# 博士論文

# The Effects of Variability on Write Stability in SRAM at Low Supply Voltage

(低電圧 SRAM における特性ばらつきの書き込み安定性への影響)

平成 28 年 06 月 01 日提出
指導教員　平本　俊郎教授

東京大学大学院工学系研究科
電気系工学専攻
37-137264
**邱　浩**

# Abstract

Static random access memory (SRAM) acts as the buffer role in pyramid-like memory hierarchy to compensate the speed gap between processors and bottom-level memories. Targeting a larger capacity of SRAM arrays with higher performance and lower cost, both designers and manufacturers are driving efforts to minimize the footprint of SRAM cells. Also, both active energy and leakage power considerations make operating voltage scaling significantly compelling for SRAM. However, continued increase in variability consisting of time-zero and time-dependent variability is perceived to be a major roadblock for future operating voltage scaling. Thus, variability analysis in SRAM becomes critical for both gaining a deeper understanding of the sources of variability and for developing more robust circuits.

By adopting the intrinsic channel, silicon-on-Thin-BOX (SOTB) – in other words, fully-depleted (FD) silicon-on-insulator (SOI) – technology eliminates large time-zero variability from random dopant fluctuations (RDF) in CMOS bulk one. The immunity to RDF also helps suppress the impact of random telegraph noise (RTN). The big innovation facilitates the experimental demonstration of low-power SOTB SRAM cells operable down to sub-0.4 V regime. Considering the limited data to date, this work presents a comprehensive variability analysis on write stability in SRAM at low $V_{DD}$ based on SOTB technology platform.

Firstly, four commonly used write stability metrics – including write static noise margin (WSNM) from write butterfly curve, $I_W$ from write N-curve, bit-line margin (BLM) from bit-line method and combined word-line margin (CWLM) from word-line method – are compared in order to select the good candidate for write yield estimation at low $V_{DD}$. The core standard is that the selected one follows good normality and can correctly predict write failure. Bit-line method and word-line method are concluded as good candidates for write yield estimation at low $V_{DD}$. On the other hand, the non-normality of WSNM and $I_W$ is clarified and ascribed to sub-$V_{th}$ operation of cell transistors at low $V_{DD}$. HSPICE simulation results help extend our conclusions up to ±6 sigma.

Besides, a new write stability metric is proposed for write yield estimation. The extended write butterfly curve extends the voltage sweeping range of conventional write butterfly curve. Due to the clearer emergence of failure mode, the extended write noise margin (E-WSNM) shows good normality and is demonstrated as a good metric for write yield estimation. More evidence is also given to support the newly proposed one.

Lastly, a statistical model is developed to evaluate the impact of time-dependent RTN in SRAM at low $V_{DD}$. $I_W$ from write N-curve is selected as the write stability metric due to its being current-based one. Based on the distribution fitting of both $I_W$ and RTN-induced fluctuation ($\delta I_W$), the degradation due to RTN on fail bit rate (FBR) is discussed. It is found that RTN degrades $V_{min}$ – the minimum voltage which guarantees stability of the whole capacity of SRAM arrays – over 10 % in sub-0.4 V regime, thus emphasizing the importance of RTN for low-power SRAM design.

Overall, our conclusions are not limited to SOTB technology but are applicable to other technologies such as FinFET SRAM, and give implications to SRAM design at low $V_{DD}$.

# Contents

List of Figures ..... iii
Chapter 1 Background ..... 1
1.1 SRAM Scaling and Challenging Issues ..... 1
1.1.1 Memory Hierarchy and SRAM Block Structure ..... 1
1.1.2 SRAM Cell and Operating Voltage Scaling ..... 3
1.1.3 Variability-limited Scaling ..... 4
1.2 SOTB Technology for Low-power SRAM ..... 6
1.2.1 SOTB versus Bulk ..... 6
1.2.2 Better Control in SOTB Transistors ..... 7
1.2.3 Sub-0.4V Operation in SOTB SRAM ..... 9
1.3 Research Goal ..... 10
1.4 Dissertation Outline ..... 11
Chapter 2 Write Stability Characterization and Time-zero Variability .. 13
2.1 Introduction ..... 13
2.1.1 Commonly Used Write Stability Metrics ..... 13
2.1.2 Contemporary Works ..... 18
2.2 Measurement Results in SOTB SRAM ..... 23
2.2.1 Test Structure ..... 23
2.2.2 Write Stability Characterization using Four Metrics ..... 24
2.2.3 Further Discussions about Unpreferred Metrics ..... 28
2.3 HSPICE Simulation Results in SOTB SRAM ..... 41
2.4 Measurement Results in Bulk SRAM ..... 45

2.4.1 Write Stability Characterization and Four Metrics' Comparison ..... 45
2.4.2 Comparison with SOTB SRAM ..... 48
2.4.3 Further Measurement in New Bulk Chip ..... 53
2.5 Summary ..... 56
Chapter 3 Proposed New Write Stability Metric for Yield Estimation .. 58
3.1 Introduction ..... 58
3.1.1 Drawback of Conventional Write Butterfly Curve ..... 58
3.1.2 Definition of Extended Write Butterfly Curve ..... 60
3.2 Measurement Results in SOTB SRAM ..... 60
3.2.1 Characterization using Extended Write Butterfly Curve ..... 60
3.2.2 Comparison with Conventional Write Butterfly Curve ..... 62
3.3 HSPICE Simulation Results in SOTB SRAM ..... 64
3.4 Summary ..... 64
Chapter 4 Time-dependent Random Telegraph Noise in SRAM ..... 66
4.1 Introduction ..... 66
4.1.1 Dynamics of Random Telegraph Noise ..... 66
4.1.2 Random Telegraph Noise in Transistors ..... 72
4.1.3 Random Telegraph Noise in SRAM ..... 73
4.1.4 Contemporary Work ..... 78
4.2 Measurement and Modeling Results in SOTB SRAM ..... 81
4.2.1 Write N-curve as Write Stability Metric ..... 81
4.2.2 Random Telegraph Noise Measurement in SRAM and Cell Transistors ..... 83
4.2.3 Statistical Distributions and Implications to SRAM Robustness ..... 85
4.3 Summary ..... 89
Chapter 5 Conclusion ..... 91
5.1 Key Parts ..... 91
5.2 Future Work ..... 93
Bibliography ..... 95
Publication List ..... 107

# List of Figures

**Fig. 1.1.** Memory hierarchy of a personal computer [2].

**Fig. 1.2.** SRAM block diagram [2].

**Fig. 1.3.** (a) Area trends with scaling [4]. (b) Size trends with scaling: SRAM cell area vs. contacted gate pitch [5].

**Fig. 1.4.** SRAM voltage design margin trends with scaling under the impact of variability from different sources [26].

**Fig. 1.5.** Schematic of (a) bulk and (b) SOTB transistor [30].

**Fig. 1.6.** Schematic of SOTB NMOS and PMOS with specified parameters [31-32].

**Fig. 1.7.** (a) Comparison of junction capacitance in bulk and SOTB NMOS/ PMOS. (b) Demonstration of $V_{th}$ adjustment by body biasing in SOTB NMOS/ PMOS. From Ref. [32].

**Fig. 1.8.** (a) $V_{th}$ distribution in both 1 M SOTB and bulk transistors. (b) Across-wafer $V_{th}$ distribution in SOTB (w/ and w/o optimized process) and bulk NMOS. From Ref. [32].

**Fig. 1.9.** RTN-induced $\Delta V_{th}$ distribution in 16 k SOTB and bulk transistors [34].

**Fig. 1.10.** (a) Layout of 6-T SOTB SRAM cell [31]. (b) Fail bit count in 2 Mb SOTB and bulk SRAM cells at different $V_{DD}$ [32].

**Fig. 2.1.** (a) Schematic of 6-T SRAM cell with swept terminals indicated by red arrows. (b) Write butterfly curves of one SRAM cell (in black) which passes and the other one (in red) which fails. Take the stable SRAM cell as an example, its write noise margin – WSNM – is

defined by grey double arrow. Voltage is normalized to $V_{DD}$. Substrate bias: $V_{bsn}$ = -1 V, $V_{bsp}$ = 1 V.

**Fig. 2.2.** (a) Schematic of 6-T SRAM cell with swept node indicated by red arrow. (b) Read N-curve of one SRAM cell. And its write ability – WTI – is defined by grey double arrow. Voltage/ current is normalized. Substrate bias: $V_{bsn}$ = -1 V, $V_{bsp}$ = 1 V.

**Fig. 2.3.** (a) Schematic of 6-T SRAM cell with swept node indicated by red arrow. (b) Write N-curves of one SRAM cell (in black) which passes and the other one (in red) which fails. Take the stable SRAM cell as an example, its write noise margin – $I_W$ – is defined by grey double arrow. Voltage/ current is normalized. Substrate bias: $V_{bsn}$ = -1 V, $V_{bsp}$ = 1 V.

**Fig. 2.4.** Schematic of 6-T SRAM cell with swept terminals indicated by red arrows. (b) Measured waveforms in bit-line method of one SRAM cell (in black) which passes and the other one (in red) which fails. Take the stable SRAM cell as an example, its write noise margin – BLM – is defined by grey double arrow. Voltage is normalized to $V_{DD}$. Substrate bias: $V_{bsn}$ = -1 V, $V_{bsp}$ = 1 V.

**Fig. 2.5.** Schematic of 6-T SRAM cell with swept terminals indicated by red arrows. (b) Measured waveforms in word-line method of one SRAM cell (in black) which passes and the other one (in red) which fails. Take the stable SRAM cell as an example, its write noise margin – CWLM – is defined by grey double arrow. Voltage is normalized to $V_{DD}$. Substrate bias: $V_{bsn}$ = -1 V, $V_{bsp}$ = 1 V.

**Fig. 2.6.** (a) Schematic of 6-T SRAM cell with "0" originally stored. Dependence of (a) WSNM, (b) BLM, and (c) CWLM on $\Delta V_{th}$ of all six cell transistors. (d) Dependence of WSNM/ BLM/ CWLM on $\Delta V_{th}$ of N3. (e) Distribution of WSNM/ BLM/ CWLM. All simulations are performed at $V_{DD}$ = 1.0 V. From Ref. [41].

**Fig. 2.7.** Die photo of the 45 nm bulk SRAM test chip [38].

**Fig. 2.8.** (a) Schematic of 6-T SRAM cell. Measured (a) write butterfly curves, (c) write N-curves of SRAM macro and (d) waveforms in bit-line method, (e) waveforms in word-line method of functional SRAM arrays. Voltage is normalized to $V_{DD}$. From Ref. [38].

**Fig. 2.9.** Cumulative pots of (a) $I_W$ of SRAM macro and (b) BWTV, (c) WWTV of functional SRAM arrays at $V_{DD}$ = 0.7 V. From Ref. [38].

**Fig. 2.10.** Schematic of DMA-TEG [43-44].

**Fig. 2.11.** Cumulative plots of $V_{th}$ of (a) TaL/ TaR, (b) TnL/ TnR, and (c) TpL/ TpR of 4 kb SOTB SRAM cells. Drain bias: $|V_{ds}|$ = 50 mV. Substrate bias: $V_{bsn}$ = -1 V, and $V_{bsp}$ = 1V.

**Fig. 2.12.** Measured (a) write butterfly curves, (b) write N-curves, (c) waveforms in bit-line method, and (d) waveforms in word-line method of 4 kb SOTB SRAM cells at $V_{DD}$ = 0.8 V. Substrate bias: $V_{bsn}$ = -1 V, and $V_{bsp}$ = 1V.

**Fig. 2.13.** Cumulative plots of (a) WSNM, (b) $I_W$, (c) BLM, and (d) CWLM of 4 kb SOTB SRAM cells at $V_{DD}$ = 0.8 V (in black), 0.6 V (in blue), and 0.4 V (in red). WSNM, $I_W$, BLM, CWLM are defined as the minimum of "0" and "1" write. $I_W$ at $V_{DD}$ = 0.6 V and 0.4 V have been magnified by 2 and 10 times, respectively. Substrate bias: $V_{bsn}$ = -1 V, $V_{bsp}$ = 1 V.

**Fig. 2.14.** Scatter plots of (a) BLM versus CWLM, (b) $I_W$ versus BLM, (c) $I_W$ versus CWLM, (d) WSNM versus BLM, (e) WSNM versus CWLM, and (f) WSNM versus $I_W$ in 4 kb SOTB SRAM cells at $V_{DD}$ = 0.8 V. Substrate bias: $V_{bsn}$ = - 1 V, $V_{bsp}$ = 1 V.

**Fig. 2.15.** Scatter plots of (a) BLM versus CWLM, (b) $I_W$ versus BLM, (c) $I_W$ versus CWLM, (d) WSNM versus BLM, (e) WSNM versus CWLM, and (f) WSNM versus $I_W$ in 4 kb SOTB SRAM cells at $V_{DD}$ = 0.4 V. Substrate bias: $V_{bsn}$ = - 1 V, $V_{bsp}$ = 1 V.

**Fig. 2.16.** Scatter plot of WTI from read N-curve versus $I_W$ from write N-curve of 4 kb SOTB SRAM cells at (a) $V_{DD}$ = 0.8 V and (b) $V_{DD}$ = 0.4 V. Red dashed line indicates where $I_W$ equals zero. Substrate bias: $V_{bsn}$ = - 1 V, $V_{bsp}$ = 1 V.

**Fig. 2.17.** (a) Histogram plot of "0" WSNM of 4 kb SOTB SRAM cells at $V_{DD}$ = 0.4V. Two peaks are indicated by Mode I and Mode II, respectively. (b) Write butterfly curves of 4 kb SOTB SRAM cells in "0" write at $V_{DD}$ = 0.4 V. Substrate bias: $V_{bsn}$ = -1 V, $V_{bsp}$ = 1 V.

**Fig. 2.18.** Two types of write butterfly curves: Write butterfly curves of (a) Cell-A and (b) Cell-B in "0" write at $V_{DD}$ = 0.4 V. Cell-A/ Cell-B corresponds to Mode I/ II. Substrate bias: $V_{bsn}$ = -1 V, $V_{bsp}$ = 1 V. $|I_{ds}|$-$|V_{gs}|$ curves of TaR and TpR in (c) Cell-A and (d) Cell-B at drain bias of 50 mV. Inset: The threshold voltage ($V_{thc}$) of TaR and TpR are listed.

**Fig. 2.19.** (a) $|I_{ds}|$-$|V_{gs}|$ curves of TaR (grey dashed line) and TpR in Cell-C at $V_{bsp}$ = 0 V (black), 0.5 V (orange), 1.0 V (red), 1.5 V (green), 2.0 V (blue), and 2.5 V (purple). (b) Write butterfly curves of Cell-C in "0" write at $V_{DD}$ = 0.4 V. Substrate bias: $V_{bsn}$ = -1 V; $V_{bsp}$ = 0 V (in black), 0.5 V (orange), 1.0 V (red), 1.5 V (green), 2.0 V (blue), and 2.5 V (purple).

**Fig. 2.20.** (a) Cumulative plots of TpR's $V_{thc}$ in 1 kb SOTB SRAM cells at $V_{bsp}$ = 0 V (black), 0.5 V (orange), 1.0 V (red), 1.5 V (green), 2.0 V (blue), and 2.5 V (purple). Drain bias: -50 mV. (b) Scatter plot of TpR's $V_{thc}$ versus TaR's $V_{thc}$. Drain bias: $|V_{ds}|$ = 50 mV. Substrate bias: $V_{bsn}$ = -1 V; $V_{bsp}$ = 0 V (black), 1.0 V (red), and 2.5 V (purple).

**Fig. 2.21.** Write butterfly curves of 1 kb SOTB SRAM cells in "0" write at $V_{DD}$ = 0.4 V. Substrate bias: $V_{bsn}$ = -1 V, (a) $V_{bsp}$ = 0 V, (b) $V_{bsp}$ = 0.5 V, (c) $V_{bsp}$ = 1.0 V, (d) $V_{bsp}$ = 1.5 V, (e) $V_{bsp}$ = 2.0 V, (f) $V_{bsp}$ = 2.5 V.

**Fig. 2.22.** Histogram plots of 1 kb SOTB SRAM cells in "0" write at $V_{DD}$ = 0.4 V. Substrate bias: $V_{bsn}$ = -1 V, (a) $V_{bsp}$ = 0 V, (b) $V_{bsp}$ = 0.5 V, (c) $V_{bsp}$ = 1.0 V, (d) $V_{bsp}$ = 1.5 V, (e) $V_{bsp}$ = 2.0 V, (f) $V_{bsp}$ = 2.5 V.

**Fig. 2.23.** (a) Write N-curves of 4 kb SOTB SRAM cells in "0" write at $V_{DD}$ = 0.4 V. (b) Histogram plot of "0" $I_W$ of 4 kb SOTB SRAM cells at $V_{DD}$ = 0.4 V. The black curve indicates fitting result of normal distribution function. Substrate bias: $V_{bsn}$ = -1 V, $V_{bsp}$ = 1 V.

**Fig. 2.24.** Two types of write N-curves: Write N-curves of Cell-A and Cell-B (the same as in Fig. 2.20) in "0" write at $V_{DD}$ = 0.4 V. Cell-A/ Cell-B corresponds to Mode i/ Mode ii. The position where "0" $I_W$ is extracted is indicated by red arrow. Substrate bias: $V_{bsn}$ = -1 V, $V_{bsp}$ = 1 V.

**Fig. 2.25.** Measured currents in TaR (in blue), TnR (in green) and TpR (in red) contributing to write N-curve (in black) of Cell-C in "0" write at $V_{DD}$ = 0.4 V. Substrate bias: $V_{bsn}$ = -1 V, (a) $V_{bsp}$ = 0 V, (b) $V_{bsp}$ = 0.5 V, (c) $V_{bsp}$ = 1.0 V, (d) $V_{bsp}$ = 1.5 V, (e) $V_{bsp}$ = 2.0 V, (f) $V_{bsp}$ = 2.5 V. Here, the positive direction of current is defined as flowing outside node VR.

**Fig. 2.26.** Write N-curves of 1 kb SOTB SRAM cells in "0" write at $V_{DD}$ = 0.4 V. Substrate bias: $V_{bsn}$ = -1 V, (a) $V_{bsp}$ = 0 V, (b) $V_{bsp}$ = 0.5 V, (c) $V_{bsp}$ = 1.0 V, (d) $V_{bsp}$ = 1.5 V, (e) $V_{bsp}$ = 2.0 V, (f) $V_{bsp}$ = 2.5 V.

**Fig. 2.27.** Histogram plots of "0" $I_W$ of 1 kb SOTB SRAM cells in "0" write at $V_{DD}$ = 0.4 V. Substrate bias: $V_{bsn}$ = -1 V, (a) $V_{bsp}$ = 0 V, (b) $V_{bsp}$ = 0.5 V, (c) $V_{bsp}$ = 1.0 V, (d) $V_{bsp}$ = 1.5 V, (e) $V_{bsp}$ = 2.0 V, (f) $V_{bsp}$ = 2.5 V.

**Fig. 2.28.** $V_{th}$ is changed from -6 sigma to 6 sigma in nFETs/ pFETs.

**Fig. 2.29.** Dependence of "0" WSNM on $\Delta V_{th}$ of six cell transistors at $V_{DD}$ = (a) 0.8 V, (b) 0.6 V, and (c) 0.4 V.

**Fig. 2.30.** Dependence of "0" $I_W$ on $\Delta V_{th}$ of six cell transistors at $V_{DD}$ = (a) 0.8 V, (b) 0.6 V, and (c) 0.4 V.

**Fig. 2.31.** Dependence of "0" BLM on $\Delta V_{th}$ of six cell transistors at $V_{DD}$ = (a) 0.8 V, (b) 0.6 V, and (c) 0.4 V.

**Fig. 2.32.** Dependence of "0" CWLM on $\Delta V_{th}$ of six cell transistors at $V_{DD}$ = (a) 0.8 V, (b) 0.6 V, and (c) 0.4 V.

**Fig. 2.33.** Dependence of (a) "0" WSNM, (b) "0" $I_W$, (c) "0" BLM, and (d) "0" CWLM on

$\Delta V_{th}$ of six cell transistors at $V_{DD}$ = 0.4 V.

**Fig. 2.34.** Cumulative plots of $V_{th}$ of (a) TaL/ TaR, (b) TnL/ TnR, and (c) TpL/ TpR of 4 kb bulk SRAM cells. Drain bias: $|V_{ds}|$ = 50 mV. Substrate bias: $V_{bsn}$ = -0.8 V, and $V_{bsp}$ = 0.25 V.

**Fig. 2.35.** Measured (a) write butterfly curves, (b) write N-curves, (c) waveforms in bit-line method, and (d) waveforms in word-line method of 4 kb bulk SRAM cells at $V_{DD}$ = 1.0 V. Substrate bias: $V_{bsn}$ = -0.8 V, and $V_{bsp}$ = 0.25 V.

**Fig. 2.36.** Cumulative plots of (a) WSNM, (b) $I_W$, (c) BLM, and (d) CWLM of 4 kb bulk SRAM cells at $V_{DD}$ = 1.0 V (in black), 0.8 V (in blue), and 0.6 V (in red). WSNM, $I_W$, BLM, CWLM are defined as the minimum of "0" and "1" write. Substrate bias: $V_{bsn}$ = -0.8 V, $V_{bsp}$ = 0.25 V.

**Fig. 2.37.** (a) Cumulative plots of TpR's $V_{thc}$ in 1 kb bulk SRAM cells at $V_{bsp}$ = -0.5 V (black), -0.25 V (orange), 0 V (red), 0.25 V (green), 0.5 V (blue), and 0.75 V (purple). Drain bias: -50 mV. (b) Scatter plot of TpR's $V_{thc}$ versus TaR's $V_{thc}$. Drain bias: $|V_{ds}|$ = 50 mV. Substrate bias: $V_{bsn}$ = -0.8 V; $V_{bsp}$ = -0.5 V (black), 0.25 V (red), and 0.75 V (purple).

**Fig. 2.38.** Write butterfly curves of 1 kb bulk SRAM cells in "0" write at $V_{DD}$ = 0.6 V. Substrate bias: $V_{bsn}$ = -0.8 V, (a) $V_{bsp}$ = -0.5 V, (b) $V_{bsp}$ = -0.25 V, (c) $V_{bsp}$ = 0 V, (d) $V_{bsp}$ = 0.25 V, (e) $V_{bsp}$ = 0.5 V, (f) $V_{bsp}$ = 0.75 V.

**Fig. 2.39.** Histogram plots of 1 kb bulk SRAM cells in "0" write at $V_{DD}$ = 0.6 V. Substrate bias: $V_{bsn}$ = -0.8 V, (a) $V_{bsp}$ = -0.5 V, (b) $V_{bsp}$ = -0.25 V, (c) $V_{bsp}$ = 0 V, (d) $V_{bsp}$ = 0.25 V, (e) $V_{bsp}$ = 0.5 V, (f) $V_{bsp}$ = 0.75 V.

**Fig. 2.40.** Write N-curves of 1 kb bulk SRAM cells in "0" write at $V_{DD}$ = 0.6 V. Substrate bias: $V_{bsn}$ = -0.8 V, (a) $V_{bsp}$ = -0.5 V, (b) $V_{bsp}$ = -0.25 V, (c) $V_{bsp}$ = 0 V, (d) $V_{bsp}$ = 0.25 V, (e) $V_{bsp}$ = 0.5 V, (f) $V_{bsp}$ = 0.75 V.

**Fig. 2.41.** Histogram plots of "0" $I_W$ of 1 kb bulk SRAM cells in "0" write at $V_{DD}$ = 0.6 V.

Substrate bias: $V_{bsn}$ = -0.8 V, (a) $V_{bsp}$ = -0.5 V, (b) $V_{bsp}$ = -0.25 V, (c) $V_{bsp}$ = 0 V, (d) $V_{bsp}$ = 0.25 V, (e) $V_{bsp}$ = 0.5 V, (f) $V_{bsp}$ = 0.75 V.

**Fig. 2.42.** Cumulative plots of (a) WSNM, (b) $I_W$, (c) BLM, and (d) CWLM in 1 kb bulk SRAM cells at $V_{DD}$ = 0.9 V. Write noise margin is the minimum between "0" and "1" write.

**Fig. 2.43.** Scatter plots of (a) BLM versus CWLM, (b) $I_W$ versus BLM, (c) $I_W$ versus CWLM, (d) WSNM versus BLM, (e) WSNM versus CWLM, and (f) WSNM versus $I_W$ of 1 kb bulk SRAM cells at $V_{DD}$ = 0.9 V. The red circle indicates the SRAM cell with smallest write noise margin.

**Fig. 3.1.** (a) Schematic of 6-T SRAM cell. (b) Scatter plots of WSNM versus $\Delta V_{th}$ of N3 at $V_{DD}$ = 1.0 V and 0.75 V. From Ref. [49].

**Fig. 3.2.** Cumulative plots of WSNM in 1 kb SOTB SRAM cells at $V_{DD}$ = 0.8 V (black), 0.6 V (blue), and 0.4 V (red). Black and blue circles correspond to Mode I. Green circles indicate WSNM of 4 kb cells at $V_{DD}$ = 0.6 V. Substrate bias: $V_{bsn}$ = -1 V, $V_{bsp}$ = 1 V.

**Fig. 3.3.** Conventional and extended BC of one SRAM cell at $V_{DD}$ = 0.4 V. WSNM/ E-WSNM is extracted as the side of black/ red square. Substrate bias: $V_{bsn}$ = -1 V, $V_{bsp}$ = 1 V.

**Fig. 3.4.** Extended BCs of 1 kb SOTB SRAM cells at $V_{DD}$ = (a) 0.8V, (b) 0.6V, and (c) 0.4 V. The black dashed lines indicate voltage sweep range of conventional BCs. Substrate bias: $V_{bsn}$ = -1 V, $V_{bsp}$ = 1 V.

**Fig. 3.5.** (a) Cumulative plots of "0" E-WSNM in 1 kb SOTB SRAM cells at $V_{DD}$ = 0.8 V (black), 0.6 V (blue), and 0.4 V (red). (b) Mode-I and Mode-II extended BCs of two cells at $V_{DD}$ = 0.6 V. WSNM/ E-WSNM is extracted from black/ red square. Substrate bias: $V_{bsn}$ = -1 V, $V_{bsp}$ = 1 V.

**Fig. 3.6.** Scatter plots of "0" E-WSNM versus "0" WSNM in 1 kb SOTB SRAM cells at $V_{DD}$ = 0.8 V (black), 0.6 V (blue) and 0.4 V (red). Substrate bias: $V_{bsn}$ = -1 V, $V_{bsp}$ = 1 V.

**Fig. 3.7.** Scatter plot of "0" E-WSNM (red)/ "0" WSNM (black) versus $V_{th}$ of TaR in 1 kb SOTB SRAM cells at $V_{DD}$ = 0.4 V. Drain bias for $V_{th}$ measurement is 50 mV. Substrate bias: $V_{bsn}$ = -1 V, $V_{bsp}$ = 1 V.

**Fig. 3.8.** Measured waveforms of word-line method in 1 kb SOTB SRAM cells at $V_{DD}$ = 0.4 V. CWLM is defined as the difference between $V_{DD}$ and $V_{WL}$ at which VR flips. Substrate bias: $V_{bsn}$ = -1 V, $V_{bsp}$ = 1 V.

**Fig. 3.9.** Scatter plot of "0" E-WSNM (red)/ "0" WSNM (black) versus "0" CWLM in 1 kb SOTB SRAM cells at $V_{DD}$ = 0.4 V. Substrate bias: $V_{bsn}$ = -1 V, $V_{bsp}$ = 1 V.

**Fig. 3.10.** (a) Dependence of "0" E-WSNM on $\Delta V_{th}$ in six cell transistors at $V_{DD}$ = 0.4 V. (b) Dependence of "0" E-WSNM (in red)/ "0" WSNM (in black) on $\Delta V_{th}$ in TaR at $V_{DD}$ = 0.4 V.

**Fig. 4.1.** Measured $I_d$ waveform in transistor along with time [51]. Three RTN-related parameters are defined as inset.

**Fig. 4.2.** One example of single-trap induced current spectral density [54].

**Fig. 4.3.** RTN-induced $\Delta V_{th}$ from (a) two traps and (b) multiple traps [55].

**Fig. 4.4.** One example of multiple-trap induced current spectral density [56].

**Fig. 4.5.** Distributions of RTN-induced and RDF-induced $V_{th}$ variations. Projected RTN-induced $V_{th}$ variations exceed RDF-induced $V_{th}$ variations at the ~ 2 sigma level in 22 nm generation [70].

**Fig. 4.6.** (a) $L_g$ dependence, (b) $W_g$ dependence, and (c) size dependence of RTN-induced $\Delta V_{th}$ at 1 or 2 sigma [70].

**Fig. 4.7.** Projected $V_{th}$ variation assuming intrinsic channel transistors [74].

**Fig. 4.8.** Extracted gate voltage dependence of $\Delta I_d/I_d$ in NFET and PFET [52].

**Fig. 4.9.** Estimated $V_{th}$ window to guarantee the stable read/ write operation in 65 nm SRAM cells [55].

**Fig. 4.10.** Measured $\Delta V_{th}$ distributions of (a) nMOS and (b) pMOS at different $V_{gs}$ in 2 kb SRAM cells. Simulated results in (c) nMOS and (d) pMOS match the measured results well. (e) Simulated read N-curves and ten of them are selected for demonstration. $I_{READ}$ is defined as the read stability metric. (f) Cumulative plots of measured (in black) and simulated (in red) $\Delta V_{min}/V_{DD}$. From Ref. [77].

**Fig. 4.11.** (a) Accelerated test results (hollow circle) and simulation results (solid circle). Each bit suffers from 32 disturbs per cycle (i.e. entire 512kbit read). Simulated noise margin degradation by RTN in (b) 40 nm and (c) 22 nm technology. From Ref. [78].

**Fig. 4.12.** (a) Measured fail probability transition in SRAM at $V_{DD}$ = 0.9 V. In the schematic, ΔVCS is defined by difference between pass and fail voltage. (b) ΔVCS distribution at different $V_{DD}$. (c) Calculated guard-band voltage in scaled SRAM assuming 1x and 3x trap density. From Ref. [79].

**Fig. 4.13.** (a) Histogram plot of $I_{write}$ and (b) Gumbel plot of RTN-induced $\delta I_{write}$ in 160 bulk SRAM cells at nominal $V_{DD}$. (c) Joint PDF plot of $\delta I_{write}$ versus $I_{write}$ at nominal $V_{DD}$. (c) FBR of SRAM at different $V_{DD}$. From Ref. [80].

**Fig. 4.14.** Measured FBC in (a) bulk and (b) FDSOI SRAM at different $V_{DD}$ and the impact of RTN is evaluated [81].

**Fig. 4.15.** Schematic of 6-T SRAM cell with $I_{VR}$ (black arrow), $I_{TaR}$ (red arrow), $I_{TpR}$ (blue arrow), and $I_{TnR}$ (green arrow) indicated during the measurement of write N-curve.

**Fig. 4.16.** Measured write N-curves in 4 kb SOTB SRAM cells at $V_{DD}$ = (a) 0.36 V, (b) 0.34 V, (c) 0.32 V, (d) 0.30 V, (e) 0.28 V, and (f) 0.26 V.

**Fig. 4.17.** Cumulative plots of measured $I_W$ of write N-curves in 4 kb SOTB SRAM cells at different $V_{DD}$ from 0.36 down to 0.26V.

**Fig. 4.18.** Histogram plot of $I_W$ in 32 kb SOTB SRAM cells at $V_{DD}$ = 0.30 V. The normal

distribution fit (black curve) shows a large deviation in tail region. (b) Quantile-quantile plot of $I_W$ with the generalized normal distribution fit (grey dashed line). $I_W$ is normalized by its standard deviation $\sigma_1$. (c) Calculated PDF plot of $I_W$ with failure region enlarged. Failure edge is defined as $I_W = 0$.

**Fig. 4.19.** (a) Quantile-quantile plots of $I_W$ in 4 kb SOTB SRAM cells at different $V_{DD}$ with generalized normal distribution fit (grey dashed line). $I_W$ is normalized by its standard deviation $\sigma_1$. (b) Calculated FBR at different $V_{DD}$.

**Fig. 4.20.** Cumulative plots of measured $\delta I_W$ and $\delta I$ of cell transistors in 32 kb SOTB SRAM cells at $V_{DD} = 0.30$ V.

**Fig. 4.21.** Measured scatter plot of $\delta I_W$ versus $I_W$ in 32 kb SOTB SRAM cells at $V_{DD} = 0.30$ V. Failure edge (red dashed line) is defined as: $I_W - \delta I_W = 0$.

**Fig. 4.22.** Measured $I_W$, as well as its two components ($I_{TaR}$ and $I_{TpR}$), at $V_{DD} = 0.26$ V along with time. The RTN-induced failure is demonstrated.

**Fig. 4.23.** Histogram plot of $\delta I_W$ in 32 kb SOTB SRAM cells at $V_{DD} = 0.30$ V. The log-normal distribution fit (black curve) shows a large deviation in tail region. (b) Quantile-quantile plot of $\delta I_W$ with lognormal-generalized pareto distribution fit (grey dashed line). $\delta I_W$ is normalized by its standard deviation $\sigma_2$. (c) Calculated PDF plot of $\delta I_W$.

**Fig. 4.24.** Joint PDF plot of $\delta I_W$ versus $I_W$ at $V_{DD} = 0.30$ V. Both $I_W$ and $\delta I_W$ are normalized by their standard deviations. The red dashed line and grey patterned area denote failure edge and failure region, respectively. The MPFP indicates where the cell fails in the largest probability.

**Fig. 4.25.** Cumulative plots of measured $\delta I_W$ in 4 kb SOTB SRAM cells at different $V_{DD}$ from 0.36 down to 0.26V. (b) Quantile-quantile plots of $\delta I_W$ with lognormal-generalized pareto distribution fit (grey dashed line). $\delta I_W$ is normalized by its standard deviation $\sigma_2$.

**Fig. 4.26.** Joint PDF plots of $\delta I_W$ versus $I_W$ at $V_{DD}$ = (a) 0.36 V, (b) 0.34 V, (c) 0.32 V, (d) 0.30 V, (e) 0.28 V, and (f) 0.26 V.

**Fig. 4.27.** FBR of SRAM at different $V_{DD}$. Linear $V_{DD}$ dependence is proposed both with and without RTN.

# Acknowledgments

I would like to begin by expressing my sincere gratitude to my thesis advisor – Prof. Toshiro Hiramoto – for his patient, yet focused, guidance throughout the course of this thesis. His broad, yet profound, knowledge of the issues and challenges faced by the low voltage CMOS devices and random variability in scaled transistors has paved ways for the breadths of projects conducted within our research group. I am particularly impressed with the level of fairness and integrity he brings to work and hope to take that with me in my future career life.

I would also like to thank Prof. Masaharu Kobayashi, for his gracious support as my co-advisor. I greatly appreciate his careful revisions during the preparations of a number of conference and journal publications. In addition, I am pleased to acknowledge many of his good suggestions to help pave my career path.

