shown in Fig. 3.1. These types of SRAM cells are attractive since by removing one access transistor and a bit-line, the total chip area per bit can be significantly reduced [15]. However, since read and write operations are performed using a single bit-line, unlike 6T cells, new methods need to be considered to enable proper operation of these types of SRAM structures.

Conventional 5T SRAM cells proposed by [14] and [15] will either require modifying transistor ratios, which will cause major reduction in the stability of cell, or do not take advantage of power reduction opportunities in SRAM design. In this chapter, a novel design is proposed which will have four major prominent features:

- All five transistors can, but not necessarily must, have equal sizes and still maintain required stability margins and performance comparable with conventional 6T design with minor decrease in speed in write ‘1’ operation. The ratios can be adjusted to optimize write ‘1’ performance comparable to 6T cell with minor power and area costs.

- A novel pre-charge method for the bit-lines is introduced which will reduce bit-line leakage in standby, simplify design structure, and will significantly improve read/write performance compared to previous research in 5T SRAM design.
- Significant power saving is achieved by using high threshold (HVT) transistors in the cross-coupled inverter pairs, and the novel pre-charge technique.
- A significant area saving compared to a conventional 6T cell is achieved.

The schematic of the 5TSDG cell including the sub-column circuitry is depicted in Fig. 3.1. Standby and Ground control circuits are required one per every sub-column while V_{SSM} control is shared in the entire memory array. TABLE 3.1 specifies some of the design parameters of 5TSDG, low-power 6T, as well as conventional 6T cells used in this thesis for comparison. An area reduction of ~13% is predicted compared to a conventional 6T cell using standard 65nm design rules [1].

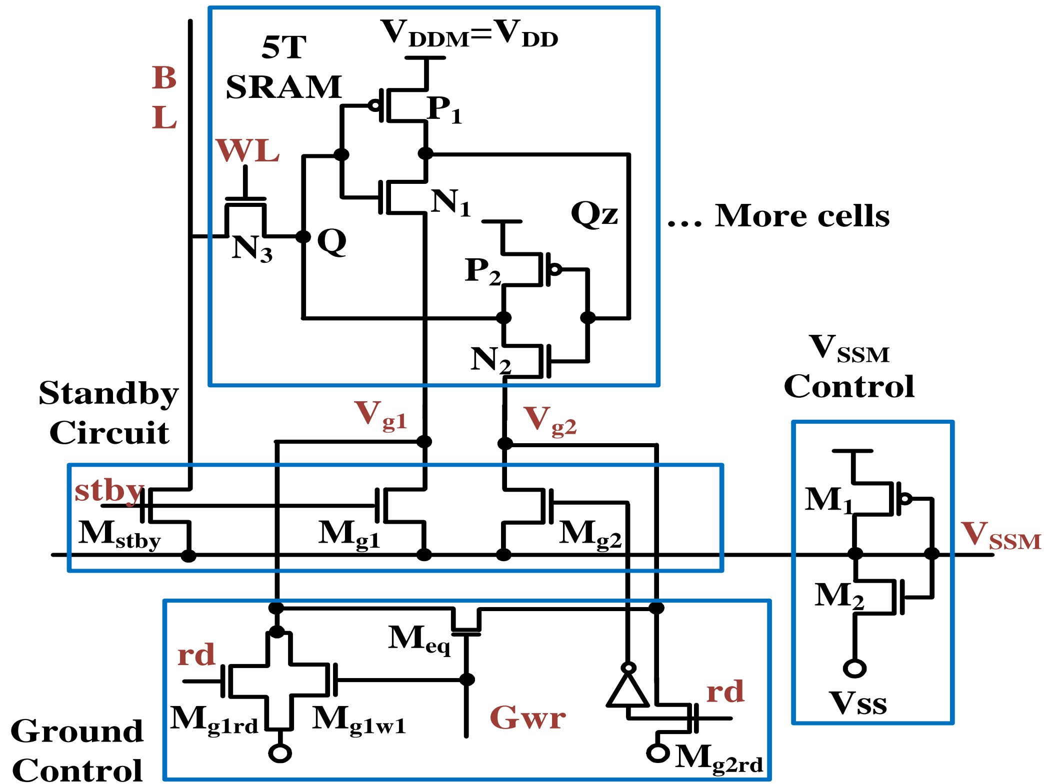


Fig. 3.1 5TSDG structure with sub-column V_{SS} control.

TABLE 3.1. Different types of SRAM cells used in this chapter, V_{DD}=V_{DDM}=1.3V, β = $(WN_2/LN_2)/(WN_3/LN_3)$.

Type	*Inverter Transistors*	*Access Transistors*	*β*	*V_{SSM}*	*BL pre-charge*
5TSDG	HVT	SVT	1.0	600mV	600mV
Low-Power 6T	HVT	SVT	1.4	600mV	V_{DD}
Conventional 6T	HVT	SVT	1.4	0mV	V_{DD}

3.1 Standby mode

One of the effective and proven methods to suppress leakage power during standby in 6T SRAMs is to use dynamic sleep design while maintaining a sufficient *Static Noise Margin* (SNM) which ultimately determines the integrity of the stored data [9][12]. The most effective way to use this method is by raising

the negative supply voltage of the memory cells, V_{SSM}, as opposed to lowering the positive one, V_{DDM}, to minimize bit-line and cell leakage power [1][9][17].

Considering this method in 5T SRAM, a prominent feature of 5TSDG is that instead of using an external on-chip power supply to raise V_{SSM} voltage above ground in standby, with existence of enough leakage sources especially sub-threshold and gate leakage currents in advanced technologies, the leaking memory array can be used as a power source to collect these charges from Vg_1 and Vg_2 via M_{g1} and M_{g2} causing a natural rise of V_{SSM} to a desired biasing level using V_{SSM} control circuit. After evaluating performance, stability and power consumptions by simulations, with various combinations of threshold voltages (V_{th}) for each single transistor in 5TSDG, it is found that V_{th} of the inverter pairs have the most significant impact on the leakage power while the access transistor, N_3 has the most significant impact on performance and stability as also discussed in section 3.2, and 3.3. Therefore, the two inverter pairs N_1-P_1 and N_2-P_2 are selected to have high threshold voltages (HVT) while the access transistor N_3 has a smaller V_{th}, (in this case Standard V_{th}, SVT). All cell transistors are selected to have equal sizes ($W_{i=}$0.15μm, L_i=0.06μm).

Using two carefully sized diode-connected transistors, M_1 and M_2, the voltage across the cell in standby can be biased to remain static for various temperatures and process corners (see also [19]). In this design, a minimum voltage across the cell, $V_{min} = V_{DDM}$-V_{SSM}, of 0.7V is selected to yield sufficient stability [13], resulting in a simulated SNM between 181-222mV in all corners and temperatures at V_{DDM}=1.3V [16]. A 64Kbit memory array arranged in 64x16

blocks was simulated in standby mode using BSIM v.4 and HSPICE at V_{DD}=V_{DDM}=1.3V. The large capacitance of V_{SSM} consisting of mostly junction and wire capacitors and sufficient available leakage current are the key factors in stability of V_{SSM} during standby/write/read modes. In case of lack of leakage especially due to HVT transistors, in some corners or temperatures, M_1 is turned on more strongly to provide the charges to V_{SSM}. During read and write operations, V_{SSM} remains close to the standby steady state value.

Another unique feature of 5TSDG that makes it different from previous research work is that V_{SSM} can also be used to pre-charge the bit-line, BL, in standby via M_{stby} as shown in Fig. 3.1 so that 1) channel and gate leakage through N_3 is reduced and minimized by up to 90% especially when a '0' is stored, 2) the cell maintains a reasonable *Read Noise Margin* (RNM) when accessed close to the optimum achievable point and 3) to accelerate read/write operation explained in the next sections.

The static noise margin of the 5TSDG cell was measured using similar methodology explained for conventional 6T cell. As shown in Fig. 3.2, SNM is defined to be the maximum value of the adverse polarity DC noise sources v_n added to the inputs of the cross-coupled inverters P_1-N_1, P_2-N_2 before initiating the state toggle (Stby signal is at V_{DD}).

Fig. 3.2 A 5T SRAM cell with two adverse polarity noise sources v_n.

As shown in Fig. 3.3, the relationship between Vmin and SNM is linear in all corners. In the proposed application-dependent design, the user must consider the fact that Vmin determines the biasing level of both the negative supply voltage, and the pre-charge voltage for the bit-lines (At a fixed V_{DDM}). Therefore, Vmin should be selected to either be around a point to maximize RNM, or not be any lower than the required RNM for all process corners. Fig. 3.4 shows simulation results indicating the relationship between positive supply voltages, V_{DDM}, at a constant Vmin=0.7V, and the read noise margin. In this experiment, the bit-line pre-charge DC voltage, Vbm, was linearly swept to find the maximum RNM while keeping V_{SSM} tied to a constant DC source. Inherently, RNM peaks at a level, and the best RNM occurs at lower bit-line pre-charge voltages as V_{DDM} is lowered while keeping the same Vmin. Since the maximum

allowable voltage in this technology node is 1.3V, a V_{DDM} of 1.3V and a V_{min} of 0.7V are chosen in this case for explanatory purposes and as reference point. The fast NMOS and slow PMOS, FS corner, is known to be the worst case RNM discussed further in the next section. Therefore, the design of 5TSDG cell has been focused to achieve reasonable RNM at the FS corner which will guarantee higher RNM results for other corners.

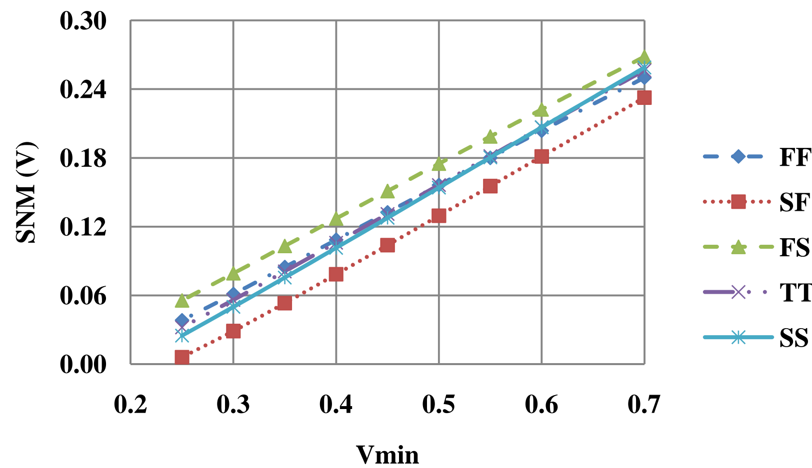


Fig. 3.3 Relationship between Vmin=V_{DDM}-V_{SSM} and SNM of 5TSDG in various corners.

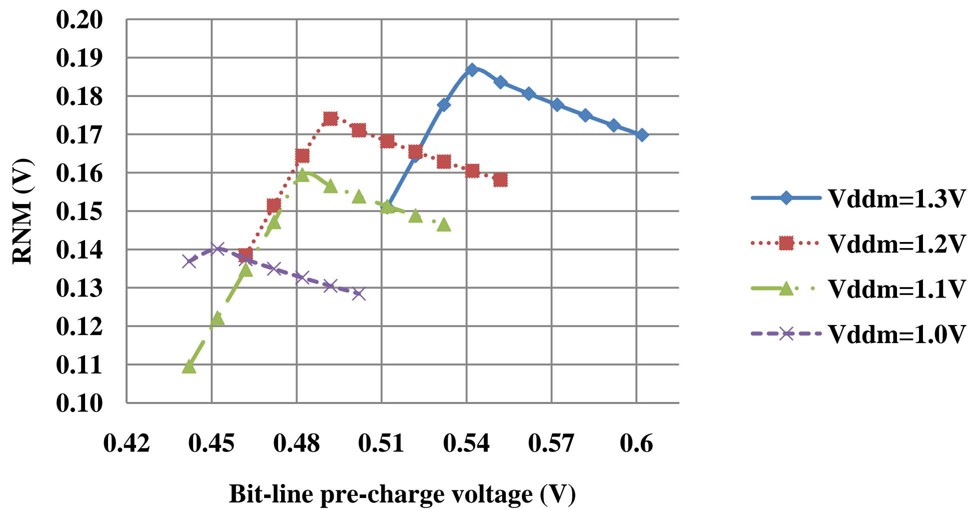


Fig. 3.4 RNM variations vs. bit-line biasing (=V_{SSM}) - Supply voltage (Worst Case, FS Corner).

As explained in Chapter 1, the sub-threshold leakage dominates other sources of leakage power in CMOS devices. In a 5T SRAM cell, the two cross-coupled inverters storing either a ‘0’ or a ‘1’ consist of four transistors in which two of them will be OFF in standby. In a case when a ‘0’ is stored in the cell (Q=’0’), as shown in Fig. 3.5, the status of each single transistor as well as the leakage sources in the cell are described in TABLE 3.2. Similarly, TABLE 3.3 describes the biasing levels and leakage sources of a cell storing a ‘1’ (Q=’1’).

Fig. 3.5 5TSDG cell: Storing a ‘0’ (Left) and a ‘1’ (Right).

TABLE 3.2. Transistor biasing voltages in standby for a cell storing a ‘0’ (Q=’0’).

Tr.	V_{GS}	V_{DS}	V_{SB}	V_{DG}	V_{DB}	**Current Sources**
P_1	V_{SSM}-V_{DDM}	0	0	V_{DDM}-V_{SSM}	0	$I_D + I_g$
N_1	0	V_{DDM}-V_{SSM}	V_{SSM}	V_{DDM}-V_{SSM}	V_{DDM}	$I_{sth} + I_j + I_{GIDL}$
P_2	0	V_{SSM}-V_{DDM}	0	V_{SSM}-V_{DDM}	V_{SSM}-V_{DDM}	$I_{sth} + I_{GIDL}$
N_2	V_{DDM}-V_{SSM}	0	V_{SSM}	V_{SSM}-V_{DDM}	V_{SSM}	$I_D + I_j + I_{GIDL} + I_g$
N_3	0	0	V_{SSM}	V_{SSM}	V_{SSM}	$I_j + I_{GIDL}$

TABLE 3.3 Transistor biasing voltages in standby for a cell storing a ‘1’ (Q=’1’).

Tr.	**V_{GS}**	**V_{DS}**	**V_{SB}**	**V_{DG}**	**V_{DB}**	**Current Sources**
P_1	0	V_{SSM}-V_{DDM}	0	V_{SSM}-V_{DDM}	V_{SSM}-V_{DDM}	I_{sth} + I_{GIDL}
N_1	V_{DDM}-V_{SSM}	0	V_{SSM}	V_{SSM}-V_{DDM}	V_{SSM}	I_D + I_j + I_{GIDL} + I_g
P_2	V_{SSM}-V_{DDM}	0	0	V_{DDM}-V_{SSM}	0	I_D + I_g
N_2	0	V_{DDM}-V_{SSM}	V_{SSM}	V_{DDM}-V_{SSM}	V_{DDM}	I_{sth} + I_j+ I_{GIDL}
N_3	0-V_{SSM}	V_{DDM}-V_{SSM}	V_{SSM}	V_{SSM}	V_{SSM}	I_{sth} + I_j + I_{GIDL}

In the design kit used in this thesis, three threshold voltages were accessible to use in a design, mainly for standard digital design purposes. The three types of V_{th} are standard (SVT), high (HVT), and low (LVT). In an experimental analysis using various combinations of these standard types, standby power was measured and the results are shown in Fig. 3.6 (All transistors are HVT unless otherwise specified). In a 5TSDG cell, as opposed to conventional 6T cell, the threshold voltage of the access transistor does not play a significant role in standby power. As shown in Fig. 3.5, the drain to source voltage of the access transistor, V_{DS}, is either V_{SSM}-V_{SSM}=0 or V_{DDM}-V_{SSM}=Vmin, depending on the cell storing a ‘0’ or a ‘1’ respectively, and therefore, sub-threshold leakage current is minimized (see equation 1.2) and in case of a ‘0’ stored in the cell, the leakage is completely eliminated. This feature in 5TSDG has an advantage over conventional 6T cell design since V_{DS} of the access transistors in a 6T cell is either V_{pc}-V_{DDM} or V_{pc}- V_{SSM} ,where V_{pc} is the pre-charge voltage of the bit-lines in a conventional 6T cell, which is higher than that of 5TSDG. Therefore, bit-line leakage in 5TSDG, unlike that in a conventional 6T

cell, is dependent on the percentage of ‘0’s and ‘1’s stored in the memory array. Comparing bit-line leakage power between 5TSDG and conventional 6T cell, it is demonstrated that the leakage is reduced by ~90%. Fig. 3.6 also shows that the threshold voltage of P_1 and P_2 (load) transistors have less effect on standby power consumption than N_1 and N_2. The reason lies within the fact that PMOS devices have lower mobility due to their physical nature than the NMOS devices. Therefore, according to equation 1.2, the sub-threshold current of an NMOS, which is the dominant leakage current source in standby, is more than that of a PMOS. Consequently, after experimenting with various combinations of V_{th} for each transistor in a 5T cell, it is found that the cross-coupled inverter’s transistors have the most significant effect on standby power while the access transistor N_3, in 5TSDG design, will cause the least amount of leakage when compared to others.