My research has been supported by The Japan Society for the Promotion of Science (JSPS) Fellowship. The chip fabrication donations were collaborated with Low-Power Electronics Association & Project (LEAP). I also acknowledge the support from The Ministry of Economy, Trade and Industry (METI) and The New Energy and Industrial Technology Development Organization (NEDO) in Japan.

I am grateful for International Multidisciplinary Engineering (IME) Graduate Program and all the help and support from the staff at the Electrical Engineering and Information Systems (EEIS) graduate office at The University of Tokyo.

I would also like to acknowledge many current and former colleagues in Hiramoto & Kobayashi Laboratory. In particular, I would like to thank Dr. Takuya Saraya for his patient and timely technical support; Ms. Tomoko Mizutani for her patient and gracious support with the lab large-scale measurement equipments; Dr. Kiyoshi Takeuchi for his instructive suggestions during the preparations of a number of conference and journal publications; Dr.

Alias Nurul Ezaila and Dr. Seung-Min Jung for their kind help at the beginning of this thesis; Mr. Akitsugu Ueda for his kind help of the basic electrical measurement guidance and Mr. Mitsuaki Nagao as my tutor to help adapt to Japan life; Ms. Miki Shiotani and Ms. Fumiko Oshita for their patient help on many administrative matters; all kinds of assistance from Dr. Waichi Nagashiro, Dr. Kazuo Ito, Dr. Toshihiko Takakura, Dr. Munetoshi Fukui, Dr. Shinichi Suzuki, Mr. Ki-Hyun Jang, Mr. Kyung-Min Jang, Mr. Nozomu Ueyama, Mr. Misumi Kawakami, Mr. Yuki Komine, Mr. Daiki Ueda, Mr. Yukihiro Tsuda, Mr. Shinta Terasawa, Mr. Katsuhisa Tanaka, Mr. Yuma Tanahashi, Mr. Hitoshi Ohno.

Finally, and certainly not the least, I would like to express my heartfelt appreciation for my parents, my wife – Xinyi Wang – and her entire family for their constant prayer and support throughout the hardest and darkest durations of my dissertation work.

# Chapter 1
# Background

## 1.1 SRAM Scaling and Challenging Issues

### 1.1.1 Memory Hierarchy and SRAM Block Structure

Memory has been the driving force behind the rapid development of CMOS technology to meet the increasing demand for higher performance and lower power consumption in many different system applications. However, along with advances in processor technology, the speed gap between processors and memories has become intolerably large [1] and makes it necessary to introduce a memory hierarchy into the processor architecture, shown in Fig. 1.1 [2]. The pyramid-like hierarchy covers various kinds of memory ranging from the large capacity but off-chip memory on bottom level to small capacity but fast on-chip memory on top level to approximate the ideal memory behavior.

The advantages of adopting memory hierarchy include reduced cost, improved performance and lower power consumption [3]. The memory hierarchy differentiates data with various levels of access frequency in different storage media. By storing infrequently accessed data in bottom-level memory which is always less expansive, the cost of overall system can be reduced. Besides, by embedding the top-level memory inside a chip, the processor only needs to process the active working set inside the embedded memory. This greatly reduces the time in which the processor would otherwise wait for the required data to

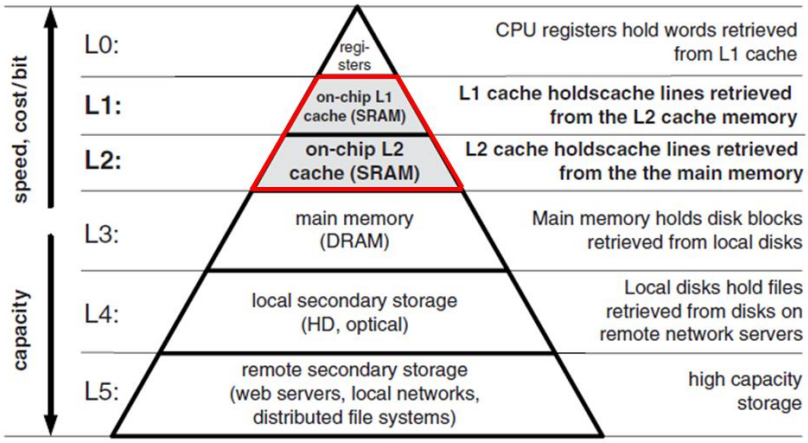


Fig. 1.1. Memory hierarchy of a personal computer [2].

be available from the bottom-level memory. Lastly, accessing the off-chip memory consumes more power than accessing on-chip memory due to larger parasitic capacitance of off-chip wires. Adopting the memory hierarchy is advantageous to reduce the number of off-chip memory transactions, thus reducing power consumption.

Among embedded memories, six-transistor (6-T) based static random access memory (SRAM) plays an important role in nearly all VLSI systems due to its superior access speed and compatibility with logic process technology, compared to other candidates. Fig. 1.2 shows an example of the basic SRAM block structure [2], which consists of SRAM core and peripheral circuits, e.g., sense amplifiers with the corresponding pre-charge and equalization circuits, write drivers, and row/column decoders. The SRAM core is commonly organized as a number of arrays of N×M×Z, where N/M/Z is the number of rows/bits/blocks. With X-, Y-, and Z-decoder, each SRAM cell can be accessed, shown inset in Fig. 1.2.

The basic storage element of an SRAM consists of the pair of inverters and access transistors connected with bit lines and word lines. The pair of inverters is cross-coupled such that the input of one is just the output of the other. As a result, either logic “0” or logic “1” state can be held as long as the SRAM cell is powered up. This is different from the requisite periodically refreshing procedure in dynamic random access memory (DRAM) and enables

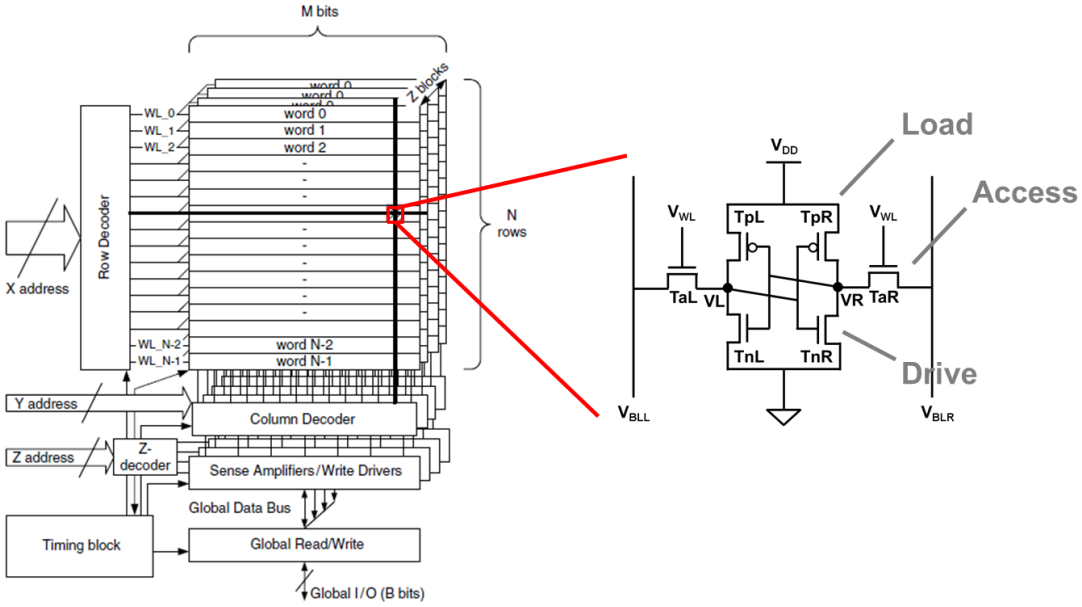


Fig. 1.2. SRAM block diagram [2].

SRAM a superior access speed.

### 1.1.2 SRAM Cell and Operating Voltage Scaling

SRAM occupies a significantly large segment of modern system-on-chips (SoCs). For example, the total percentage of occupied SRAM area of overall chip is estimated to reach over 70% in the near future, shown in Fig. 1.3(a) [4]. In order to incorporate large capacity of SRAM arrays into a chip to balance the requirements of boosting performance and reducing cost, both designers and manufacturers are driving efforts to minimize the footprint of SRAM cells. Fig. 1.3(b) shows scaling trend of contacted gate pitch and SRAM cell size [5]. According to Moore's Law that on-chip functionality doubles every two years, technology node scaling of 0.7 in linear size and 0.5 in area has to be carried out every two years.

Besides, power and energy consumption is the other critical factor for SRAM design. For active switching, operating voltage scaling acts effectively since active energy has a square dependence on supply voltage ($V_{DD}$) [6-10]. On the other hand, leakage power also benefits from operating voltage scaling. Here, leakage power is more important than active energy

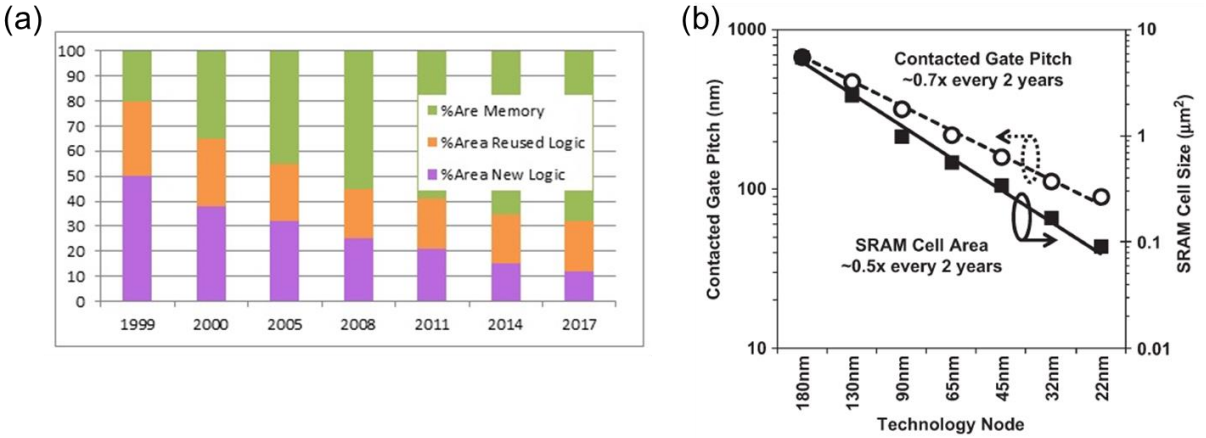


Fig. 1.3. (a) Area trends with scaling [4]. (b) Size trends with scaling: SRAM cell area vs. contacted gate pitch [5].

[11], considering that leakage in SRAM during retention dominates and its retention time is unrelated with operation or access delay. In addition, the reduction in leakage power can be greatly large, since a reduction of $V_{DD}$ implies a smaller drain bias in nano-scale cell transistors and significantly alleviates drain-induced barrier lowering (DIBL). For example, in SRAM design with 65 nm technology, operating voltage scaling from 1 V to 0.3 V can reduce the leakage power by over a factor of 15 [12]. Therefore, both active energy and leakage power considerations make operating voltage scaling significantly compelling for SRAM.

### 1.1.3 Variability-limited Scaling

However, maintaining an acceptable stability in embedded SRAM while scaling size and operating voltage becomes increasingly challenging. The continuous delay of lithography in extreme ultraviolet lithography (EUV) forces the industry to keep pushing with double and quad processes, which results in device patterning challenges [13-14]. As a result, precise control of process parameters becomes extremely difficult and the increased process variability is translated into a wider distribution of characteristics. This part is determined during fabrication and is called time-zero variability, whose sources include random dopant fluctuation (RDF) in channel [15-16], line edge roughness (LER) in channel [17], work

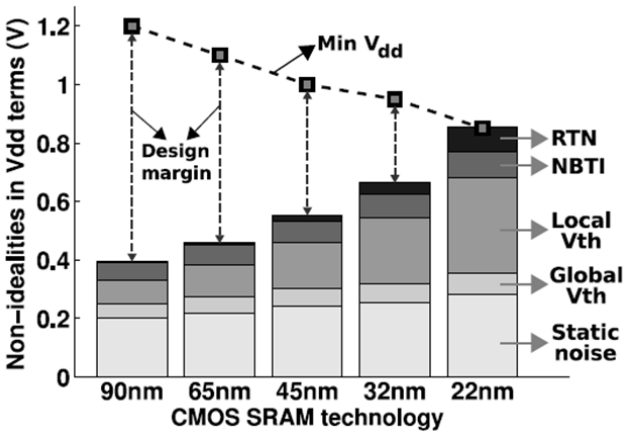


Fig. 1.4. SRAM voltage design margin trends with scaling under the impact of variability from different sources [26].

function variation (WFV) in gate [18-19], and so on. On the other hand, time-dependent or post-fabrication variability – e.g., random telegraph noise (RTN) [20-21], bias temperature instability (BTI) [22-23], hot carrier injection (HCI) [24] and soft error induced by high-energy radiation [25] – introduces a growing concern about reliability in the design/ test community. That indicates that the SRAM cell even designed stable with process optimization can fail after a long term.

In addition, both time-zero and time-dependent variability increases with size scaling [26]. It causes large variation of stability in large capacity of SRAM arrays and pushes the cell with smallest margin towards failure edge, shown in Fig. 1.4. With an increased $V_{min}$ in advanced node technology, the shrinking design margin makes operating voltage scaling much more difficult. Therefore, meeting SRAM design target with both size and operating voltage scaling requires a deep understanding of variability issue.

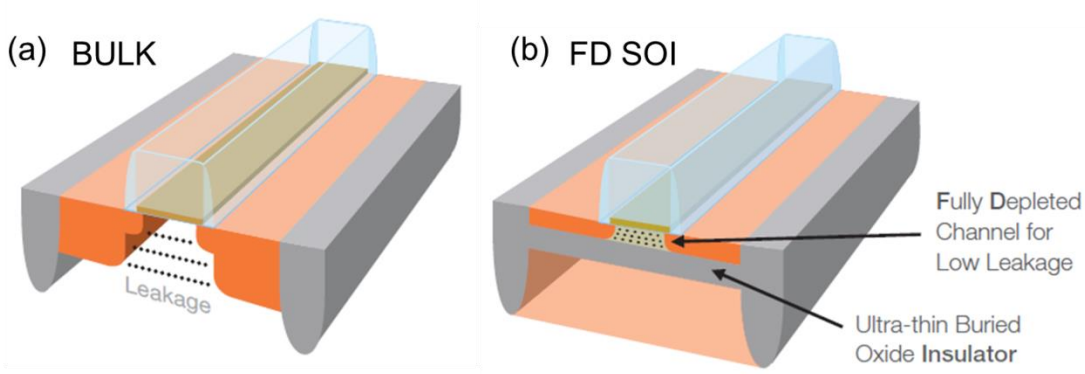


Fig. 1.5. Schematic of (a) bulk and (b) SOTB transistor [30].

## 1.2 SOTB Technology for Low-power SRAM

### 1.2.1 SOTB versus Bulk

CMOS Bulk technology has played a long-term role on the scaling road up to 20/28 nm [27]. There are various issues regarding bulk technology [28], such as large RDF-induced variability and performance-degrading leakage between source and drain. Compared with bulk technology, the fully-depleted (FD) silicon-on-insulator (SOI) technology was proposed and promoted by the SOI industry consortium [29], aimed at leveraging the established planar process while ensuring a continuation of the efficient improvements projected by Moore's Law.

Fig. 1.5 gives the comparison between schematic of bulk and FD SOI transistor [30]. The main innovations in FD SOI technology compared to bulk one consist of a fully-depleted channel and an ultra-thin buried oxide insulator. Firstly, RDF-induced variability can be largely eliminated by adopting intrinsic channel. Besides, the buried oxide layer reduces the parasitic capacitance between the source and drain exhibited by bulk technology, allowing for reduction of active switching energy. It also constrains carriers flowing between the source and drain through body to significantly reduce performance-degrading leakage currents. Another advantage with the buried oxide layer is an efficient body biasing. Due to isolation

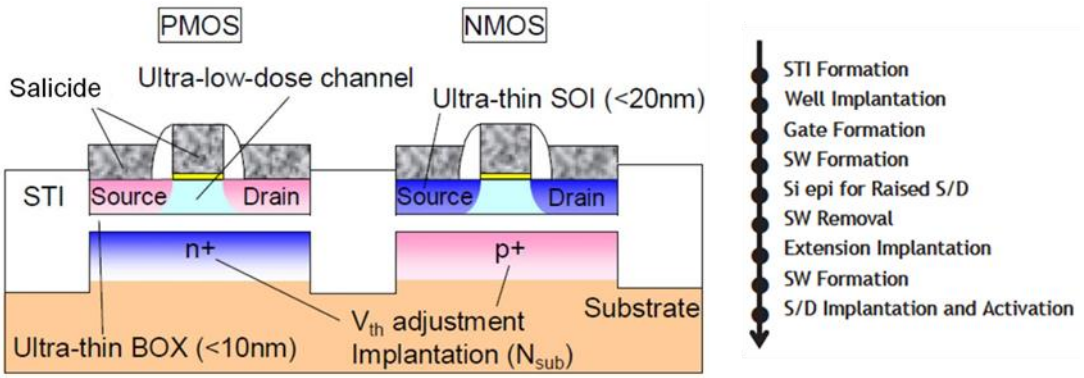


Fig. 1.6. Schematic of SOTB NMOS and PMOS with specified parameters [31-32].

between body and source/drain, strong biasing voltages can be applied to dynamically switch FD SOI transistor between modes of speed and power efficiency. Other than technological innovations, FD SOI process is an important evolutionary step from bulk CMOS process and can leverage much of the design tools, manufacturing infrastructure, and IP ecosystem already in place.

### 1.2.2 Better Control in SOTB Transistors

Silicon-on-Thin-BOX (SOTB) [31-32] is one of the FD SOI technologies. The above two innovations in FD SOI give a better control in SOTB transistors compared to bulk ones. Schematic of SOTB NMOS and PMOS with optimized process in 65 nm technology is shown in Fig. 1.6 [31-32]. As is discussed before, the buried oxide layer reduces junction capacitance both in NMOS and PMOS in Fig. 1.7(a). Also, due to its thin thickness ~ 10 nm, an efficient body biasing is demonstrated. For example, a reverse bias of 1 V realizes a large $V_{th}$[1] shift ~ 150 mV in both NMOS and PMOS in Fig. 1.7(b).

Fig 1.8(a) shows distribution of both $V_{th}$ and $I_{on}$ of 1 M SOTB transistors. Firstly, both follow a normal distribution up to ±5 sigma, contrary to the pessimistic predictions by simulation [33]. Due to the adoption of intrinsic channel, variability of both $V_{th}$ and $I_{on}$ is

[1] All through the text in this thesis, $V_{th}$ is defined as threshold voltage at constant current ($I_d=10^{-7}\times W/L$, W and L is gate width and gate length, respectively).

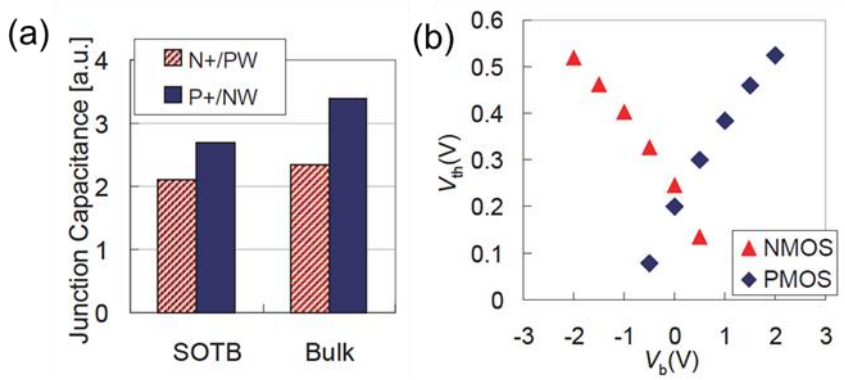


Fig. 1.7. (a) Comparison of junction capacitance in bulk and SOTB NMOS/ PMOS. (b) Demonstration of $V_{th}$ adjustment by body biasing in SOTB NMOS/ PMOS. From Ref. [32].

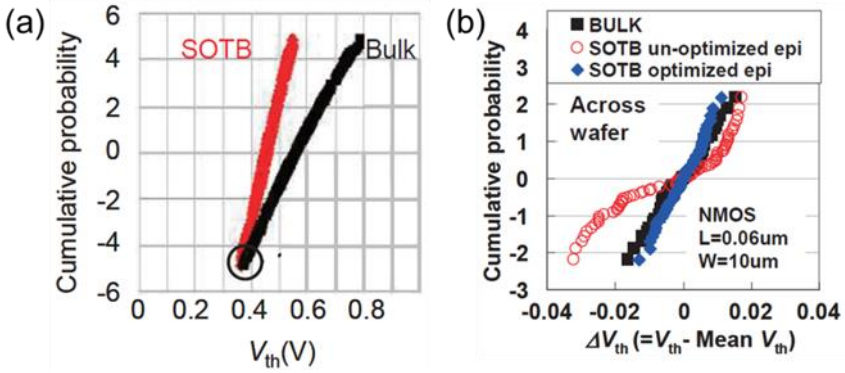


Fig. 1.8. (a) $V_{th}$ distribution in both 1 M SOTB and bulk transistors. (b) Across-wafer $V_{th}$ distribution in SOTB (w/ and w/o optimized process) and bulk NMOS. From Ref. [32].

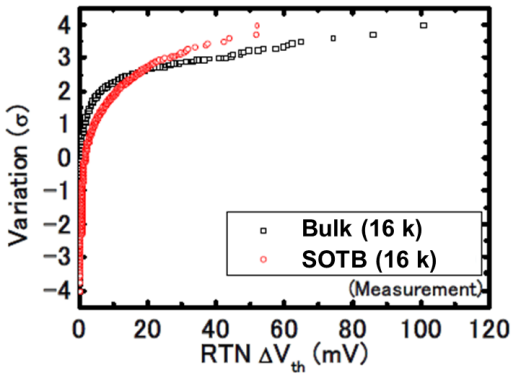


Fig. 1.9. RTN-induced $\Delta V_{th}$ distribution in 16 k SOTB and bulk transistors [34].

much smaller than in bulk transistors. Here, the bulk transistor with the largest leakage is tuned the same as SOTB one for comparison. And it is found that the drive current in SOTB transistor exhibits twice as in bulk transistor (not shown here). In addition, Fig. 1.8(b) shows

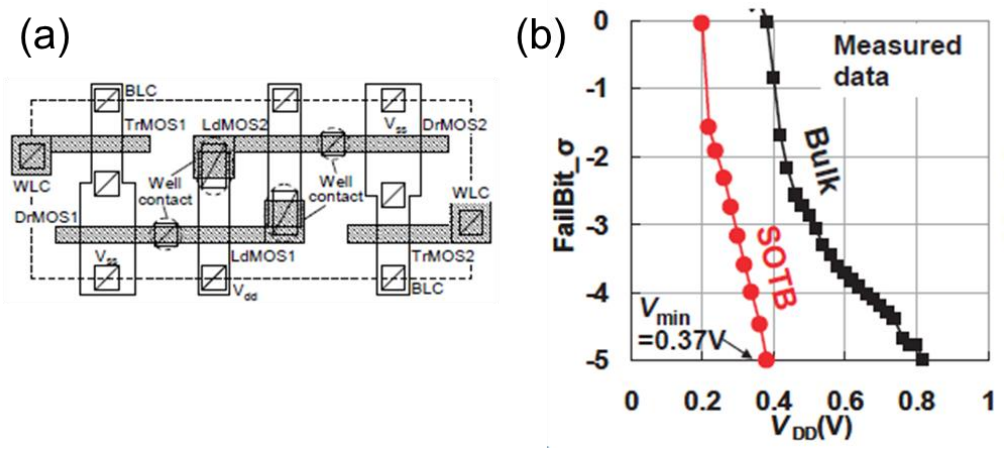


Fig. 1.10. (a) Layout of 6-T SOTB SRAM cell [31]. (b) Fail bit count in 2 Mb SOTB and bulk SRAM cells at different $V_{DD}$ [32].

across-wafer $V_{th}$ distribution in bulk and SOTB NMOS for comparison. With optimized S/D-epitaxial process, the global variability is shown as good as bulk technology, implying that thickness variability of both top and buried oxide layer is not a big issue in SOTB technology.

Besides time-zero variability, time-dependent variability is also compared between bulk and SOTB technology [34]. Take RTN as an example. Fig. 1.9 shows distributions of RTN-induced $\Delta V_{th}$ in 16 k bulk and SOTB transistors. Different from the normal distribution of $V_{th}$, they both show long-tailed distributions. But the distribution in SOTB transistors has a smaller tail, ascribed to intrinsic channel with smaller number of traps. Therefore, both time-zero and time-dependent variability is reduced in SOTB transistors compared to bulk ones.

### 1.2.3 Sub-0.4V Operation in SOTB SRAM

Consisting of six transistors, in order for a better pattern reproducibility as well as good transistor $V_{th}$ matching, the layout of SRAM cell beyond 90 nm technology preferred "wide" instead of "tall" one [35-36]. Fig. 1.10(a) shows the layout of 6-T SOTB SRAM which has a uniform orientation of all cell transistors [31]. Due to a much smaller variability compared to bulk technology, a larger margin is tolerable in SOTB SRAM design, thus making it possible to operate under low $V_{DD}$. Experimentally, $V_{min}$ down to 0.37 V in active mode has been

demonstrated in 2 Mb SOTB SRAM cells [32]. Also, it can be seen that $V_{min}$ in SOTB SRAM degrades linearly with increasing capacity. Differently, degradation of $V_{min}$ in bulk SRAM becomes much worse beyond 3 sigma's capacity, which makes it difficult to predict yield in bulk SRAM. Thus, SOTB SRAM has demonstrated itself as a good candidate for low-$V_{DD}$ operation in sub-0.4 V regime.

## 1.3 Research Goal

Incorporation of large capacity of SRAM arrays into a chip calls for tightly-packed minimum-size cell transistors. With aggressive technology scaling, variability including time-zero and time-dependent variability increases and results in limited operating voltage scaling. On the other hand, by adoption of intrinsic channel which shows immunity to RDF, SOTB SRAM can operate down to sub-0.4 V regime. However, limited data has been published for the statistical variability analysis in SRAM at low $V_{DD}$, which is critical for gaining a deeper understanding of the sources of variability and for developing robust low-power SRAM.

This dissertation facilitates the design of low-power embedded SRAM design in the presence of variability in the following ways:

1) Comparing different write stability metrics and selecting the good candidates for write yield estimation in SRAM at low $V_{DD}$.

Write stability characterization is performed at large capacity of SOTB SRAM at low $V_{DD}$ using several write stability metrics. Write noise margin is extracted for statistical analysis and the one that follows a normal distribution is preferred. The correlation between SRAM and cell transistors is established and the reason for improper write stability metrics is also clarified. In addition, on the basis of our results, one new write stability metric is proposed for write yield at low $V_{DD}$. Having a deeper understanding of various write stability metrics can help designers to better evaluate the guard-band for stable SRAM operation at low $V_{DD}$.

2) Evaluating the impact of time-dependent RTN on write yield in SRAM at low $V_{DD}$.

Besides time-zero variability, the time-dependent variability due to RTN is also considered. Through data analysis based on large-capacity SRAM measurement as well as modeling, the impact of RTN on write yield is evaluated. Also, degradation of $V_{min}$ at specific SRAM capacity can be estimated which can help designers to enlarge design margin for robust SRAM in a long-term view.

## 1.4 Dissertation Outline

Chapter 2 targets the preferred write stability metric for write stability characterization in SOTB SRAM at low $V_{DD}$ with our specially developed test-element-group (TEG). Several commonly used write stability metrics are introduced in details and four of them are compared. Based on statistical results, two of them are selected as preferred candidates for write yield estimation at low $V_{DD}$ and further discussions are given regarding the other two. In addition, write stability characterization using four write stability metrics is also performed in bulk SRAM and compared with SOTB SRAM in low-$V_{DD}$ regime. Lastly, HSPICE simulations are performed to help extend our conclusions up to ±6 sigma.

Chapter 3 proposes a new write stability metric for write yield estimation in SRAM at low $V_{DD}$. On the basis of understanding of conventional write butterfly curve [37], an extended write stability metric is proposed. Write stability characterization results using this metric are presented in SOTB SRAM, where direct correlations between the extended and conventional write butterfly curve is established. In addition, more experimental evidence as well as HSPICE simulations supporting proposed metric for yield estimation at low $V_{DD}$ is given.

Chapter 4 evaluates the effects of time-dependent RTN on write stability in SOTB SRAM at low $V_{DD}$. Selecting $I_W$ from write N-curve [38] as write stability metric, RTN measurement is performed in SRAM, which is also correlated with RTN in cell transistors. Considering the

complex distribution of RTN, in order for yield estimation, statistical model is utilized to extend measurement results to a larger SRAM capacity and the effects of RTN are evaluated.

Chapter 5 presents a summary of this dissertation – highlighting the key parts of this work, along with future research directions.

# Chapter 2

# Write Stability Characterization and Time-zero Variability

## 2.1 Introduction

### 2.1.1 Commonly Used Write Stability Metrics

Butterfly curve was proposed by Seevink [39] and has been a popular method for read stability characterization in SRAM, since it considers SRAM cell as a pair of cross-couple inverters and is easy to understand. Afterwards, write butterfly curve [37] was utilized as the method for write stability characterization in a similar way. Since write operation relates to state flip in SRAM and is more complex than read operation, several more methodss are also proposed with their own advantages over write butterfly curve. Then, this section gives an introduction of commonly used write stability metrics, including the measurement method, cell pass/failure, and definition of write stability. Either logic "0"[2] or logic "1" state can be stored in one SRAM cell and the whole write stability is defined as the minimum of the two.

[2] All through the text in this thesis, logic "0" write is focused if there is no special note.

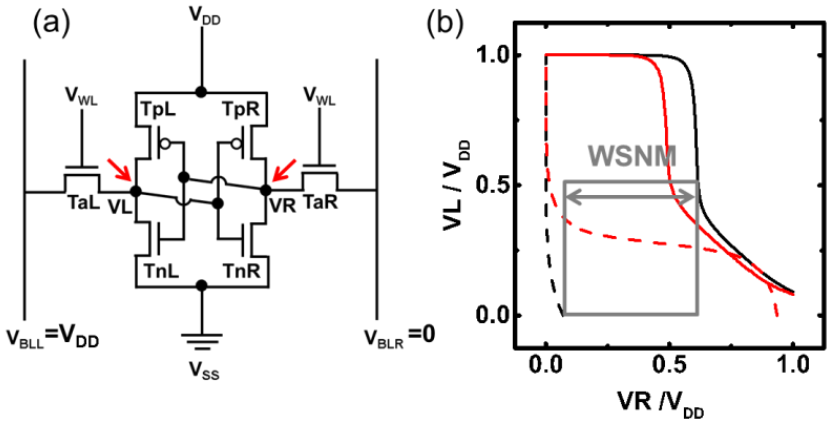


Fig. 2.1. (a) Schematic of 6-T SRAM cell with swept terminals indicated by red arrows. (b) Write butterfly curves of one SRAM cell (in black) which passes and the other one (in red) which fails. Take the stable SRAM cell as an example, its write noise margin – WSNM – is defined by grey double arrow. Voltage is normalized to $V_{DD}$. Substrate bias: $V_{bsn}$ = -1 V, $V_{bsp}$ = 1 V.

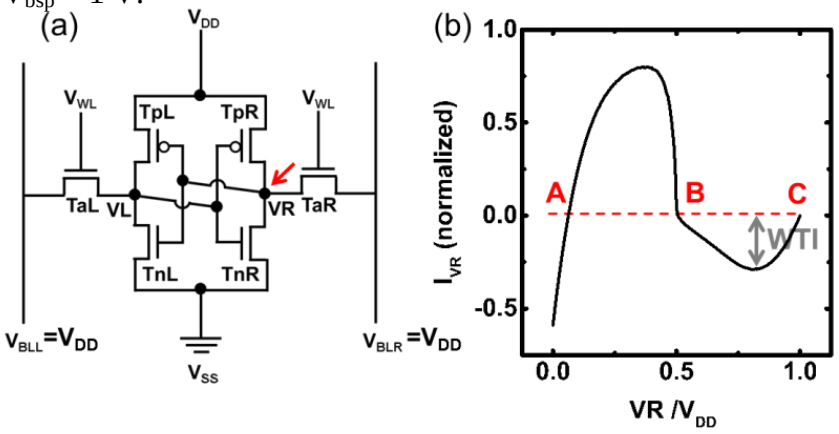


Fig. 2.2. (a) Schematic of 6-T SRAM cell with swept node indicated by red arrow. (b) Read N-curve of one SRAM cell. And its write ability – WTI – is defined by grey double arrow. Voltage/ current is normalized. Substrate bias: $V_{bsn}$ = -1 V, $V_{bsp}$ = 1 V.

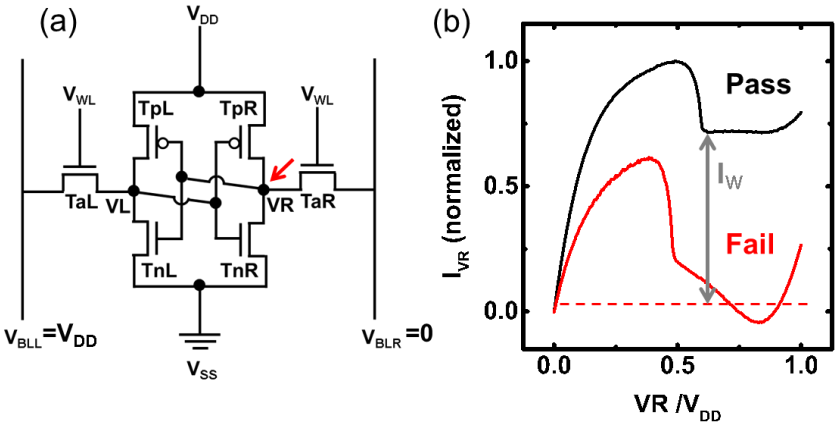


Fig. 2.3. (a) Schematic of 6-T SRAM cell with swept node indicated by red arrow. (b) Write N-curves of one SRAM cell (in black) which passes and the other one (in red) which fails. Take the stable SRAM cell as an example, its write noise margin – $I_W$ – is defined by grey double arrow. Voltage/ current is normalized. Substrate bias: $V_{bsn}$ = -1 V, $V_{bsp}$ = 1 V.