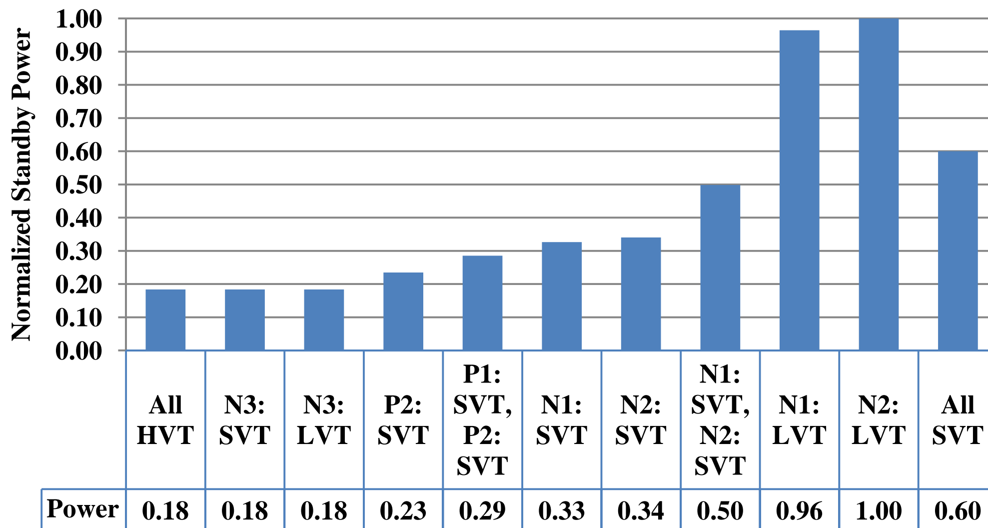


Fig. 3.6 Normalized standby power consumption of various V_{th} combinations in a 5T cell.

Due to various sources of leakage in standby shown in TABLE 3.2 and TABLE 3.3, mostly consisting of channel, sub-threshold and gate leakages (I_{ch},

I_{sth}, I_g), the negative supply voltage can be raised by storing these charges on V_{SSM}. Since V_{SSM} is a wire connected to the sources of driver transistors (N_1, N_2) as well as the standby transistor connected to the bit-lines, its large capacitance is capable of holding these charges.

The charges due to the leaking transistors in standby, when stored on a capacitive wire, V_{SSM}, cause V_{SSM} to raise above V_{SS}. This rise must be regulated to stabilize V_{SSM} during different operations of the memory. The circuit that controls this voltage is called V_{SSM} control circuit in this thesis. With adequately strong voltage regulation gained by diode-connected pull up and pull down transistors M_1 and M_2 (see Fig. 3.1 and Fig. 3.7), the biasing level of V_{SSM} can be adjusted depending on the desirable voltage across the cell voltage during standby ($Vmin=V_{DDM}-V_{SSM}$). Higher SRAM array sizes will result in more stability of V_{SSM} and less extra power consumption caused by adding M_1 pull-up. The amount of extra power consumption caused by M_1 varies from ~5-35% depending on the memory array size, temperature, and process corner. The 35% increase of power is for very low temperatures (-40℃) and SS corner, where the leakage power is already very low (see Fig. 3.9). The desirable voltage across the cell is dependent on performance, power and reliability requirements. Only one V_{SSM} control circuit unit is needed for the entire memory as opposed to the ground control circuit needed one per sub-column.

At higher temperatures, the leakage in memory cells is increased resulting in the flow of more charges on V_{SSM} in an attempt to raise it. This will cause M_2 to turn on more strongly since both of gate-to-source voltage, V_{GS}, and drain-to-

source voltage, V_{DS}, are increased, and thereby cancelling this effect. As soon as the leakage current is reduced in lower temperatures, V_{SSM} is decreased resulting in higher V_{GS} and V_{DS} for M_1, and causing it to turn on more effectively, and pulling V_{SSM} high. The total effect of M_1 and M_2, therefore, needs to be balanced to keep V_{SSM} stable at all temperatures. Fig. 3.7 depicts the simulation of V_{SSM} biasing level gained by leakage storage vs. different temperatures for different corners with and without using a pull-up diode connected transistor (M_1). If instead of HVT transistors for the cross-coupled inverters, SVT transistors were to be used, the amount of leakage of the cells would increase resulting in less necessity of requiring a pull-up charge compensator (M_1) to recover missing charges at lower temperatures.

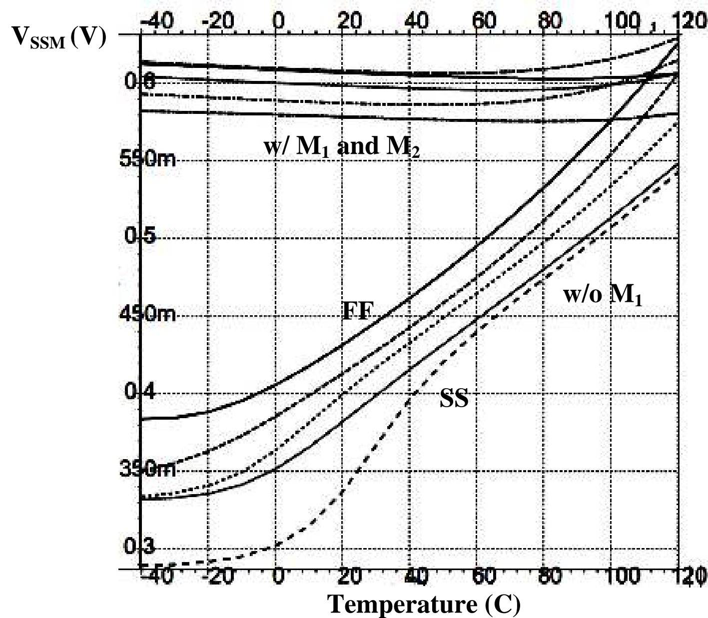


Fig. 3.7 Effect of Temperature on V_{SSM} with and without voltage regulation for different corners.

TABLE 3.4 shows leakage current and worst case RNM for various types of SRAM cells introduced in TABLE 3.1. Traditional 5T designs as in [14][15],

where V_{SSM} is held at V_{SS} level, require lower V_{th} for internal cell transistors in 65nm technology, such as N_1, and P_2 (see [7]), to enable write '1' operation discussed in section 3.3. Thus, even though some leakage power is saved by cutting a bit-line and biasing the other to a lower voltage, the overall leakage is quite high, being about half of the conventional 6T cell value in TABLE 3.4.

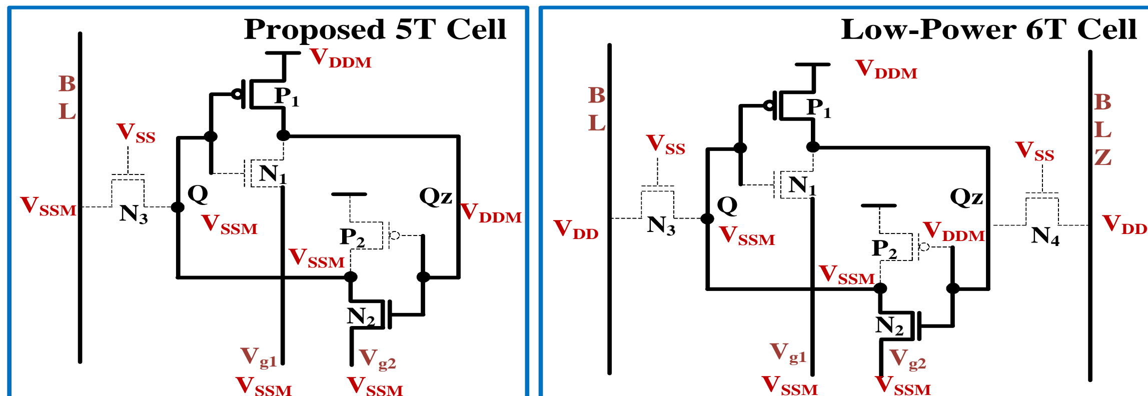


Fig. 3.8 5TSDG vs. a low-power 6T cell comparison when storing a '0'.

TABLE 3.4. Worst-case standby leakage current and RNM comparison in different SRAM types (not including peripheral circuits).

64 cells	*5TSDG*	*Low-Power 6T*	*Conventional 6T*
Leakage (nA)	*80.6*	*88.7*	*2020.0*
RNM (mV)	*172.3*	*123.2*	*123.2*

The statistical distribution of data ('0's and '1's) in a memory array can result in minor differences in power consumption in 5TSDG. As shown in Fig. 3.8, in 5TSDG, when a '0' is stored, the bit-line channel leakage through N_3 is completely eliminated, and when a '1' is stored, since V_{SSM} is raised, a small amount of leakage exists. On the other hand, the power consumption of a conventional 6T cell does not vary with the statistical distribution of data since

whether a ‘0’ or ‘1’ is stored in the cell, total bit-line leakage is constant due to symmetry.

Fig. 3.9 compares the power consumption of 5TSDG including peripheral circuits with a low-power 6T design in various process corners. The peripheral circuits include sense amplifiers, write drivers, bit-line drivers, column control circuitry, and timing control circuitry. FF corner is the most critical case in terms of power consumption as opposed to the SS corner, which consumes the least amount of power and is also slowest in performance.

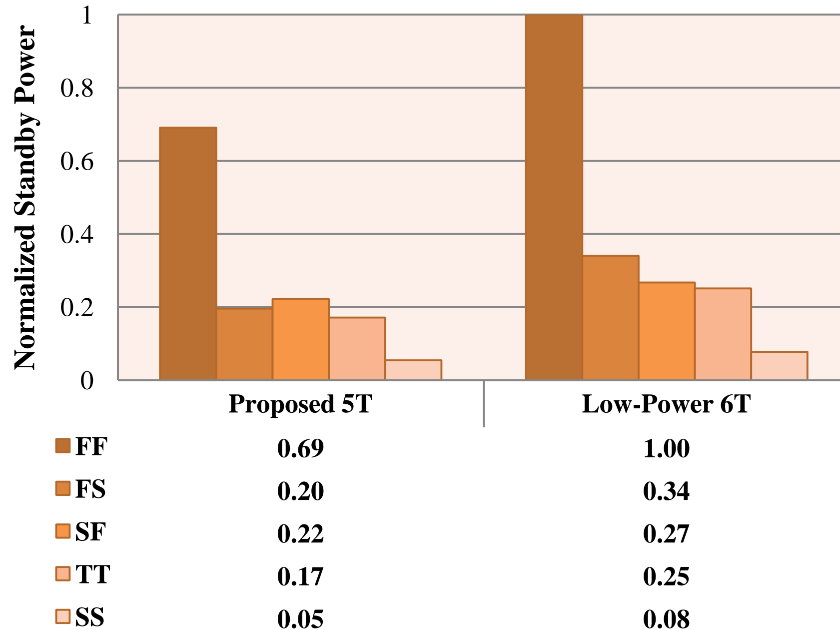


Fig. 3.9 5TSDG vs. low-power 6T SRAM designs: (64Kb SRAM array and peripheral circuits, FF corner on the left, 120℃, 50% ‘0’s 50% ‘1’s stored in the array).

Fig. 3.10 shows SNM determinant VTC curves for 5TSDG with V_{SSM} at 0.6V, and V_{DDM} at 1.3V for various corners, and shows the reliability of 5TSDG in all process corners. Measurements for SNM were based on the method explained in section 1.6.1.

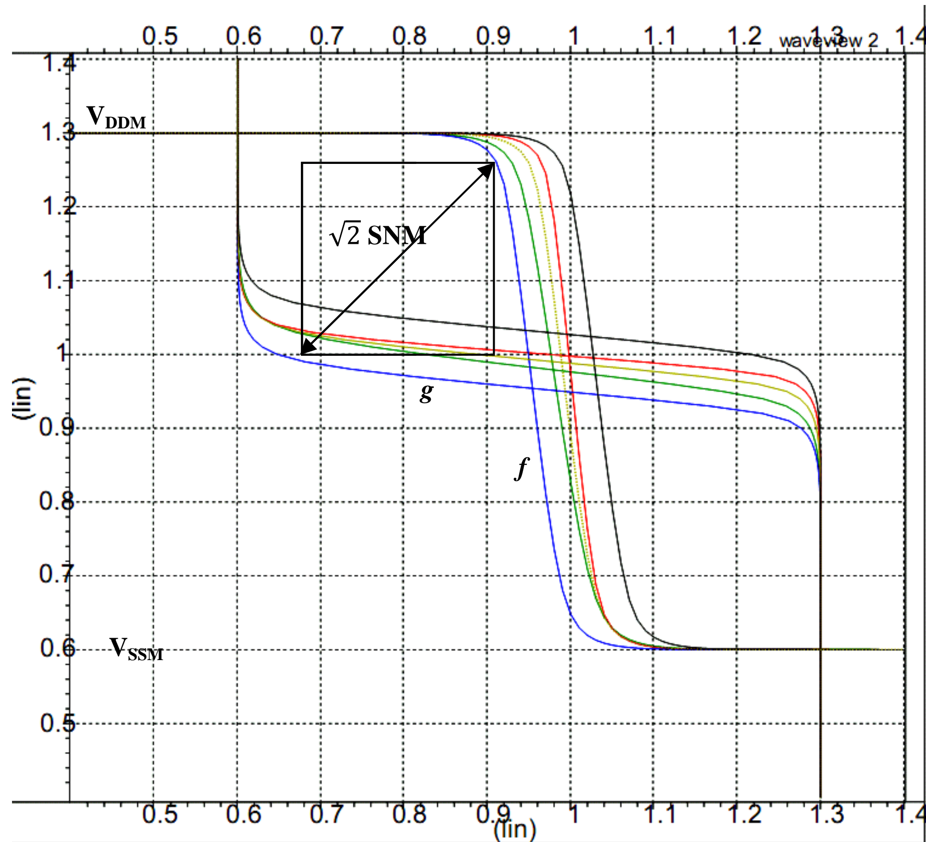


Fig. 3.10. Butterfly diagrams for different corners for 5T SRAM in standby (SF, SS, TT, FF, FS corners from bottom up).

Fig. 3.11 shows an example of the sub-column architecture of 5TSDG including the ground and V_{SSM} control circuitry required for different operations of the memory discussed further in the next sections. In a case where the total column size is 128 bits, the column can be divided into 16, 32 or 64 cell segments (sub-columns). Each sub-column and the adjacent one will require a shared sense amplifier (see Fig. 3.12) for read purposes, which will be discussed in section 3.2. In this thesis, a column size of 128 bits and a sub-column size of 64 bits were selected to result in reasonable performance. Shorter bit-lines (cells/bit-line) will result in an improved performance especially in the read operation, and will save on power consumption due to bit-line voltage swing on a capacitive wire. However, fewer cells per bit-line will result in additional sub-

columns, which will add to the number of sense amplifiers hence the area and power. Therefore, a reasonable balance between number of cells per sub-column and hence, the area, the required performance, and power limit needs to be considered.

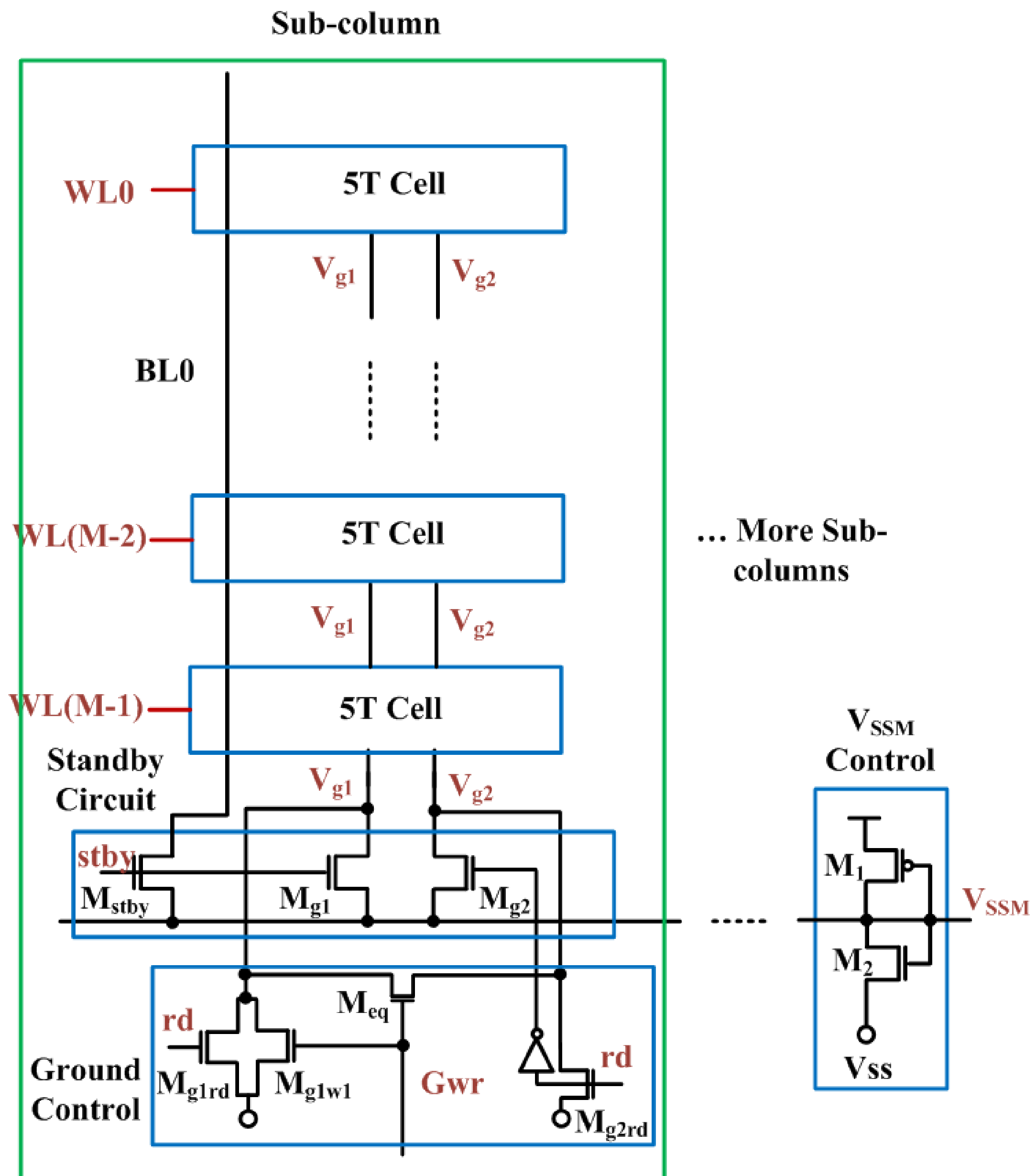


Fig. 3.11 Sub-column structure of 5TSDG.

3.2 Read operation

The read operation is similar to a 6T SRAM cell except that only one bit-line is used. In 5TSDG, the bit-line is pre-charged in standby by V_{SSM}, which is near the optimum point to maximize RNM in the worst case (FS), and cuts down the bit-line leakage in standby by up to ~90%. Another advantage of this pre-charge method compared to [15] is that it does not require an additional power supply on the chip such as a DC-DC converter or a level shifter, which will add to the chip area and power consumption. A simple sense amplifier circuit used in 5TSDG is shown in Fig. 3.12. Although not the fastest type, it is attractive due to its simplicity and that it does not need a clock signal [18]. During read, rd signal in Fig. 3.11 is raised causing V_{g1} and V_{g2} to be pulled down to V_{SS} by M_{g1rd} and M_{g2rd}, which will maximize RNM and read performance. The global bit-line, Gbit, is the output of the sense amplifier and is pre-charged to V_{SSM} through M_8 in standby, and is pulled down to V_{SS} by M_7 during a read '0'. Therefore, a read '1' is always implied unless Gbit is pulled down. Inverter M_5-M_6 should have a sufficiently low sensitivity to prevent a false trigger. This sense amplifier can be shared by two bit-lines from two adjacent sub-columns. For instance, in a 128-cell column composed of two 64-cell sub-columns, the sense amplifier is placed in between bit-lines BitL and BitR. SelL and SelR signals should be selected by a row decoder to select the appropriate bit-line to read from. M_3 and M_4 are used to pre-charge the input of the inverter M_5-M_6 in standby. For similar bit-line capacitances, read speed in 5T and 6T SRAMs is comparable.

Fig. 3.13 shows simulation results for a read operation of 5TSDG in a 64Kbit memory array arranged in 64x16 bit blocks for two neighbouring cells sharing the same word-line, WL, storing a ‘0’ and a ‘1’ on Q0 and Q1 nodes, and having two bit-lines BL0 and BL1 respectively (similar to Fig. 1.4). A dynamic increase in Q0 node occurs while reading a ‘0’ due to the current flow from the bit-line to N_3 and N_2 as the word-line is raised (see Fig. 3.13). This rise in voltage is denoted by Q_{max} Fig. 3.13. During read ‘1’, a drop of voltage in Q1 node is observed for a similar reason (Q_{min}). Q_{max} and Q_{min} should not cause a read upset i.e. they should be less and more than the tripping voltage of the inverter pairs, respectively, to avoid turning a read into a write, especially in FS corner. In order to further reduce the probability of read-upset in 5T cell, it is possible to increase word-line rise time and make the bit-lines shorter to reduce their capacitances [2]. The latter will also improve read speed by reducing bit-line swing delay.

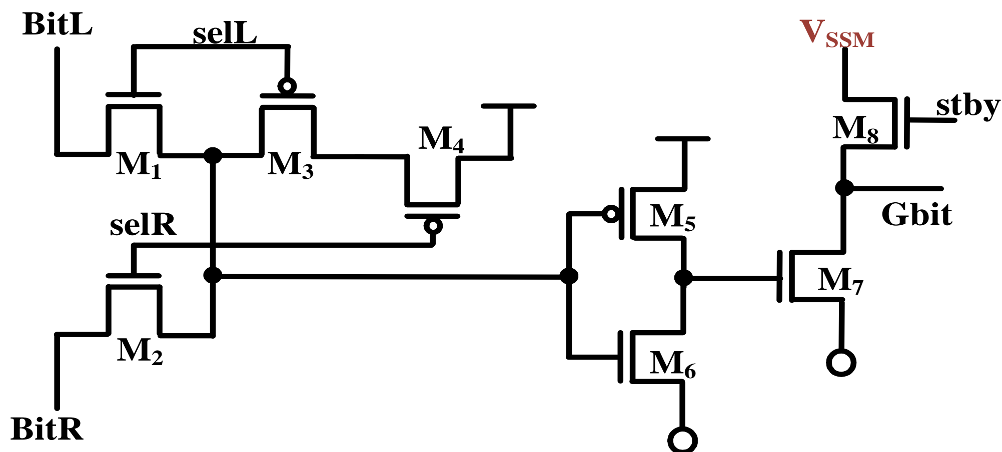


Fig. 3.12 Sense amplifier used in 5TSDG read operation.

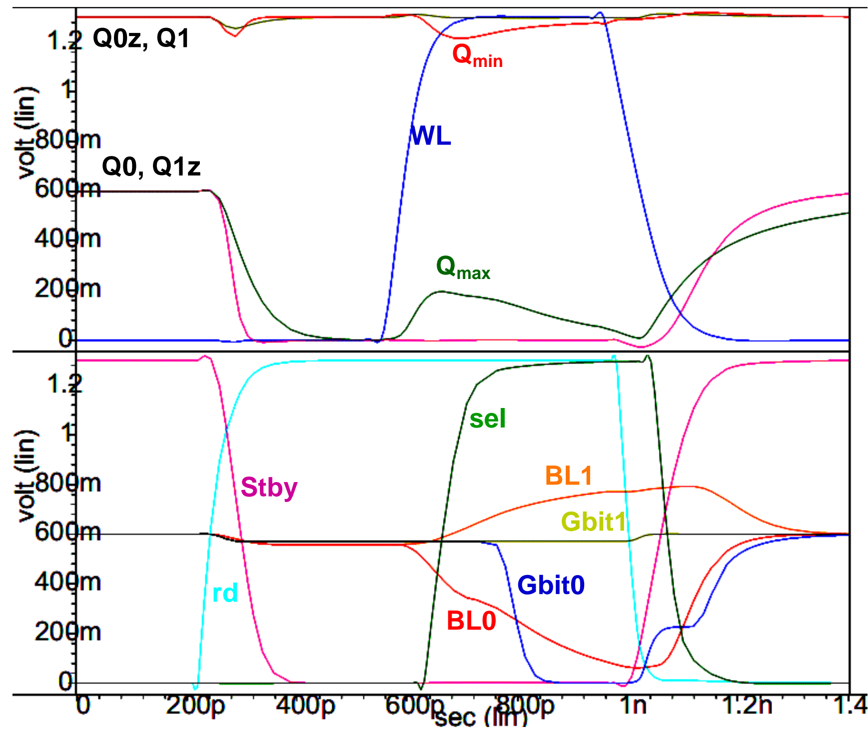


Fig. 3.13 Read operation in a 5TSDG cell, typical corner (TT), 120℃, Q0-Q0z-BL0-Gbit0 and Q1-Q1z-BL1-Gbit1 are related to a cell storing a '0' and '1' respectively and sharing the same word-line (WL).

TABLE 3.5 shows the trip voltage vs. Q_{max} and Q_{min} while reading in different corners in 5TSDG. Trip points were measured by injecting current into the cell node Q, using a variable current source, and measuring when the caused voltage increase trips the cell to the opposite value. These results when compared with RNM measurements for various mismatch cases in worst case corner FS, show the stability of 5TSDG. RNM is measured with V_{SSM} on the bit-line similar to [15].

Fig. 3.15 shows the effect of V_{th} variations on the RNM of 5TSDG. The best biasing value for V_{SSM} maximizes RNM for the worst-case (FS corner). Fig. 3.15 is based on simulation results in the FS corner. A robust and reliable design will only need to have acceptable noise margins up to $\sigma_N = \sigma_P = \pm 3$

(~99.9999998% of statistical possibilities for V_{th}). Therefore, in Fig. 3.15, since in FS corner $\sigma_N = -3$ and $\sigma_P = 3$ by default, it is only required to consider RNM values for $0 < \sigma_N < +6$, and $-6 < \sigma_P < 0$, and the rest of the values can be neglected.

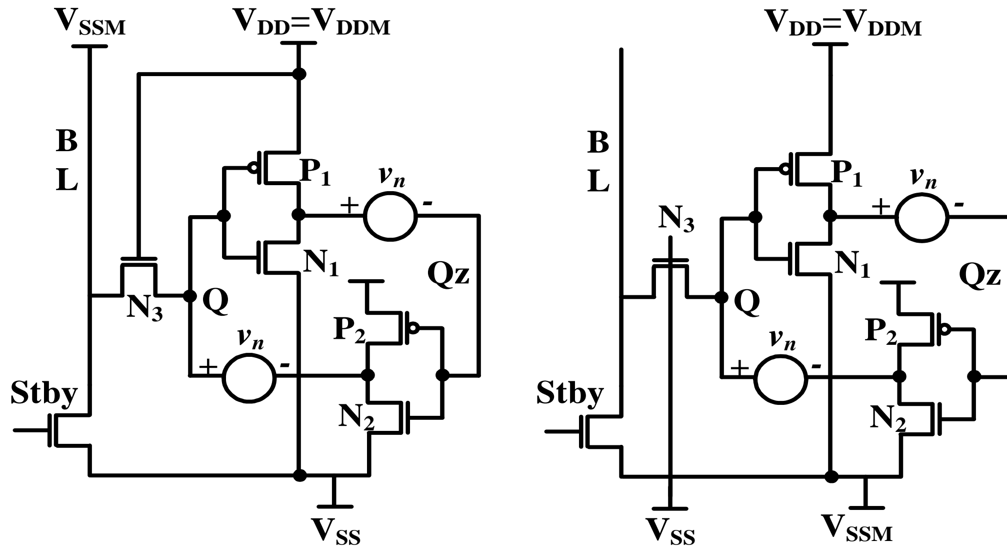


Fig. 3.14 Comparison between 5TSDG when accessed (Left) with the cell in standby with the presence of the noise sources, v_n.

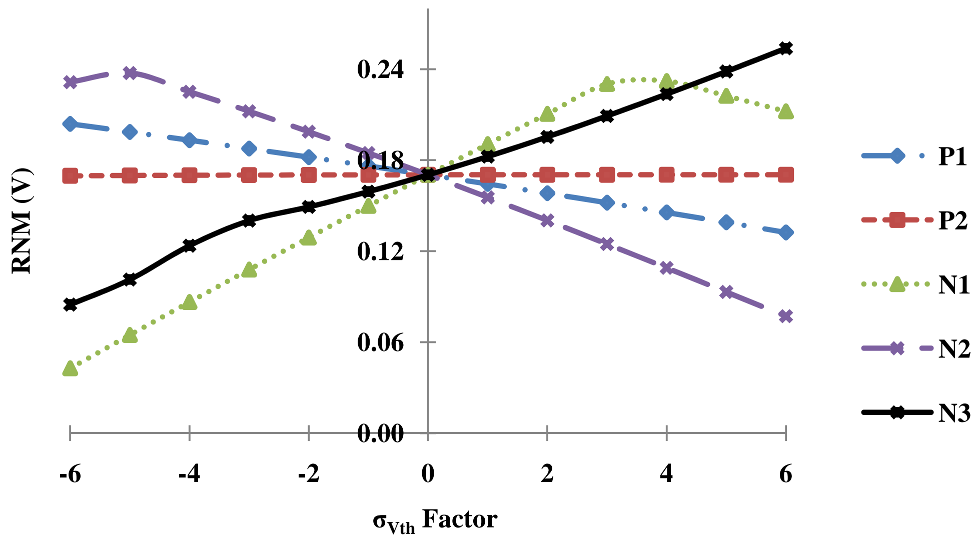


Fig. 3.15 Effect of V_{th} variations on RNM varying each single transistor in a 5T cell (Worst case: Relative to FS corner, 120℃).

TABLE 3.5 5TSDG: Trip point voltage vs. Q_{max} and Q_{min}, 120℃).

5T Cell Trip Voltage (mV) / Q_{max} (mV)				
FF	**FS**	SF	TT	SS
883/193	**866/185**	934/202	899/195	905/196
5T Cell Q_{min} (V)				
1.19	**1.18**	1.24	1.22	1.24

Fig. 3.16 simulation results demonstrate how the read noise margin, the optimum RNM, and optimum bit-line pre-charge voltage is affected by the threshold voltage of the access transistor in 5TSDG. In section 3.1, it was demonstrated that the access transistor does not play a significant role in standby power consumption in 5TSDG as opposed to conventional 6T cell design. Therefore, assuming that the required safe Vmin is 0.7V requiring V_{SSM} to be at 0.6V for a V_{DDM} of 1.3V, in case of an LVT, SVT and HVT type threshold voltage for the access transistor, the best RNM varies from about 140mV, 170mV and 220mV respectively in the worst case (FS corner).

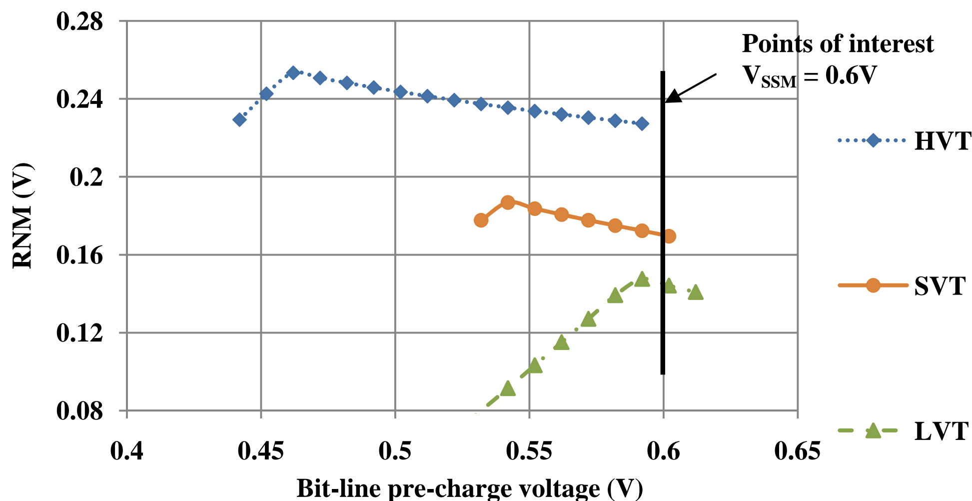


Fig. 3.16. RNM variations vs. bit-line biasing voltage (=V_{SSM}) for three threshold voltages of the access transistor (N_3) in FS corner.

Fig. 3.17 shows RNM variations vs. bit-line biasing voltage for various process corners and illustrates how FS corner has the smallest whereas SF corner results in the highest value for RNM. This is due to the fact that faster NMOS device will cause stronger current drive in the access transistor, and for similar reasons discussed regarding Fig. 3.16, faster access transistor will result in smaller RNM. Decreasing threshold voltage for P_2 will result in a left shift in the best RNM value shown in Fig. 3.18. This effect does not help with the RNM, provided that a fixed V_{SSM} is required, since the RNM values for both P_2 having HVT or SVT would be the same at the particular point of interest. On the other hand, lower threshold for P_2 is interesting since it will improve write '1' (Discussed in the next section), and as discussed in standby section, it has little effect on power consumption. Layout complications, power increase, and no effect on RNM, are reasons to choose HVT for P_2.

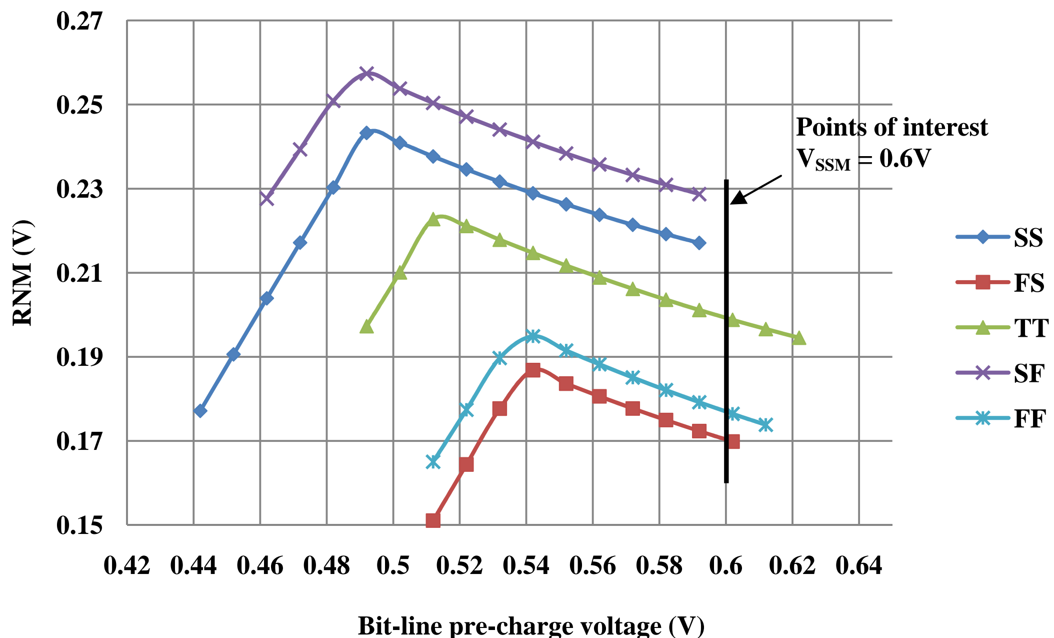


Fig. 3.17 RNM variations vs. bit-line biasing (=V_{SSM}) for various corners in 5TSDG.