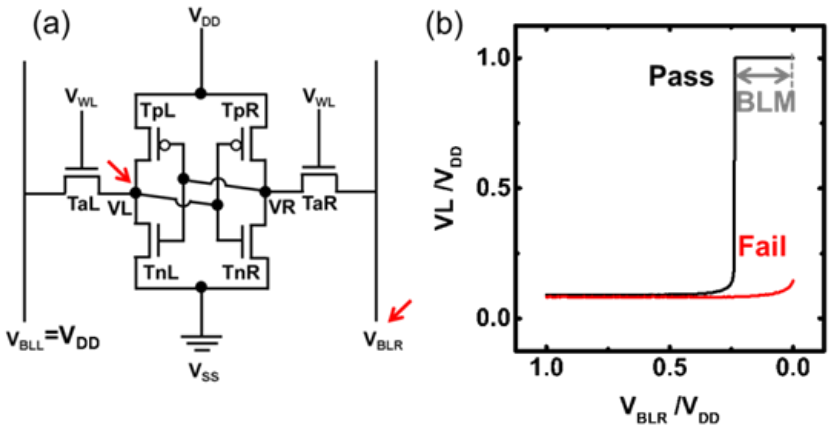


Fig. 2.4. Schematic of 6-T SRAM cell with swept terminals indicated by red arrows. (b) Measured waveforms in bit-line method of one SRAM cell (in black) which passes and the other one (in red) which fails. Take the stable SRAM cell as an example, its write noise margin – BLM – is defined by grey double arrow. Voltage is normalized to $V_{DD}$. Substrate bias: $V_{bsn}$ = -1 V, $V_{bsp}$ = 1 V.

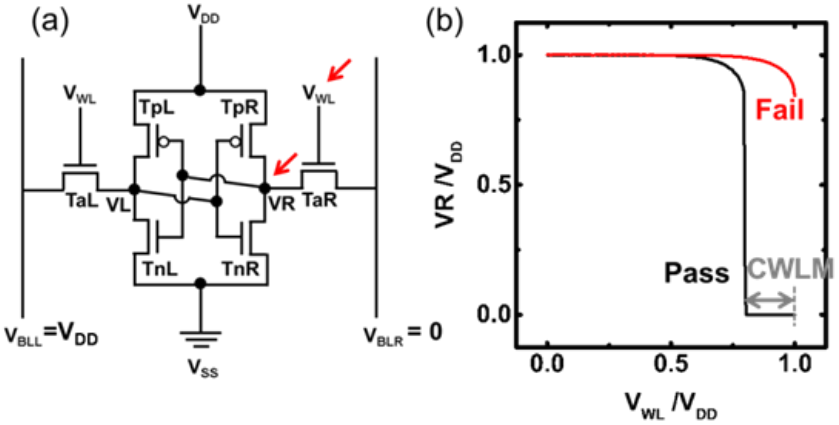


Fig. 2.5. Schematic of 6-T SRAM cell with swept terminals indicated by red arrows. (b) Measured waveforms in word-line method of one SRAM cell (in black) which passes and the other one (in red) which fails. Take the stable SRAM cell as an example, its write noise margin – CWLM – is defined by grey double arrow. Voltage is normalized to $V_{DD}$. Substrate bias: $V_{bsn}$ = -1 V, $V_{bsp}$ = 1 V.

**Write Butterfly Curve (WSNM)**

In write butterfly curve [37], voltage of one node is swept and that of the other one is monitored. Fig. 2.1(a) shows the schematic of 6-T SRAM cell with two internal nodes indicated. Write butterfly curve consists of the voltage transfer curves (VTCs) of two

consisting cross-coupled inverters. Take one SRAM cell [plotted in black in Fig. 2.1(b)] as an example. One branch (black solid line) is the same as that in read butterfly curve and can be named “read VTC”. However, write butterfly curve is characteristic of the other one (black dashed line), due to the different bias condition that $V_{BLR}$ is biased at zero instead of $V_{DD}$ in read operation. And this branch can be named “write VTC”. Also, different from read butterfly curve where the existence of three intersection points is representative of stable read operation, two branches of write butterfly curve only intersect at one logic state (here, it is “0”) if the SRAM cell is stable in write operation. When read VTC and write VTC intersect at more than one point, the SRAM cell [plotted in red in Fig. 2.1(b)] fails. And write stability – “0” write static noise margin (WSNM) – can be extracted as the side of smallest square nested inside write butterfly curves, as indicated for the stable cell in Fig. 2.1(b).

**N-curve ($I_W$)**

Read N-curve, in which $V_{BLR}$ is biased at $V_{DD}$, is usually used for characterization of read stability in the SRAM cell. Fig. 2.2(a) indicates the node in which voltage is swept as well current is monitored. Besides, Ref. 37 has extended its use as an alternative for write ability and write-trip current (WTI) is defined as the negative value of minimum current of write part in read N-curves, as shown in Fig. 2.2(b).

However, due to the bias condition in read N-curve where $V_{BLR} = V_{DD}$, the meta-stable point in read N-curve cannot be the write-trip point – where voltage of internal node flips – in real write operation. Instead, $V_{BLR}$ is proposed to be biased at zero for write N-curve [38], as shown in Fig. 2.3(a). Write N-curve is well correlated with read VTC in write butterfly curve, since in both methods voltage of VR is swept from zero to $V_{DD}$ during measurement. But current information is the unique feature in write N-curve. Fig. 2.3(b) gives the write N-curves of two SRAM cells. The black one passes, whereas the red one fails since it has the

part below $I_{VR}$ = 0. Besides, these two represent two typical types of write N-curves which will be discussed in detail in section 2.2.3. Write stability in write N-curve – "0" $I_W$ – can be defined as the local minimum current for the stable cell as indicated in Fig. 2.3(b).

**Bit-line Method (BLM)**

Different from write butterfly curve and write N-curve, two more write stability methods – bit-line method [41] and word-line method [42] – have been proposed. In these two methods, since no voltage is forced in internal nodes, the positive feedback loop is not disrupted and facilitates the change of internal nodes' states. Thus, they can both monitor write-trip point in SRAM cells. In bit-line method, voltage of node (here, it is VL) is monitored while $V_{BLR}$ is discharged from $V_{DD}$ to zero, as shown in Fig. 2.4(a). The measured waveform of one SRAM cell (plotted in black) is shown in Fig. 2.4(b), where voltage of VL is zero in the beginning and flips up to $V_{DD}$ around $V_{BLR} = 0.25 \times V_{DD}$. If VL never flips until $V_{BLR}$ is fully discharged, the SRAM cell [plotted in red in Fig. 2.4(b)] fails. Write stability in bit-line method – "0" bit line margin (BLM) – is defined as $V_{BLR}$ where voltage of VL flips for the stable cell indicated in Fig. 2.4(b).

**Word-line Method (CWLM)**

In word-line method, voltage of node (here, it is VR) is monitored while $V_{WL}$ is swept from zero to $V_{DD}$, as shown in Fig. 2.5(a). Similar to bit-line method, if state flip occurs, the SRAM cell [plotted in black in Fig. 2.5(b)] passes. If not, failure occurs [plotted in red in Fig. 2.5(b)]. And write stability in word-line method – "0" combined word line margin (CWLM) – is defined as voltage difference between $V_{DD}$ and $V_{WL}$ where voltage of VR flips for the stable cell indicated in Fig. 2.5(b).

**Performance Factor for Write Operation**

The SRAM cell which has larger write stability corresponds to a more writeable cell.

During the real write process, one bit line ($V_{BLL}$) is charged to $V_{DD}$ while the other one ($V_{BLR}$) is driven to ground potential by a write driver, followed by a generated voltage pulse to word-line. Meanwhile, voltage of node VR is also driven down through TaR. If the voltage of VR is driven down below the trip point of inverter TpL – TnL, a successful write operation takes place. Thus, the write operation can be facilitated by increasing the strength of access transistor relative to load transistor – that is the SRAM α ratio. This can be achieved by selection of $V_{th}$ or adjusting size ratio (gate width/gate length) of SRAM write performance factor (TaR and TpR).

Above all, four of the mentioned write stability metrics, including WSNM of write butterfly curve, $I_W$ of write N-curve, BLM of bit-line method and CWLM of word-line method are to be compared in the following part. The primary standard is that write noise margin follows a normal distribution. That means write yield can be easily predicted according to a small number of measured samples, which helps save test duration if more than 6 standard deviations of margin is required.

### 2.1.2 Contemporary Works

**Simulation Work**

Makino et al [41] compared three write stability metrics – WSNM, BLM and CWLM – and investigated the dependence of each metric on cell transistors' $V_{th}$ in 45 nm bulk SRAM using SPICE simulation. The basic assumption is that $V_{th}$ of each cell transistors is independent from each other and the variability of write noise margin is correlated with the variability of cell transistors' $V_{th}$. Since the variability is dominant by random fluctuation, if the differential coefficients between write noise margin and $V_{th}$ are constant, write noise margin is linear for $V_{th}$ over a wide range of variability, thus obeying a normal distribution.

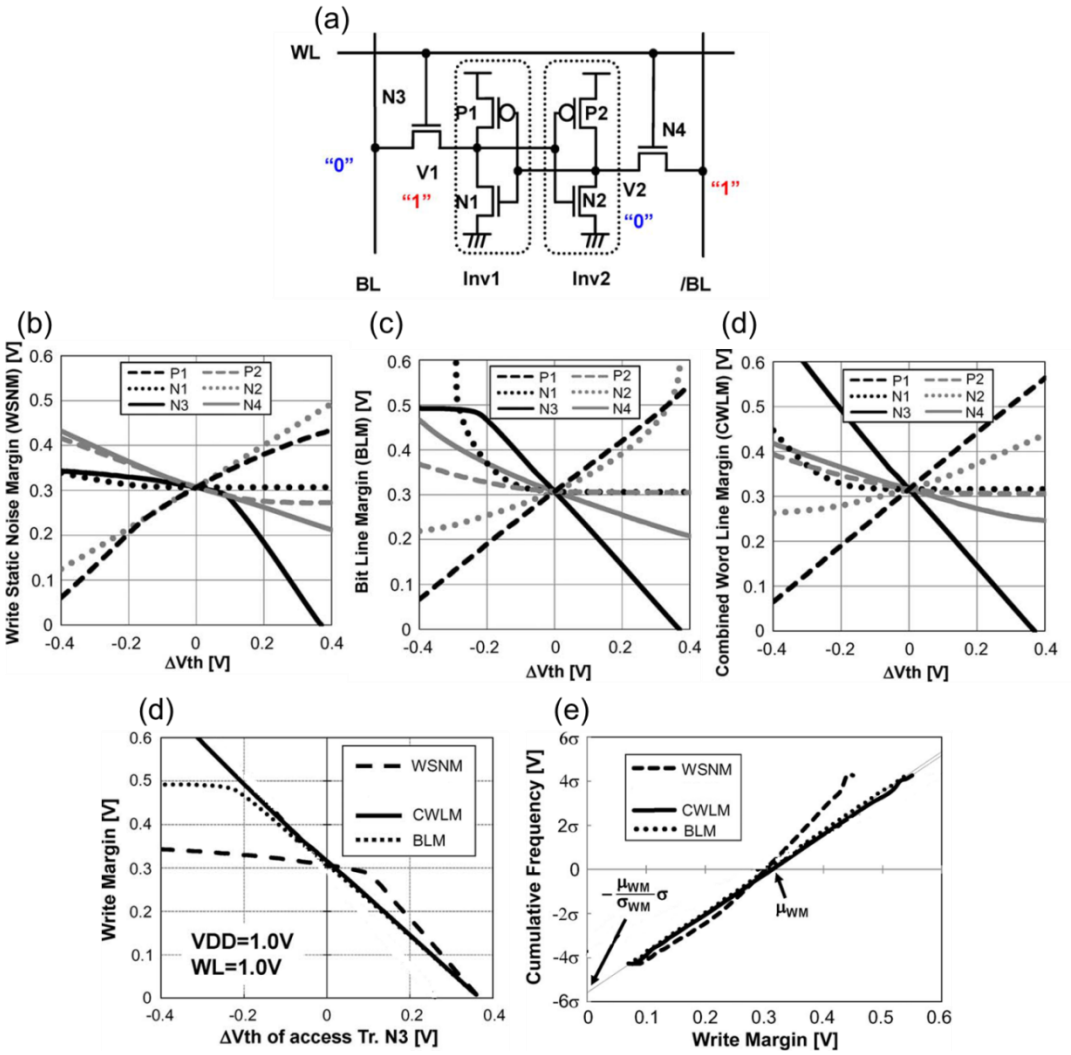


Fig. 2.6. (a) Schematic of 6-T SRAM cell with "0" originally stored. Dependence of (a) WSNM, (b) BLM, and (c) CWLM on $\Delta V_{th}$ of all six cell transistors. (d) Dependence of WSNM/ BLM/ CWLM on $\Delta V_{th}$ of N3. (e) Distribution of WSNM/ BLM/ CWLM. All simulations are performed at $V_{DD}$ = 1.0 V. From Ref. [41].

Fig. 2.6(a) shows the schematic of SRAM cell, in which logic "1" write operation is focused. Simulation results at $V_{DD}$ = 1 V are given in Fig. 2.6(b)-(d) for write butterfly curve, bit-line method and word-line method, respectively. As is discussed in section 2.1.1, the access transistor (N3) and load transistor (P1) are the dominant cell transistors in write operation. Since the correlation coefficient (CC) between write noise margin and $\Delta V_{th}$ of P1 is constant in all figures, Fig. 2.6(d) highlights the dependence of each metric on $\Delta V_{th}$ of N3. In

terms of write butterfly curve, the slope is almost zero below $\Delta V_{th}$ = 0.1 V, which means WSNM is not sensitive to $V_{th}$ change in access transistor and cannot reflect the real write noise margin in SRAM. Besides, the slope abruptly changes around $\Delta V_{th}$ = 0.1 V and results in non-normal distribution of WSNM all over the $\Delta V_{th}$ range. On the contrary, BLM/CWLM has linear relationship with $V_{th}$ change in N3 and P1, except the deviation of CC between BLM and $V_{th}$ in N3 below $\Delta V_{th}$ = 0.2 V.

For further confirmation, Monte Carlo simulation is performed to calculate the distributions of three write stability metrics at $V_{DD}$ = 1 V, as shown in Fig. 2.6(e). As is predicted, WSNM deviates from a normal distribution indicating write butterfly curve not as a good candidate and CWLM from word-line method has good normality. In terms of bit-line method, BLM also follows a normal distribution though the CC between BLM and $\Delta V_{th}$ of N3 does not keep constant all over $V_{th}$ range. Thus, it is concluded that bit-line method and word-line method are good candidates for write yield estimation at $V_{DD}$ = 1 V through simulation.

**Experimental Work**

Guo et al. [38] presented a comprehensive stability characterization work based on 45 nm bulk SRAM. Fig. 2.7 shows the test chip consisting of large functional SRAM arrays (by square in blue) and small SRAM macros (by square in red). In SRAM macros, all internal nodes including CL and CH in Fig. 2.8(a) can be accessed. And Fig. 2.8(a) and Fig. 2.8(b) show the measured waveforms of write butterfly curve and write N-curve at $V_{DD}$ = 0.7 V in SRAM macros. With regard to bit-line method and word-line method, the measurement methods are different from described in section 2.1.1 in that bit-line current instead of node voltage is monitored to anchor the state flip. Figs. 2.8(c)-(d) show measured waveforms of bit-line method and word-line method at $V_{DD}$ = 0.7 V in functional SRAM arrays.

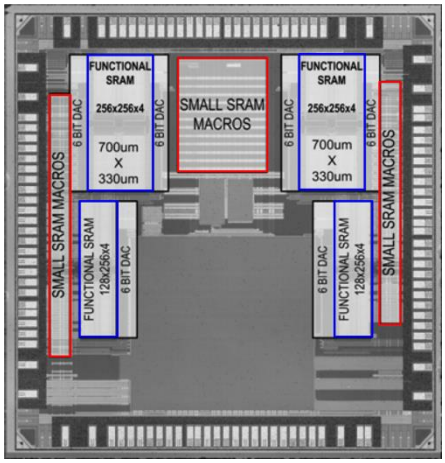


Fig. 2.7. Die photo of the 45 nm bulk SRAM test chip [38].

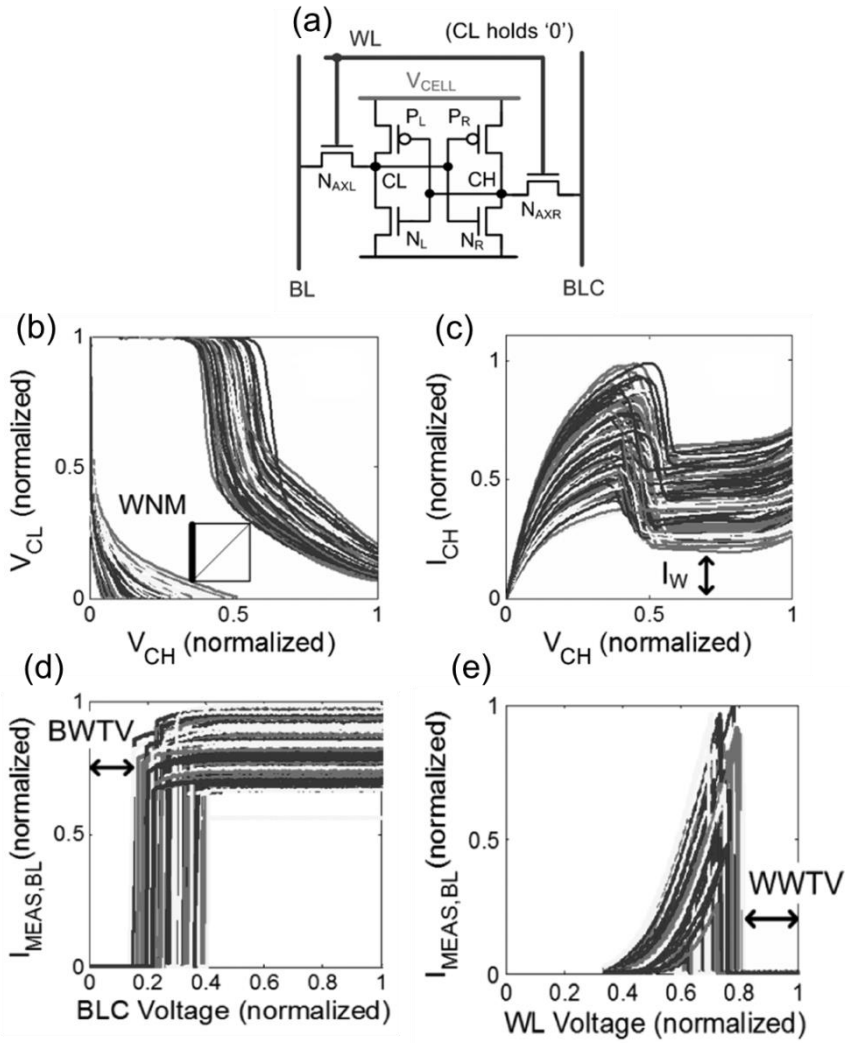


Fig. 2.8. (a) Schematic of 6-T SRAM cell. Measured (a) write butterfly curves, (c) write N-curves of SRAM macro and (d) waveforms in bit-line method, (e) waveforms in word-line method of functional SRAM arrays. Voltage is normalized to $V_{DD}$. From Ref. [38].

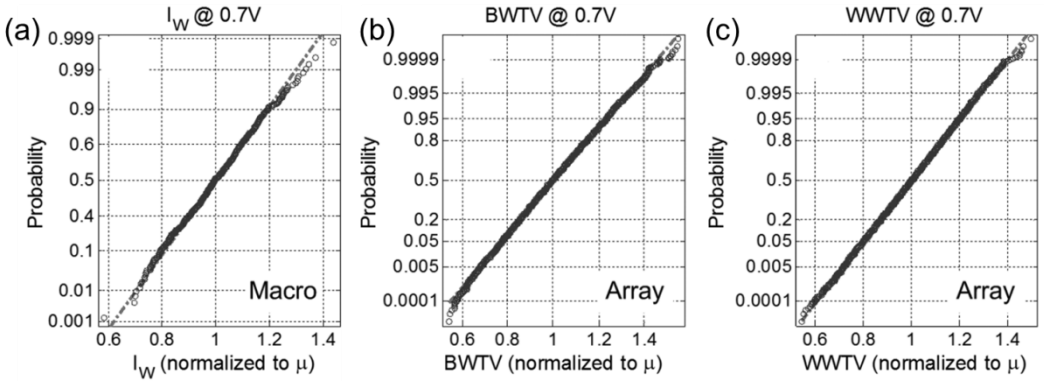


Fig. 2.9. Cumulative pots of (a) $I_W$ of SRAM macro and (b) BWTV, (c) WWTV of functional SRAM arrays at $V_{DD}$ = 0.7 V. From Ref. [38].

Based on measured data, the distributions of these metrics are plotted in Fig. 2.9 (no data of write butterfly curve is shown here). Due to small number of SRAM cells in SRAM macros, the data of write N-curve is limited to around ±3 sigma. But no big deviation from a normal distribution is demonstrated in Fig. 2.9(a). 64 kb SRAM cells are accessed for bit-line method and word-line method and support the data analysis up to ±4 sigma. These two metrics show good normality, thus confirming them as good candidates for write yield at $V_{DD}$ = 0.7 V, similar to simulation results in Ref. [41].

But for statistical analysis, there are two main limitations in this experimental work. One is the limited number of SRAM cells in SRAM macros and the other is the unconvincing comparison of different metrics between SRAM macros and functional SRAM arrays. In addition, both simulation and experimental work up to now focused on SRAM stability characterization at high $V_{DD}$ and low-$V_{DD}$ characterization [42] is called for by designers for robust low-power SRAM design.

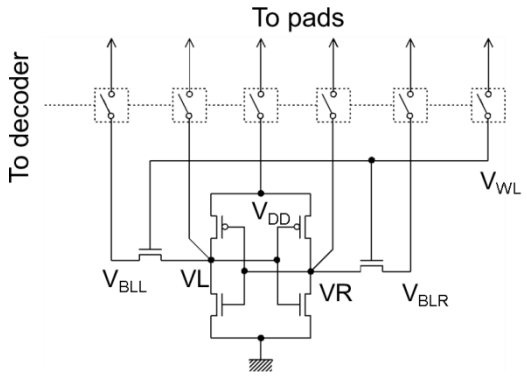


Fig. 2.10. Schematic of DMA-TEG [43-44].

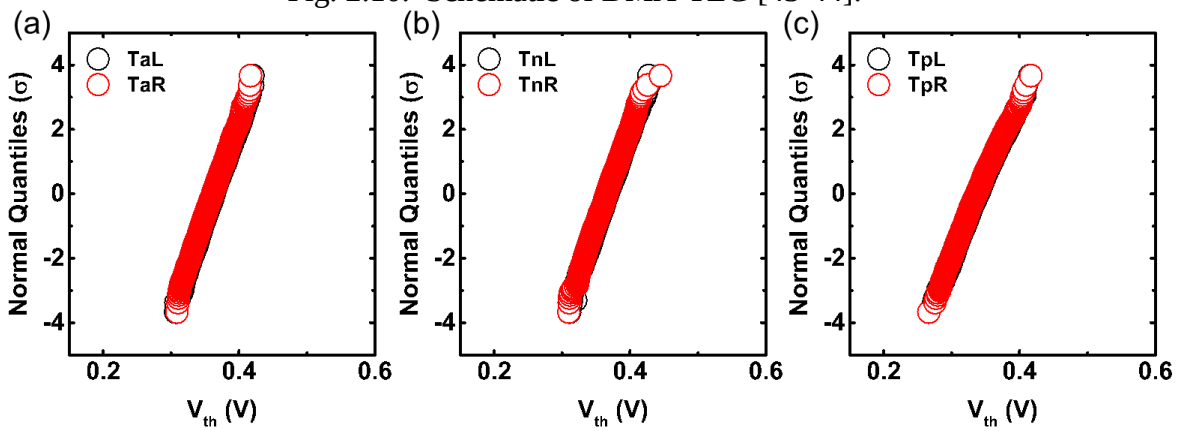


Fig. 2.11. Cumulative plots of $V_{th}$ of (a) TaL/ TaR, (b) TnL/ TnR, and (c) TpL/ TpR of 4 kb SOTB SRAM cells. Drain bias: $|V_{ds}|$ = 50 mV. Substrate bias: $V_{bsn}$ = -1 V, and $V_{bsp}$ = 1V.

# 2.2 Measurement Results in SOTB SRAM

## 2.2.1 Test Structure

A SRAM die is implemented in a 65 nm SOTB CMOS process and includes 16 kb (128 word-line × 128 bit-line) SRAM cells for large-scale read/write stability. Fig. 2.10 shows the schematic of specially developed device-matrix-array (DMA) TEG [43-45], in which two internal nodes and each pin including voltage supply, two bit lines, word lines as well as body substrate for NMOS/PMOS can be accessed. In that way, I-V measurements of individual cell transistors can be performed in order to establish a direct correlation between SRAM and cell transistors. Also, the peripheral decoder makes it easy to access each SRAM cell in a large capacity of SRAM arrays.

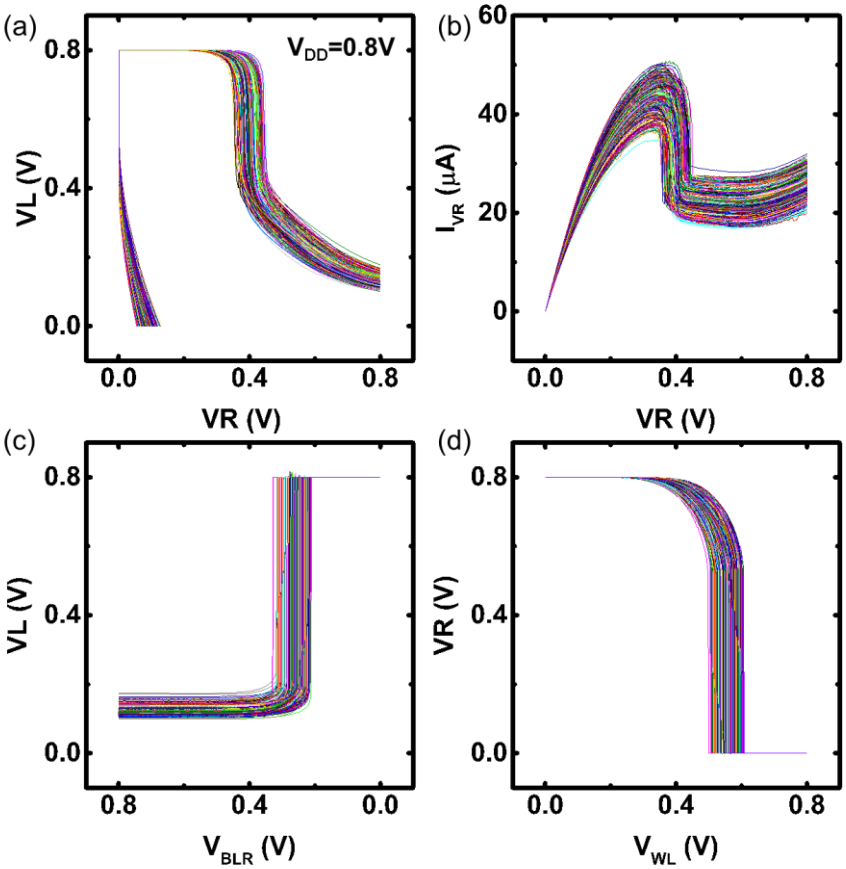


Fig. 2.12. Measured (a) write butterfly curves, (b) write N-curves, (c) waveforms in bit-line method, and (d) waveforms in word-line method of 4 kb SOTB SRAM cells at $V_{DD}$ = 0.8 V. Substrate bias: $V_{bsn}$ = -1 V, and $V_{bsp}$ = 1V.

### 2.2.2 Write Stability Characterization using Four Metrics

Fig. 2.11 shows $V_{th}$ distribution of cell transistors in 4 kb SOTB SRAM cells. Though unaffected by RDF, the remained variability can be mainly ascribed to gate work function variation (WFV) [19]. Write stability characterization is performed at $V_{DD}$ = 0.8 V, 0.6 V, and 0.4 V. Take $V_{DD}$ = 0.8 V as an example, Fig. 2.12 shows measured waveforms of (a) write butterfly curve, (b) write N-curve, (c) bit-line method, and (d) word-line method.

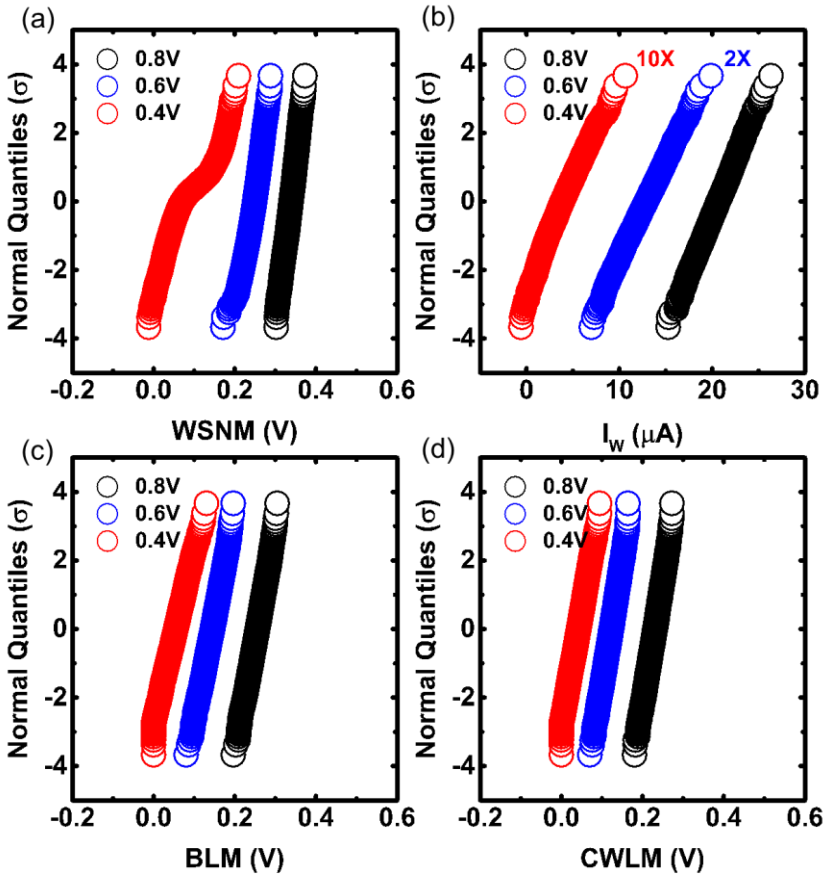


Fig. 2.13. Cumulative plots of (a) WSNM, (b) $I_W$, (c) BLM, and (d) CWLM of 4 kb SOTB SRAM cells at $V_{DD}$ = 0.8 V (in black), 0.6 V (in blue), and 0.4 V (in red). WSNM, $I_W$, BLM, CWLM are defined as the minimum of "0" and "1" write. $I_W$ at $V_{DD}$ = 0.6 V and 0.4 V have been magnified by 2 and 10 times, respectively. Substrate bias: $V_{bsn}$ = -1 V, $V_{bsp}$ = 1 V.

Fig. 2.13 shows the cumulative plots of extracted (a) WSNM, (b) $I_W$, (c) BLM, and (d) CWLM. They have good normality at high $V_{DD}$ = 0.8 V (in black) up to ±4 sigma. BLM and CWLM keep normal distributions at low $V_{DD}$ = 0.6 V (in blue) and 0.4 V (in red). Thus, BLM and CWLM can be used to extrapolate the present distribution to give write yield in large capacity of SRAM cells, even at low $V_{DD}$. On the contrary, when $V_{DD}$ goes down to 0.4 V, WSNM shows a "two-mode" distribution with the transition region at the center of the distribution and the lower tail of $I_W$'s distribution shows slight deviation from a normal distribution. The non-normality of WSNM and $I_W$ at low $V_{DD}$ will be discussed later. It also indicates that these two metrics are not good candidates for yield estimation at low $V_{DD}$.

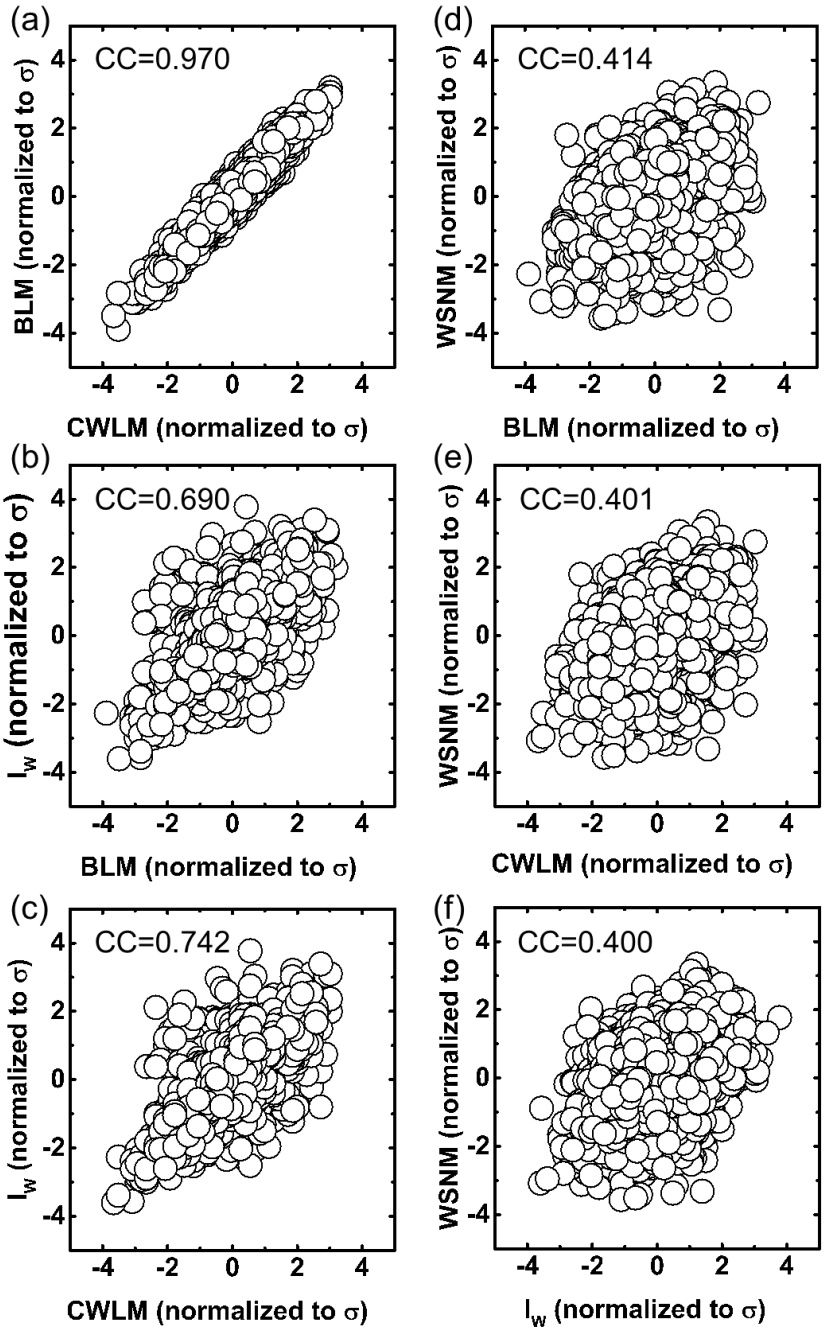


Fig. 2.14. Scatter plots of (a) BLM versus CWLM, (b) $I_W$ versus BLM, (c) $I_W$ versus CWLM, (d) WSNM versus BLM, (e) WSNM versus CWLM, and (f) WSNM versus $I_W$ in 4 kb SOTB SRAM cells at $V_{DD}$ = 0.8 V. Substrate bias: $V_{bsn}$ = - 1 V, $V_{bsp}$ = 1 V.