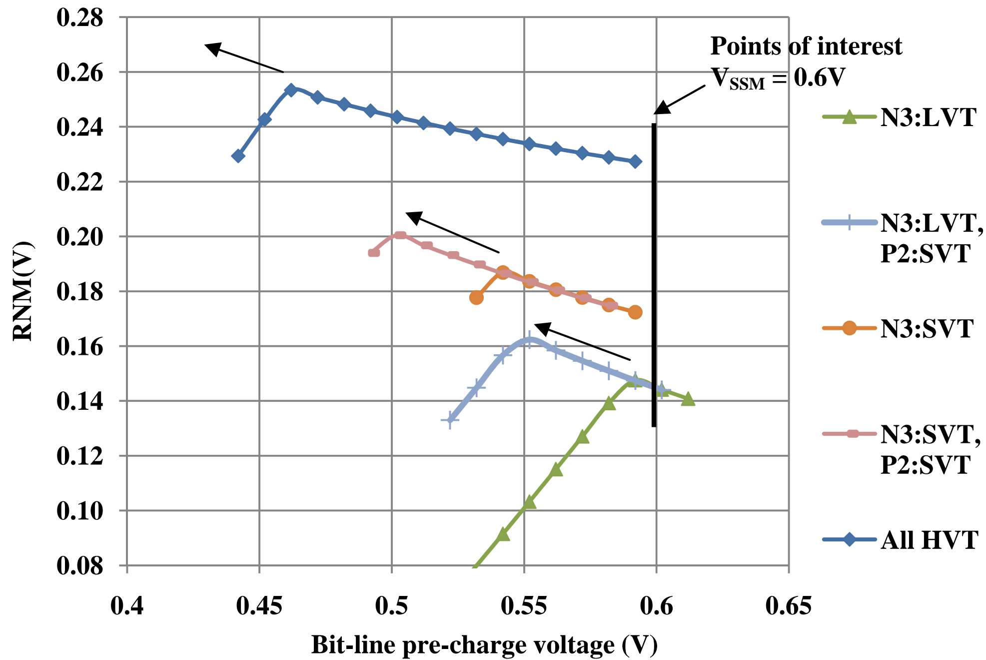


Fig. 3.18 Effect of the threshold of P_2 transistor on RNM. All transistors are HVT unless otherwise specified.

A reduction in the threshold voltage of P_1 transistor may be attractive to improve read speed, but it causes a shift-right effect on the RNM vs. Vbm curve (see Fig. 3.19). By strengthening the current drive of the P_1, it will be harder for a noise source to toggle the state of the cross-coupled inverter pairs. Furthermore, reducing the threshold voltage of P_1 adds to the leakage power and affects write '1' performance since it will oppose Qz to be driven to '0' by N_1. The combinational effect of lowering the threshold voltage of the P_1 and P_2 will cause a shift up effect on the RNM curves (Arrow 3 in Fig. 3.19) while each individual one causes a shift left or right to the RNM vs. Vbm curve (Arrow 1 and 2).

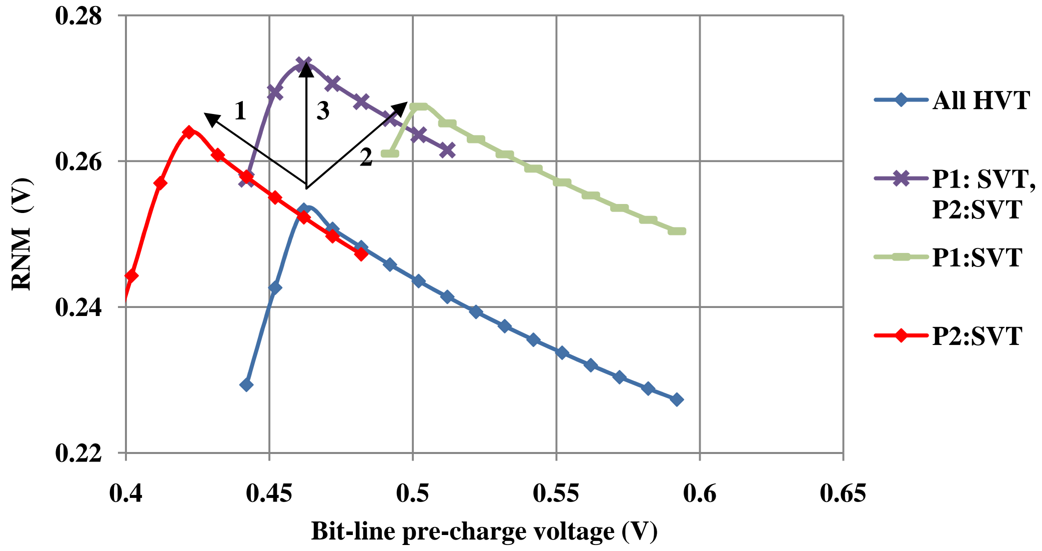


Fig. 3.19 Effect of changing the threshold voltage of the load transistors (P_1 and P_2) on optimum RNM (All transistors are HVT unless otherwise specified.

Reading with 5TSDG varies in delay time in different process corners by up to ~80ps as shown in Fig. 3.20 (Gbit0 is the global output in the sense amplifier for a read '0').

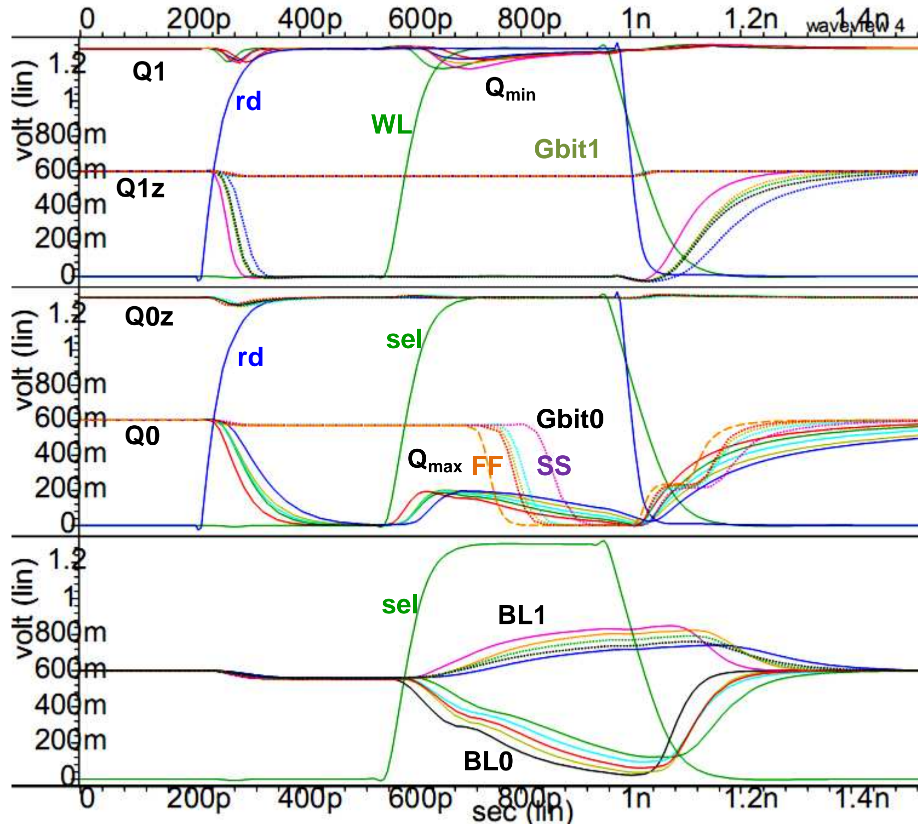


Fig. 3.20 Read '0' and read '1' operations in 5TSDG in different corners (superimposed). BL0-Gbit0, and BL1-Gbit1 are the bit-lines and sense amplifier outputs for cells storing a '0' and '1' respectively.

It is required for any SRAM design to have sufficient stability margins in all process corners. Static noise margin simulation results for 5TSDG indicating the stability of the cell in standby was shown in Fig. 3.3 and Fig. 3.10 for various process corners. Read noise margin is a more severe case of static noise margin when the memory is accessed for read. Two methods of measuring RNM were previously introduced being graphical measurement of maximum diagonal of the square within the VTC curves, and the SPICE implementation of the method using rotated coordinate systems. RNM measurement simulations results using both of these techniques are illustrated in Fig. 3.21. The left figure shows the conventional butterfly diagrams and the right one shows the results of the rotated

coordinate system method. RNM’ is equal to RNM times a scale factor equal to $^1/_{\sqrt{2}}$.

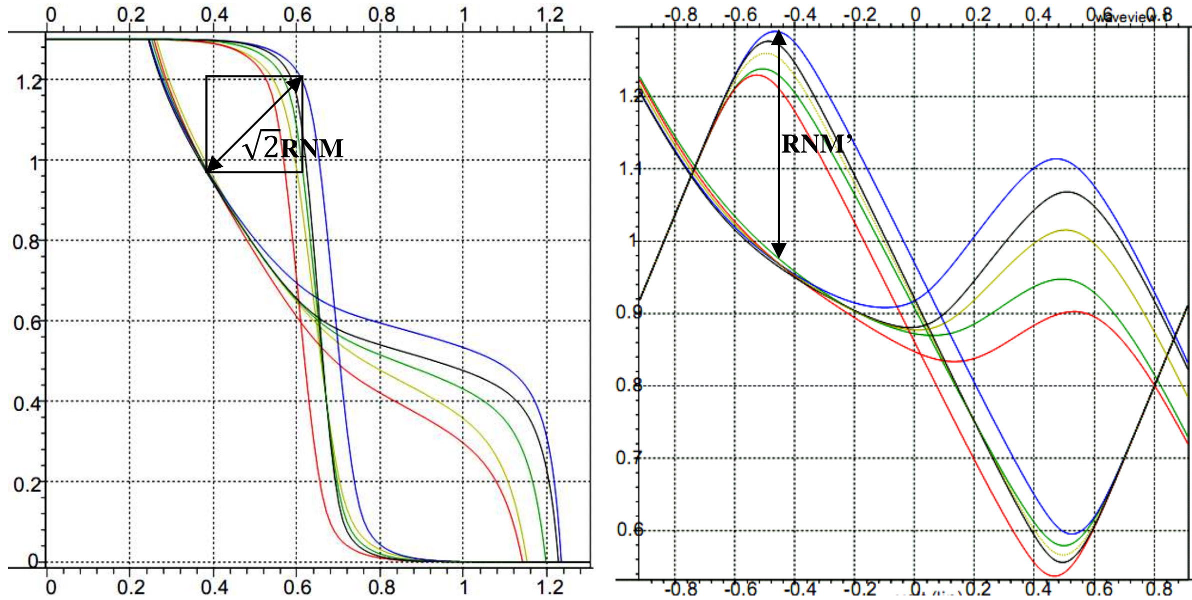


Fig. 3.21 Standard (Left) and rotated (Right) butterfly diagrams of 5TSDG when accessed (Read mode) in different corners.

The VTC curves of 5TSDG in comparison with a low-power 6T cell, in standby and access modes are depicted in Fig. 3.22 for a typical-typical corner (TT). As shown, the butterfly diagrams of 5TSDG are quite similar to that of the 6T cell, except that since a 5T cell only has one bit-line, and it is tied to V_{SSM} (see Fig. 3.14), there is an asymmetry in each half side of the butterfly diagrams in read mode. Therefore, the smaller diagonal of the largest fitting square will determine the RNM. In addition, the RNM of a 6T cell is smaller than 5TSDG partly since the bit-lines are charged less in the latter one due to smaller potential voltage and thus have less power in destructive read especially with the presence of environmental disturbances.

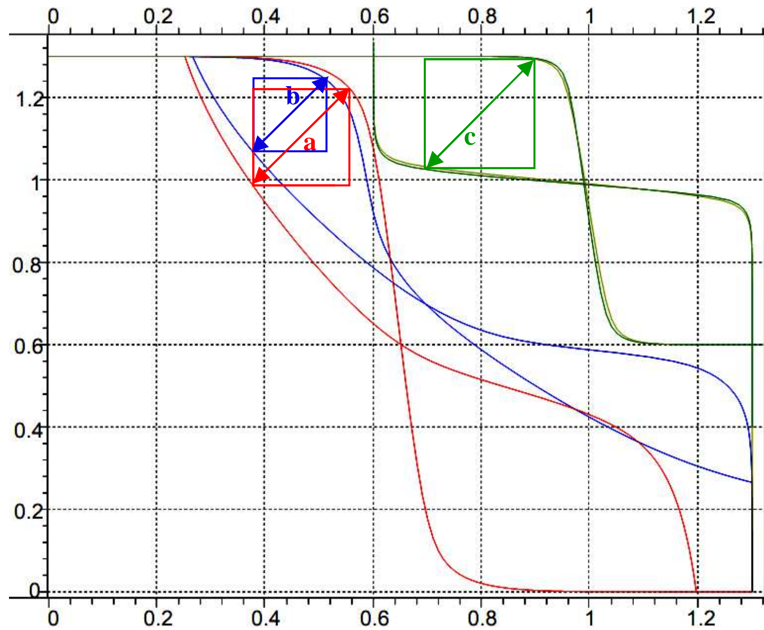


Fig. 3.22 Butterfly diagrams of 5TSDG and low-power 6T cells in read and standby modes: a: RNM_{5T} , b: RNM_{6T}, c: $SNM_{5T} \approx SNM_{6T}$.

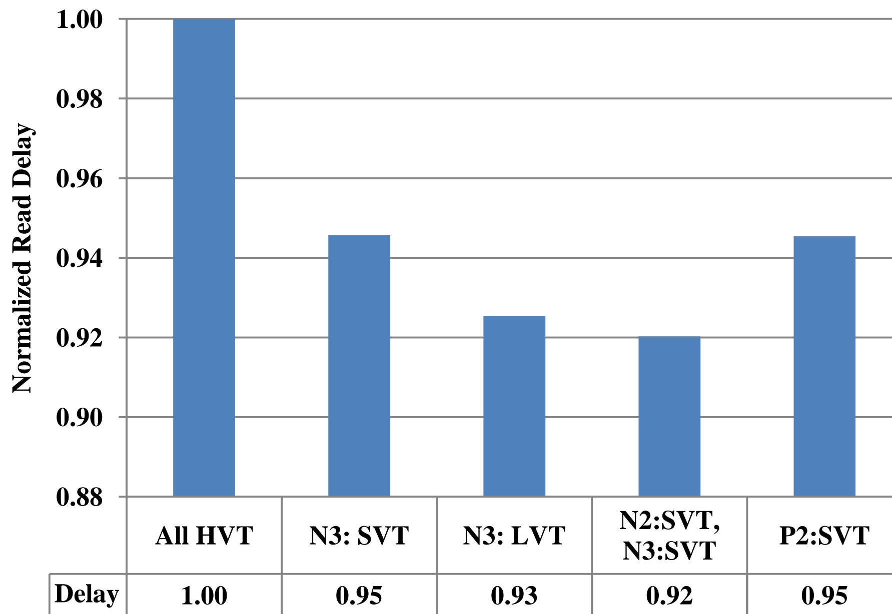


Fig. 3.23 Read delay comparison in 5TSDG for different interesting threshold voltages in the cell (All transistors are assumed to be HVT unless otherwise specified).

3.2.1 V_{SSM} stability in dynamic mode of operation

During standby mode of 5TSDG, V_{SSM} is used as a power supply to raise V_{g1} and V_{g2} above V_{SS} and pre-charge the bit-line. During read operation, V_{g1} and

V_{g2} are driven to V_{SS} to maximize RNM and the read speed. On the other hand, after a read operation is completed, V_{g1}, V_{g2} and the bit-line are driven back to V_{SSM} since the memory cell will be in standby again. This voltage swing of V_{g1}, V_{g2} and the bit-line affects voltage level of V_{SSM} since each re-charge of these voltages takes charges away from V_{SSM} causing it to drop by an amount of ΔV_i, where i is the index of consecutive read operations. ΔV_i is maximized in FF corner. In case of reading a '1', the bit-line will actually discharge the extra charges on V_{SSM}, however, that amount is much less than the effect of the ground lines taking away charges. Fortunately, V_{SSM} is highly capacitive and its capacitance is much higher than that of V_{g1} and V_{g2}, and many memory cells in standby provide electric charges to it. Therefore, V_{SSM} changes very little during read operation especially when it has large capacitance (attached to large memory arrays), and even if it does, it will actually help the read operation in terms of performance and read noise margin (see Fig. 3.17). In addition, V_{SSM} does not decrease beyond a steady-state value, and when reading is complete, it is pulled back towards its standby level due to an increase in memory cell leakage (see Fig. 3.24 and Fig. 3.25). Fig. 3.26 shows how V_{SSM} reaches a steady-state value after several read operations for different SRAM array sizes (64Kb, 1Mb, and 2Mb) in FF corner. This figure demonstrates that when larger number of memory cells are attached to V_{SSM}, the initial values of ΔV_i which are instantaneous voltage decays after each read, and the total decay to reach the steady-state value, ΔV_{tot}, will be smaller than that of smaller arrays. In contrast, in larger memory arrays, the dynamic steady-state voltage is closer to the

standby level than that in smaller arrays. After each read cycle, ΔV_i is reduced until it reaches 0V. At this point (steady state), the memory leakage is sufficiently increased such that it can fully replenish the lost charges between every read cycle. V_{SSM} voltage after each read cycle (*i*) can be described by equation 3.1.

$$V_{SSM}(i+1) \cong \varphi(i).V_{SSM}(i) + \frac{i_{m_{avg}}(i).\Delta t}{C_{VSSM}(stby)} \quad (3.1)$$

$$\text{where } \varphi(i) = \frac{C_{V_{SSM}}(read) + (\frac{V_{BL_0}}{V_{SSM}(i)} N_{B_0} + \frac{V_{BL_1}}{V_{SSM}(i)} N_{B_1})C_{BL}}{C_{VSSM}(stby)},$$

$$C_{VSSM}(stby) = C_{V_{SSM}}(read) + C(SubCol),$$

$$C(SubCol) = (N_{B_0} + N_{B_1})(C_{BL} + C_{V_{g1}} + C_{V_{g2}}),$$

C_{Vg1}, C_{Vg2}, C_{BL}, and are the capacitances of V_{g1}, V_{g2}, and the bit-line respectively. It is assumed that V_{g1} and V_{g2} have been driven to 0V initially. V_{BL_0} and V_{BL_1} are the bit-line voltages after a read '0' and a read '1' respectively. N_{B_0} and N_{B_1} are the number of '0' and '1' bits in a word respectively (16 bits/word in the simulation results of this paper). In standby, $C_{VSSM}(stby)$ includes the capacitance of all sub-column connections, and V_{SSM} interconnections. During read, a single sub-column with capacitance of $C(SubCol)$ is removed. $i_{m_{avg}}(i)$ is the average memory leakage current over the i-th read cycle period, Δt, and also includes a small contribution from V_{SSM} control circuit in Fig. 3.1. It is increased as V_{SSM} is reduced. Similarly, $V_{SSM}(i)$ is the V_{SSM} voltage at the end of the i-th read cycle. As the memory array size is increased, $\varphi(i)$ approaches to one since $C_{V_{SSM}}(read)$ approaches C_{VSSM}(stby). Part of ΔV_i is caused by a small amount of overlap between rd and stby signals in Fig. 3.13.

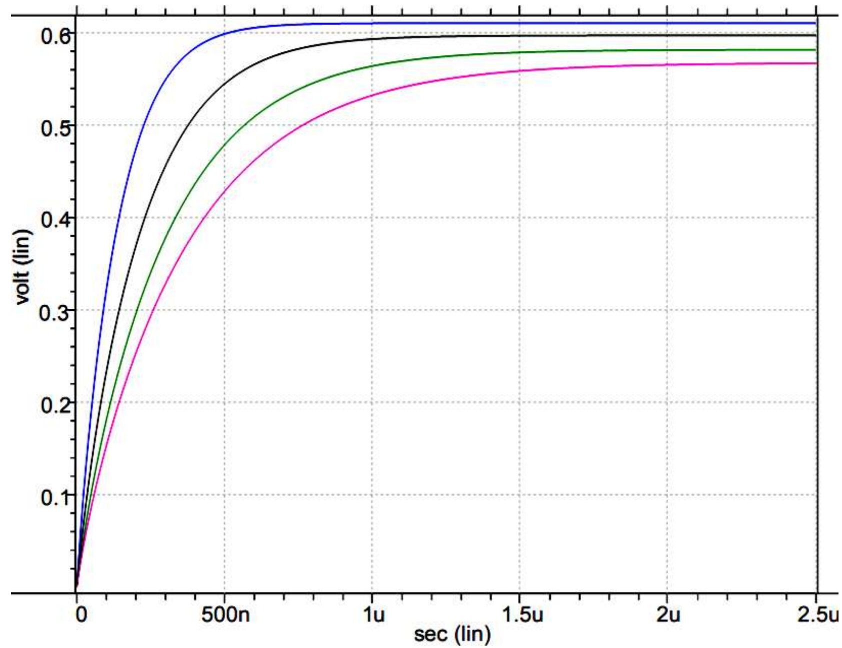


Fig. 3.24. The response of V_{SSM} when forced to '0' volts and left floating in standby for 64Kb, 128Kb, 256Kb, and 512Kb from left to right for 5T SRAM array (FF corner, 120℃).

This effect on V_{SSM} occurs also in write operation when the bit-lines are charged and discharged. However, for explanatory purposes, read operation, which is the most severe among other operations, is selected to be demonstrated for explanatory purposes. The number of cells per bit-line and number of bits per word contribute to this effect.

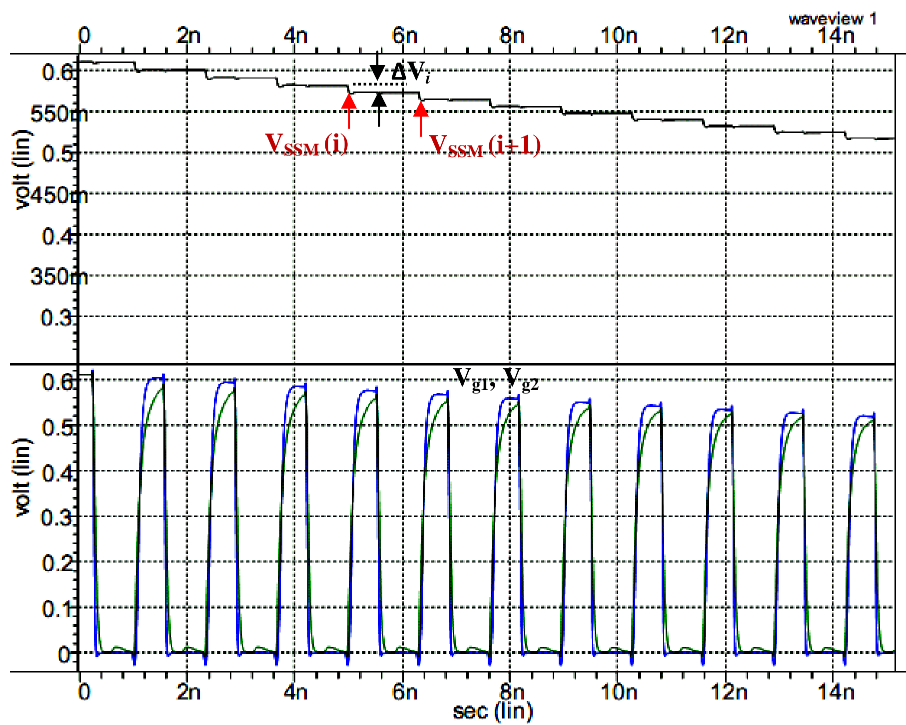


Fig. 3.25 Worst case effect of read operation on V_{SSM} (64Kb array 16 bits/word (all '0's), FF, 120℃).

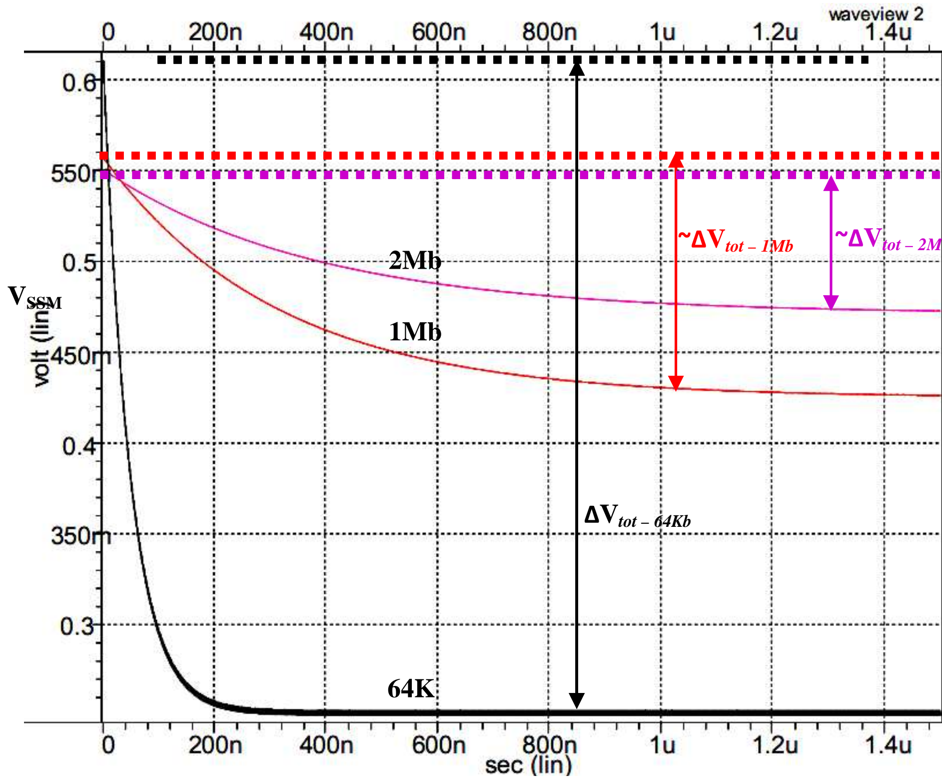


Fig. 3.26 V_{SSM} Saturation during several read operations for 64Kb, 1Mb and 2Mb 5T SRAM arrays (64 bits/word, FF, 120℃).

3.2.2 Read power consumption

The power consumption during the read operation is a function of Vmin, which determines V_{SSM} biasing level. During several consecutive reads, V_{g1} and V_{g2} in Fig. 3.1 are driven to V_{SS} and V_{SSM} frequently. Active power consumption is changed as supply voltage is changed due to the square law dependency. This power is also dependent on the frequency of V_{SSM} swing during read. Equation 3.2 shows the dynamic power consumed due to the voltage swing of the ground lines of 5TSDG, where C_L is the summation of V_{g1} and V_{g2} capacitances, ΔV is V_{SSM}-V_{SS} and f is the frequency of voltage swing.

$$P_{dyn} = C_L \Delta V^2 f \tag{3.2}$$

Reading a ‘0’ consumes more power than a read ‘1’ since in a read ‘0’, the bit-line is pulled sufficiently low to trigger the sense amplifier, and the global bit-line of the sense amplifier is also pulled down. In a read ‘1’, bit-line is only required to be pulled high enough to avoid the sense amplifier to turn on, and the global bit-line stays at V_{SSM}. Therefore, in an application where statistically more ‘0’s are to be stored in the cell, it is more efficient to store ‘1’s instead and complement the output data. Fig. 3.27 compares read power with standby power for various V_{DDM} values while keeping Vmin=V_{DDM}-V_{SSM} constant at 0.7V for a 64x16 bit block of 5TSDG. As V_{DDM} is increased, V_{SSM} also increases accordingly causing ΔV in equation 3.2 to increase during read operation. Therefore, read power is increased quadratically with higher V_{DDM}.

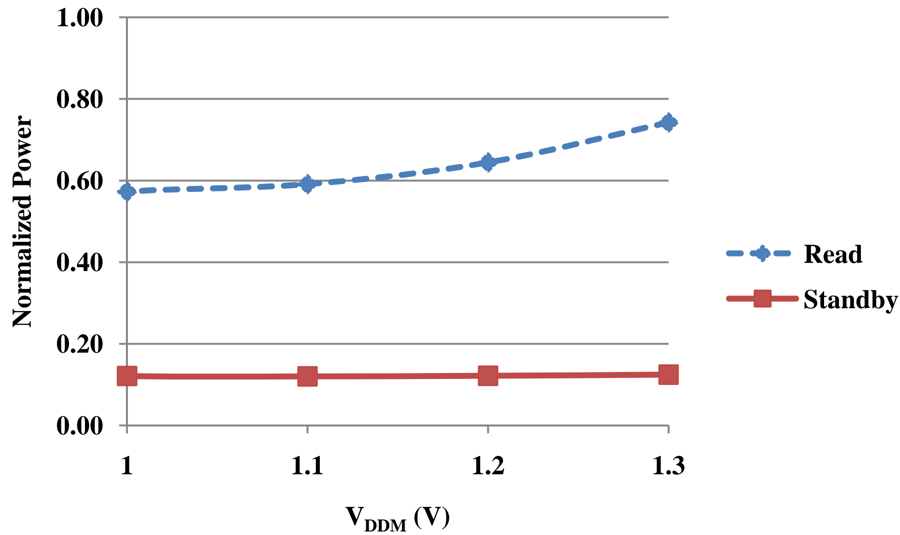


Fig. 3.27 Comparison of normalized read and standby power vs. V_{DDM} for 5TSDG, 64x16 bit block, reading 16 '0's continuously from 16-bit words (FF corner, 120℃).

Fig. 3.28 shows the case study results of the worst-case (FF, 120℃) normalized power consumption for standby mode and read operation of 5TSDG in comparison with low-power 6T design. Other corners have similar results comparable to Fig. 3.9. Read power consists of standby power of the idle memory cells, and the dynamic power described by equation 3.2. In this case study where a 64Kbit array consisting of 64x16 bit blocks was studied (reading from a 16-bit word), 5TSDG could achieve up to ~30% power reduction in read mode compared to that of the low-power 6T structure. In this example, reading a '1' consumes ~7% less power in 5TSDG compared to a read '0' as explained earlier. Obviously, larger number of read operations will result in a linearly higher power consumption difference in comparison with standby power due to larger values of f in equation 3.2. Read operation of 5TSDG and low-power 6T cells in this experiment were similar to Fig. 3.13 and Fig. 2.5 respectively.

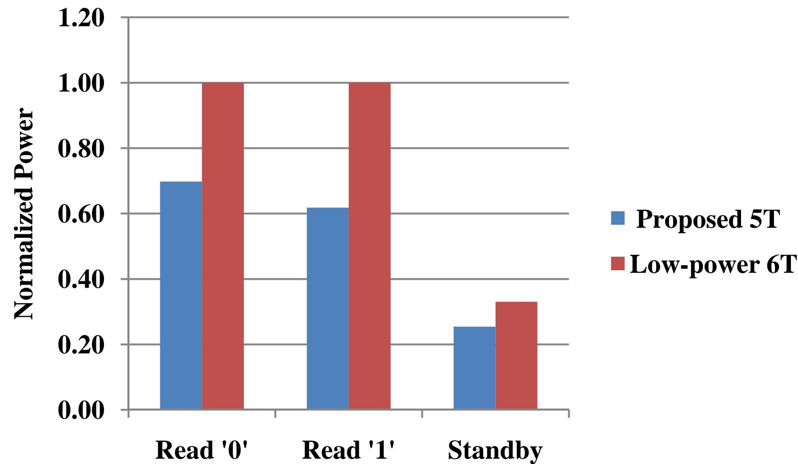


Fig. 3.28 Case study results of the worst-case normalized power consumption for standby mode and read operation of 5TSDG in comparison with low-power 6T design (FF, 120℃), 1.00 ≈ 25.8 mW.

3.3 Write operation

Since a 5T SRAM cell only has a single bit-line, writing either a ‘0’ (W0) or a ‘1’ (W1) into the cell is performed using the same bit-line. This is different from the 6T structure where there is technically no difference between a W0 or a W1, i.e. by selectively pulling down one of the bit-lines depending on the data status, a W0 operation is applied on one side of the cell and the feedback will recover the opposite storage node to the complement value. In 5TSDG, W0 is performed in a similar way. On the other hand, in W1, the bit-line is pulled high by global write signal, Gwr, so that when the word-line is selected, state toggle is initiated. Gwr is driven high by the write circuit in W1 and is driven to V_{SS} otherwise. Using conventional 6T transistor ratios and sizing, it is almost impossible to write a ‘1’ in a 5T cell because in a 6T cell: 1) N_2 needs to be stronger than N_3 by *cell ratio* (CR) factor β, typically between (1.2~1.5) to maintain read stability [9]. 2) P_1 and P_2 need to be weak enough, usually minimum size for *write-ability* purposes. 3)

The access transistor is an NMOS, which does not pull up strongly due to its physical nature. These constraints will oppose raising Q if applied in a 5T memory cell for a W1 using a single bit-line. To combat this problem, [15] suggests using different (W/L) sizes for the transistors such as, using a CR of ~0.45, weakening P_1, strengthening P_2 and N_1 with the cost of noise margin. As opposed to 5TSDG in this paper, design in [15] will cause a 50% reduction of RNM when compared to conventional 6T cell and therefore is more susceptible to parameter mismatch and performance fluctuations in more advanced technologies, especially due to process variations.

On the other hand, to make W1 possible, [14] suggests disconnecting Vg_2 from V_{SS} and letting it float near a biasing voltage by using a capacitor while keeping Vg_1 at V_{SS} during write. This method will weaken N_2 by lowering its V_{DS}, which will facilitate W1. However, this method does not take advantage of leakage-power reduction opportunities.

As shown in Fig. 3.1, and also discussed earlier, in 5TSDG, V_{SSM} is connected to Vg_1, Vg_2 and the bit-lines in standby mode. In W0, Vg_2 stays connected to V_{SSM} via Mg_2 while Vg_1 floats near V_{SSM}. In W1, Vg1 is pulled down to V_{SS} through M_{g1w1}. M_{equ} is turned on by Gwr signal which is high when W1 and is at V_{SS} otherwise. The role of this transistor is to limit $\Delta Vg=Vg_2-Vg_1$ as shown in Fig. 3.30 to improve SNM of the disturbed cells in the same sub-column. The strength of M_{equ} is chosen through simulation to limit write disturb for all process corners especially for fast NMOS corner cases [1][2]. This disturbance can also be minimized by reducing the write pulse period to its limit. In summary, in W1,

N_1 will have a stronger current drive than N_2 since its V_{DS} is maximized i.e. increased by V_{SSM}.