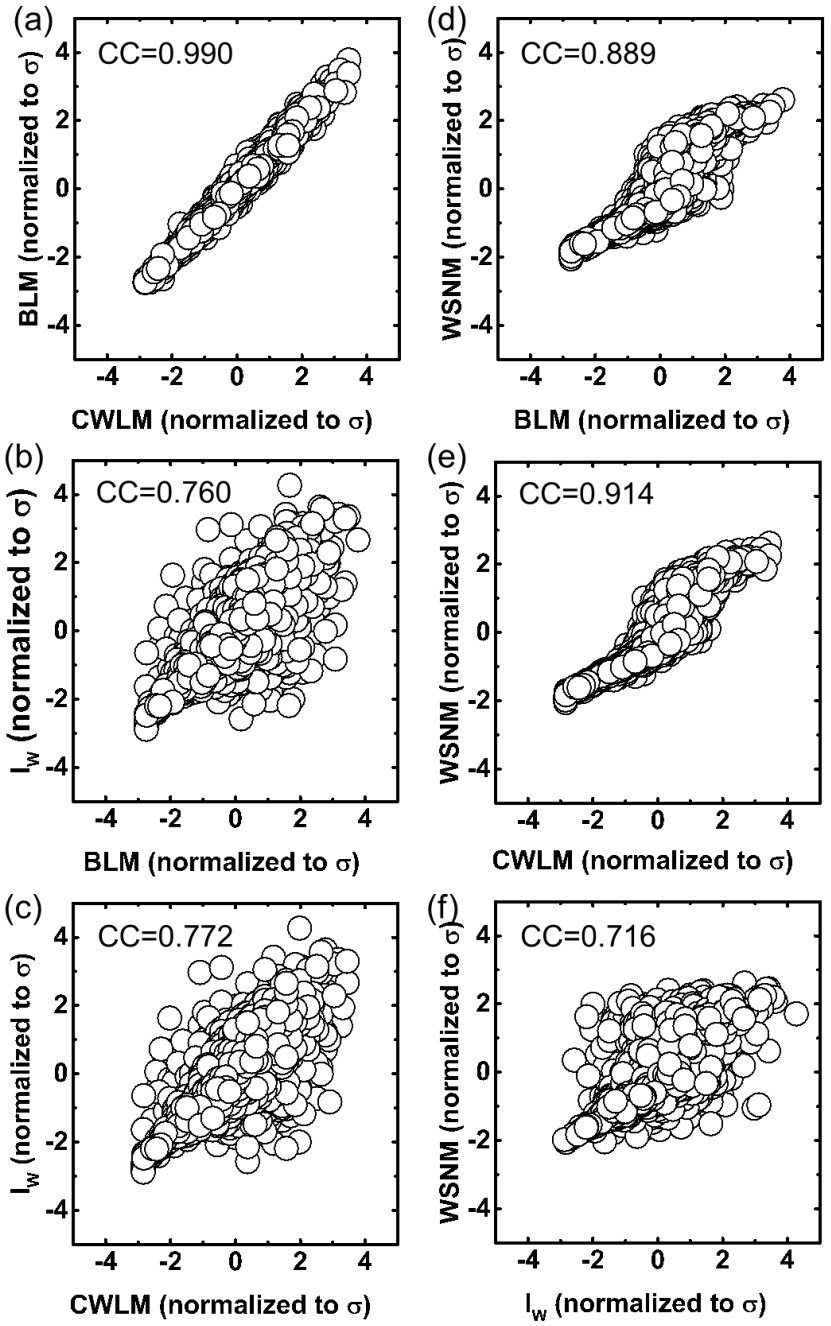


Fig. 2.15. Scatter plots of (a) BLM versus CWLM, (b) $I_W$ versus BLM, (c) $I_W$ versus CWLM, (d) WSNM versus BLM, (e) WSNM versus CWLM, and (f) WSNM versus $I_W$ in 4 kb SOTB SRAM cells at $V_{DD}$ = 0.4 V. Substrate bias: $V_{bsn}$ = - 1 V, $V_{bsp}$ = 1 V.

Fig. 2.14 shows the scatter plots of each two write stability at high $V_{DD}$ = 0.8 V. Both x-axis and y-axis have been normalized to its standard deviation relative to average value. And the CC is calculated in the upper left corner of each plot. The best correlation is found in scatter plot of BLM versus CWLM. This is due to the fact that, in both bit-line method and word-line method, BLM and CWLM are extracted on the basis of write-trip point of the SRAM cell. Setting BLM and CWLM as the standard, $I_W$ shows much better correlation than WSNM. And the worst correlation is found between WSNM and $I_W$. Fig. 2.15 shows the scatter plots at low $V_{DD}$ = 0.4 V. Generally, better correlation is demonstrated in each metric pairs than at $V_{DD}$ = 0.8 V. The CC of BLM versus CWLM is almost near to 1. In terms of WSNM and $I_W$, the dispersion also becomes much smaller at low $V_{DD}$ but still cannot be neglected. These four write stability metrics indicate the same failure point, suggesting that all metrics can be used for write failure detection in SRAM cells.

### 2.2.3 Further Discussions about Unpreferred Metrics

Three unpreferred metrics are to be discussed, including WTI, WSNM and $I_W$. And the origin of non-normality of write butterfly curve and write N-curve at low $V_{DD}$ (Fig. 2.13(a) and Fig. 2.13(b)) is also clarified.

**Read N-curve (WTI)**

Correlation between WTI from read N-curve and $I_W$ from write N-curve is analyzed in 4 kb SOTB SRAM cell at $V_{DD}$ = 0.8 V [Fig. 2.16(a)] and $V_{DD}$ = 0.4 V [Fig. 2.16(b)]. Here, write noise margin is defined as the minimum between logic “0” and “1” write. Large dispersion is found at $V_{DD}$ = 0.8 V and the bad correlation is ascribed to different bias conditions in terms of bit lines in SRAM cells. Moreover, the cells demonstrated write failure by write N-curve in Fig. 2.16(b) show positive WTI, indicating that read N-curve cannot be used for write failure detection. Thus, WTI from read N-curve is not a correct metric.

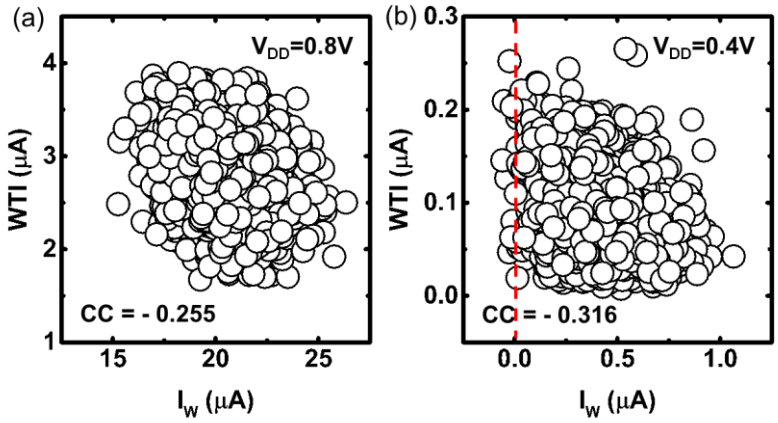


Fig. 2.16. Scatter plot of WTI from read N-curve versus $I_W$ from write N-curve of 4 kb SOTB SRAM cells at (a) $V_{DD}$ = 0.8 V and (b) $V_{DD}$ = 0.4 V. Red dashed line indicates where $I_W$ equals zero. Substrate bias: $V_{bsn}$ = - 1 V, $V_{bsp}$ = 1 V.

**Write Butterfly Curve (WSNM)**

Fig. 2.17 (a) shows the histogram plot of "0" WSNM of 4 kb SOTB SRAM cells at $V_{DD}$ = 0.4 V, with two peaks clearly distinguished. The peak at higher value is defined as Mode I and the other one at lower value is defined as Mode II. Fig. 2.17(b) shows write butterfly curves of these SRAM cells, which can be classified into two types.

To clarify the difference of the two types, Fig. 2.18(a) and Fig. 2.18(b) give write butterfly curves of Cell-A and Cell-B at $V_{DD}$ = 0.4 V. These two cells are characteristic of Mode-I and Mode-II write butterfly curve, respectively. Compared to Cell-A, Cell-B shows a smaller eye enclosed by write butterfly curve and a smaller value of "0" WSNM is extracted in Mode II. Besides, the value of VR on write VTC when VL equals zero approaches $V_{DD}$, resulting in the appearance of tail in the write VTC of Cell-B. That is the main difference from write butterfly curve of Cell-A.

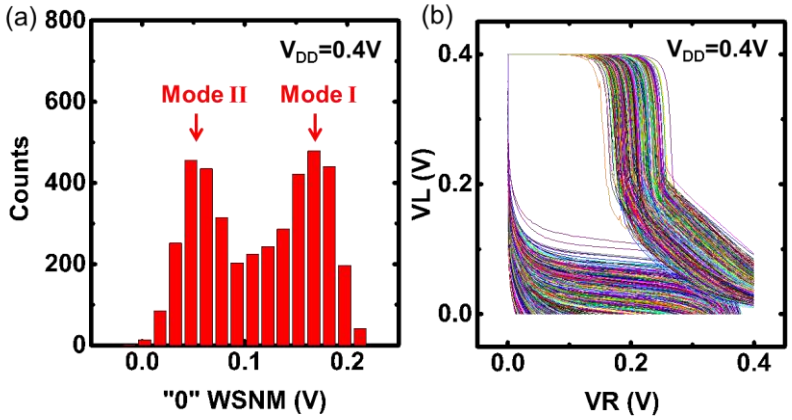


Fig. 2.17. (a) Histogram plot of "0" WSNM of 4 kb SOTB SRAM cells at $V_{DD}$ = 0.4V. Two peaks are indicated by Mode I and Mode II, respectively. (b) Write butterfly curves of 4 kb SOTB SRAM cells in "0" write at $V_{DD}$ = 0.4 V. Substrate bias: $V_{bsn}$ = -1 V, $V_{bsp}$ = 1 V.

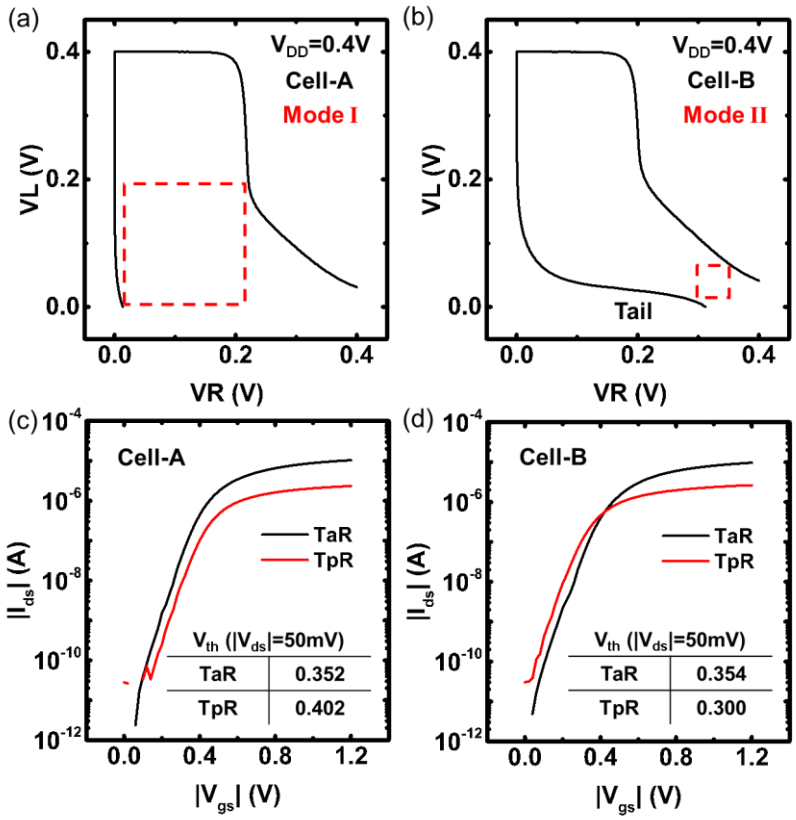


Fig. 2.18. Two types of write butterfly curves: Write butterfly curves of (a) Cell-A and (b) Cell-B in "0" write at $V_{DD}$ = 0.4 V. Cell-A/ Cell-B corresponds to Mode I/ II. Substrate bias: $V_{bsn}$ = -1 V, $V_{bsp}$ = 1 V. $|I_{ds}|$-$|V_{gs}|$ curves of TaR and TpR in (c) Cell-A and (d) Cell-B at drain bias of 50 mV. Inset: The threshold voltage ($V_{thc}$) of TaR and TpR are listed.

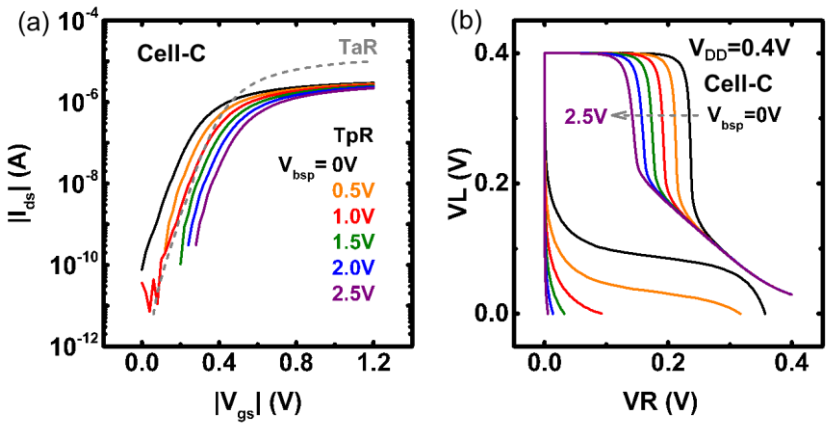


Fig. 2.19. (a) $|I_{ds}|$-$|V_{gs}|$ curves of TaR (grey dashed line) and TpR in Cell-C at $V_{bsp}$ = 0 V (black), 0.5 V (orange), 1.0 V (red), 1.5 V (green), 2.0 V (blue), and 2.5 V (purple). (b) Write butterfly curves of Cell-C in "0" write at $V_{DD}$ = 0.4 V. Substrate bias: $V_{bsn}$ = -1 V; $V_{bsp}$ = 0 V (in black), 0.5 V (orange), 1.0 V (red), 1.5 V (green), 2.0 V (blue), and 2.5 V (purple).

Since the value of VR when VL equals zero is determined by resistive voltage divider consisting of TaR and TpR, $|I_{ds}|$-$|V_{gs}|$ curves of TaR and TpR in Cell-A and Cell-B are plotted in Fig. 2.18(c) and Fig. 2.18(d), respectively. The near-$V_{th}$ or sub-$V_{th}$ region is focused on. Different from super-$V_{th}$ region, $V_{th}$ is dominant over other parameters in transistors' conductance. Here, $V_{th}$ is defined as threshold voltage extracted at constant sub-threshold current of W/L × 100 nA (W is gate width and L is gate length). Compared to Cell-A, TpR has a lower $V_{th}$ and becomes stronger than TaR in Cell-B in near-$V_{th}$ or sub-$V_{th}$ region. That is the reason why the tail appears in Cell-B at $V_{DD}$ = 0.4 V, which is the characteristic of Mode-II write butterfly curve.

More clear evidence is shown in another Cell-C. By applying substrate bias ($V_{bsp}$) for TpR, its $V_{th}$ can be adjusted in a wide range, due to advantageous design to isolate source/drain from substrate in SOTB transistors [45]. Fig. 2.19(a) shows $|I_{ds}|$-$|V_{gs}|$ curves of TpR in Cell-C at different $V_{bsp}$. And the $I_{ds}$-$V_{gs}$ curve of TaR is plotted in grey dashed line for reference. Originally, Cell-C has a balanced pair of TpR and TaR when $V_{bsp}$ = 1 V. So TpR becomes stronger than TaR when $V_{bsp}$ is decreased to 0 V, whereas becomes weaker when $V_{bsp}$ is

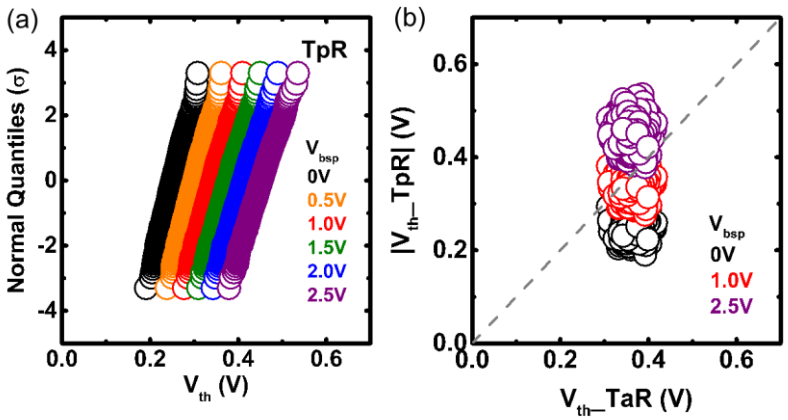


Fig. 2.20. (a) Cumulative plots of TpR's $V_{thc}$ in 1 kb SOTB SRAM cells at $V_{bsp}$ = 0 V (black), 0.5 V (orange), 1.0 V (red), 1.5 V (green), 2.0 V (blue), and 2.5 V (purple). Drain bias: -50 mV. (b) Scatter plot of TpR's $V_{thc}$ versus TaR's $V_{thc}$. Drain bias: $|V_{ds}|$ = 50 mV. Substrate bias: $V_{bsn}$ = -1 V; $V_{bsp}$ = 0 V (black), 1.0 V (red), and 2.5 V (purple).

increased to 2.5 V. Fig. 2.19(b) shows the write butterfly curves of Cell-C at $V_{DD}$ = 0.4 V when different $V_{bsp}$ is applied. A clear transition from Mode-I to Mode-II write butterfly curve is demonstrated when TpR becomes much stronger than TaR.

Statistically, Fig. 2.20(a) shows cumulative plots of TpR's $V_{th}$ in 1 kb SOTB SRAM cells at different $V_{bsp}$. Also, the scatter plots of TpR's $V_{th}$ versus TaR's $V_{th}$ at $V_{bsp}$ = 0 V, 1.0 V and 2.5 V are plotted in Fig. 2.20(b). Fig. 2.21 and Fig. 2.22 show write butterfly curves of 1 kb SOTB SRAM cells and histogram plots of "0" WSNM, respectively, at $V_{DD}$ = 0.4 V when different $V_{bsp}$ is applied. According to Fig. 2.20(b), TpR's $V_{th}$ centers on the same value as TaR's $V_{th}$ at $V_{bsp}$ = 1.0 V. When $V_{bsp}$ = 0 V is applied, TpRs in all cells become stronger than TaRs, generating Mode-II write butterfly curves in Fig. 2.21(a). Whereas, TpRs in all cells become much weaker at $V_{bsp}$ = 2.5 V. Thus, the curves in Fig. 2.21(f) are classified to Mode-I write butterfly curves. That is the reason why "0" WSNM in Fig. 2.22(a) and Fig. 2.22(f) both show normal distributions with single peak. Between these two particular cases, TpRs of some cells are stronger than TaR and vice versa. As a result, both Mode-I and Mode-II write butterfly curves coexist in Fig. 2.21(c) and "two-mode" distribution of "0" WSNM is found in Fig. 2.22(c). That is the origin of WSNM's non-normality at low $V_{DD}$.

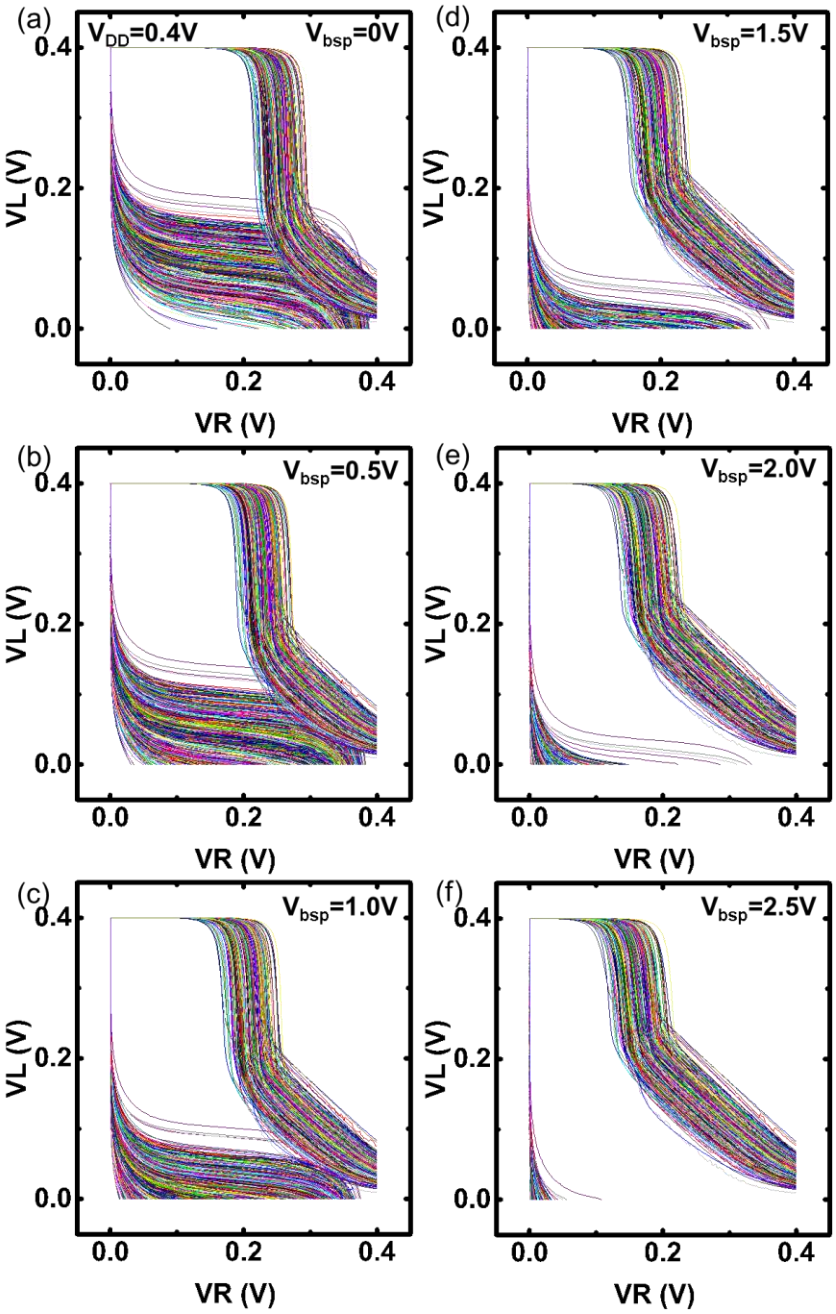


Fig. 2.21. Write butterfly curves of 1 kb SOTB SRAM cells in "0" write at $V_{DD}$ = 0.4 V. Substrate bias: $V_{bsn}$ = -1 V, (a) $V_{bsp}$ = 0 V, (b) $V_{bsp}$ = 0.5 V, (c) $V_{bsp}$ = 1.0 V, (d) $V_{bsp}$ = 1.5 V, (e) $V_{bsp}$ = 2.0 V, (f) $V_{bsp}$ = 2.5 V.

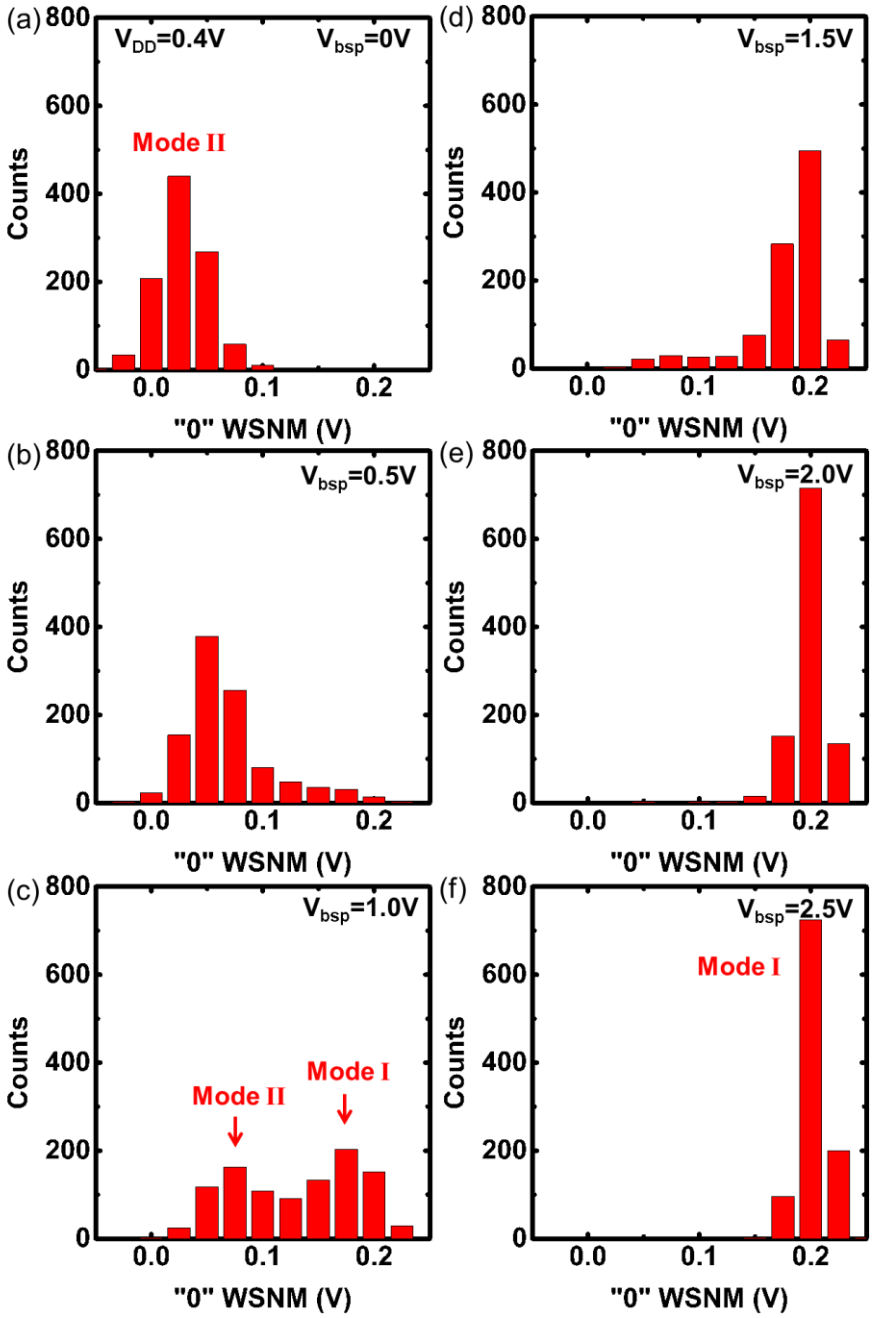


Fig. 2.22. Histogram plots of 1 kb SOTB SRAM cells in "0" write at $V_{DD}$ = 0.4 V. Substrate bias: $V_{bsn}$ = -1 V, (a) $V_{bsp}$ = 0 V, (b) $V_{bsp}$ = 0.5 V, (c) $V_{bsp}$ = 1.0 V, (d) $V_{bsp}$ = 1.5 V, (e) $V_{bsp}$ = 2.0 V, (f) $V_{bsp}$ = 2.5 V.

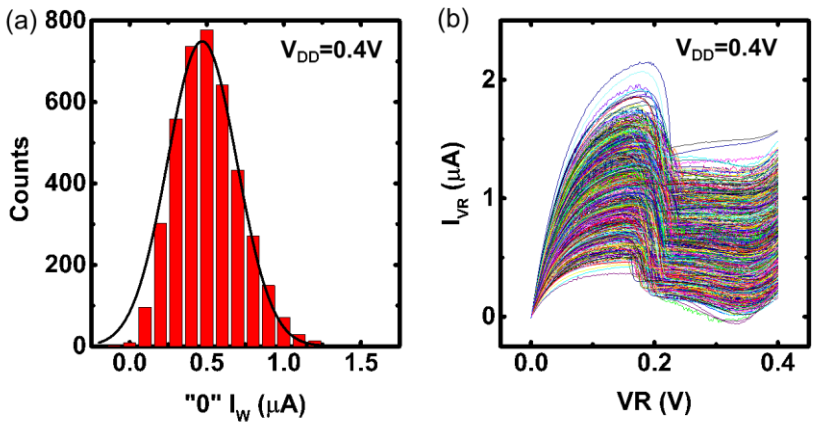


Fig. 2.23. (a) Write N-curves of 4 kb SOTB SRAM cells in "0" write at $V_{DD}$ = 0.4 V. (b) Histogram plot of "0" $I_W$ of 4 kb SOTB SRAM cells at $V_{DD}$ = 0.4 V. The black curve indicates fitting result of normal distribution function. Substrate bias: $V_{bsn}$ = -1 V, $V_{bsp}$ = 1 V.

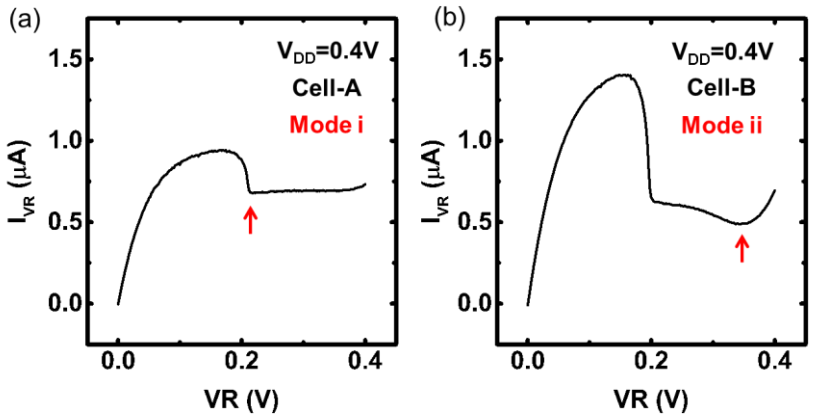


Fig. 2.24. Two types of write N-curves of Cell-A and Cell-B (same as in Fig. 2.20) in "0" write at $V_{DD}$ = 0.4 V. Cell-A/ Cell-B corresponds to Mode i/ Mode ii. The position where "0" $I_W$ is extracted is indicated by red arrow. Substrate bias: $V_{bsn}$ = -1 V, $V_{bsp}$ = 1 V.

### Write N-curve ($I_W$)

Fig. 2.23(a) shows the histogram plot of "0" $I_W$ of 4 kb SOTB SRAM cells at $V_{DD}$ = 0.4 V, which obviously deviates from a normal distribution (black curve). Fig. 2.23(b) shows write N-curves of these SRAM cells. Similar to write butterfly curves, write N-curves can also be classified into two types. Mode-i write N-curve of Cell-A and Mode-ii write N-curve of Cell-B in "0" write at $V_{DD}$ = 0.4 V are plotted in Fig. 2.24(a) and Fig. 2.24(b), respectively.

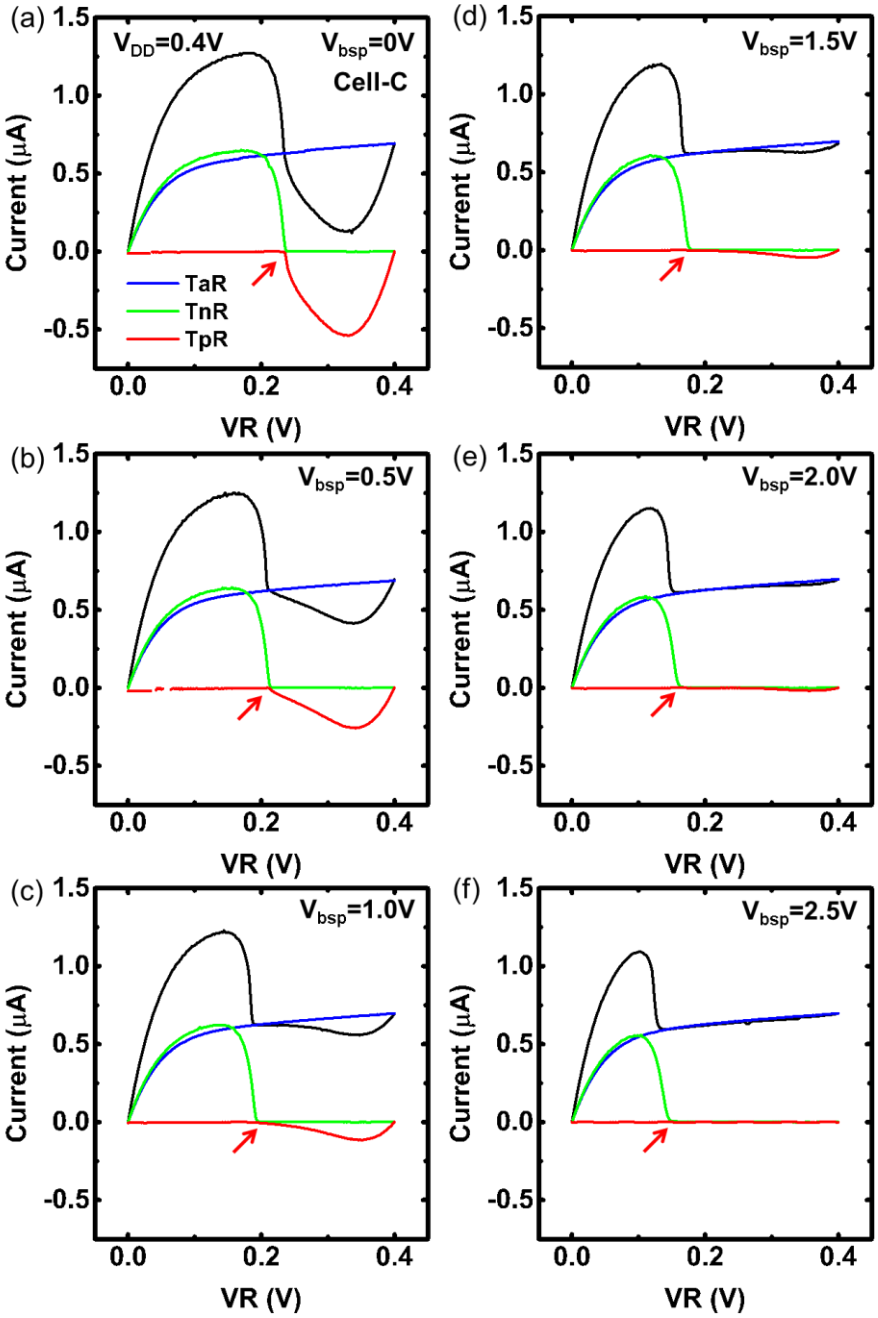


Fig. 2.25. Measured currents in TaR (in blue), TnR (in green) and TpR (in red) contributing to write N-curve (in black) of Cell-C in "0" write at $V_{DD}$ = 0.4 V. Substrate bias: $V_{bsn}$ = -1 V, (a) $V_{bsp}$ = 0 V, (b) $V_{bsp}$ = 0.5 V, (c) $V_{bsp}$ = 1.0 V, (d) $V_{bsp}$ = 1.5 V, (e) $V_{bsp}$ = 2.0 V, (f) $V_{bsp}$ = 2.5 V. Here, the positive direction of current is defined as flowing outside node VR.