The threshold voltage of access transistor, N_3, plays a key role in W1 performance. Simulation results reveal that standby power varies less than 2% using high, standard or low V_{th} (HVT, SVT, LVT) for N_3. In order to improve W1 performance, the V_{th} of N_3 can be reduced with some loss of RNM. In 5TSDG, V_{th} of N_3 can be between the HVT and LVT to maintain a reasonable RNM/W1 performance as shown in TABLE 3.6.

TABLE 3.6. RNM and W1 comparison for different V_{th} for N_3, at worst case RNM (FS corner, 120℃).

5T Cell RNM (mV) / W1 Delay (ps) for Various N3 V_{th}		
LVT (~230 mV)	***SVT(~440 mV)***	***HVT(~600 mV)***
144.1/96.8	**172.3/116.4**	225.1/170.6

As discussed earlier in Chapter 2, there are various types of write circuits which can be used in a design and they can vary depending on performance, power and area limitations. The write circuitry used in this thesis for 5TSDG driving each sub-column is depicted in Fig. 3.29 showing N bit-lines consisting of a word. When the write enable signal is active (WE), depending on the data values ($D_0 \ldots D_{N-1}$), the global write signals ($Gwr_0 \ldots Gwr_{N-1}$) are driven high or low by the drivers. When the transmission gates ($TG_0 \ldots TG_{N-1}$) are selected, the bit-lines are either pulled high or low depending on whether the data value is a ‘0’ or a ‘1’. The write enable signal will deactivate the standby signal (Stby, see Fig. 3.1) and will configure the ground control circuitry to complete the write operation into the cell.

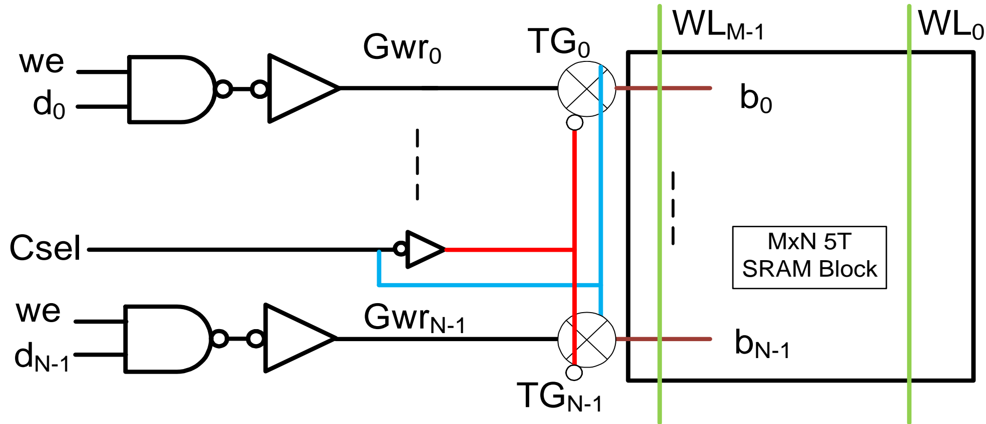


Fig. 3.29 The write circuitry for 5TSDG (Not including the ground control circuit needed in write operation).

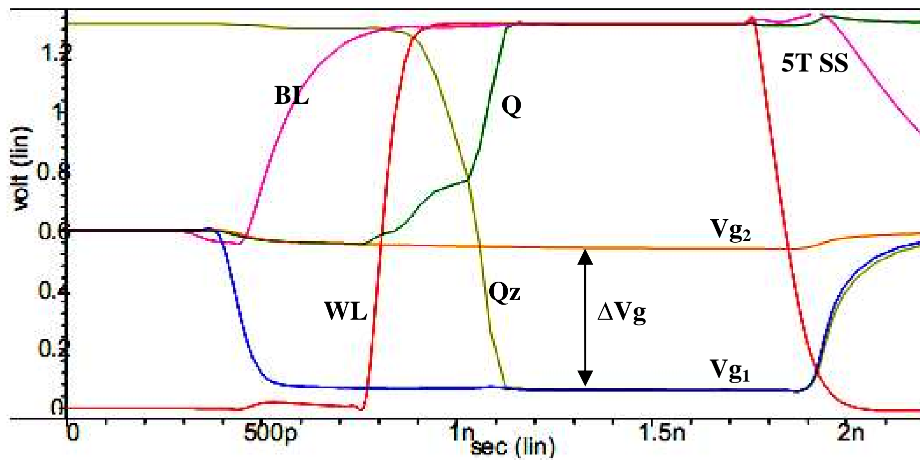


Fig. 3.30 W1 operation of 5TSDG in slow corner (SS) (120℃).

Therefore, to improve read stability and write-ability (particularly W1), the solution is to find a reasonable mid-point considering the fact that N_3 does not play a key role in standby power consumption. Limited to three choices for V_{th} selection, SVT for N_3 is reasonable as shown in TABLE 3.6. However, in chip foundries, even a lower threshold somewhere between LVT and SVT can be achieved by changing gate oxide thickness. Fig. 3.31 shows simulation results for W1 under various V_{th} variations. In this experiment, mismatch was applied to the threshold voltage of each single transistor in the slowest corner for W1 (SS, σ_N=σ_P=2). Fig. 3.31 also shows the write-disturbed cell voltages Qd and Qdz.

Fig. 3.32 compares W0 and W1 performance of 5TSDG with a low-power 6T SRAM described in TABLE 3.1. For both cases, W1 delay is measured from when WL = 50%V_{DDM} to when Q0 = 80%V_{DDM}, and W0 delay is measured similarly but when Q=20%V_{DDM} above V_{SS}. This measurement is different from what was reported in TABLE 3.6 (word-line to Q-Qz cross point). W1 can be ~11-31% slower than a conventional 6T design and can be improved by reducing V_{th} of N_3. W0 performance is similar to conventional 6T cell.

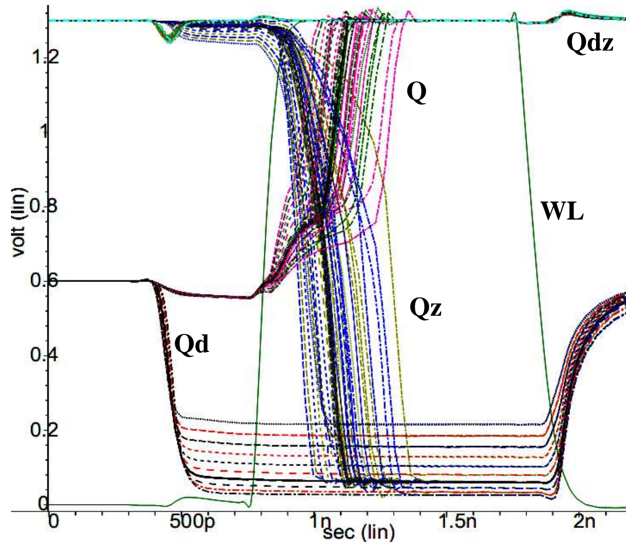


Fig. 3.31 W1 operation and effect of various V_{th} variations applied on each transistor at a time (-6σ to 4σ) relative to SS corner in a writing and a disturbed cell (Q, Qz and Qd, Qdz), 120℃, Note that Qdz is not disturbed.

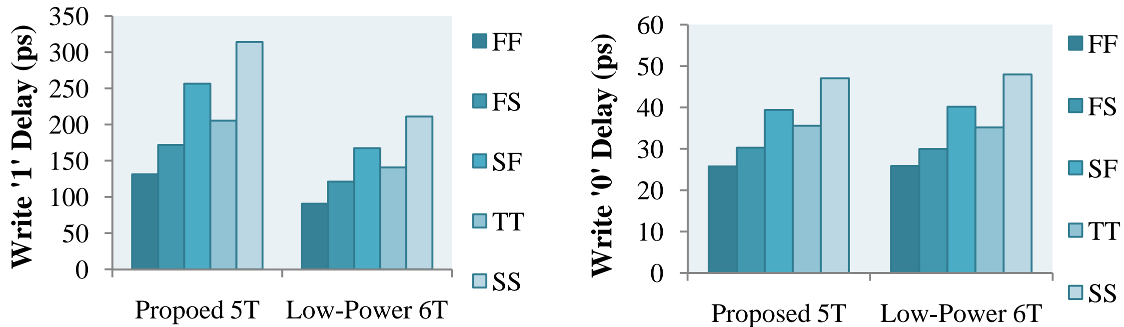


Fig. 3.32 Write '1' (left) and Write '0' (right) delay comparison in different corners between a 5TSDG and a low-power 6T cell (120℃).

Numerical results for comparison of W1 performance under threshold voltage variations in the slowest corner (SS) are demonstrated in Fig. 3.33. V_{th} of N_3 and N_1 have the most significant factor in W1 delay. N_3 is the only gateway to pass charges from the bit-line to the pull up transistor P_2 to drive Q high and therefore plays a significant role in W1 speed. When the bit-line is driven high by the write circuit, and the word-line is raised, N_1 is turned on, and it will pull down the complementary node, QZ. This will turn P_2 on which will pull Q high to complete the write operation.

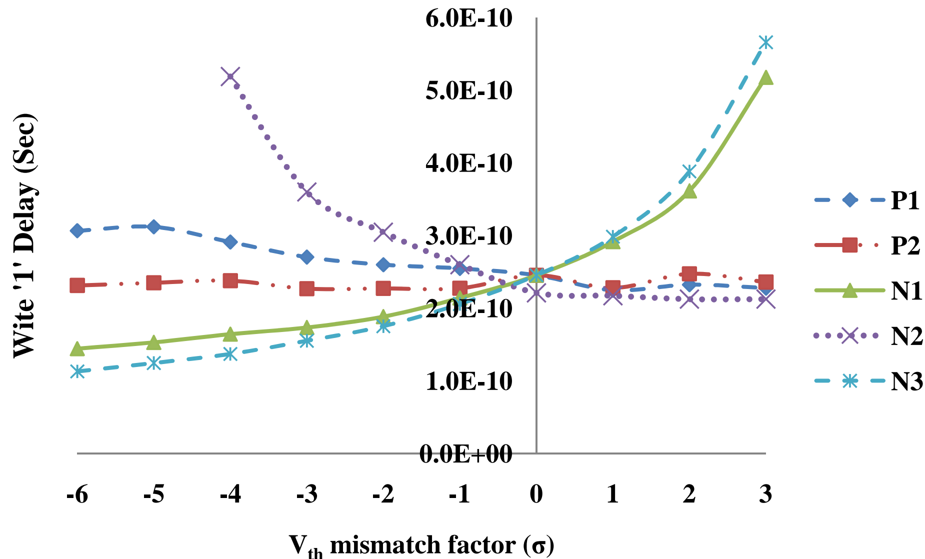


Fig. 3.33 Effect of Vth mismatch on write '1' delay in worst case corner (SS), V_{g1}=0.07, V_{g2}=0.47.

Since in 5TSDG V_{g1} is pulled down during W1 by M_{g1w1} (see Fig. 3.1), cells in the same sub-column sharing same V_{g1} and V_{g2} are disturbed. As explained earlier in this section, the role of M_{equ} transistor is to limit this disturbance by also pulling V_{g2} down weakly and limit the voltage drop of V_{g1} all the way to V_{SS}. This will prevent the SNM of the disturbed cells to drop below the safe limit. Fig. 3.34 (a) shows how the voltage of V_{g1} can affect SNM on disturbed

cells while driving Vg2 at a fixed voltage (at V_{SSM}) mimicking that there is no M_{eq}. It is immediately noticed that as V_{g1} is lowered, SNM drops and can even reach 0V. Fig. 3.34 (b) demonstrates the reverse scenario where V_{g1} is fixed at 0V, and V_{g2} varies from 0V to V_{SSM}. Similarly, this figure shows that with no weak equalization between V_{g1} and V_{g2}, the disturbed cells are susceptible to data corruption due to environmental disturbances. The strength of M_{equ} will determine the limitation on this disturbance by both lowering V_{g2} from V_{SSM} and not allowing V_{g1} to be pulled down so much. In 5TSDG, M_{equ} was ratioed to maintain V_{g1} at 0.07V and pulling V_{g2} down to $V_{SSM} - 0.13V = 0.47V$ for a V_{SSM}=0.6V in W1 resulting a minimum disturbed SNM of ~50mV. Fig. 3.35 shows the SNM values for the W1 disturbed cell in different corners for two cases of M_{eq} strengths. Widening M_{eq} will result in a higher disturbed SNM but a higher W1 delay.

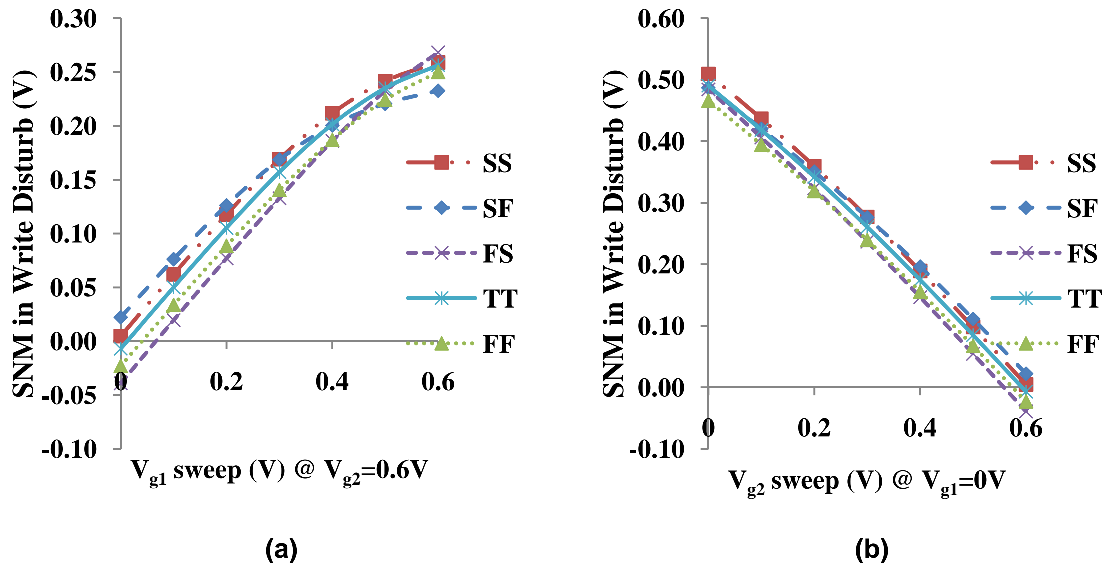


Fig. 3.34 SNM in W1 disturbed cells

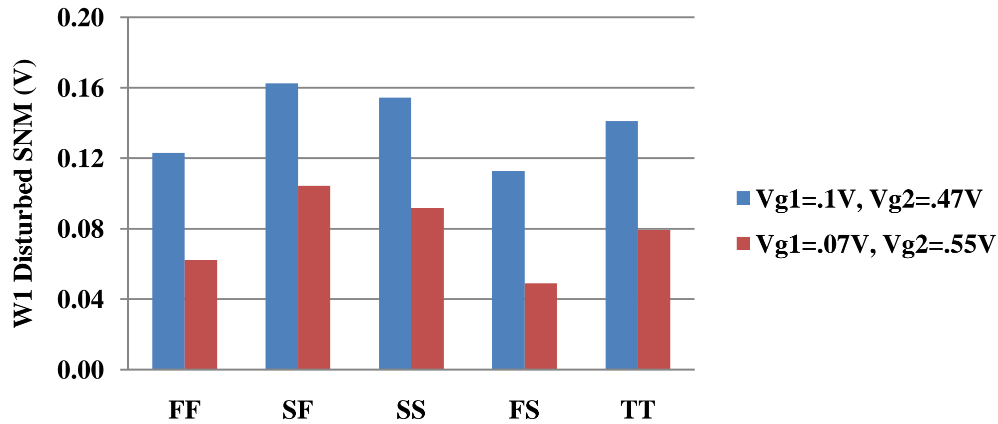


Fig. 3.35 The SNM of W1 disturbed cells in different corners and different M_{eq} strengths.

Although parity check circuitry is necessary for soft error recovery, and will restore data in an unlikely case of data corruption, some of the ways to improve SNM disturbance in W1 to the neighbouring cells are:

- Increasing the strength of M_{equ} transistor to pull down V_{g2} more strongly and hence, reducing ΔV_g especially for fast NMOS;
- Reducing W1 write pulse to limit the period the neighbouring cells are disturbed, and hence reducing the probability of environmental disturbances to flip the cell;
- Increasing transistor sizes accordingly to improve SNM in general.

3.3.1 Write Margin (WM)

The write margin of 5TSDG can be divided into W0 margin (W0M), and W1 margin W1M since as opposed to the 6T cell counterpart, W0 and W1 have different WMs. One of the common methods to measure WM in conventional 6T SRAMs is by measuring the maximum BL voltage able to flip the cell state [22].

For 5TSDG, W1M is defined to be the difference between the positive supply voltage, V_{DDM}, and the minimum BL voltage able to write a ‘1’ into the cell while W0M is defined to be the maximum BL voltage able to write a ‘0’ into the cell. In 5TSDG (V_{DDM}=1.3V), for a typical-typical corner (TT), W1M is ~0.501V, and W0M is ~0.400V.