Compared to Cell-B, the position of local minimum current changes in Cell-A. The value of VR where "0" $I_W$ is extracted is smaller in Mode-i write N-curve. Here, considering the bad correlation between WSNM and $I_W$ in Fig. 2.15(f), two modes in write N-curves are named Mode-i and Mode-ii, different from in write butterfly curves. Also, the different mechanism is discussed as follows.

Fig. 2.25 shows write N-curves of Cell-C and currents in cell transistors (TaR, TnR and TpR) at $V_{DD}$ = 0.4 V when different $V_{bsp}$ is applied. A clear transition is shown from Mode-ii write N-curve to Mode-i write N-curve when $V_{bsp}$ increases from 0 V to 2.5 V. According to Kirchhoff's current law, the current flowing into node VR is the sum of all currents in TaR, TnR and TpR. Current in TaR (in blue) is the drain current with gate biased at $V_{DD}$. Currents in TnR (in green) and TpR (in red) are both zero at intersection point indicated by red arrow, which corresponds to voltage trip point of left half cell. In Fig. 2.25(a), since TpR is strong with a large current (negative) beyond the voltage trip point, the local minimum current of write N-curve is determined by maximum current in TpR. With the increase of $V_{bsp}$, TpR becomes much weaker and finally operates in sub-$V_{th}$ region in Fig. 2.25(f). In this case, since current in TaR reaches saturation after voltage trip point, the local minimum current in write N-curve is extracted near voltage trip point. Different from in write butterfly curve, Mode-i and Mode-ii write N-curve are distinguished mainly according to the strength of TpR itself. When TpR goes into sub-$V_{th}$ region, Mode-ii N-curve transitions to Mode-i N-curve.

Statistically, Fig. 2.26 and Fig. 2.27 show write N-curves of 1 kb SOTB SRAM cells and histogram plots of 0 "$I_W$", respectively, at $V_{DD}$ = 0.4 V when different $V_{bsp}$ is applied. According to Fig. 2.20(a), when $V_{bsp}$ = 0 V is applied, TpRs in all cells operate in super-$V_{th}$ region, generating Mode-ii write N-curves in Fig. 2.26(a). That is the reason why "0" $I_W$ in Fig. 2.27(a) follows a normal distribution. Whereas, TpRs in all cells goes into sub-$V_{th}$ region

at $V_{bsp}$ = 2.5 V. Thus, the curves in Fig. 2.26(f) are classified to Mode-i write N-curves. Different from "0" WSNM in Fig. 2.22(f), Fig. 2.27(f) shows a right-skewed distribution, rather than a symmetrical one, of "0" $I_W$ in Mode i. Also, the value of "0" $I_W$ in Mode i cannot be clearly distinguished from that in Mode ii. Under substrate bias of 1.0 V, Mode-ii and Mode-i write N-curves co-exist in Fig. 2.26(c). As a result, two modes entangle with each other and it induces the non-normal distribution of "0" $I_W$ at low $V_{DD}$ shown in Fig. 2.27(c).

It has been concluded that two modes exist in both WSNM and $I_W$ at low $V_{DD}$. When it comes to high-$V_{DD}$ distributions, the above conclusion can explain well. In terms of WSNM, both two modes follow normal distributions but Mode I centers on a larger value than Mode II. When SRAM cells operate at low $V_{DD}$, TpRs in some cells are stronger than TaRs in near-$V_{th}$ or sub-$V_{th}$ region, such as Cell-B in Fig. 2.18(d). But at high $V_{DD}$, due to a lower hole mobility as well as a smaller gate width/length ratio, TpRs are usually weaker than TaRs in super-$V_{th}$ reigon. Thus, Mode-I write butterfly curves dominate at high $V_{DD}$ [Fig. 2.12(a)], contributing to a normal distribution of WSNM in Fig. 2.13(a).

On the other hand, $I_W$ in Mode ii shows good normality whereas that in Mode i deviate from a normal distribution. And these two modes cannot be distinguished as in WSNM. When SRAM cells operate at low $V_{DD}$, TpRs are in near-$V_{th}$ or sub-$V_{th}$ region. But at high $V_{DD}$, all cell transistors including load transistors operate in super-$V_{th}$ region. As a result, Mode-ii write N-curves dominate at high $V_{DD}$ [Fig. 2.12(b)], also contributing to a normal distribution of $I_W$ in Fig. 2.13(b).

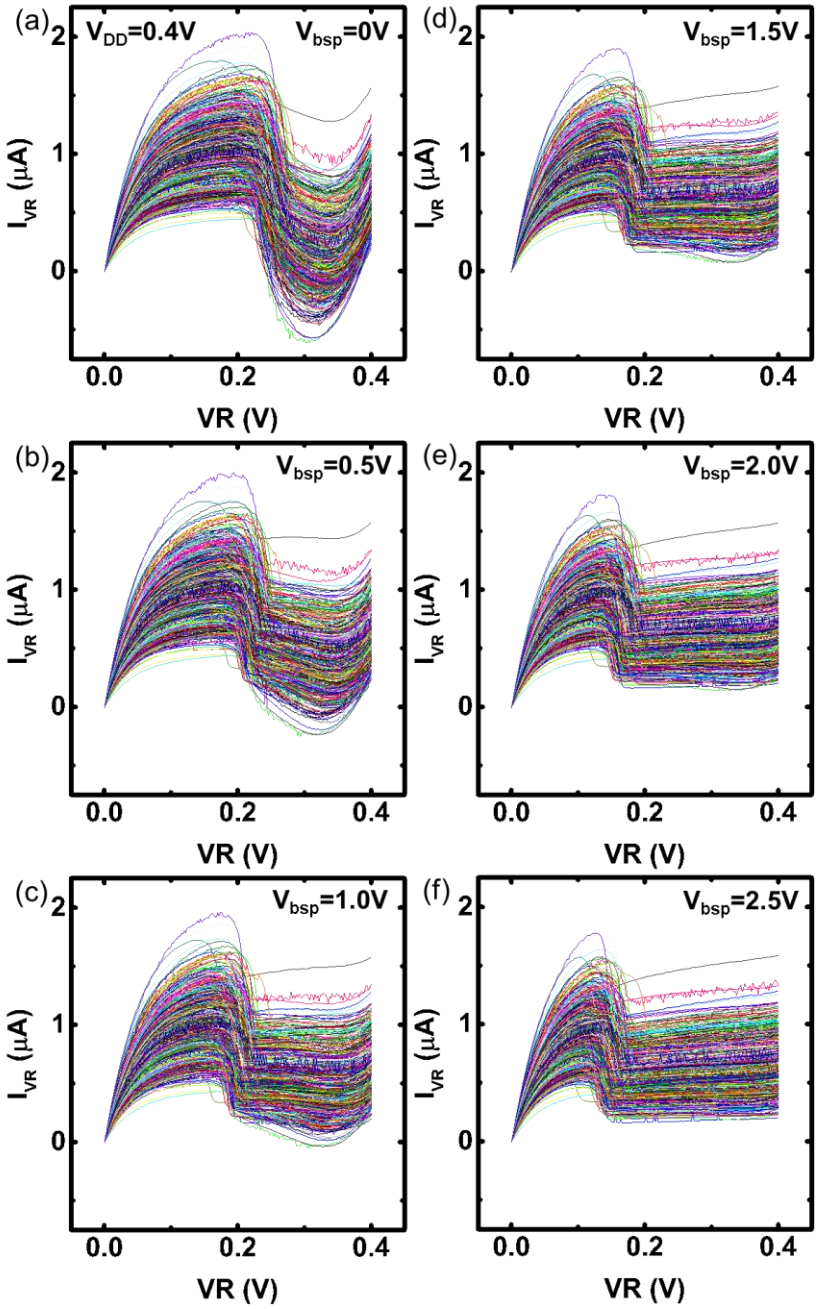


Fig. 2.26. Write N-curves of 1 kb SOTB SRAM cells in "0" write at $V_{DD}$ = 0.4 V. Substrate bias: $V_{bsn}$ = -1 V, (a) $V_{bsp}$ = 0 V, (b) $V_{bsp}$ = 0.5 V, (c) $V_{bsp}$ = 1.0 V, (d) $V_{bsp}$ = 1.5 V, (e) $V_{bsp}$ = 2.0 V, (f) $V_{bsp}$ = 2.5 V.

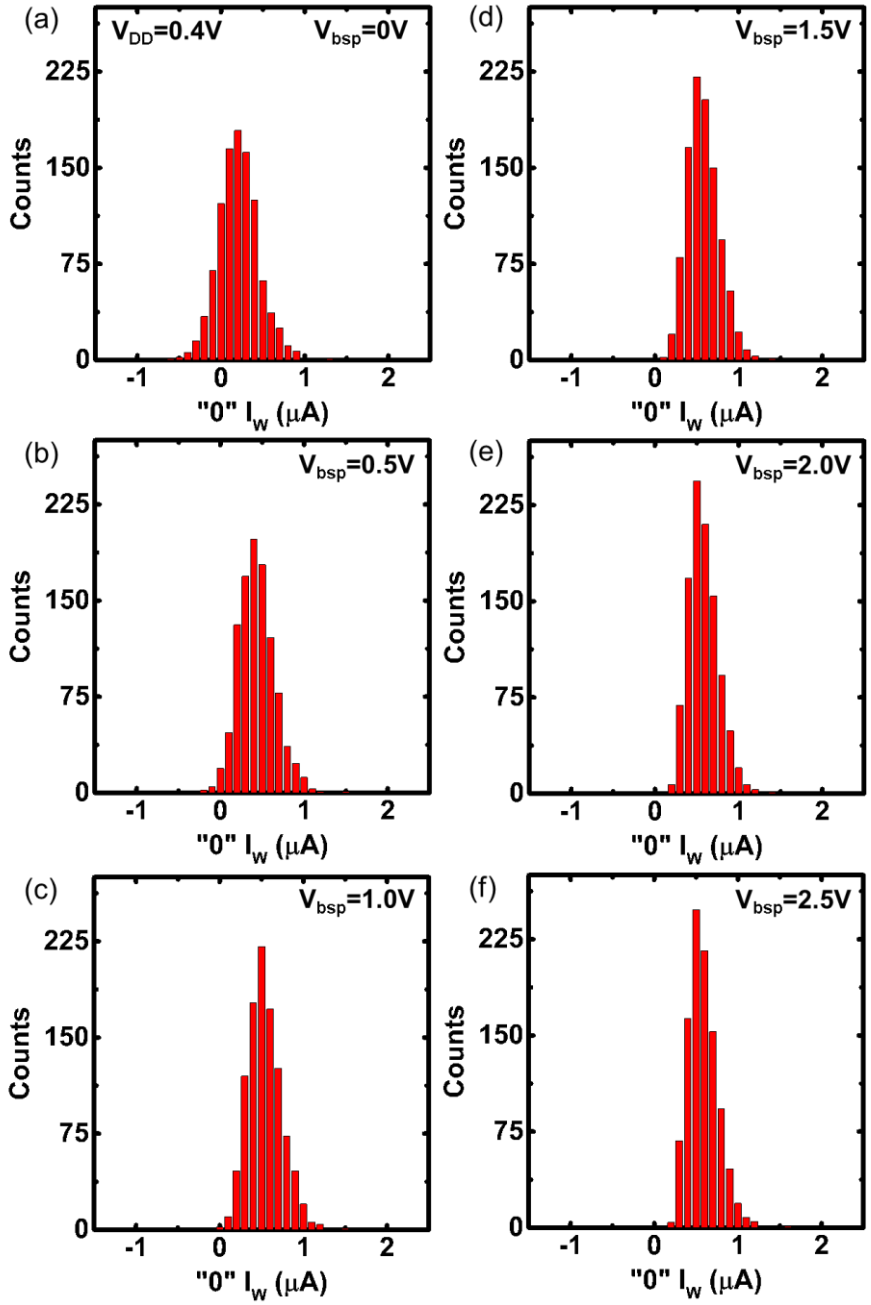


Fig. 2.27. Histogram plots of "0" $I_W$ of 1 kb SOTB SRAM cells in "0" write at $V_{DD}$ = 0.4 V. Substrate bias: $V_{bsn}$ = -1 V, (a) $V_{bsp}$ = 0 V, (b) $V_{bsp}$ = 0.5 V, (c) $V_{bsp}$ = 1.0 V, (d) $V_{bsp}$ = 1.5 V, (e) $V_{bsp}$ = 2.0 V, (f) $V_{bsp}$ = 2.5 V.

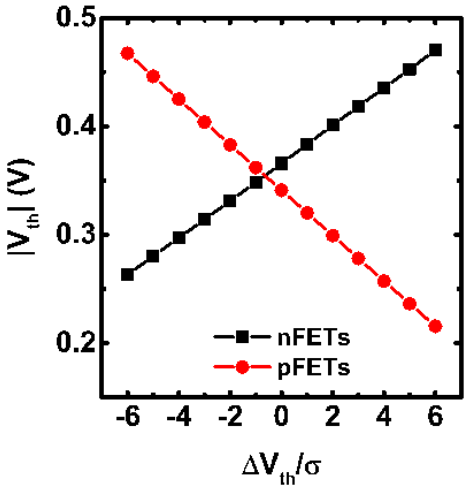


Fig. 2.28. $V_{th}$ is changed from -6 sigma to 6 sigma in nFETs/ pFETs.

## 2.3 HSPICE Simulation Results in SOTB SRAM

Simulations are presented to further discuss the distributions of the above four write stability metrics. We did an analysis similar to H. Makino et al. [41] to discuss the distribution of WNM in SOTB SRAM cells. Based on the equation $\sigma_{WNM}{}^2 = \sum_{i=1}^{6}(\frac{\partial WNM}{\partial V_{thi}})^2 \sigma_{V_{thi}}{}^2$, the variance of WNM distribution is correlated with the variance of $V_{th}$ by each differential coefficient. We assume $V_{th}$ of each cell transistor is independent with each other. And if the differential coefficient $\partial WNM/\partial V_{thi}$ is constant, WNM is expected to obey a normal distribution.

Dependence of WNM on each cell transistor's $V_{th}$ is simulated by HSPICE with 65 nm SOTB transistor parameters. $\Delta V_{th}$ in nFETs/ pFETs is set between -6 sigma and 6 sigma, shown in Fig. 2.28. Typical values of nFETs and pFETs as well as each variance are selected according to experimental parameters. In terms of WNM in SRAM, three $V_{DD}$ are selected as 0.8 V, 0.6 V, and 0.4 V for comparison with experimental results.

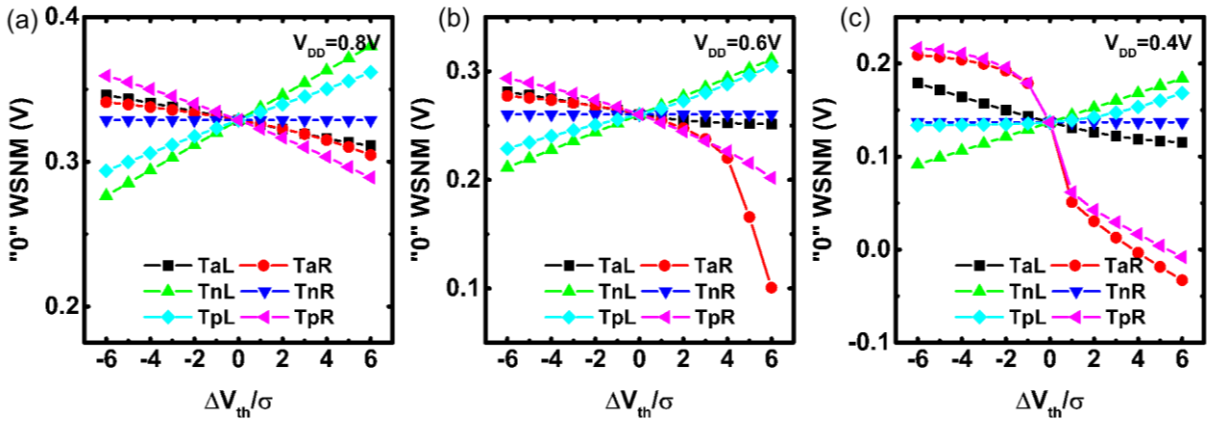


Fig. 2.29. Dependence of "0" WSNM on $\Delta V_{th}$ at $V_{DD}$ = (a) 0.8 V, (b) 0.6 V, and (c) 0.4 V.

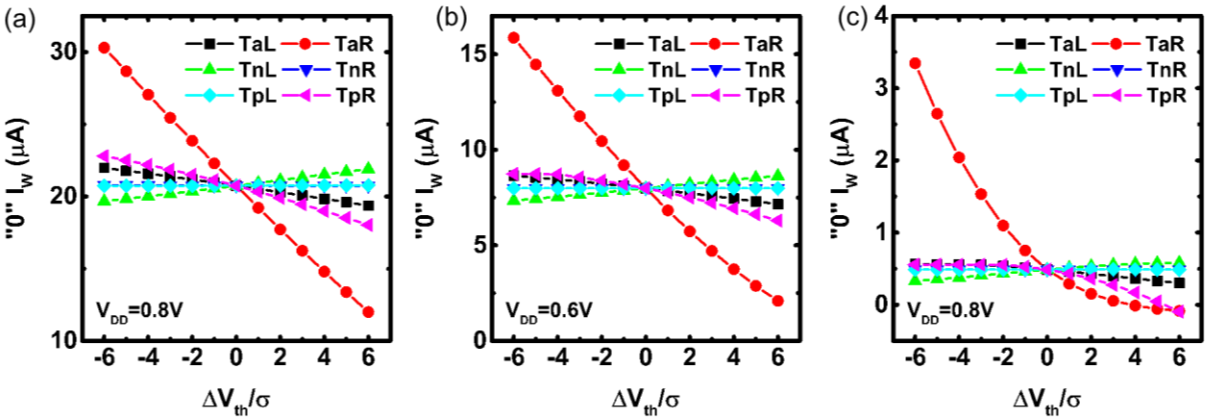


Fig. 2.30. Dependence of "0" $I_W$ on $\Delta V_{th}$ at $V_{DD}$ = (a) 0.8 V, (b) 0.6 V, and (c) 0.4 V.

Fig. 2.29 shows the dependence of "0" WSNM on $\Delta V_{th}$ in each cell transistor at $V_{DD}$ = (a) 0.8 V, (b) 0.6 V and (c) 0.4 V. Among all six cell transistors, TaR/ TpR dominates the region with small WNM[3]. It can be seen that WSNM has good linearity with $\Delta V_{th}$ in all cell transistors at high $V_{DD}$ = 0.8 V, indicating that WSNM follows a normal distribution. But when $V_{DD}$ goes down, like at 0.6 V, the slope of WSNM on TaR/ TpR changes towards failure edge, indicating WSNM starts to deviate from a normal distribution. The deviation is more serious at 0.4 V. In particular, WSNM shows linearity with TaR/ TpR in -6 sigma < $\Delta V_{th}$ < -2 sigma and 1 sigma < $\Delta V_{th}$ < 6 sigma, but abruptly changes in -2 sigma < $\Delta V_{th}$ <

[3] In Fig. 2.31(a), TnL is shown to dominate WSNM in -6 sigma < $\Delta V_{th}$ < -5 sigma. However, we expect TaR/ TpR dominates when the range of $\Delta V_{th}$ is beyond 6 sigma.

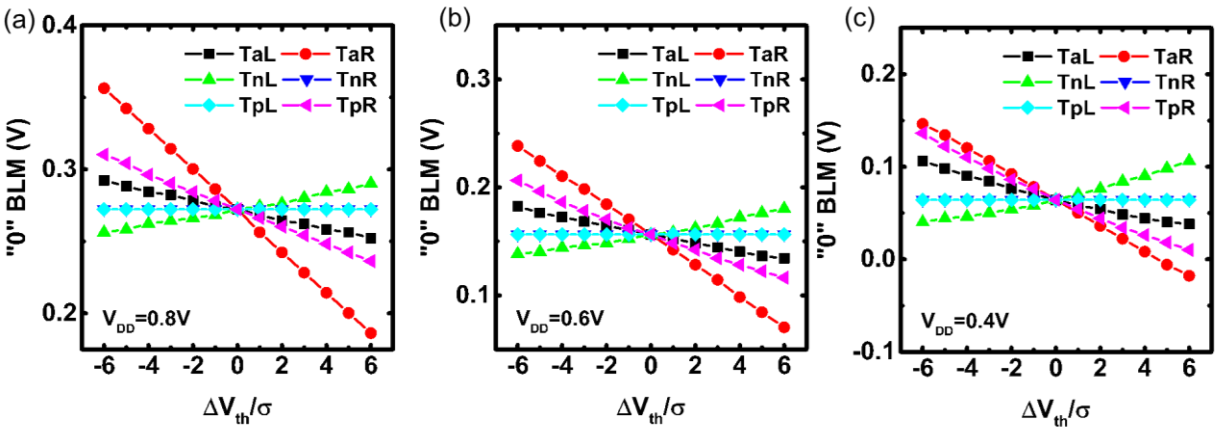


Fig. 2.31. Dependence of "0" BLM on $\Delta V_{th}$ at $V_{DD}$ = (a) 0.8 V, (b) 0.6 V, and (c) 0.4 V.

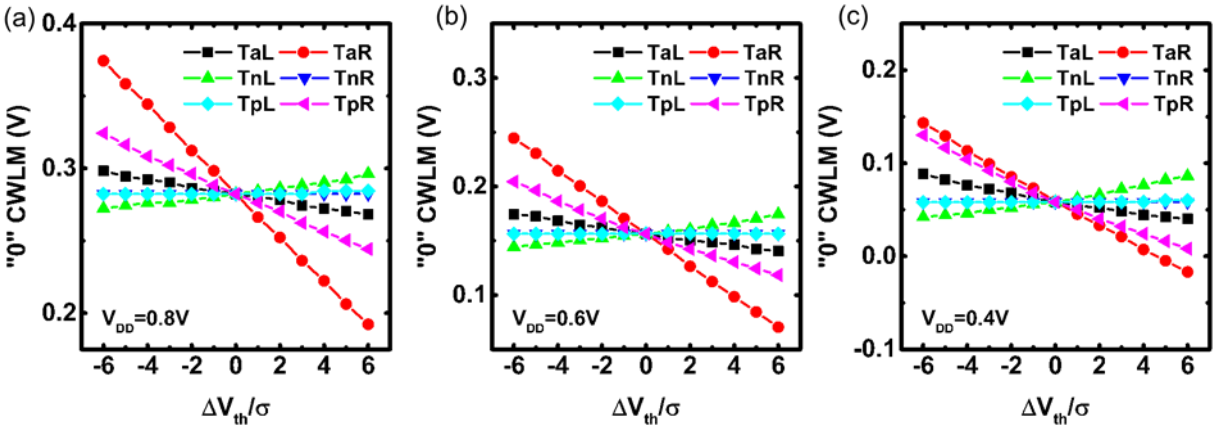


Fig. 2.32. Dependence of "0" CWLM on $\Delta V_{th}$ at $V_{DD}$ = (a) 0.8 V, (b) 0.6 V, and (c) 0.4 V.

sigma. This is a clear evidence for "two-mode" distribution of WSNM at low $V_{DD}$, which has been demonstrated in previous discussion of experimental results.

The distribution of $I_W$ is also discussed at different $V_{DD}$ in Fig. 2.30. Same as "0" WSNM, TaR and TpR are the dominant cell transistors in region with small WNM. The good linearly with all cell transistors at $V_{DD}$ = 0.8 V is consistent with the normal distribution of measured $I_W$. But the slope of $I_W$ on TaR/ TpR gradually changes when $V_{DD}$ goes down but does not show a two-stair step as in Fig. 2.29(c). Thus, $I_W$ was found to deviate from a normal distribution but does not show "two-mode" distribution as in WSNM at low $V_{DD}$.

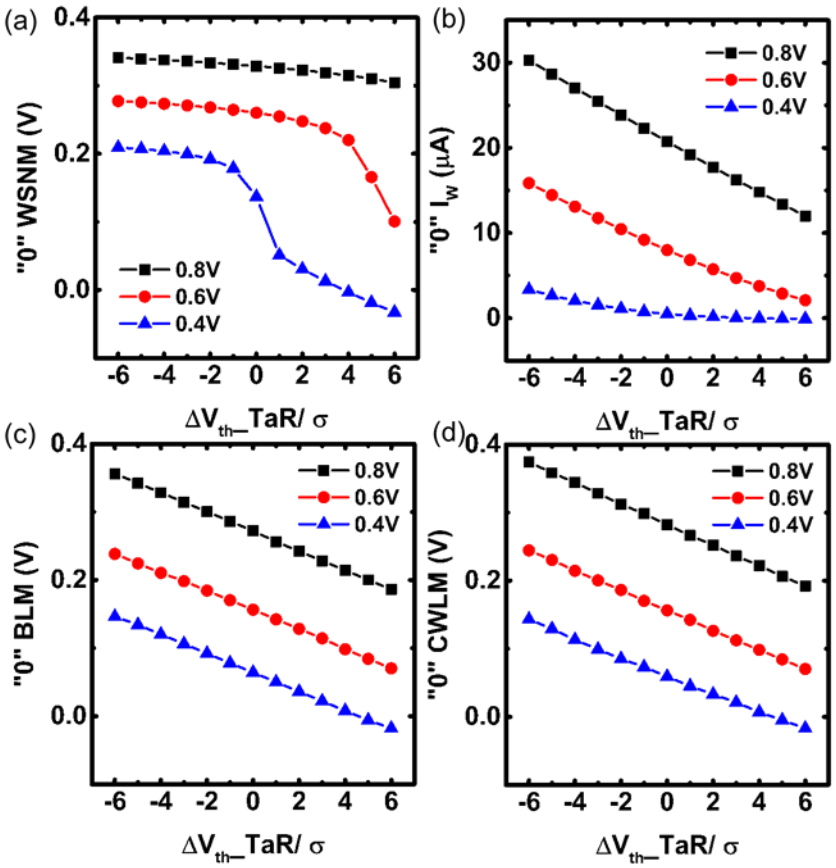


Fig. 2.33. Dependence of (a) "0" WSNM, (b) "0" $I_W$, (c) "0" BLM, and (d) "0" CWLM on $\Delta V_{th}$ in TaR at $V_{DD}$ = 0.4 V.

The demonstration of normal distributions of BLM and CWLM is more difficult than the discussion of non-normality in WSNM and $I_W$. Therefore, here, we limit our discussions in -6 sigma < $\Delta V_{th}$ < -6 sigma, which is also enough for real memory yield estimations. In both Fig. 2.31 and Fig. 2.32, BLM and CWLM keep good linearity with $\Delta V_{th}$ of all six cell transistors down to $V_{DD}$ = 0.4 V[4] and is consistent with our experimental demonstrations.

Finally, to give a clear comparison among these four write stability metrics, selecting one performance factor in write operation – TaR, we plotted the dependences of each defined WNM on $\Delta V_{th}$ at different $V_{DD}$ in Fig. 2.33.

[4] Here, we define the negative WNM in BLM/ CWLM at low $V_{DD}$. Take "0" BLM as an example. In order t give further information of unstable SRAM cells, $V_{BLR}$ is discharged from $V_{DD}$ down to $-V_{DD}/2$, instead of zero. And "0" CWLM is defined as $V_{BLR}$ at which voltage of VL flips.

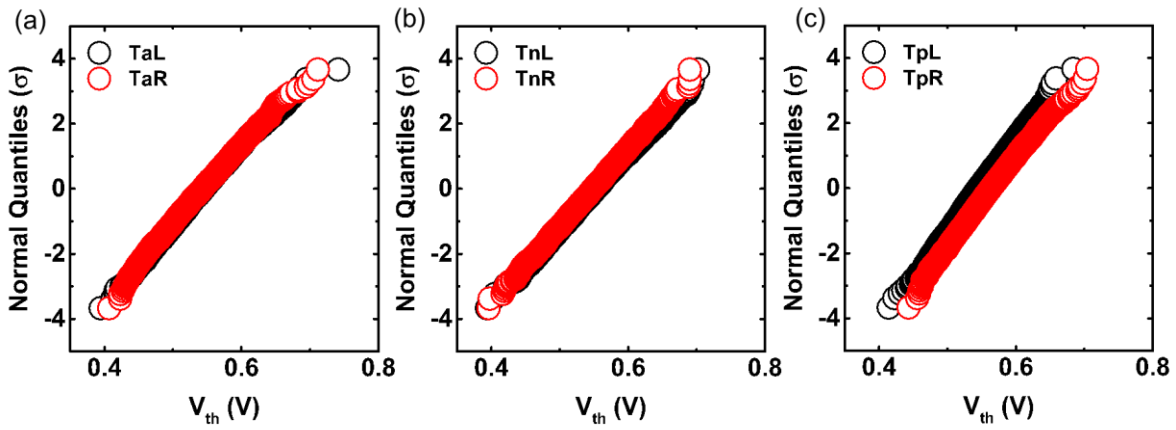


Fig. 2.34. Cumulative plots of $V_{th}$ of (a) TaL/ TaR, (b) TnL/ TnR, and (c) TpL/ TpR of 4 kb bulk SRAM cells. Drain bias: $|V_{ds}|$ = 50 mV. Substrate bias: $V_{bsn}$ = -0.8 V, and $V_{bsp}$ = 0.25 V.

## 2.4 Measurement Results in Bulk SRAM

### 2.4.1 Write Stability Characterization and Four Metrics' Comparison

Write stability characterization is performed in 4 kb bulk SRAM cells. Fig. 2.34 shows $V_{th}$ distribution of cell transistors in 4 kb bulk SRAM cells and a large variability is demonstrated. Considering the average value of $V_{th}$ in bulk transistors is 0.2 V larger than that of $V_{th}$ in SOTB transistors, for further comparison with SOTB SRAM, bulk SRAM cells are measured at a higher $V_{DD}$ = 1.0 V, 0.8 V, and 0.6 V, respectively. Take $V_{DD}$ = 1.0 V as an example, Fig. 2.35 shows measured waveforms of (a) write butterfly curve, (b) write N-curve, (c) bit-line method, and (d) word-line method. Write noise margin is defined as the minimum between logic "0" and logic "1", and their distributions are shown in Fig. 2.36. BLM and CWLM show good normal distribution even when $V_{DD}$ goes down to 0.6 V, demonstrating that bit-line method and word-line method are good candidates for write yield estimation for bulk SRAM at low $V_{DD}$. On the other hand, WSNM and $I_W$ clearly deviates from normal distribution at $V_{DD}$ = 0.6 V, arriving at the same conclusion in SOTB SRAM.

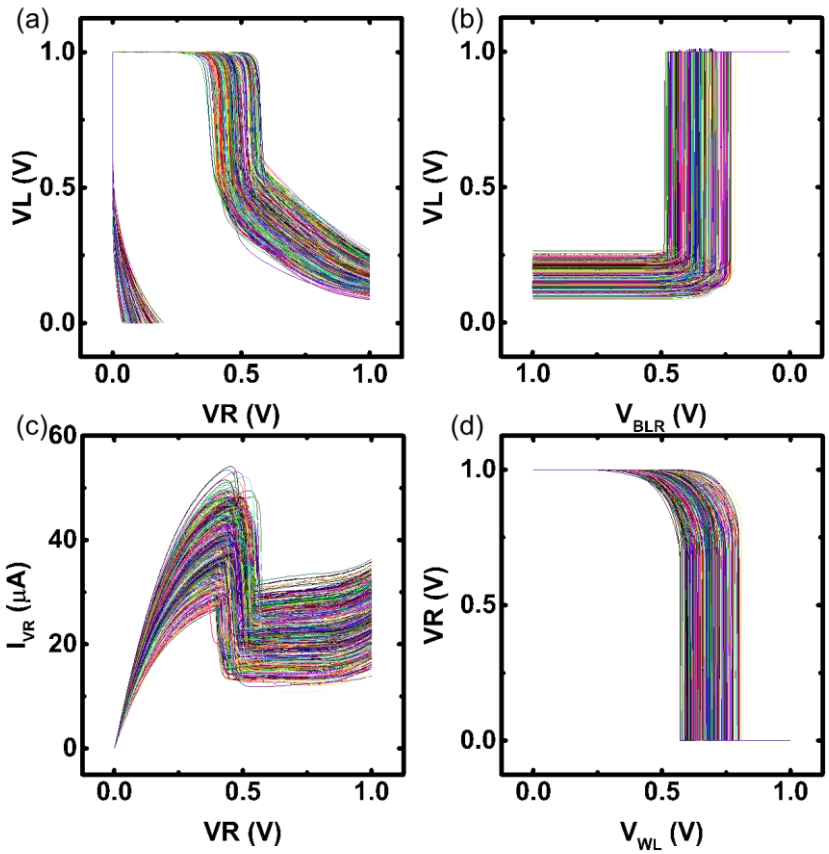


Fig. 2.35. Measured (a) write butterfly curves, (b) write N-curves, (c) waveforms in bit-line method, and (d) waveforms in word-line method of 4 kb bulk SRAM cells at $V_{DD}$ = 1.0 V. Substrate bias: $V_{bsn}$ = -0.8 V, and $V_{bsp}$ = 0.25 V.

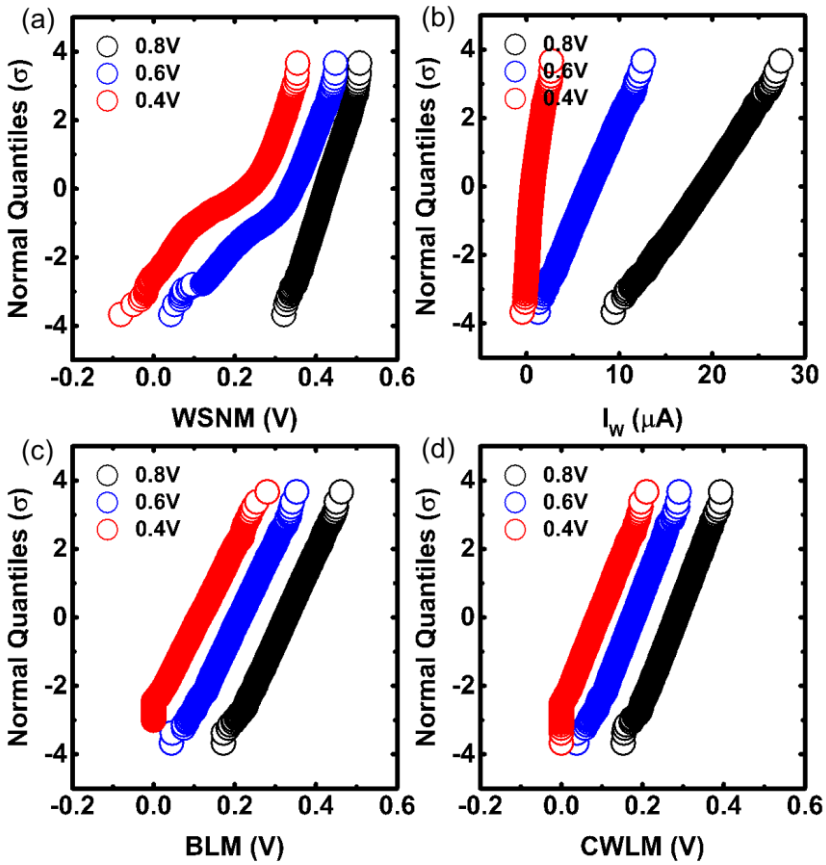


Fig. 2.36. Cumulative plots of (a) WSNM, (b) $I_W$, (c) BLM, and (d) CWLM of 4 kb bulk SRAM cells at $V_{DD}$ = 1.0 V (in black), 0.8 V (in blue), and 0.6 V (in red). WSNM, $I_W$, BLM, CWLM are defined as the minimum of "0" and "1" write. Substrate bias: $V_{bsn}$ = -0.8 V, $V_{bsp}$ = 0.25 V.

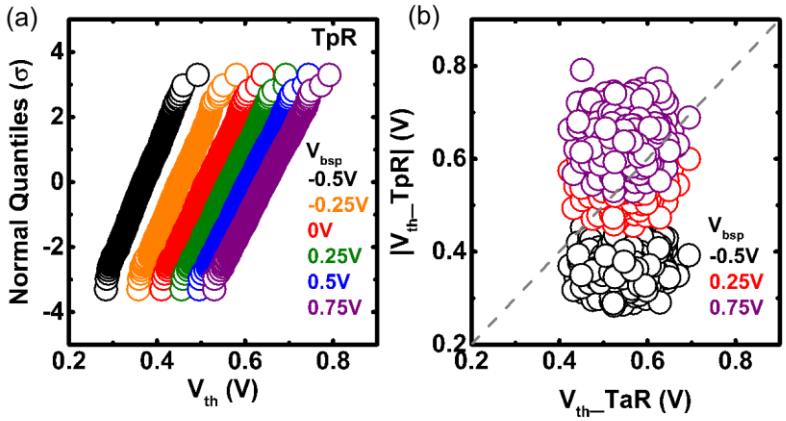


Fig. 2.37. (a) Cumulative plots of TpR's $V_{thc}$ in 1 kb bulk SRAM cells at $V_{bsp}$ = -0.5 V (black), -0.25 V (orange), 0 V (red), 0.25 V (green), 0.5 V (blue), and 0.75 V (purple). Drain bias: -50 mV. (b) Scatter plot of TpR's $V_{thc}$ versus TaR's $V_{thc}$. Drain bias: $|V_{ds}|$ = 50 mV. Substrate bias: $V_{bsn}$ = -0.8 V; $V_{bsp}$ = -0.5 V (black), 0.25 V (red), and 0.75 V (purple).

### 2.4.2 Comparison with SOTB SRAM

The conclusion about comparison among four write stability metrics is the same whether in SOTB or bulk SRAM at low $V_{DD}$. However, due to a larger variability in bulk technology, the distribution patterns in write butterfly curve and write N-curve at low $V_{DD}$ are different. And the following part mainly focuses on discussions about non-normality in write butterfly curve and write N-curve in bulk SRAM cells at low $V_{DD}$.

Similar to SOTB SRAM, substrate bias is applied to pFETs in bulk SRAM. Fig. 2.37(a) shows cumulative plots of TpR's $V_{th}$ in 1 kb bulk SRAM cells at different $V_{bsp}$. Here, $V_{bsp}$ is selected to shift $V_{th}$ with the same value as in Fig. 2.20(a) for comparison. The scatter plots of TpR's $V_{th}$ versus TaR's $V_{th}$ at $V_{bsp}$ = -0.5V, 0.25 V, and 0.75 V are plotted in Fig. 2.37(b). Fig. 2.38 and Fig. 2.39 show write butterfly curves of 1 kb bulk SRAM cells and histogram plots of "0" WSNM, respectively, at $V_{DD}$ = 0.6 V when different $V_{bsp}$ is applied. According to Fig. 2.37(b), TpR's $V_{th}$ centers on the same value as TaR at $V_{bsp}$ = 0.25 V. When $V_{bsp}$ = -0.5 V is applied, due to a larger variability in bulk technology, TpRs in some cells are still weaker than TaRs, generating Mode-I write butterfly curves in Fig. 2.38(a). Whereas, TpRs in some cells are still stronger than TaRs at $V_{bsp}$ = 0.75 V. Thus, some curves in Fig. 2.38(f) correspond to Mode II. That is the reason why "0" WSNM in Fig. 2.39(a) and Fig. 2.39(f) both show a tailed distribution highlighted by black dashed circle, which is different from SOTB case in Fig. 2.22(a) and Fig. 2.22(f). And the entanglement of Mode I and Mode II can be seen more clear in Fig. 2.39(b) and Fig. 2.39(c).

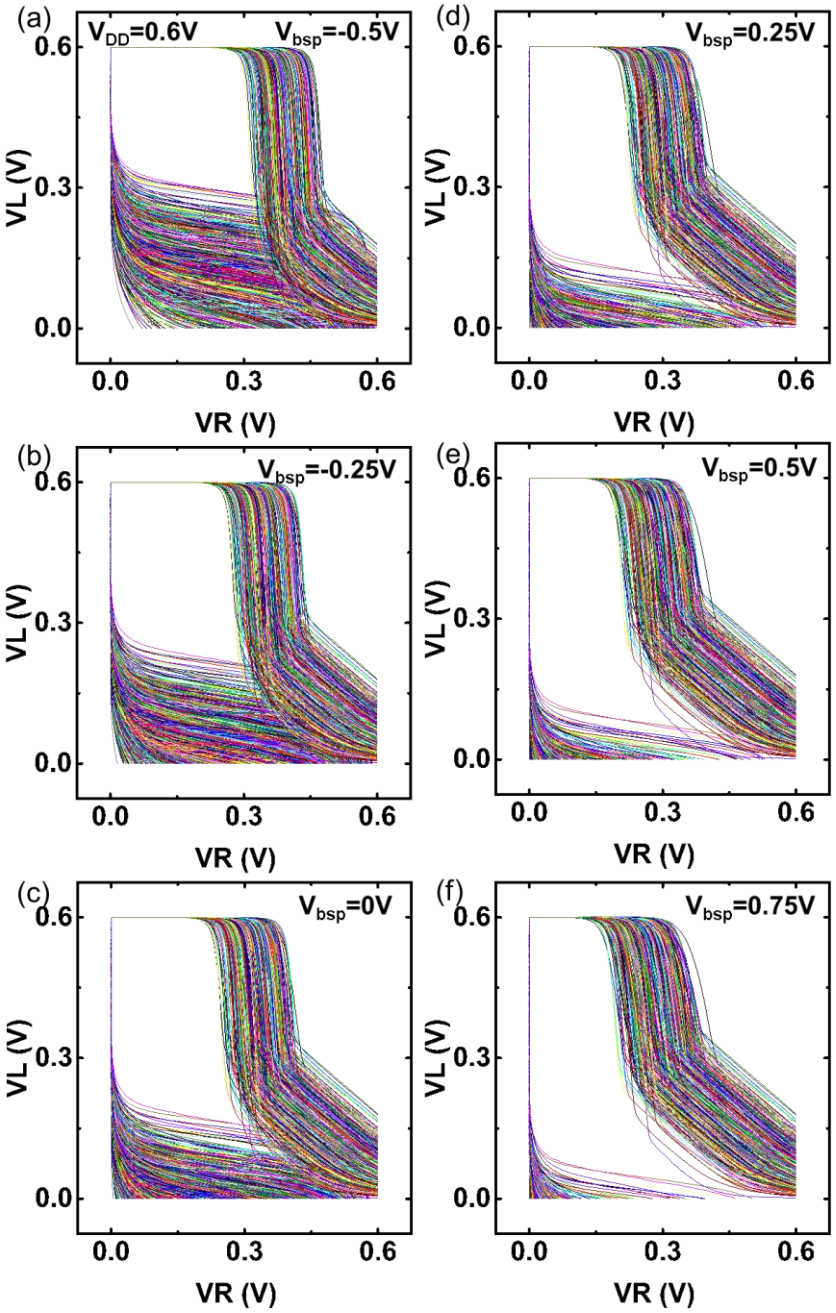


Fig. 2.38. Write butterfly curves of 1 kb bulk SRAM cells in "0" write at $V_{DD}$ = 0.6 V. Substrate bias: $V_{bsn}$ = -0.8 V, (a) $V_{bsp}$ = -0.5 V, (b) $V_{bsp}$ = -0.25 V, (c) $V_{bsp}$ = 0 V, (d) $V_{bsp}$ = 0.25 V, (e) $V_{bsp}$ = 0.5 V, (f) $V_{bsp}$ = 0.75 V.

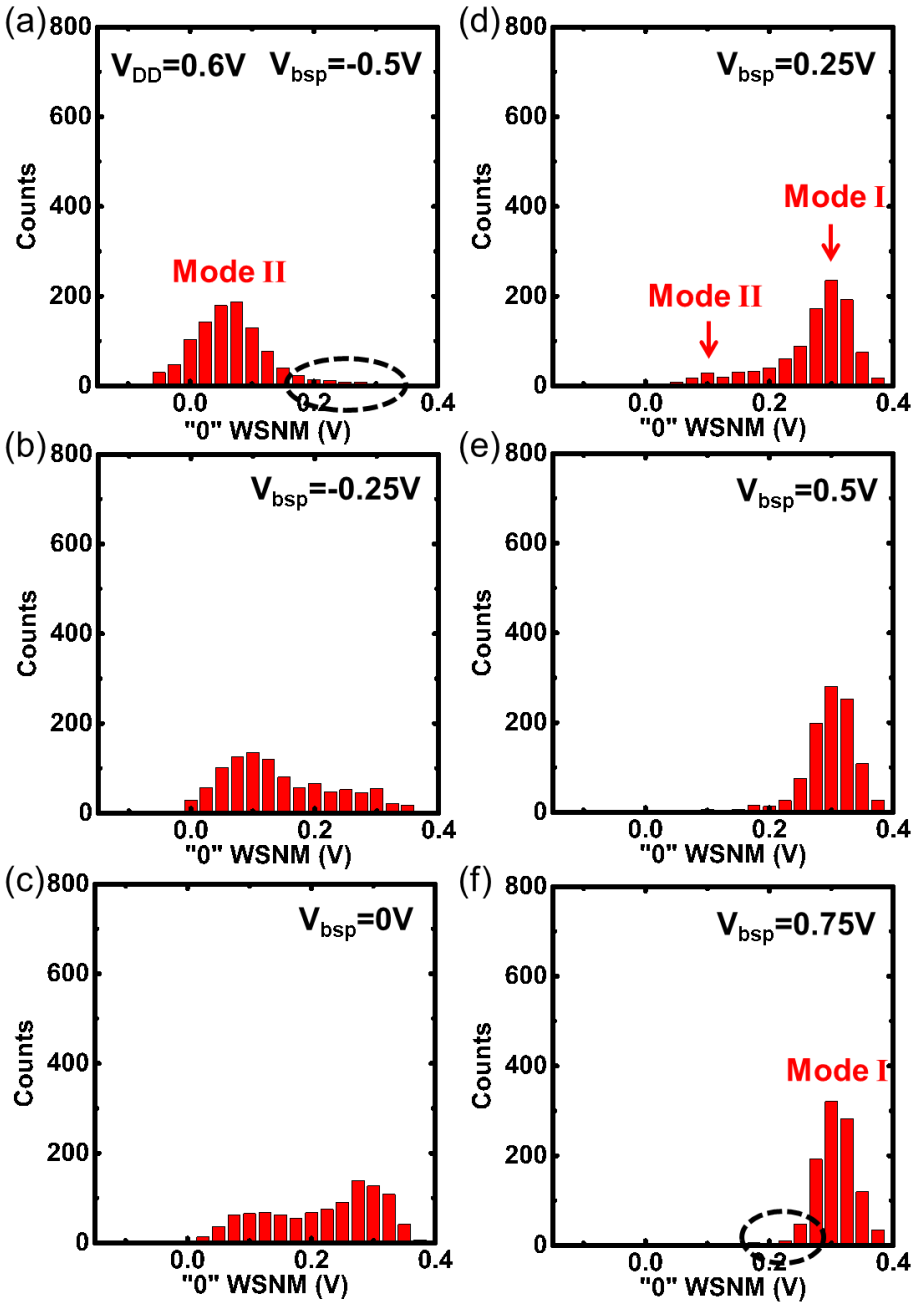


Fig. 2.39. Histogram plots of 1 kb bulk SRAM cells in "0" write at $V_{DD}$ = 0.6 V. Substrate bias: $V_{bsn}$ = -0.8 V, (a) $V_{bsp}$ = -0.5 V, (b) $V_{bsp}$ = -0.25 V, (c) $V_{bsp}$ = 0 V, (d) $V_{bsp}$ = 0.25 V, (e) $V_{bsp}$ = 0.5 V, (f) $V_{bsp}$ = 0.75 V.

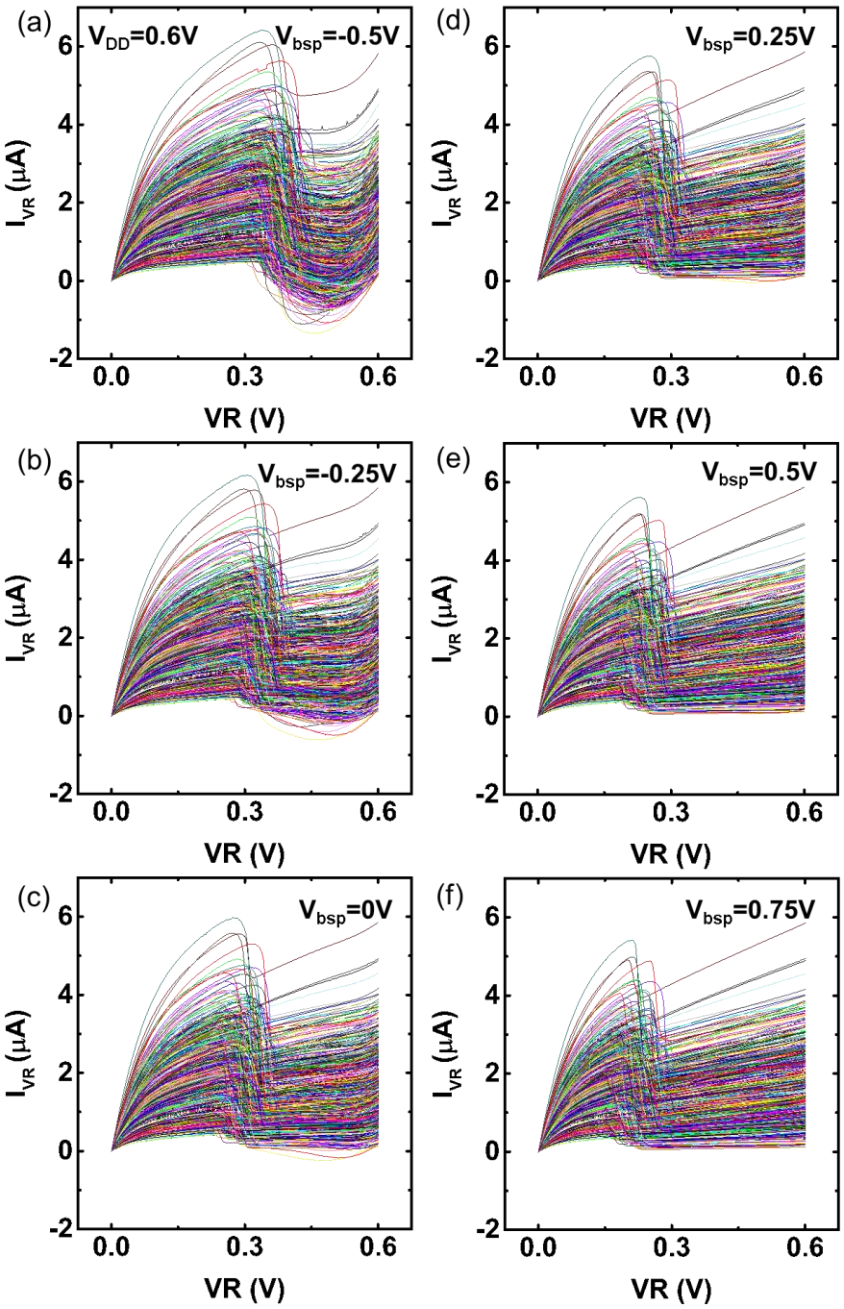


Fig. 2.40. Write N-curves of 1 kb bulk SRAM cells in "0" write at $V_{DD}$ = 0.6 V. Substrate bias: $V_{bsn}$ = -0.8 V, (a) $V_{bsp}$ = -0.5 V, (b) $V_{bsp}$ = -0.25 V, (c) $V_{bsp}$ = 0 V, (d) $V_{bsp}$ = 0.25 V, (e) $V_{bsp}$ = 0.5 V, (f) $V_{bsp}$ = 0.75 V.

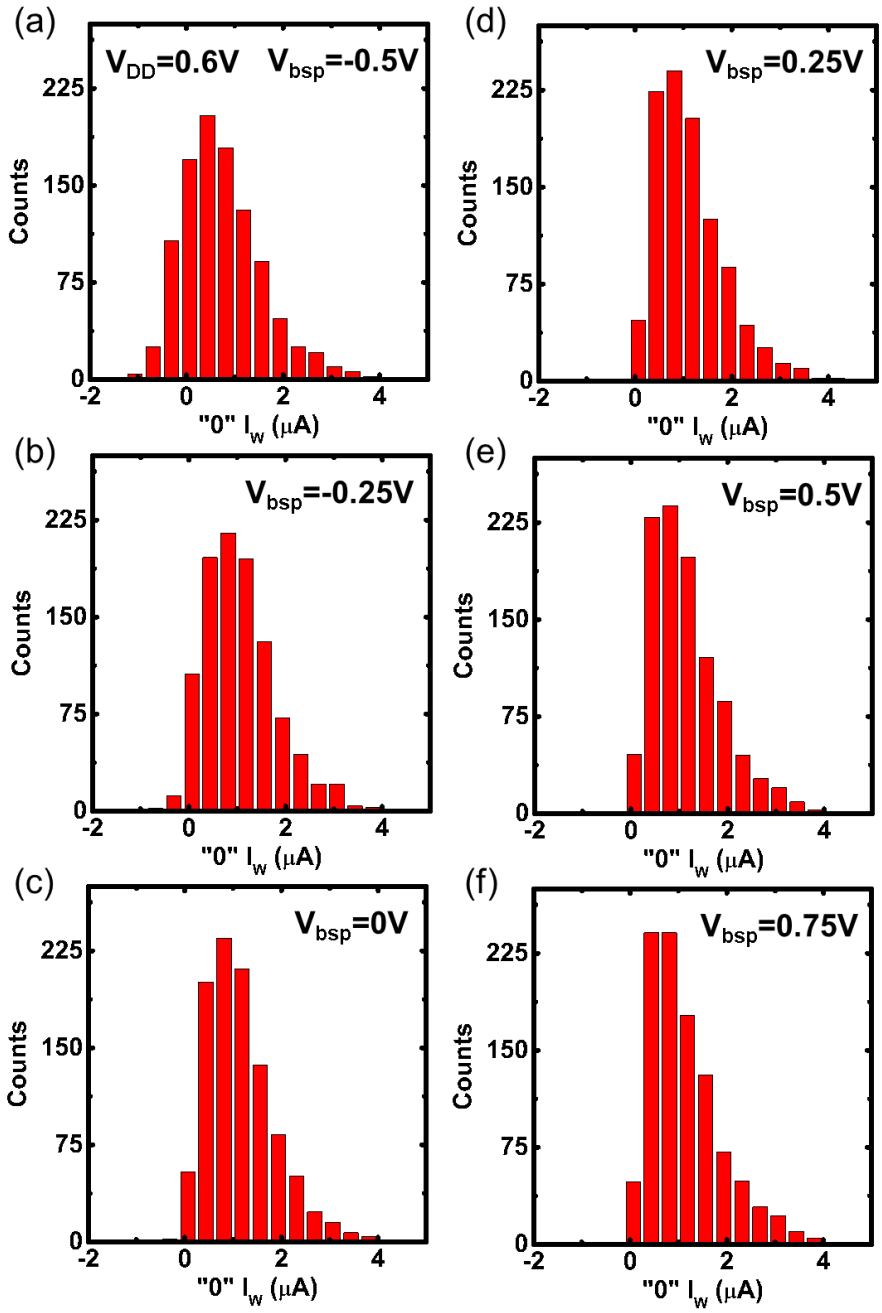


Fig. 2.41. Histogram plots of "0" $I_W$ of 1 kb bulk SRAM cells in "0" write at $V_{DD}$ = 0.6 V. Substrate bias: $V_{bsn}$ = -0.8 V, (a) $V_{bsp}$ = -0.5 V, (b) $V_{bsp}$ = -0.25 V, (c) $V_{bsp}$ = 0 V, (d) $V_{bsp}$ = 0.25 V, (e) $V_{bsp}$ = 0.5 V, (f) $V_{bsp}$ = 0.75 V.

Things are also different in write N-curve. Fig. 2.40 and Fig. 2.41 show write N-curves of 1 kb bulk SRAM cells at histogram plots of "0" $I_W$, respectively, at $V_{DD}$ = 0.6 V when different $V_{bsp}$ is applied. Due to a larger variability in cell transistors' $V_{th}$, whether at $V_{bsp}$ = -0.5 V or 0.75 V, there is always a mixture of Mode-i and Mode-ii write N-curve from Fig. 2.40(a) to Fig. 2.40(f). That is the reason why "0" $I_W$ deviates from normal distribution in Fig. 2.41(a)-(f). Therefore, compared to SOTB case, a larger $V_{th}$ shift is necessary to compensate the larger variability in $V_{th}$ distribution of bulk SRAM cell transistors to distinguish two modes at low $V_{DD}$.

### 2.4.3 Further Measurement in New Bulk Chip

We also did measurements in another SRAM chip [46-47] based on different bulk technology at high $V_{DD}$. Fig. 2.42 shows cumulative plots of four metrics at $V_{DD}$ = 0.9 V. Here, write noise margin is defined as the minimum between logic "0" and logic "1". WSNM from write butterfly curve and $I_W$ from write N-curve deviate from normal distributions while the other two metrics have good normality up to ±4 sigma.

In addition, Fig. 2.43 gives scatter plots of each two metrics in order to understand the correlation among the four metrics. The best correlation can be found in the scatter plot between BLM and CWLM in Fig. 2.43(a), due to the fact that both metrics can monitor the state flip in SRAM cells. The worst correlation is found between WSNM and $I_W$ in Fig. 2.43(f). For all the plots, the red circle indicates the same SRAM cell with worst write noise margin, meaning that all four write stability metrics are good indicators to predict the write failure in SRAM arrays.

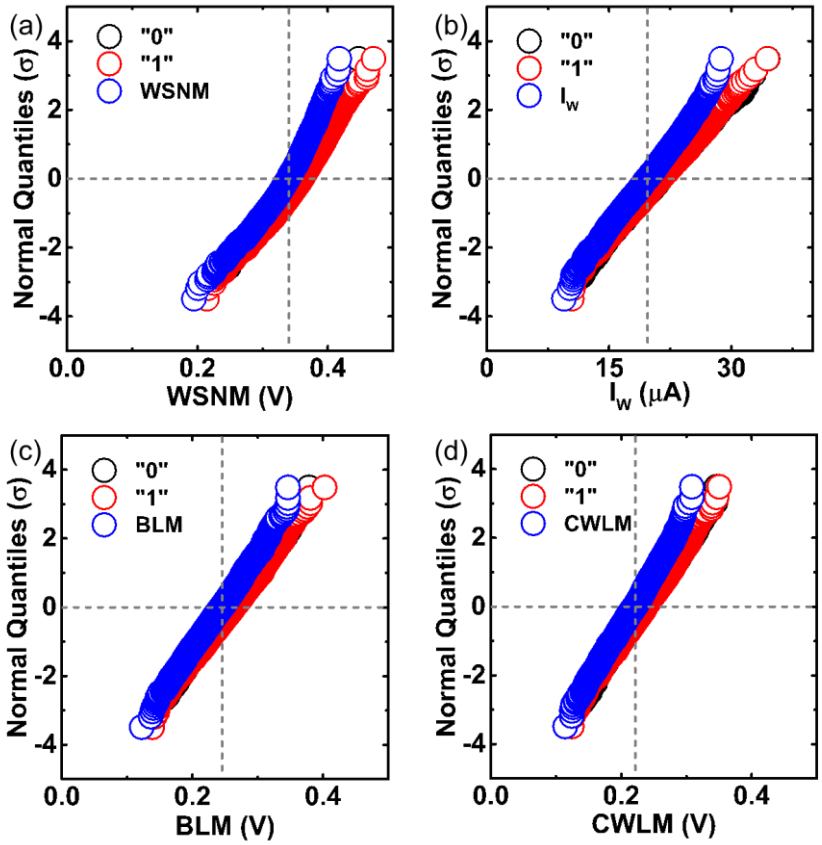


Fig. 2.42. Cumulative plots of (a) WSNM, (b) $I_W$, (c) BLM, and (d) CWLM in 1 kb bulk SRAM cells at $V_{DD}$ = 0.9 V. Write noise margin is the minimum between "0" and "1" write.

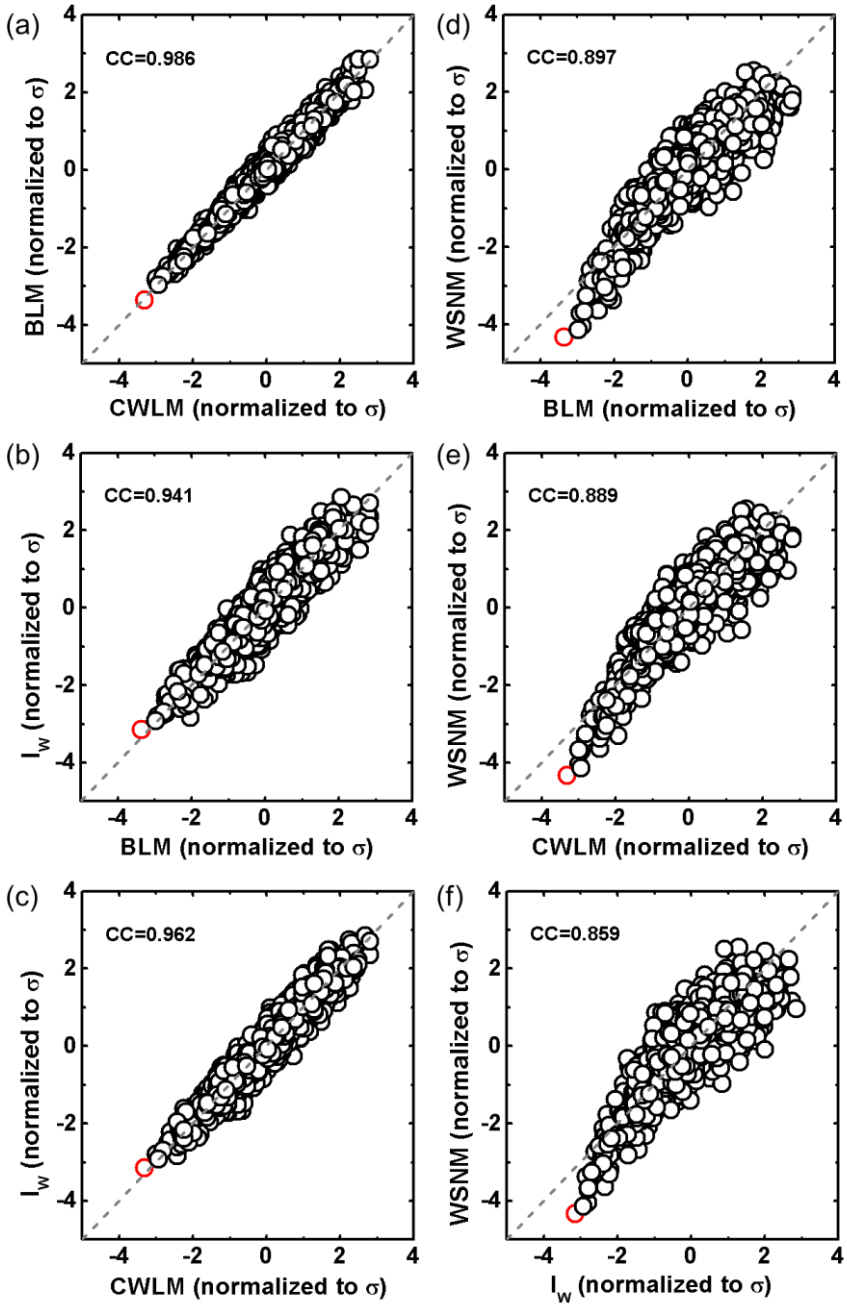


Fig. 2.43. Scatter plots of (a) BLM versus CWLM, (b) $I_W$ versus BLM, (c) $I_W$ versus CWLM, (d) WSNM versus BLM, (e) WSNM versus CWLM, and (f) WSNM versus $I_W$ of 1 kb bulk SRAM cells at $V_{DD}$ = 0.9 V. The red circle indicates the SRAM cell with smallest write noise margin.

## 2.5 Summary

This chapter presents the results for write stability characterization using several commonly used write stability metrics in both SOTB and bulk SRAM cells at low $V_{DD}$. Due to large time-zero variability, such as RDF, LER, WFV and so on, write performance always varies in different SRAM cells. Thus, sufficient margin with 6 standard deviations is necessary for robust SRAM applications. However, a direct measurement up to ±6 sigma costs too long a time and is not practical on the way of test for volume production. Therefore, a proper write stability metrics is of great importance since it enables a reasonable extrapolation for yield estimation based on only a small number of samples, which helps save test duration.

Four write stability metrics – including WSNM from write butterfly curve, $I_W$ from write N-curve, BLM from bit-line method and CWLM from word-line method – are compared firstly in SOTB SRAM. Due to a large elimination from RDF with intrinsic channel, SOTB SRAM cells have been demonstrated to operate down to $V_{DD}$ = 0.4 V [32]. BLM/ CWLM keeps good normality from high to low $V_{DD}$. In addition, both methods can monitor write-trip point in write operation. Thus, BLM and CWLM are good metrics for write yield estimation even at low $V_{DD}$. On the contrary, WSNM/ $I_W$ is found to deviate from normal distribution when $V_{DD}$ goes down to 0.4 V. In particular, WSNM shows two-mode distribution, which has been discovered for the first time. By establishing the correlation between WSNM in SRAM cells and $V_{th}$ in cell transistors, we find that it is the relative strength of the load transistor over the access transistor that dominates two modes in write butterfly curve. At high $V_{DD}$, with a lower hole mobility as well as a smaller gate width/length ratio, the load transistor is usually weaker than the access transistor resulting in Mode I. When $V_{DD}$ goes down, in some SRAM cells, the load transistor with a smaller $V_{th}$ can be stronger than the access transistor and induces Mode II. Due to a mixture of Mode I and Mode II, its write noise margin shows a

two-mode distribution which is not proper for yield estimation at low $V_{DD}$. A similar explanation can be applied to write N-curve, considering the origin of two modes as strength of the load transistor.

Considering the limited range of measurement, HSPICE simulations using SOTB model are performed to help predict the distribution of each write stability metric up to ±6 sigma. Assuming $V_{th}$ of nFETs/ pFETs follows a normal distribution and each cell transistors' $V_{th}$ can be considered independently, the metric is predicted to follow a normal distribution if its variance is linearly correlated with the variance of $V_{th}$. We found that the both BLM and CWLM depends linearly on $\Delta V_{th}$ of each cell transistor down to 0.4 V, indicating that BLM and CWLM are good metrics up to ±6 sigma at low $V_{DD}$. On the contrary, the slopes of WSNM versus $\Delta V_{th}$ and $I_W$ versus $\Delta V_{th}$ do not keep constant at 0.4 V. Thus, the non-normality of WSNM and $I_W$ can be predicted at low $V_{DD}$.

Other than SOTB technology, write stability characterization is also performed in bulk SRAM cells at low $V_{DD}$ to demonstrate the universality of our findings. Considering a larger variability in bulk technology, for a reliable SRAM operation, $V_{DD}$ = 0.6 V instead of 0.4 V is selected. Same as in SOTB case, BLM/ CWLM shows good normality while WSNM/ $I_W$ deviates from normal distribution in bulk SRAM cells at $V_{DD}$ = 0.6 V. A detailed explanation of two modes in both write butterfly curve and write N-curve is given. Therefore, our findings in SOTB SRAM cells are also applicable to bulk technology. The only difference is that, due to a larger variability in bulk technology, the two modes in write butterfly curve/write N-curve cannot be easily distinguished as in SOTB technology resulting in a more limited range of applicability for these two write stability metrics.

# Chapter 3

# Proposed New Write Stability Metric for Yield Estimation

## 3.1 Introduction

### 3.1.1 Drawback of Conventional Write Butterfly Curve

Write butterfly curve (BC), which we call conventional write BC here in order to distinguish from the proposed new one [48], consists of two VTCs of cross-coupled inverters. The main difference from read BC is that one bit line is grounded for write operation. WSNM from conventional write BC has been a popular write stability metric since it is easy to understand. As well, its analytical model can be obtained on the basis of the basic MOS model neglecting the second-order effects [2].

However, conventional write BC has two main drawbacks. The first one is the bad correlation between its write noise margin and $V_{th}$ of performance-dominant cell transistors. Take access transistor as an example. A stable write operation requires a strong write current through access transistor down to adjacent bit line. According to Monte Carlo simulation results with 32,000 SRAM cells in Fig. 3.1(b), WSNM only shows a very weak correlation with $V_{th}$ of access transistor (N3) only after $\Delta V_{th}$ exceeds 0.14 V [49]. The second one is about its distribution. According to our discussions in chapter 2, WSNM from conventional

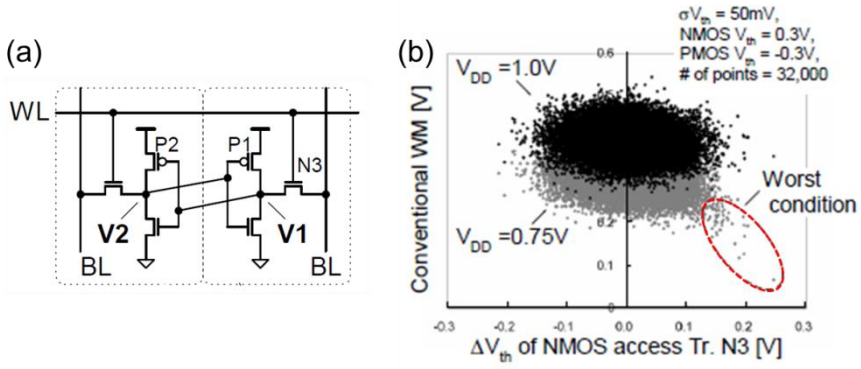


Fig. 3.1. (a) Schematic of 6-T SRAM cell. (b) Scatter plots of WSNM versus $\Delta V_{th}$ of N3 at $V_{DD}$ = 1.0 V and 0.75 V. From Ref. [49].