3.3.2 Write power consumption

The write power in 5TSDG can be divided into write ‘0’ and write ‘1’ power, each consisting of the standby power of the idle cells, plus the dynamic power. In a case study in which the results are shown in Fig. 3.36, a 64Kbit array consisting of 64x16 bit blocks was studied while writing into a 16-bit word. In this example, write ‘0’ consumes ~80% less power, and write ‘1’ consumes ~9% less power compared low-power 6T structure in worst-case scenario (FF corner, 120℃). Fig. 3.36 depicts a comparison between the worst-case power consumption of different modes of operation comparing 5TSDG with a low-power design. Other corners have similar results relatively comparable to Fig. 3.9.

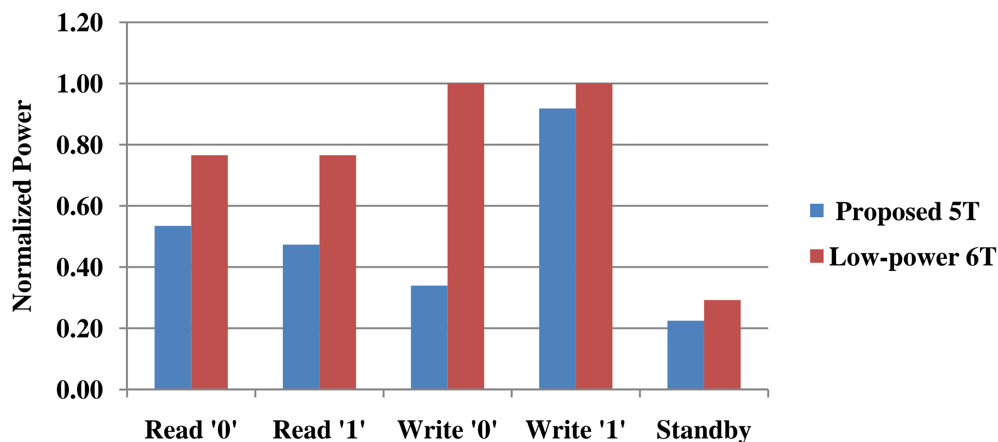


Fig. 3.36 Case study results of the worst-case write power consumption in comparison with read and standby power for 5TSDG vs. low-power 6T design (FF, 120℃), 1.00 ≈ 33.8 mW

4: LAYOUT DESIGN

Layout structure of a 6T SRAM cell and the cross-coupled inverters are categorized into four types of possible combinations of transistors in terms of physical orientation and connections discussed in [21]. Type-4 category in [21] is selected in this thesis due to its area efficiency advantages and that it enables the differential ground capability needed in this design, and is drawn using 65nm design rules. Fig. 4.1 compares a single cell layout of a 5TSDG cell (a) with a conventional 6T cell (b), drawn using standard digital logic design rules. Of course, memory design companies do not use standard design rules for digital circuits and develop their own ones. For example well-to-active spacing is quite large and can be reduced in SRAM cells since they are low-current devices as opposed to some regular digital circuits. The area saving opportunity is due to the lack of one bit-line and one less access NMOS device, which enables the two neighboring cells sharing the same word-line to be placed more closely on the top of each other than the 6T cells. A demonstration of this area saving opportunity (not taking the area saving due to smaller transistor widths into account) is shown by the yellow box (labeled “Saved Area”) in Fig. 4.2. An area saving of ~13% is achieved using this method assuming that a 5TSDG cell with $\beta=\beta_p=1$, is compared with a conventional 6T cell with $\beta=1.58$, and $\beta_p=1.13$. It is interesting to observe that even if a 5TSDG cell is compared with a conventional 6T cell with $\beta=\beta_p=1$, the area saving due to the removal of a bit-line and an

access transistor is ~11%.5TSDG layout is shorter in height and is the same width as the conventional 6T cell. Fig. 4.3 shows an array of 5TSDG in 4x4 (a) and 64x16 (b) arrays including substrate contacts at two ends. Depending on the technology used, the maximum distance allowed to place memory cells without a substrate contact is a good indication of the number of cells per sub-column since that length can also be used to place the sense amplifiers shared between two adjacent sub-columns.

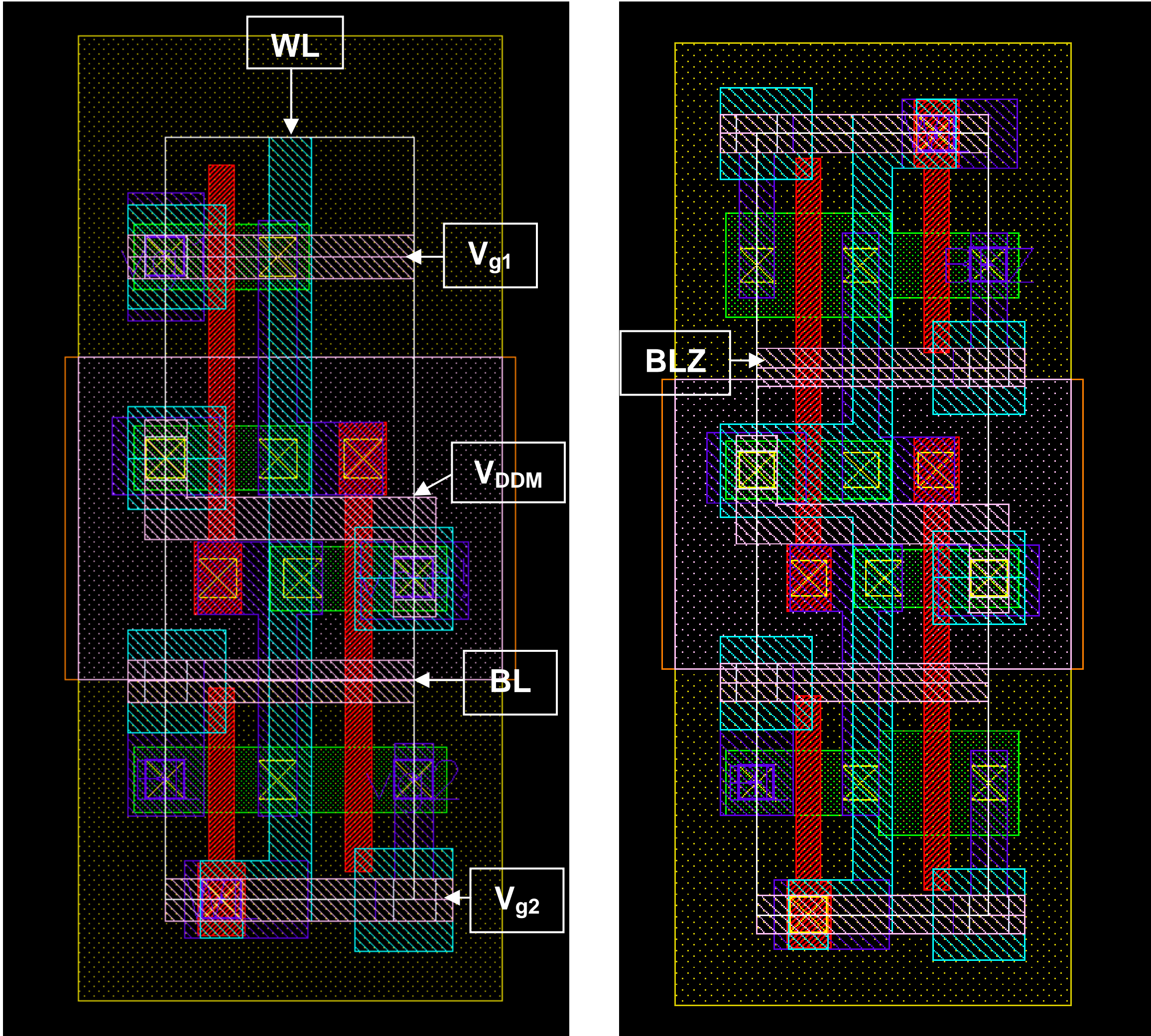
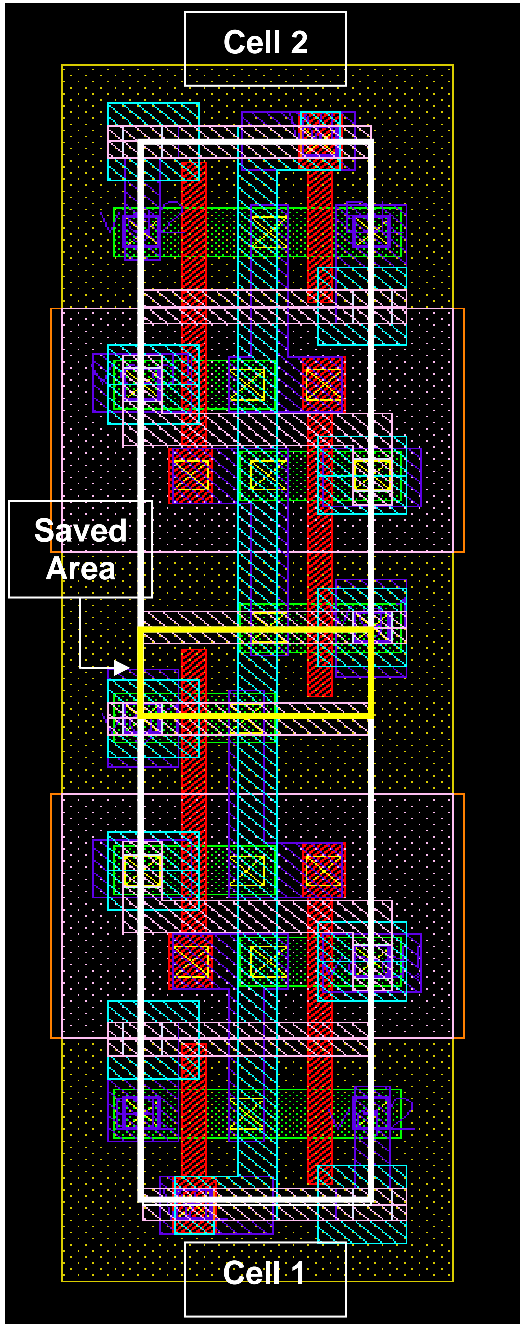
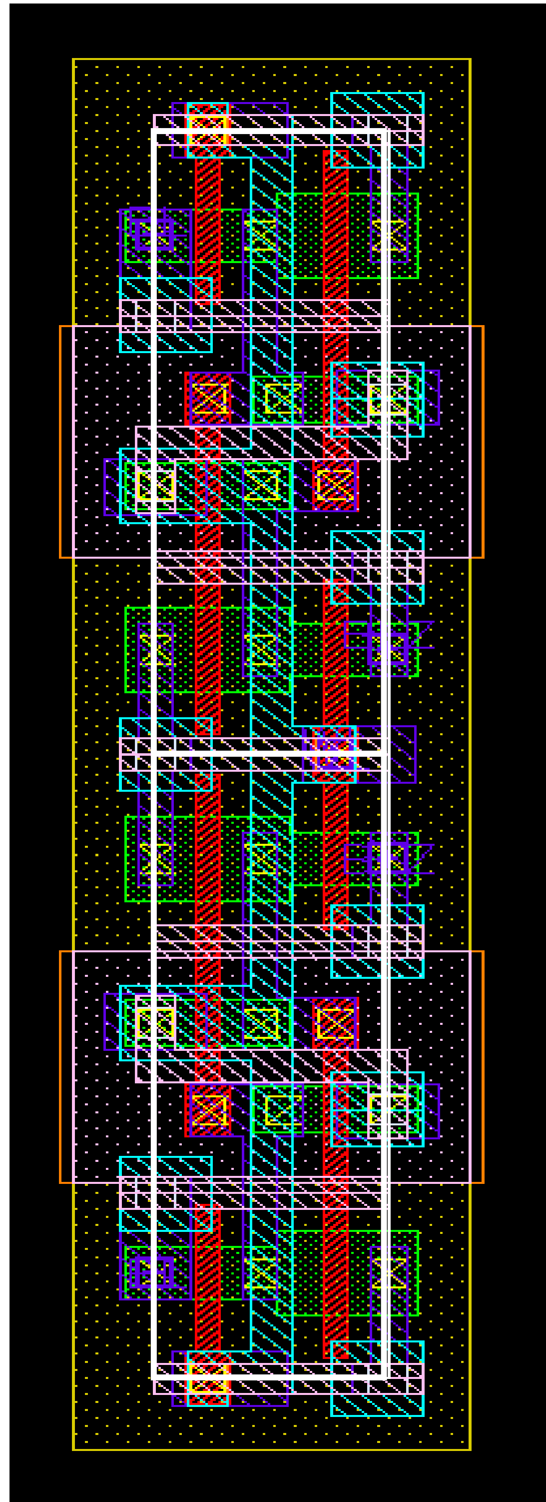


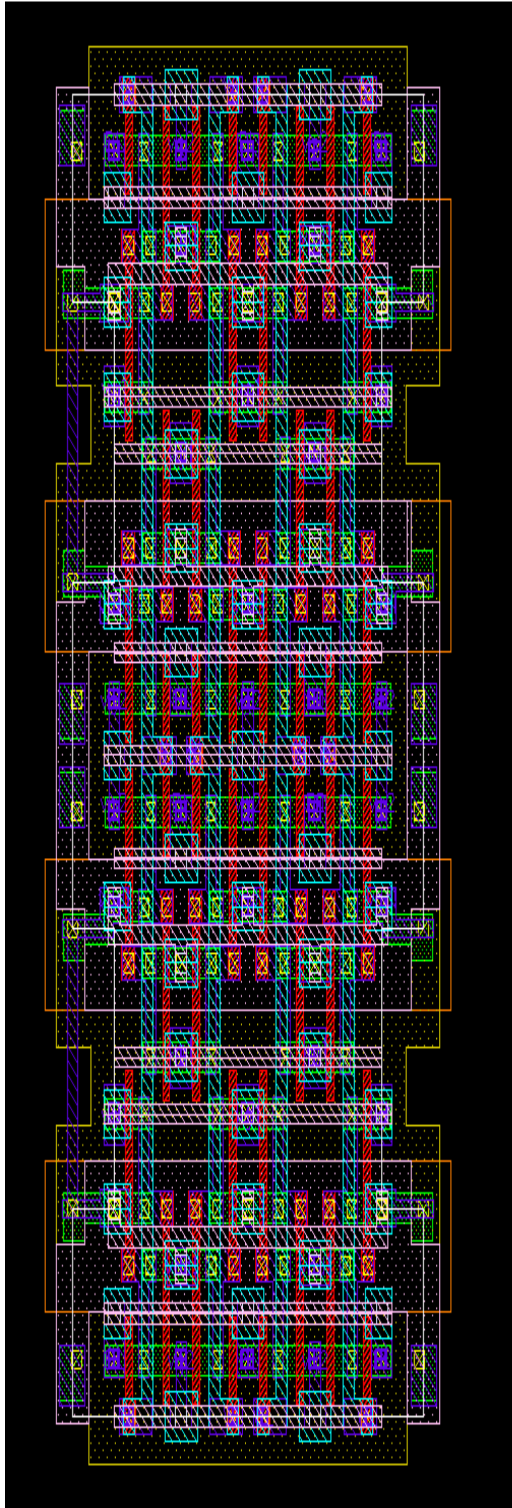
Fig. 4.1 5TSDG (a) and conventional 6T (b) cell layouts.



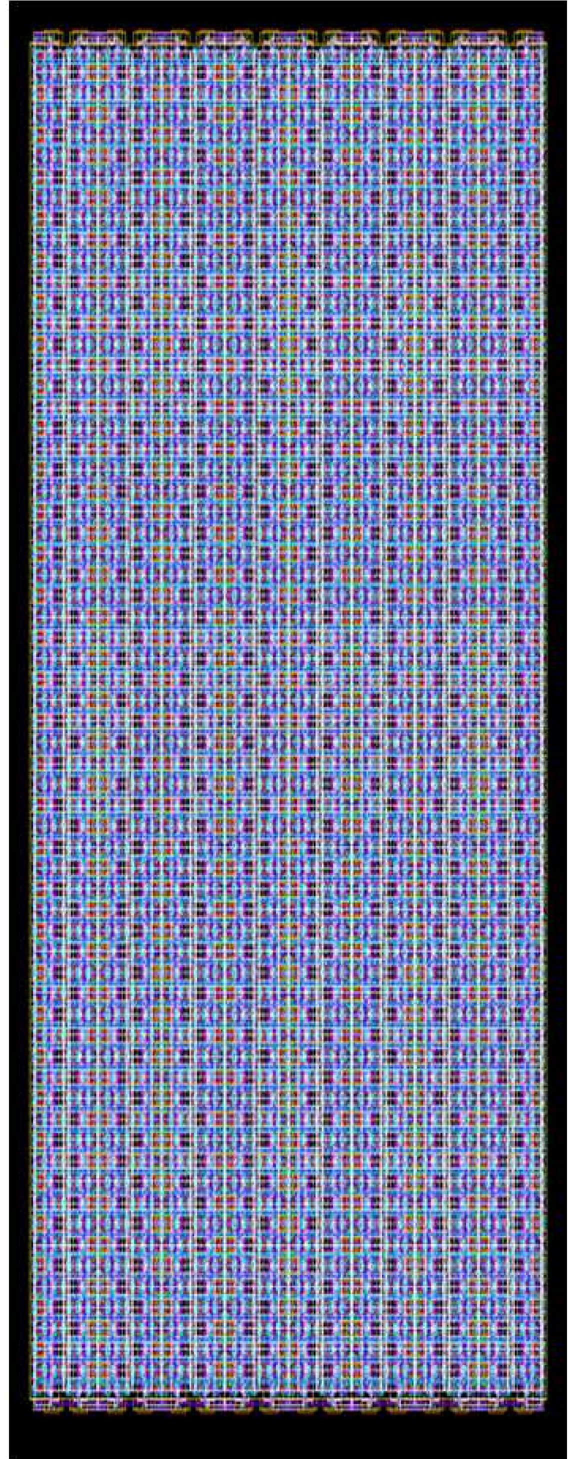


(a) **(b)**

Fig. 4.2 5TSDG (a) and conventional 6T (b) pair of cells sharing same word-line (WL), Total lengths and widths not to precise scale.