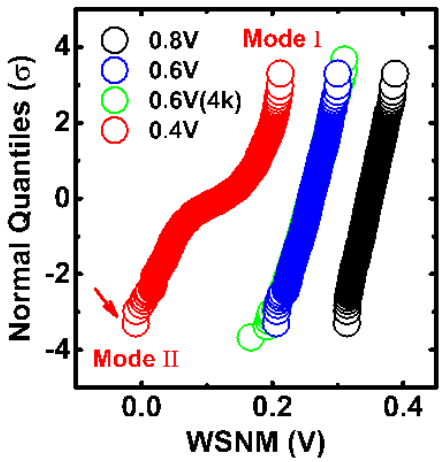


Fig. 3.2. Cumulative plots of WSNM in 1 kb SOTB SRAM cells at $V_{DD}$ = 0.8 V (black), 0.6 V (blue), and 0.4 V (red). Black and blue circles correspond to Mode I. Green circles indicate WSNM of 4 kb cells at $V_{DD}$ = 0.6 V. Substrate bias: $V_{bsn}$ = -1 V, $V_{bsp}$ = 1 V.

write butterfly curve deviates from normal distribution at low $V_{DD}$. Again, Fig. 3.2 shows the experimental results of 1 kb SOTB SRAM cells. It can be seen that extrapolating Mode-I distribution to e.g. 6 sigma will result in serious overestimation of the worst case margin. For example, a trace of Mode-II is seen at $V_{DD}$ = 0.6 V, if the number of measured cells is increased to 4 kb (in green).

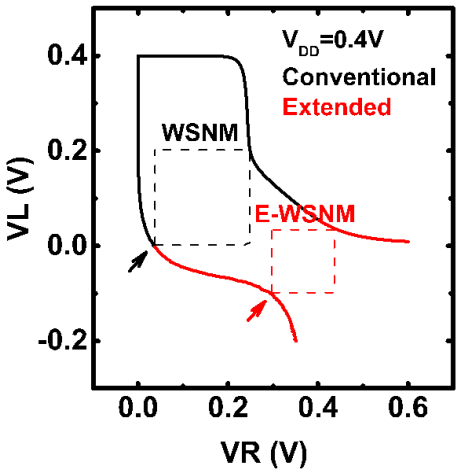


Fig. 3.3. Conventional and extended BC of one SRAM cell at $V_{DD}$ = 0.4 V. WSNM/ E-WSNM is extracted as the side of black/ red square. Substrate bias: $V_{bsn}$ = -1 V, $V_{bsp}$ = 1 V.

### 3.1.2 Definition of Extended Write Butterfly Curve

On the other hand, it is found that write failure occurs when WSNM shows Mode-II behavior. Therefore, Mode-II distribution can be used for write yield estimation and is also named as failure mode. However, it is not always possible to measure sufficient number of cells to fully reveal Mode-II part of the distribution using DMA test structures. In order to effectively detect Mode-II behavior of the cells, we propose the extended write BC. "0" write is focused.

In the conventional write BC, the voltage sweeping range is from zero to $V_{DD}$. In the extended write BC, the voltage sweeping range is extended. By extending the voltage sweeping range of VL below zero and that of VR beyond $V_{DD}$, Fig. 3.3 gives one example at $V_{DD}$ = 0.4 V. In the conventional write BC, due to the limited voltage sweeping range, the smallest fitting square between the curves is the one drawn in black, whereas in extended write BC (the extended part is plotted in red), its write noise margin (E-WSNM) is extracted from the red square.

## 3.2 Measurement Results in SOTB SRAM

### 3.2.1 Characterization using Extended Write Butterfly Curve

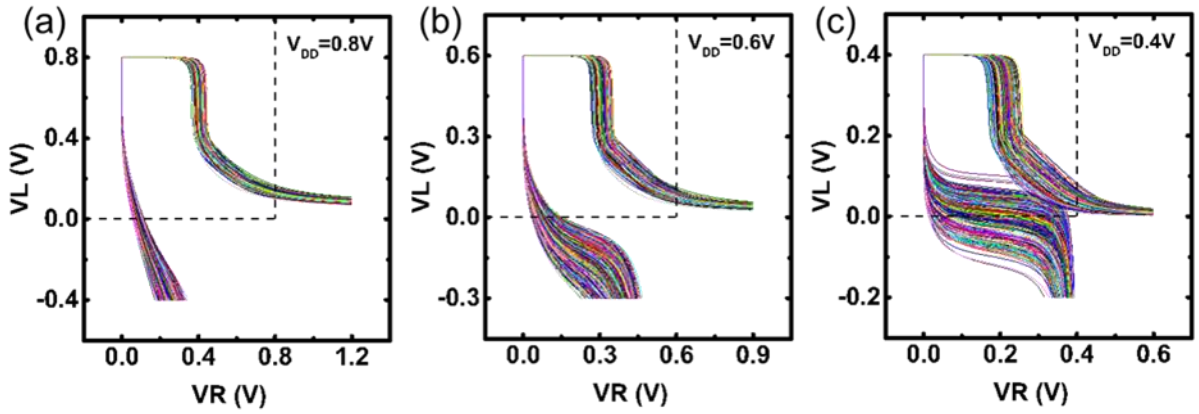


Fig. 3.4. Extended BCs of 1 kb SOTB SRAM cells at $V_{DD}$ = (a) 0.8V, (b) 0.6V, and (c) 0.4 V. The black dashed lines indicate voltage sweep range of conventional BCs. Substrate bias: $V_{bsn}$ = -1 V, $V_{bsp}$ = 1 V.

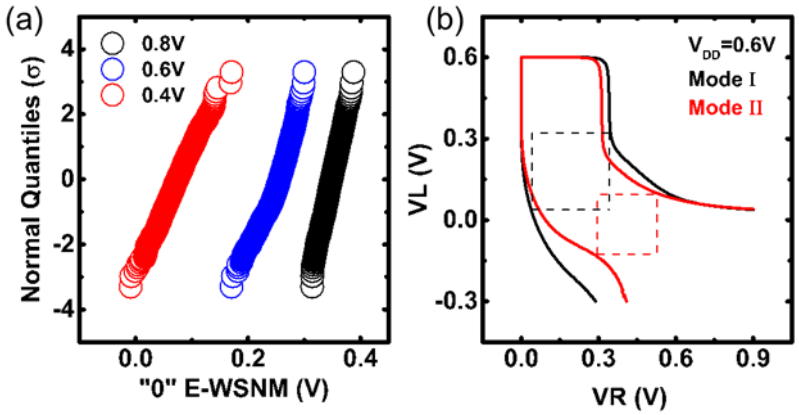


Fig. 3.5. (a) Cumulative plots of "0" E-WSNM in 1 kb SOTB SRAM cells at $V_{DD}$ = 0.8 V (black), 0.6 V (blue), and 0.4 V (red). (b) Mode-I and Mode-II extended BCs of two cells at $V_{DD}$ = 0.6 V. WSNM/ E-WSNM is extracted from black/ red square. Substrate bias: $V_{bsn}$ = -1 V, $V_{bsp}$ = 1 V.

Fig. 3.4 shows extended write BCs of 1 kb SOTB SRAM cells at $V_{DD}$ = 0.8 V, 0.6 V, and 0.4 V. The black dashed lines indicate voltage sweeping range of conventional write BCs. Through the measured waveforms, Mode-II behavior becomes more clear. Corresponding cumulative plots of E-WSNM at each $V_{DD}$ are plotted in Fig. 3.5(a). It can be seen that, at $V_{DD}$ = 0.4 V, E-WSNM is fully in Mode-II behavior and shows good normality. At $V_{DD}$ = 0.6 V, Mode-II part clearly emerges, though Mode I is still dominant. Fig. 3.5(b) shows typical extended write BCs for Mode I and Mode II at $V_{DD}$ = 0.6 V.

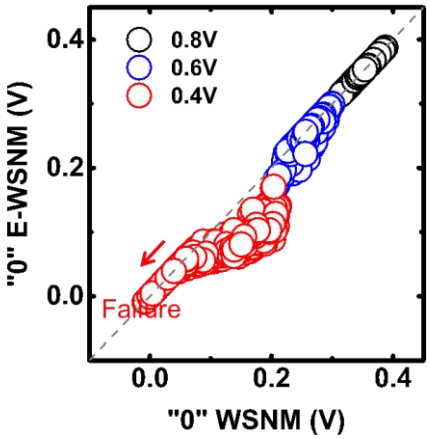


Fig. 3.6. Scatter plots of "0" E-WSNM versus "0" WSNM in 1 kb SOTB SRAM cells at $V_{DD}$ = 0.8 V (black), 0.6 V (blue) and 0.4 V (red). Substrate bias: $V_{bsn}$ = -1 V, $V_{bsp}$ = 1 V.

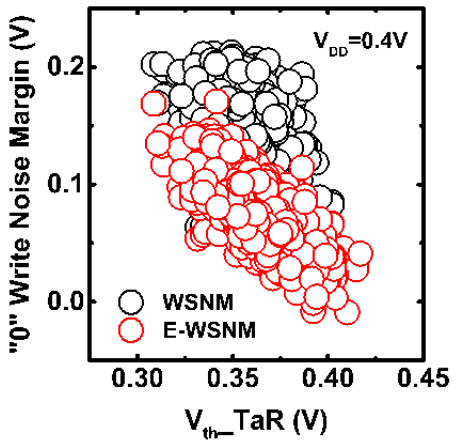


Fig. 3.7. Scatter plot of "0" E-WSNM (red)/ "0" WSNM (black) versus $V_{th}$ of TaR in 1 kb SOTB SRAM cells at $V_{DD}$ = 0.4 V. Drain bias for $V_{th}$ measurement is 50 mV. Substrate bias: $V_{bsn}$ = -1 V, $V_{bsp}$ = 1 V.

### 3.2.2 Comparison with Conventional Write Butterfly Curve

To compare extended write BC and conventional write BC, Fig. 3.6 shows the scatter plot between E-WSNM and WSNM at $V_{DD}$ = 0.8 V, 0.6 V, and 0.4 V. Good correlation is found at $V_{DD}$ = 0.8 V and towards failure edge at $V_{DD}$ = 0.4 V. That means E-WSNM and WSNM share the same write failure at $V_{DD}$ = 0.4 V. Also considering its normal distribution, E-WSNM is demonstrated as a good write stability metric for yield estimation at low $V_{DD}$.

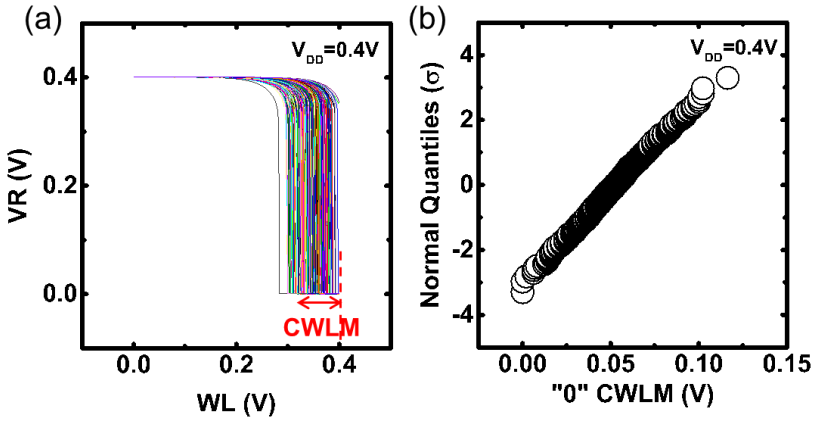


Fig. 3.8. Measured waveforms of word-line method in 1 kb SOTB SRAM cells at $V_{DD}$ = 0.4 V. CWLM is defined as the difference between $V_{DD}$ and $V_{WL}$ at which VR flips. Substrate bias: $V_{bsn}$ = -1 V, $V_{bsp}$ = 1 V.

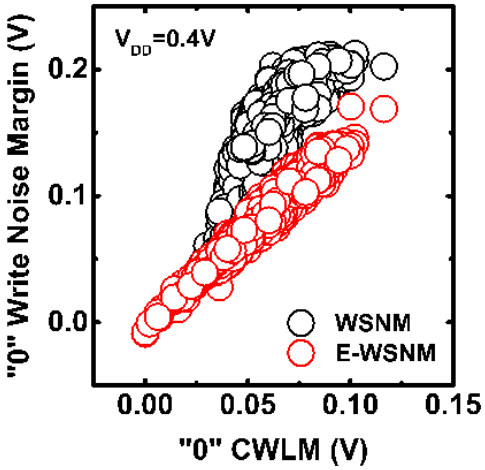


Fig. 3.9. Scatter plot of "0" E-WSNM (red)/ "0" WSNM (black) versus "0" CWLM in 1 kb SOTB SRAM cells at $V_{DD}$ = 0.4 V. Substrate bias: $V_{bsn}$ = -1 V, $V_{bsp}$ = 1 V.

Further evidence is given as follows. As mentioned in the introduction section, write noise margin in the SRAM cell should have good correlation with $V_{th}$ in the load transistor. Fig. 3.7 shows the scatter plots between E-WSNM/WSNM and $V_{th}$ of access transistors (TaR). Compared to WSNM, E-WSNM shows a much better correlation with TaR's $V_{th}$, supports that extended write butterfly curve can measure correct write noise margin in SRAM cells. Besides, as confirmed in Chapter 2, CWLM from word-line method is a good write stability metric at low $V_{DD}$. Fig. 3.8 shows the measured waveforms and extracted CWLM's

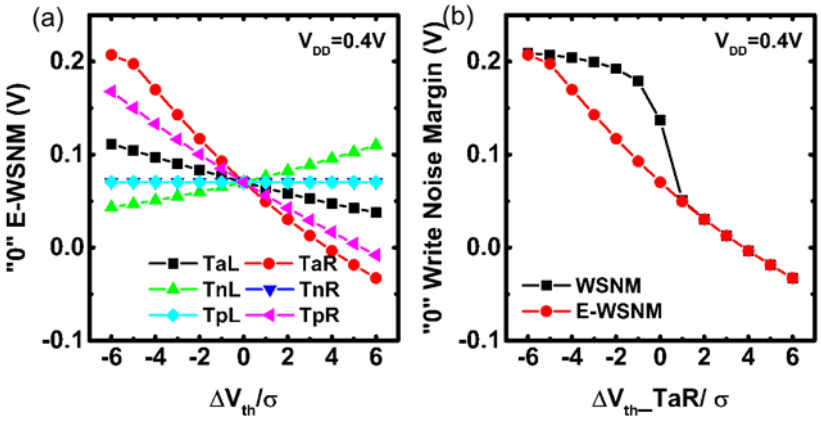


Fig. 3.10. (a) Dependence of "0" E-WSNM on $\Delta V_{th}$ in six cell transistors at $V_{DD}$ = 0.4 V. (b) Dependence of "0" E-WSNM (in red)/ "0" WSNM (in black) on $\Delta V_{th}$ in TaR at $V_{DD}$ = 0.4 V.

distribution of 1 kb SOTB SRAM cells at $V_{DD}$ = 0.4 V. Same as previous results [42], CWLM shows good normality. Setting CWLM as reference metric, Fig. 3.9 shows the scatter plots between E-WSNM/WSNM and CWLM at $V_{DD}$ = 0.4 V and a good correlation between E-WSNM and CWLM is demonstrated.

## 3.3 HSPICE Simulation Results in SOTB SRAM

HSPICE simulations are presented in this sub-section to discuss distribution of E-WSNM at low $V_{DD}$. Simulation method is the same as in section 2.3. Fig. 3.10(a) shows the dependence of "0" E-WSNM on $\Delta V_{th}$ of six cell transistors at $V_{DD}$ = 0.4 V. Different from Fig. 2.40(c), "0" E-WSNM shows a good linearity with all cell transistors' $\Delta V_{th}$ in -6 sigma < $\Delta V_{th}$ < 6 sigma. The dependences of "0" E-WSNM and "0" WSNM on $\Delta V_{th}$ in TaR are given and compared in Fig. 3.10(b). It can be seen that "0" E-WSNM accurately matches "0" WSNM in region with small WNM, indicating E-WSNM gives the same write failure point as WSNM.

## 3.4 Summary

This chapter proposes a new write stability metric to compensate two main drawbacks of WSNM from conventional write BC for yield estimation at low $V_{DD}$. The extended write BC

extends the voltage sweeping range of conventional write BC. Write stability characterization is performed using extended write BC at $V_{DD}$ = 0.4 V. Due to the existence of only one mode – failure mode, E-WSNM shows good normal distribution. Also, sharing the same failure SRAM cell with conventional write BC, extended write BC shows itself a good candidate for yield estimation at low $V_{DD}$. Besides, E-WSNM is demonstrated to be well correlated with $V_{th}$ of write performance-dominant cell transistors. In addition, a good correlation between E-WSNM and reference metric – CWLM – supports our conclusion.

Furthermore, HSPICE simulations are performed to give more details. Using the same method as in Chapter 2, we found that by extending the voltage range in conventional write BC, E-WSNM of extended write BC is shown to have a linear dependence on $\Delta V_{th}$ of each cell transistors even at 0.4 V. It also helps strengthen our conclusion.

Lastly, considering the proposed write BC is not limited to SOTB SRAM, our conclusion is also applicable for low-$V_{DD}$ yield estimation in other technologies, such as FinFET SRAM.

# Chapter 4

# Time-dependent Random Telegraph Noise in SRAM

## 4.1 Introduction

### 4.1.1 Dynamics of Random Telegraph Noise

Time-dependent random telegraph noise (RTN) has been studied for a long term since the era of vacuum electronics, when it was often called burst or popcorn noise. The first observation of RTN in transistors was reported by Ralls et al. in 1984 [50]. As the simplest case shown in Fig. 4.1, RTN causes the fluctuation between a high current level and a low current level in time domain [51]. And three important parameters – amplitude of current fluctuation ($\Delta I_d$), capture time constant ($\tau_c$), and emission time constant ($\tau_e$) – are defined inset. It covers a wide range of time scale from $10^{-6}$ s [52] to $10^3$ s [53]. Besides the time-domain analysis, another effective way is analysis in frequency domain by applying a Fourier transformation of the two-level signal. Fig. 4.2 shows one example of corresponding Lorentzian spectrum [54] described by:

$$S_I = \frac{2(\Delta I_d)^2 \tau_0}{4 + (\omega \tau_0)^2}$$

with $S_I$ the current spectral density, $\Delta I_d$ the amplitude of current fluctuation, $\tau_0$ the characteristic time constant, and $\omega = 2\pi f$ the radio frequency. From the spectrum, $S_I$ is

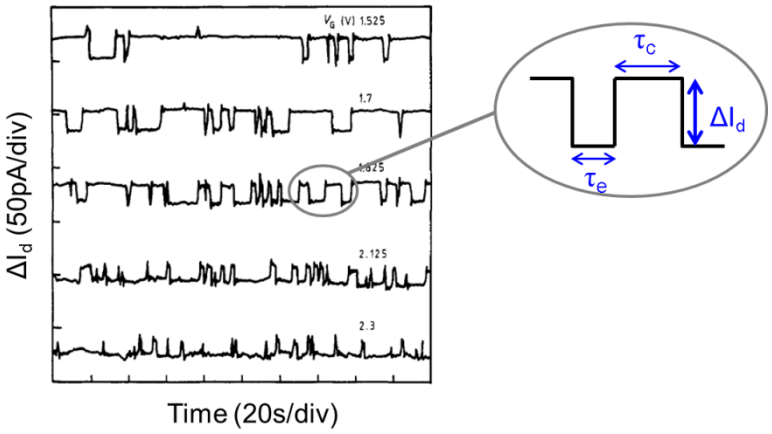


Fig. 4.1. Measured $I_d$ waveform in transistor along with time [51]. Three RTN-related parameters are defined as inset.

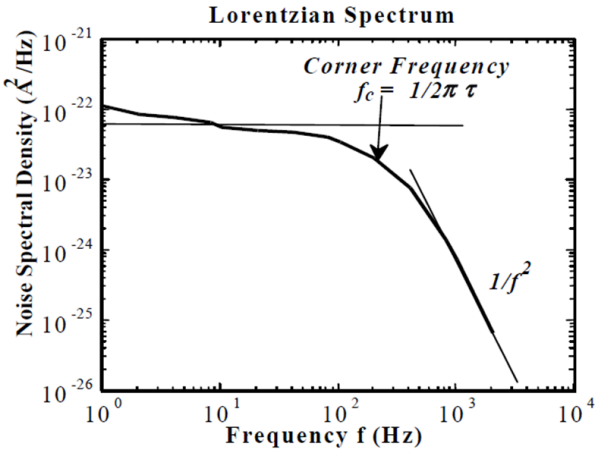


Fig. 4.2. One example of single-trap induced current spectral density [54].

constant at low-frequency regime $f << f_c$ and rolls off with $1/f^2$ at high-frequency regime. $\tau_0$ is derived as:

$$2\pi f_c = \frac{1}{\tau_0} = \frac{1}{\tau_e} + \frac{1}{\tau_c}$$

with $f_c$ the corner frequency.

This is the single-trap case, while in transistors two or more traps may contribute to current fluctuation [56], which makes current waveform more complex. Fig. 4.3 shows the complex current waveforms induced by two traps [Fig. 4.3(a)] and multiple traps [Fig. 4.3(b)]. Considering the whole contribution from different traps, the current spectral density is

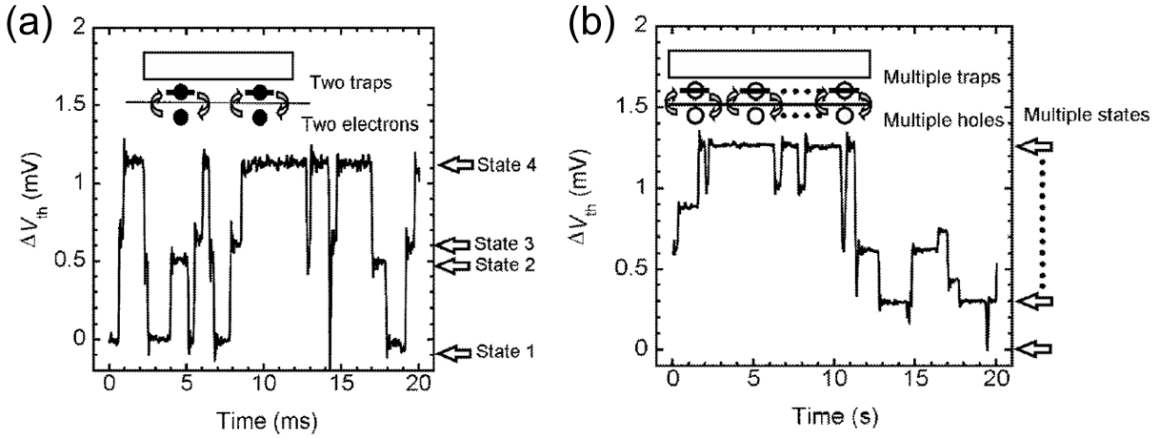


Fig. 4.3. RTN-induced $\Delta V_{th}$ from (a) two traps and (b) multiple traps [55].

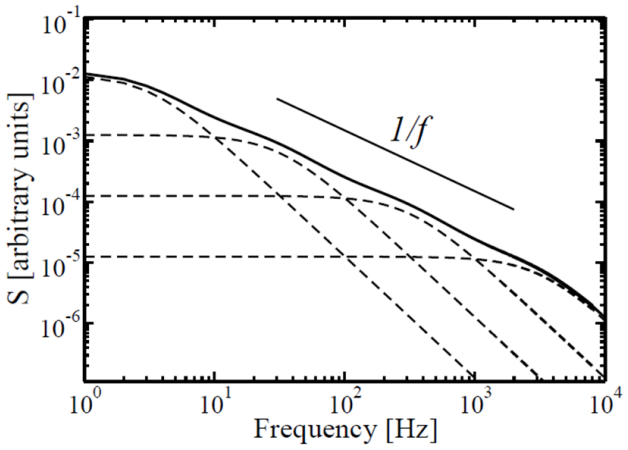


Fig. 4.4. One example of multiple-trap induced current spectral density [56].

a superposition of each component [56] and rolls off with 1/f at high-frequency regime, shown in Fig. 4.4. Due to its inverse proportional dependence on frequency, the superposed noise is called 1/f noise.

For further understanding the physical origin of 1/f noise, two main different theories have been proposed. One is number-fluctuation model and the other is mobility-fluctuation model, based on the fluctuation of conductivity in transistors:

$$\sigma = \mu n q$$

with μ and n are the mobility and carrier density, respectively.

**Number-fluctuation (ΔN) Model**

The carrier number-fluctuation model proposed by McWhorter [57-60], attributes origin of RTN to the exchange of carriers between channel and oxide traps. And the trapping/detrapping process is dominated by tunneling from channel to oxide or vice versa. Each trap is characterized by its own time constant τ and the occupation function N(t) is defined as: N(t) = 1 when trap is occupied and N(t) = 0 when trap is empty. Thus, the power spectral density (PSD) of N(t) is given by:

$$S_N(\tau) = (\Delta N)^2 \frac{4\tau}{1 + (2\pi f)^2 \tau^2}$$

When several traps are present with time constant τ distributed as:

$$g(\tau) = \begin{cases} \frac{K}{\tau} & if\ \tau_1 < \tau < \tau_2 \\ 0 & otherwise \end{cases}$$

The superposition of each single-trap gives PSD as:

$$S(f) = \int_0^\infty g(\tau) S_N(\tau) d\tau = \frac{4K(\Delta N)^2}{2\pi f} [\arctan(2\pi f \tau)]_{\tau_1}^{\tau_2}$$

**Mobility-fluctuation (Δμ) Model**

This model was firstly proposed by Hooge in 1969 [61] for homogeneous semiconductors or metals and extended to explain 1/f noise in transistors. The noise of homogeneous layers can be described by Hooge's empirical formula:

$$\frac{S_I}{I^2} = \frac{\alpha_H}{fN}$$

with I the current flowing through the sample, $S_I$ the spectral density of noise affecting current, N the number of carriers, and $\alpha_H$ the Hooge's parameter usually in the range between $10^{-6}$ and $10^{-4}$ [62].

The mobility in transistors is determined by the scattering of free carriers. Several scattering mechanisms are present, such as 1) bulk phonon scattering, 2) surface acoustic

phonon scattering, 3) impurity scattering by charged or neutral centers, and so on. By assuming the different scattering mechanisms are independent from each other and have the same energy dependence, the effective mobility $\mu_{eff}$ in transistors can be calculated using Matthiessen's rule:

$$\frac{1}{\mu_{eff}} = \sum_j \frac{1}{\mu_j}$$

with $\mu_j$ each component limited by different scattering mechanisms. But in Hooge model, only phonon scattering is considered for 1/f noise.

**A Unified Model**

Both number-fluctuation and mobility-fluctuation models try to explain experimental results to support their own applicability. In transistors, considering the charge transport is always confined near the channel surface, number-fluctuation model apparently provides a better explanation of the physical origin of 1/f noise. However, experimental results don't show consistent conclusion. In one comprehensive work by Chang et al. [63], a systematic study of 1/f noise in CMOS transistors from twelve different fabricators is performed. And measurement results suggest that 1/f noise in n-channel transistors can be well explained by number-fluctuation model while in p-channel transistors the noise is dominated by mobility fluctuation.

On the other hand, mobility-fluctuation model only takes phonon scattering into consideration and neglects other scattering mechanisms. Considering the drawbacks of these two models, Hung et al. [64-65] proposed a unified model to correlate the number-fluctuation model which dominates at low bias and the mobility-fluctuation model which is mostly effective at high bias. Though it uses non-physical fitting parameters, this model can explain most of the experimental results and has been the popular one adopted in circuit design tools to simulate 1/f noise.

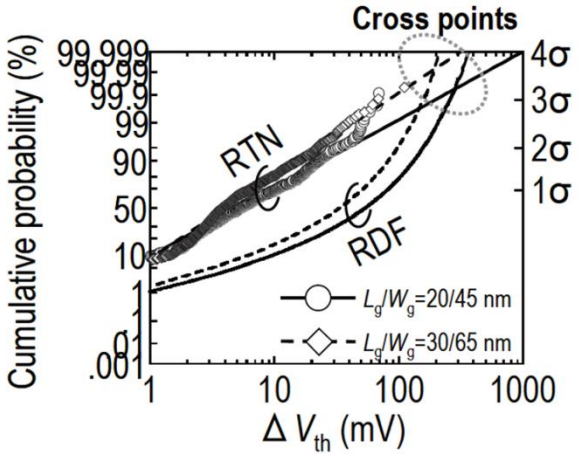


Fig. 4.5. Distributions of RTN-induced and RDF-induced $V_{th}$ variations. Projected RTN-induced $V_{th}$ variations exceed RDF-induced $V_{th}$ variations at the ~ 2 sigma level in 22 nm generation [70].

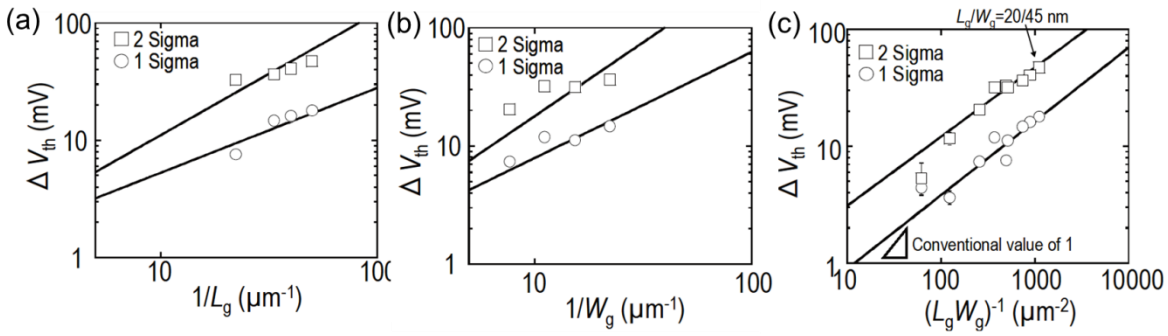


Fig. 4.6. (a) $L_g$ dependence, (b) $W_g$ dependence, and (c) size dependence of RTN-induced $\Delta V_{th}$ at 1 or 2 sigma [70].

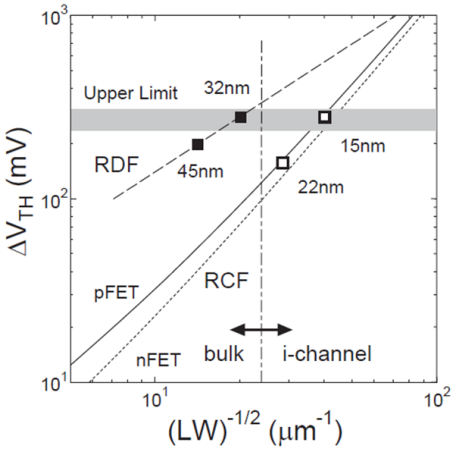


Fig. 4.7. Projected $V_{th}$ variation assuming intrinsic channel transistors [74].

### 4.1.2 Random Telegraph Noise in Transistors

The understanding of dynamics of RTN is the key to developing noise-resilient devices. However, RTN characteristics including its amplitude of fluctuation and capture/emission time constant distribute in a wide range. Therefore, a statistical analysis is of great importance to clarify its statistical phenomenon. In this section, we focus on RTN-induced $\Delta V_{th}$ in transistors and its dependences on size/gate voltage.

Different from $V_{th}$ variability due to RDF, which follows normal distribution up to ±5 sigma [66], RTN-induced $\Delta V_{th}$ variability always shows a long-tailed distribution [67-69]. Fig. 4.5 [70] shows distribution of RTN-induced $\Delta V_{th}$ in transistors with L/W = 20/45 and 30/65 nm, respectively. RDF-induced $V_{th}$ distribution is also plotted for comparison. It can be seen that, due to the long-tailed distribution, the projected RTN-induced variability exceeds that by RDF at ~ 3 sigma in 22 nm node technology.

To investigate its size dependence, RTN-induced $\Delta V_{th}$ at 1 or 2 sigma of transistors with different sizes are plotted in Fig. 4.6 [70]. $\Delta V_{th}$ shows a stronger dependence on gate width [Fig. 4.6(b)] than gate length [Fig. 4.6(a)], which can be explained by the percolation path model [71]. In addition, $\Delta V_{th}$'s dependence on gate area is studied by plotting $\Delta V_{th}$ versus $(L \times W)^{-1}$ in log-log scale [Fig. 4.6(c)]. It can be seen that $\Delta V_{th}$ is almost inversely proportional to gate area but with the power law component as ~ 0.6. The component is important since it determines the size scaling trend of RTN. The reason for a smaller component value less than conventional 1 is still not clear. But compared to that due to RDF ~ 0.5 [72-73], RTN shows a much severer size scaling, indicating that RTN will exceed RDF in future node technology. Particularly, in ultra-scaled node technology where bulk cannot apply to, RTN dominates the limitation of size scaling. Fig. 4.7 [74] considers intrinsic channel beyond 28 nm technology and assumes RTN-induced $\Delta V_{th}$ be in inverse proportion to gate area. According to the

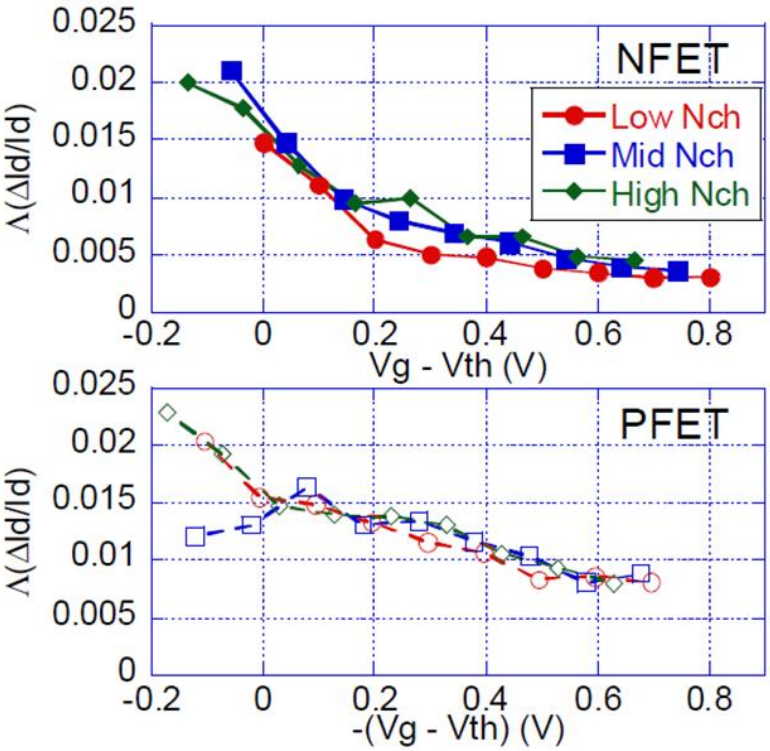


Fig. 4.8. Extracted gate voltage dependence of $\Delta I_d/I_d$ in NFET and PFET [52].

results, though intrinsic channel in such as SOTB technology help extending the scaling, rapid increase of RTN due to the 1/(L×W) term will hinder miniaturization beyond 15 nm.