(a)



(b)

Fig. 4.3 A 4x4 (a), and a 64x16 cell array of 5TSDG. Total lengths and widths not to precise scale.

4.1 Layout extraction and Post-Layout Simulations (PLS)

In order to simulate the behaviour of the designs using SPICE model as accurately as possible, some layout factors needed to be manually inserted in the simulation code, which would not have been considered by default by the simulation software program otherwise. These extra parameters include wire capacitances (such as bit-line, word-line, etc) and drains/source capacitances of the transistors with shared drains/sources (which are less than two separate transistors). After finalizing the design and drawing the layouts, their compliance with the design rules were verified. The layouts were then extracted including parasitic capacitances. Parameters such as bit-line and word-line capacitances were compared with the simulated ones, and therefore the simulation results were verified. The only discrepancy is due to the software limitation and that is, since in this particular type of cell layout, the NMOS devices partly share same drains, the total drain area of two side-by-side transistors is not twice the drain area of one transistor (see Fig. 4.1) although the extraction software program considers it as separate transistors.

5: DISCUSSIONS, CONCLUSIONS AND FUTURE WORK

In this thesis, concepts, methodologies, and related research on the design of SRAM cells, particularly 6T type, were first reviewed. Methods of stability measurement, identifying power consumption sources such as leakage, standby and dynamic power, and power-saving opportunities were discussed. Performance fluctuation issues due to the effect of process variations, and solutions to mitigate these effects were addressed. Previous research results on 5T SRAMs were reviewed, and problems, performance issues, and power reduction opportunities in them were addressed. A new low-power, low-area, reliable, high performance and process variations tolerant 5T SRAM architecture (5TSDG) was proposed which has improvements in standby and dynamic power consumptions, stability margins, and performance when compared to previous research results in this area. Up to ~30% standby, and a significant amount of dynamic power reduction are gained in 5TSDG compared to a low-power 6T design. In 5TSDG, W1 can be slower by ~11-31% where process and layout choices can be made to make W1 performance acceptable. Post-layout simulations from extracted views of 5TSDG were performed to verify the validity of the simulated results based on SPICE models. Significant area saving (~13%) is predicted compared to the conventional and low-power 6T cells. Column control methods to improve the SNM of W1 disturbed cells are part of on-going research. Another possible future work can be to study the feasibility of applying

5TSDG, in ultra-low voltage applications. The block structure and memory organization of 5TSDG design creates opportunity for future research. However, an example of such organization is demonstrated in Fig. B.1 for explanatory purposes. This figure shows a simplified diagram of a memory block in an array of 5T cells, arranged into two-sub-column sub-blocks, and includes decodes, write drivers, and ground control circuits (*VgCtrl*). The sense amplifiers for read purposes are placed between two adjacent sub-columns (not shown).

APPENDICES

Appendix A

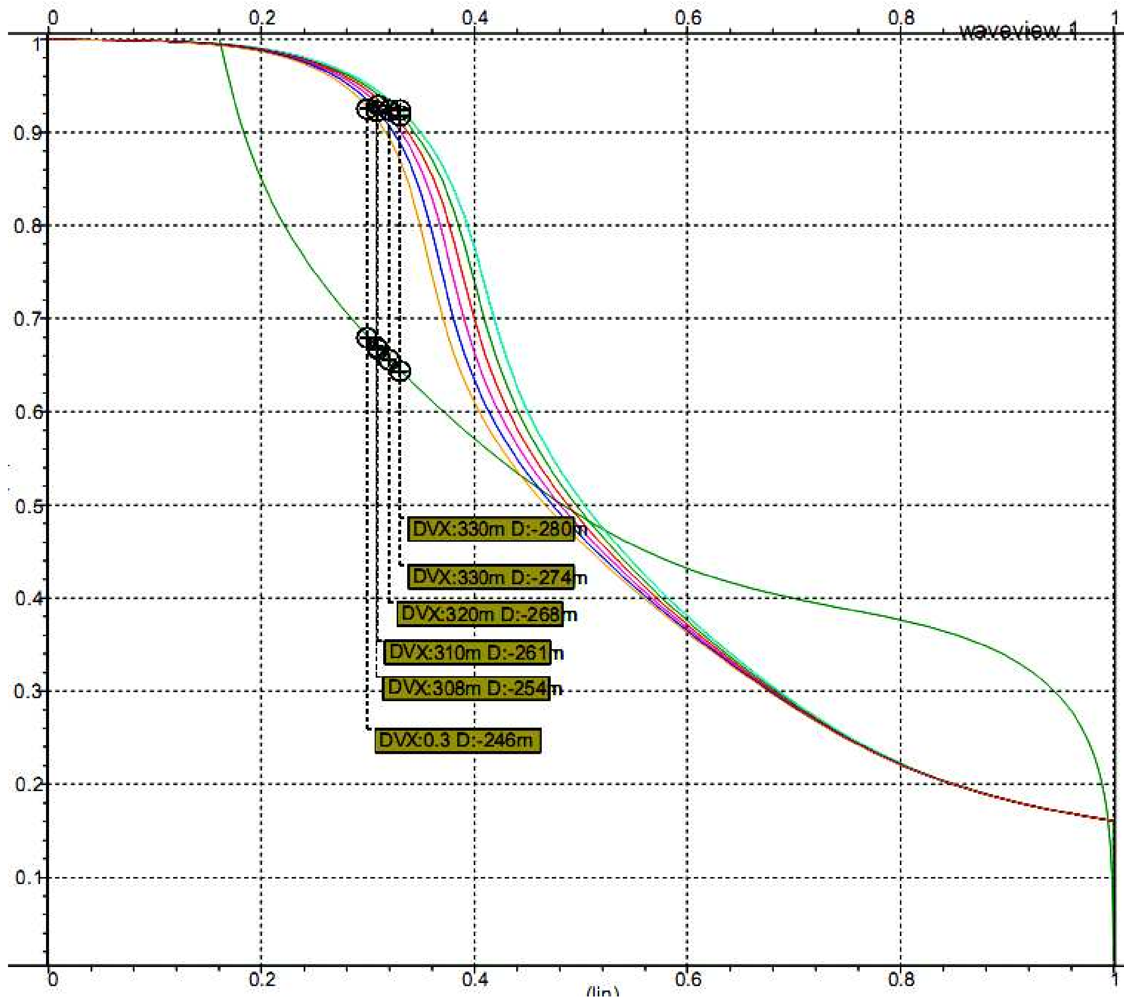


Fig. A.1 Measurements of vertical distances in the original non-rotated butterfly diagrams showing the effect of P-type mismatch on the RNM of a conventional 6T cell.

Appendix B

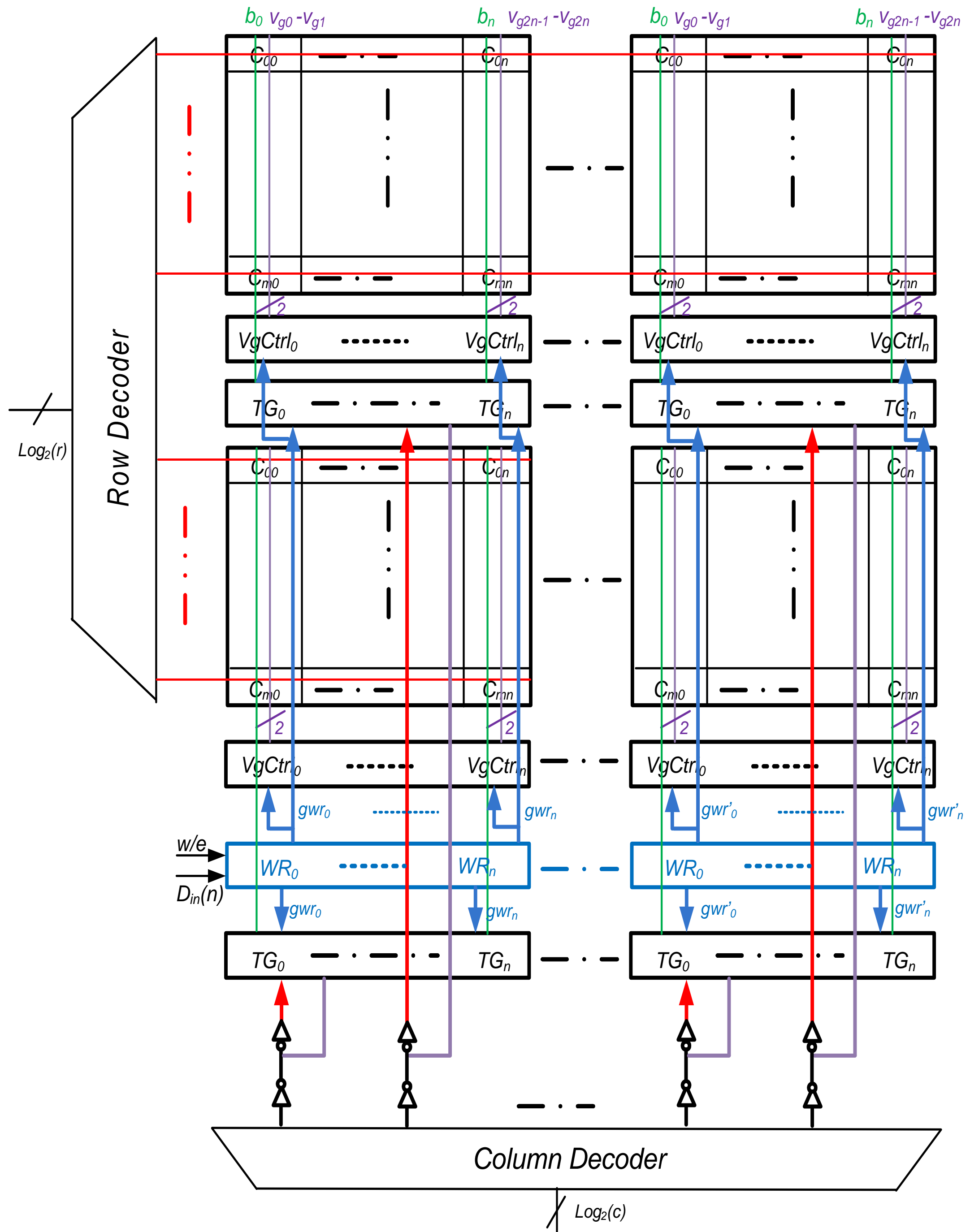


Fig. B.1 5TSDG block structure including write circuitry.

REFERENCE

[1] Jarollahi, Hooman, Hobson, R.F., "Power and Area Efficient 5T-SRAM with Improved Performance for Low-Power SoC in 65nm CMOS ", *IEEE International Midwest Symposium on Circuits and Systems, MWSCAS 2010.*, Aug. 2010, In Press.

[2] Hobson, R.F., "A 5T SRAM Cell with 4 Power Terminals for Read/Write/Standby Assist", *Proceedings of the 2009 International Conference on Computer Design, CDES 2009*, pp.10-16, 13-16 Jul. 2009

[3] Andrei Pavlov, Manoj Sachdev, "CMOS SRAM Circuit Design and Parametric Test in Nano Scaled Technologies", Springer, 2008

[4] Ching-Te Chuang, Saibal Mukhopadhyay, Jae-Joon Kim, Keunwoo Kim, Rahul Rao, "High-performance SRAM in nanoscale CMOS: Design challenges and techniques," mtdt, pp.4-12, 2007 *IEEE International Workshop on Memory Technology, Design and Testing*, 2007

[5] H. Yamauchi et al., "A Differential Cell Terminal Biasing Scheme Enabling A Stable Write Operation Against A Large Random Threshold Voltage (Vth) Variation", *IEICE EC* , Volume 11, 2006

[6] Chung-Kuan Tsai, Marek-Sadowska, M., "Analysis of process variation's effect on SRAM's read stability," *Quality Electronic Design, 2006. ISQED '06. 7th International Symposium on* , pp.8 pp.-610, 27-29 March 2006

[7] H.Qin, Y.Cao, D.Markovic, A.Vladimirescu, J. Rabaey, "Standby supply voltage minimization for deep sub-micron SRAM", *Microelectronics Journal*, vol 36, pp. 789-800, March, 2005.

[8] Ding-Ming Kwai, "Review of 6t SRAM cell" Intellectual Property Library Company, June 3, 2005

[9] Sui Huang, Zhangcai Huang, Kurokawa, A., Inoue, Y., "A novel SRAM structure for leakage power suppression in 45nm technology," *Communications, Circuits and Systems, 2008. ICCCAS 2008. International Conference on*, pp.1070-1074, 25-27 May 2008

[10] Hook, T.B., Breitwisch, M., Brown, J., Cottrell, P., Hoyniak, D., Chung Lam, Mann, R., "Noise margin and leakage in ultra-low leakage SRAM cell design," *Electron Devices, IEEE Transactions on* , vol.49, no.8, pp. 1499-1501, Aug 2002

[11] Fu-Liang Yang, Jiunn-Ren Hwang, Yiming Li, "Electrical Characteristic Fluctuations in Sub-45nm CMOS Devices," *Custom Integrated Circuits Conference, 2006. CICC '06. IEEE* , pp.691-694, 10-13 Sept. 2006

[12] Hamzaoglu, F., Kevin Zhang, Yin Wang, Ann, H.J., Bhattacharya, U., Zhanping Chen, Yong-Gee Ng, Pavlov, A., Smits, K., Bohr, M., "A 153Mb-SRAM Design with Dynamic Stability Enhancement and Leakage Reduction in 45nm High-K Metal-Gate

CMOS Technology," *Solid-State Circuits Conference, 2008. ISSCC 2008. Digest of Technical Papers. IEEE International* , pp.376-621, 3-7 Feb. 2008

[13] Khellah, M., Somasekhar, D., Ye, Y., Kim, N. S., Howard, J., Ruhl, G., Sunna, M., Tschanz, J., Borkar, N., Hamzaoglu, F., Pandya, G., Farhang, A., Zhang, K., De, V., "A 256-Kb Dual-VCC SRAM Building Block in 65-nm CMOS Process With Actively Clamped Sleep Transistor," *Solid-State Circuits, IEEE Journal of* , vol.42, no.1, pp.233-242, Jan. 2007

[14] Tran, H., "Demonstration of 5T SRAM and 6T dual-port RAM cell arrays," *VLSI Circuits, 1996. Digest of Technical Papers., 1996 Symposium on*, pp.68-69, 13-15 Jun. 1996

[15] Carlson, I., Andersson, S., Natarajan, S., Alvandpour, A., "A high density, low leakage, 5T SRAM for embedded caches," *Solid-State Circuits Conference*, 2004. ESSCIRC 2004. Proceeding of the 30th European, pp. 215- 218, 21-23 Sept. 2004

[16] Seevinck, E., List, F.J., Lohstroh, J., "Static-noise margin analysis of MOS SRAM cells," *Solid-State Circuits, IEEE Journal of* , vol.22, no.5, pp. 748- 754, Oct. 1987

[17] Romanovsky, S., Achyuthan, A., Natarajan, S., Wing Leung, "Leakage reduction techniques in a 0.13 um SRAM cell," *VLSI Design, 2004. Proceedings. 17th International Conference on*, pp. 215- 221, 2004

[18] Hobson, R.F., "A New Single-Ended SRAM Cell With Write-Assist," *Very Large Scale Integration (VLSI) Systems, IEEE Transactions on* , vol.15, no.2, pp.173-181, Feb. 2007

[19] Lysinger, M., Roche, P., Zamanian, M., Jacquet, F., Sahoo, N., McClure, D., Russell, J., "A Radiation Hardened Nano-Power 8Mb SRAM in 130nm CMOS," *Quality Electronic Design, 2008. ISQED 2008. 9th International Symposium on*, pp.23-29, 17-19, Mar. 2008

[20] Fallah, F., Pedram, M., "Standby and Active Leakage Current Control and Minimization in CMOS VLSI Circuits," *IEICE Trans. on Electronics, Special Section on Low-Power LSI and Low-Power IP,* Vol. E88-C, No. 4 Apr. 2005, pp. 509-519.

[21] Ishida, M., Kawakami, T., Tsuji, A., Kawamoto, N., Motoyoshi, M., Ouchi, N., "A novel 6T-SRAM cell technology designed with rectangular patterns scalable beyond 0.18 μm generation and desirable for ultra high speed operation," *Electron Devices Meeting, 1998. IEDM '98 Technical Digest., International*, pp.201-204, 6-9 Dec. 1998

[22] Zheng Guo, Carlson, A., Liang-Teck Pang, Duong, K., Tsu-Jae King Liu, Nikolic, B., "Large-scale read/write margin measurement in 45nm CMOS SRAM arrays," *VLSI Circuits, 2008 IEEE Symposium on* , pp.42-43, 18-20, Jun. 2008

[23] Wieckowski, M., Patil, S., Margala, M., "Portless SRAM—A High-Performance Alternative to the 6T Methodology," *Solid-State Circuits, IEEE Journal of* , vol.42, no.11, pp.2600-2610, Nov. 2007