Fig. 4.8 shows dependence on gate overdrive of singe trap RTN-induced $\Delta I_d/I_d$ in both NMOS and PMOS [52]. A clear tendency is found that $\Delta I_d/I_d$ increases with decreasing gate overdrive indicating a more important role of RTN at low gate overdrive, which is in agreement with TCAD predictions [75].

### 4.1.3 Random Telegraph Noise in SRAM

RTN in SRAM was firstly reported by Agostinelli et al. [76] in 2005, in which the erratic fluctuations of SRAM $V_{min}$ were observed at the 90 nm process technology node. And it suggested that a combination of process and circuit solutions be needed to enable continued SRAM cell scaling and voltage scaling. From then on, both experimental and simulation works have paid great attention to RTN in SRAM.

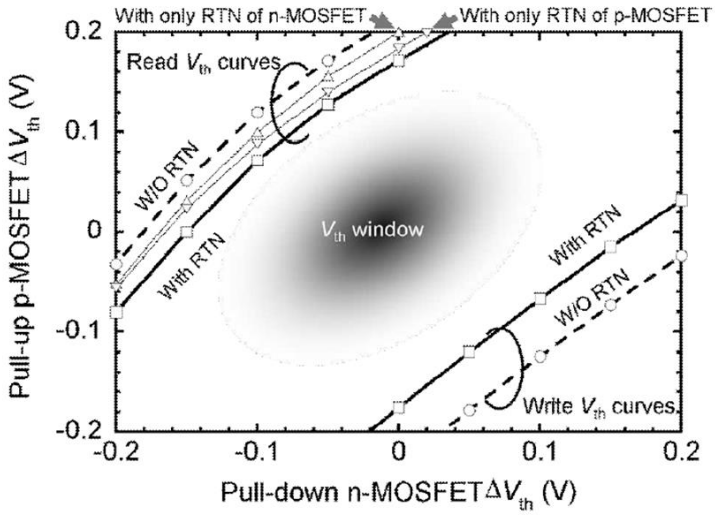


Fig. 4.9. Estimated $V_{th}$ window to guarantee the stable read/ write operation in 65 nm SRAM cells [55].

An SRAM cell consists of six cell transistors and its read noise margin generally decreases when $V_{th}$ of NMOS is smaller and the absolute value of $V_{th}$ of PMOS is larger than typical value. The write operation is opposite to read case. In that way, one graph can be plotted using $V_{th}$ of NMOS as horizontal axis and $V_{th}$ of PMOS as vertical axis. Fig. 4.9 gives simulation results of read and write noise margin of one SRAM cell at 65 nm technology [55]. The upper line (read $V_{th}$ curve) indicates the read boundary along which the read noise margin is zero while the lower line (write $V_{th}$ curve) indicates the write boundary. And the region enclosed by read and write $V_{th}$ curve is $V_{th}$ window, in which the SRAM is stable in both read and write operation. In addition, when RTN in both NMOS and PMOS is included, the $V_{th}$ window is found to shrink. That means RTN degrades both read and write noise margin.

The above $V_{th}$ window gives a good correlation between SRAM and cell transistors. However, a statistical analysis of noise margin in large capacity of SRAM arrays is also important to reliable embedded memory applications. In this section, the impact of RTN on read noise margin is focused, while in contemporary work section the write noise margin is to be discussed.

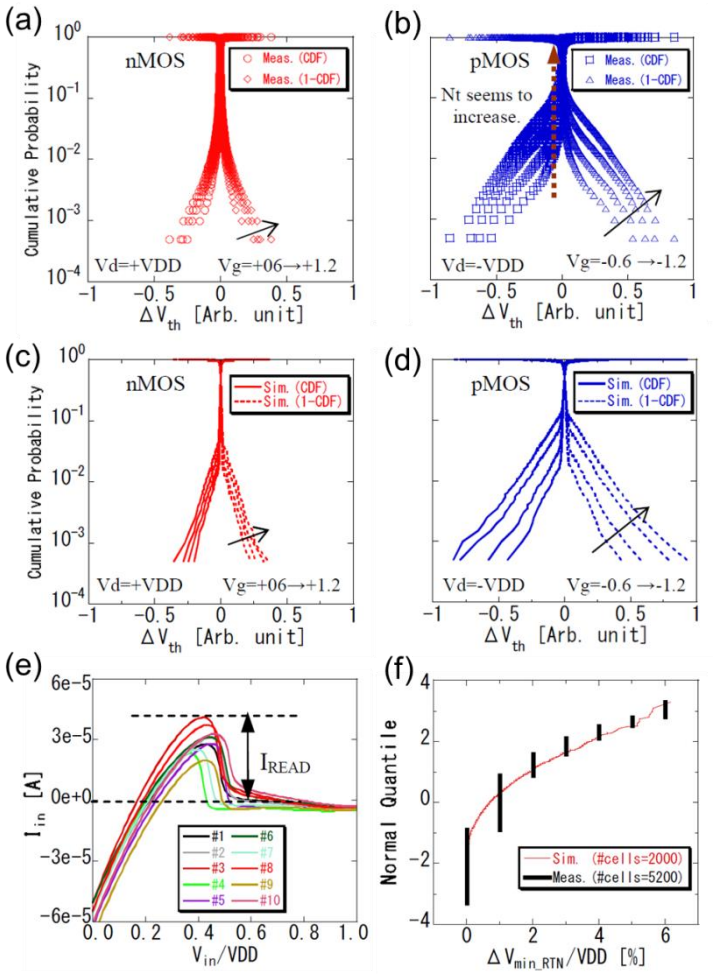


Fig. 4.10. Measured $\Delta V_{th}$ distributions of (a) nMOS and (b) pMOS at different $V_{gs}$ in 2 kb SRAM cells. Simulated results in (c) nMOS and (d) pMOS match the measured results well. (e) Simulated read N-curves and ten of them are selected for demonstration. $I_{READ}$ is defined as the read stability metric. (f) Cumulative plots of measured (in black) and simulated (in red) $\Delta V_{min}/V_{DD}$. From Ref. [77].

Tanizawa et al. [77] developed a statistical compact RTN model to investigate the impact of RTN on read $V_{min}$ in SRAM at 45 nm bulk technology. It starts from the reproduction of experimental observation of $V_{th}$ fluctuation due to RTN in both NMOS and PMOS, as shown in Fig. 4.10(a)-(d). On the basis of fixed parameters for RTN-induced $\Delta V_{th}$ distribution, the RTN model is applied to SRAM read stability analysis. $\Delta V_{min}$ is defined as the difference between maximum and minimum $V_{min}$ value. $I_{READ}$ from read N-curve is selected as the metric for read noise margin. Fig. 4.10(c) selects read N-curves of 10 SRAM cells for an

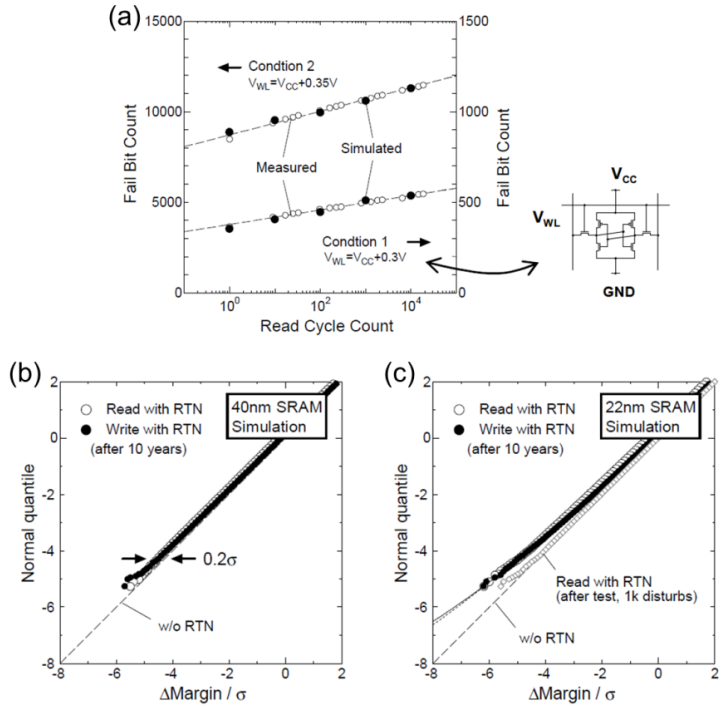


Fig. 4.11. (a) Accelerated test results (hollow circle) and simulation results (solid circle). Each bit suffers from 32 disturbs per cycle (i.e. entire 512kbit read). Simulated noise margin degradation by RTN in (b) 40 nm and (c) 22 nm technology. From Ref. [78].

illustration and Fig. 4.10(f) shows the comparison between experimental and simulation results. It can be seen that RTN-induced $\Delta V_{min}$ follows a log-tailed distribution. And the accuracy match with experimental results demonstrates the proposed RTN model as a useful tool for estimating the impact of RTN in SRAM design.

Besides, two other representative works perform direct read and count the fail bits instead of read noise margin analysis. One is by Takeuchi et al. [78], in which the fail bit count (FBC) in 40 nm SRAM is repeatedly monitored along with time. Fig. 4.11(a) shows the results at reduced margin operation regime by raising the word-line voltage compared to $V_{DD}$. The increasing FBC trend linearly with logarithm of time is ascribed to RTN. Also, combining Mont Carlo simulation, the impact of RTN in SRAM after 10 years at both 40 nm and 22 nm technology is estimated in Fig. 4.11(a) and Fig. 4.11(b). Though RTN causes only slight shift of the distribution ~ 0.2 sigma in 40 nm SRAM, RTN shows a bigger impact in 22 nm SRAM.

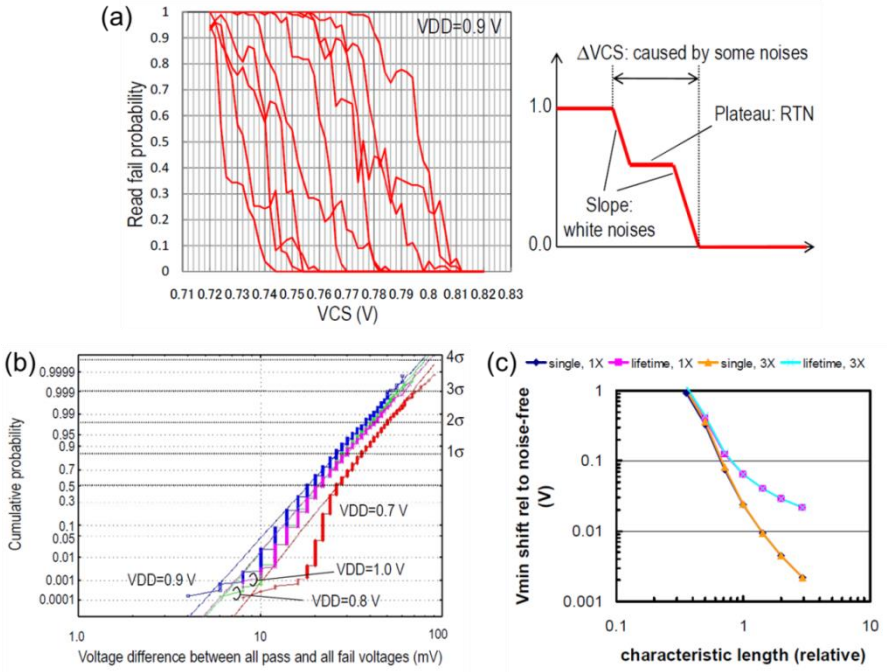


Fig. 4.12. (a) Measured fail probability transition in SRAM at $V_{DD}$ = 0.9 V. In the schematic, ΔVCS is defined by difference between pass and fail voltage. (b) ΔVCS distribution at different $V_{DD}$. (c) Calculated guard-band voltage in scaled SRAM assuming 1x and 3x trap density. From Ref. [79].

Particularly, the distribution in 22 nm SRAM clearly deviates normal distribution, which makes the margin approaches zero faster. Thus, RTN is suggested to be carefully taken into consideration for scaled SRAM.

The other one representative work is by Yamaoka et al. [79], in which a novel method is proposed to evaluate the impact of RTN in partially-depleted (PD) SOI SRAM. Fig. 4.12(a) shows selected cell failure probability transitions versus voltage cell supply (VCS). Due to some noise including RTN and others, the failure probability does not change abruptly from 1.0 to 0.0 but shows a plateau. And ΔVCS is selected as the parameter for further analysis. Fig. 4.12(b) shows distribution of ΔVCS at different $V_{DD}$. Below ~ 20 mV, it follows normal distribution and is ascribed to thermal or shot noise within cells. Above ~ 20 mV, it shows log-normal behavior and is mainly ascribed to RTN. Considering the impact of RTN, an enough guard band is needed for safe SRAM operation. Simulation results in Fig. 4.12(c)

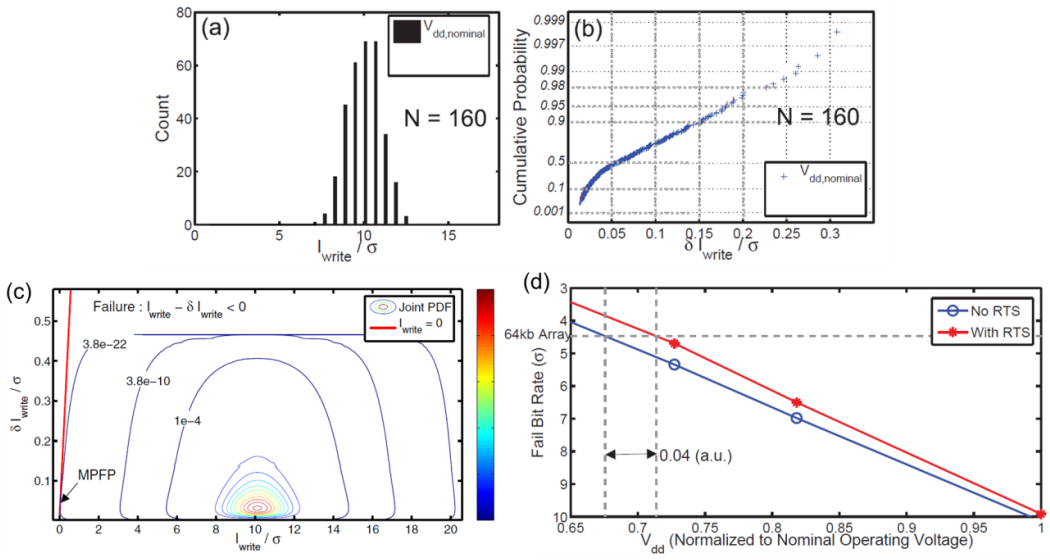


Fig. 4.13. (a) Histogram plot of $I_{write}$ and (b) Gumbel plot of RTN-induced $\delta I_{write}$ in 160 bulk SRAM cells at nominal $V_{DD}$. (c) Joint PDF plot of $\delta I_{write}$ versus $I_{write}$ at nominal $V_{DD}$. (c) FBR of SRAM at different $V_{DD}$. From Ref. [80].

shows that read $V_{min}$ for 4 Mb SRAM arrays is increased by ~ 25 mV from its ideal noise-free value, and that a further increase of 40 mV is needed for an error-free lifetime. Also, the projected results show a dramatic increase of guard band with size scaling in future node generations.

Above all, RTN plays an important role in read operation in SRAM arrays, especially at advanced node technologies.

### 4.1.4 Contemporary Work

Besides read stability, this section discusses about the impact of RTN on write stability in SRAM cells. Two main experimental works are introduced. One is by Toh et al. [80], in which a statistical model is developed to estimate the FBR with RTN in 45 nm bulk SRAM arrays. On the basis of 160 SRAM cells, Fig. 4.13(a) and Fig. 4.13(b) show the distributions of $I_W$ and RTN-induced $\delta I_W$ at nominal $V_{DD}$ using $I_W$ as the write stability metric. $I_W$ follows normal distribution while $\delta I_W$ shows a long-tailed distribution. Since the distribution functions

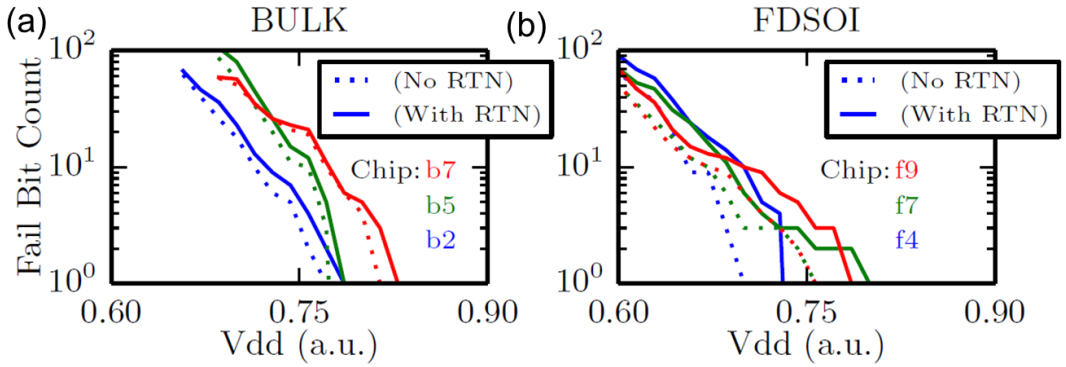


Fig. 4.14. Measured FBC in (a) bulk and (b) FDSOI SRAM at different $V_{DD}$ and the impact of RTN is evaluated [81].

for $I_W$ and $\delta I_W$ have been known and $\delta I_W$ is uncorrelated with $I_W$, the joint probability density function (PDF) plot is generated in Fig. 4.13(c). Through integrating PDF over failure region in Fig. 4.13(c), the failure probability as well as FBR can be calculated at different $V_{DD}$, shown in Fig. 4.13(d). It is shown that write $V_{min}$ degradation due to RTN is less than 50 mV, which is negligible at nominal $V_{DD}$. In addition, degradation due to RTN becomes less significant in larger SRAM arrays. Thus, RTN is concluded as unimportant in bulk SRAM at high $V_{DD}$.

On the other hand, Zimmer et al. [81] directly measured write FBC in 32 kb bulk and FD-SOI SRAM arrays at 28 nm technology. Fig. 4.14 compares results for bulk and FD-SOI SRAM at different $V_{DD}$. $V_{min}$ is extracted as the $V_{DD}$ at which FBC = 0. It is found that decreased random variability in FD-SOI enables an approximately 7% reduction in $V_{min}$, but also exacerbates the effect of RTN on $V_{min}$. But the effect of RTN is suppressed for entire array because the cells with the largest RTN are not necessarily the ones that limit $V_{min}$. Thus, it is concluded that RTN is not important in FD-SOI SRAM arrays at high $V_{DD}$.

However, all previous works focused on impact of RTN on write stability in SRAM at high $V_{DD}$ and low-$V_{DD}$ RTN analysis [82] is called for by designers for robust low-power SRAM design.

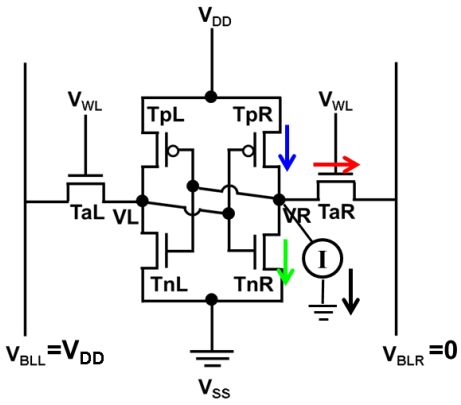


Fig. 4.15. Schematic of 6-T SRAM cell with $I_{VR}$ (black arrow), $I_{TaR}$ (red arrow), $I_{TpR}$ (blue arrow), and $I_{TnR}$ (green arrow) indicated during the measurement of write N-curve.

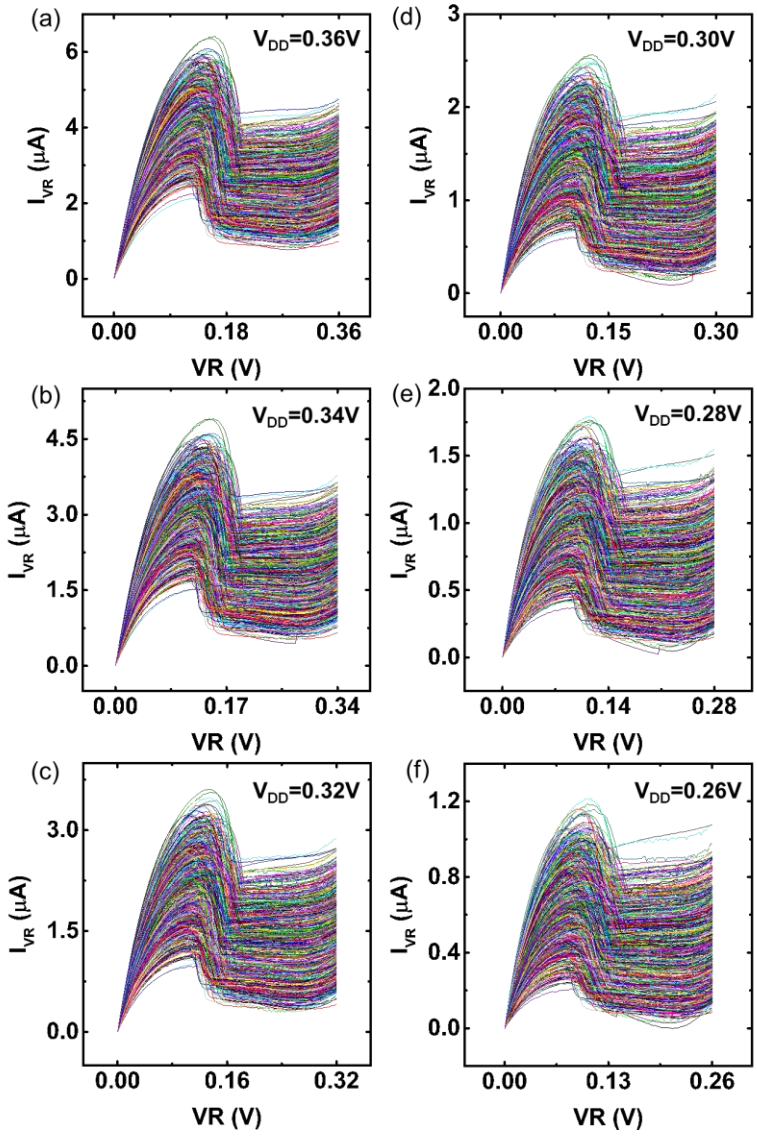


Fig. 4.16. Measured write N-curves in 4 kb SOTB SRAM cells at $V_{DD}$ = (a) 0.36 V, (b) 0.34 V, (c) 0.32 V, (d) 0.30 V, (e) 0.28 V, and (f) 0.26 V.

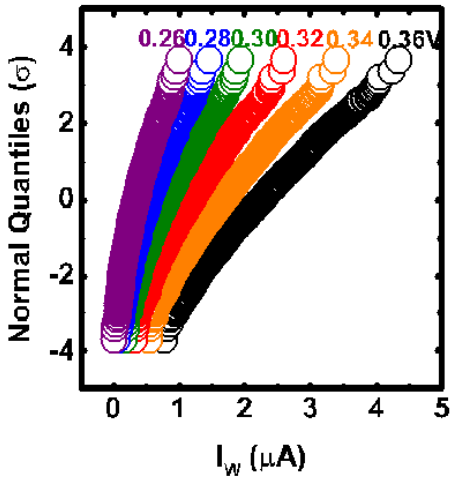


Fig. 4.17. Cumulative plots of measured $I_W$ of write N-curves in 4 kb SOTB SRAM cells at different $V_{DD}$ from 0.36 down to 0.26V.

# 4.2 Measurement and Modeling Results in SOTB SRAM

## 4.2.1 Write N-curve as Write Stability Metric

Fig. 4.15 shows the schematic of SRAM cell with bias conditions set for logic "0" write. Though write N-curve has been excluded as good candidate for yield estimation due to its non-normality in section 2.2.3, it is preferred for RTN measurement due to its being current-based metric. As indicated by arrows, the current outside node VR consists of three current components in access transistor (TaR), load transistor (TpR) and drive transistor (TnR).

Fig. 4.16 shows measured write N-curves of 4 kb SOTB SRAM cells at different $V_{DD}$ from 0.36 V to 0.26 V with the step of 0.02 V. Since these waveforms consist of write N-curves of two modes at low $V_{DD}$, its write noise margin $I_W$ deviates from normal distribution, as shown in Fig. 4.17. In addition, it shows a sever skewness with $V_{DD}$ scaling.

Take $V_{DD}$ = 0.3 V as an example, the non-normal distribution can be more clearly seen in histogram plot in Fig. 4.18(a). A clear deviation between experiment data (red column) and fitted line using normal distribution function (black line) is demonstrated in tail region.

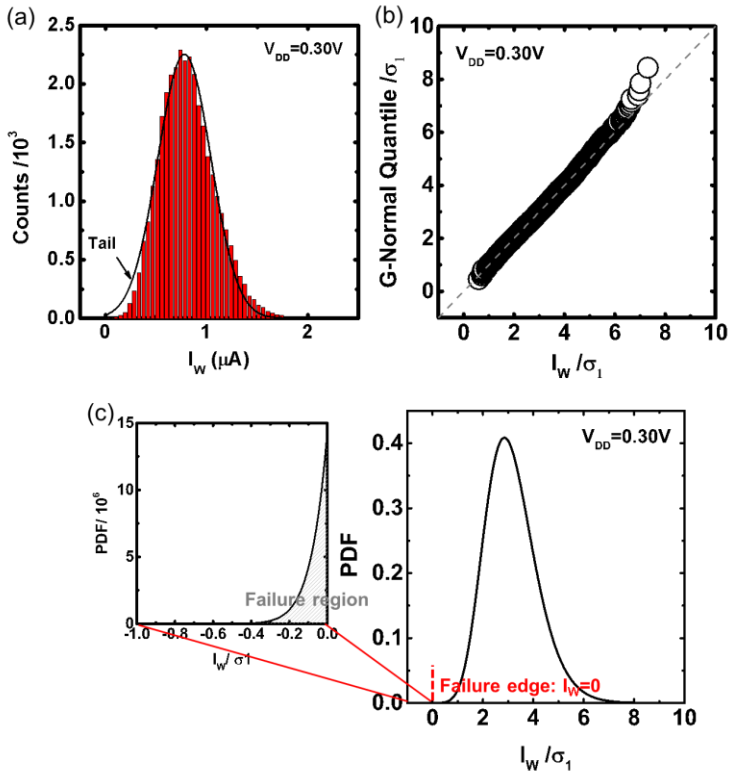


Fig. 4.18. Histogram plot of $I_W$ in 32 kb SOTB SRAM cells at $V_{DD}$ = 0.30 V. The normal distribution fit (black curve) shows a large deviation in tail region. (b) Quantile-quantile plot of $I_W$ with generalized normal distribution fit (grey dashed line). $I_W$ is normalized by its standard deviation $\sigma_1$. (c) Calculated PDF plot of $I_W$ with failure region enlarged. Failure edge is defined as $I_W$ = 0.

Considering its non-normality, we cannot directly extrapolate its yield but propose a function to fit its distribution with which the yield can be analytically calculated. Here, we use the generalized normal distribution function:

$$f_1(x) = -1/k \cdot \log(1 - k/\alpha \cdot (x - \beta))$$

with $f_1(x)$ its PDF, k the shape parameter, α the scale parameter, and β the location parameter. It is used to fit the data at $V_{DD}$ = 0.30 V and the result is shown in Fig. 4.18(b). In the quantile-quantile (Q-Q) plot [83], both x- and y-axis is normalized to its standard deviation and good fitting is obtained if the data points follow the straight line. After obtaining the fitting parameters k, α and β, the PDF plot is derived in Fig. 4.18(c). By integrating PDF in the failure region where $I_W < 0$ in inset figure, the failure probability can be calculated.

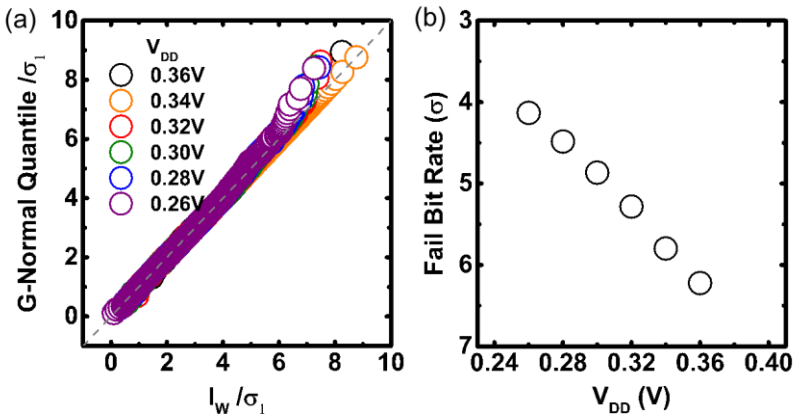


Fig. 4.19. (a) Quantile-quantile plots of $I_W$ in 4 kb SOTB SRAM cells at different $V_{DD}$ with generalized normal distribution fit (grey dashed line). $I_W$ is normalized by its standard deviation $\sigma_1$. (b) Calculated FBR at different $V_{DD}$.

In the same way, good fitting results with adjusted parameters are demonstrated at different $V_{DD}$ in Fig. 4.19(a). By calculating failure probability in PDF plot, Fig. 4.19(b) gives the further calculated FBR which means the capacity in which SRAM arrays can operate safely at each $V_{DD}$. We can see that yield dramatically degrades linearly with scaled $V_{DD}$, showing a similar trend as in Ref. [80].

### 4.2.2 Random Telegraph Noise Measurement in SRAM and Cell Transistors

In the above section, only time-zero variability is considered. Then, this section takes RTN into consideration and evaluates the impact of RTN in SRAM.

Fig. 4.20 shows RTN-induced fluctuation of $I_W$ ($\delta I_W$) as well as that of current in TaR ($\delta I_{TaR}$) and in TpR ($\delta I_{TpR}$) in 32 kb SOTB SRAM cells at $V_{DD}$ = 0.30 V.[5] They show a long-tailed distribution rather than random variation-induced normal distribution. Though a larger cross section of hole compared to electron generally enhances RTN-induced amplitude in PMOS, $\delta I_{TaR}$ is much larger than $\delta I_{TpR}$ due to a larger absolute value of current in TaR than in

[5] Here, another component of $\delta I_{TnR}$ is not plotted due to the negligible value of current in TnR at bias conditions where $I_W$ is defined. More details can refer to Fig. 2.27, though different substrate bias is applied.

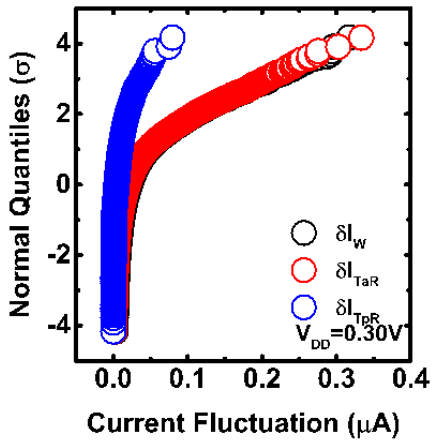


Fig. 4.20. Cumulative plots of measured $\delta I_W$ and $\delta I$ of cell transistors in 32 kb SOTB SRAM cells at $V_{DD}$ = 0.30 V.

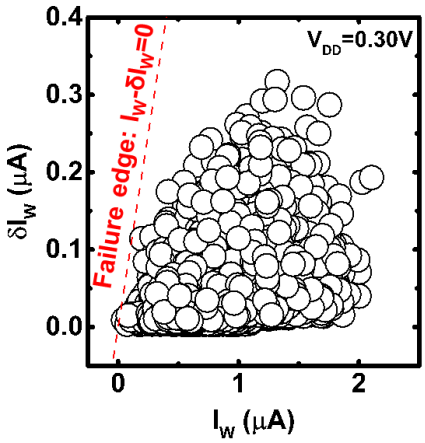


Fig. 4.21. Measured scatter plot of $\delta I_W$ versus $I_W$ in 32 kb SOTB SRAM cells at $V_{DD}$ = 0.30 V. Failure edge (red dashed line) is defined as: $I_W - \delta I_W = 0$.

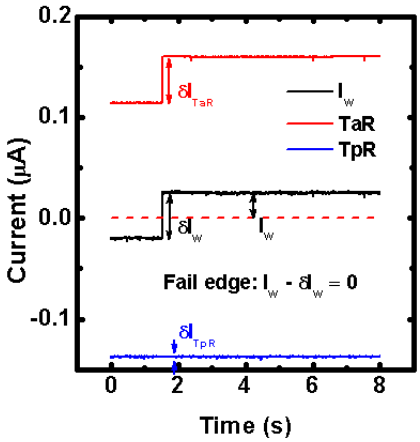


Fig. 4.22. Measured $I_W$, as well as its two components ($I_{TaR}$ and $I_{TpR}$), at $V_{DD}$ = 0.26 V along with time. The RTN-induced failure is demonstrated.

TpR. And $\delta I_W$ is found to be dominated by $\delta I_{TaR}$, further strengthen the importance of access transistor in write performance of SRAM cells.

To clarify the write failure at $V_{DD}$ = 0.30 V, Fig. 4.21 shows the scatter plot of $\delta I_W$ versus $I_W$ of 32 kb SOTB SRAM cells. And the failure edge $I_W$ - $\delta I_W$ = 0 is highlighted in red dashed line, beyond which the SRAM cell will be identified. No failure occurs in 32 kb SOTB SRAM cells at $V_{DD}$ = 0.30 V. However, a clear tendency towards failure region is clear and it is estimated that RTN-induced failure would occur as the number of cells increases at $V_{DD}$ = 0.30 V or $V_{DD}$ is decreased. This is demonstrated at a lower $V_{DD}$ = 0.26 V, as shown in Fig. 4.22. A clear abrupt change of $I_W$ across the failure edge is seen at time ~ 2 s. This is also the first ever example to demonstrate RTN-induced write failure in SRAM cells at low $V_{DD}$.

### 4.2.3 Statistical Distributions and Implications to SRAM Robustness

In this section, a statistical model is developed to evaluate the impact of RTN in SRAM, considering that no correlation is shown between $I_W$ and $\delta I_W$ in Fig. 4.21. Firstly take $V_{DD}$ = 0.30 V as an example. Fig. 4.23(a) shows $\delta I_W$ histogram plot of 4 kb SOTB SRAM cells, in which a clear deviation from commonly used lognormal distribution function (black line) [70, 84-85] is demonstrated in tail region. Also, another exponential distribution function [86-88] cannot give good fitting results (data not shown here). Thus, a composite function called lognormal-generalized pareto distribution function is proposed:

$$f_2(\mathrm{x}) = \begin{cases} r \cdot \dfrac{1}{\frac{1}{2} \cdot \left[1 + \mathrm{erf}\left(\dfrac{\log(x_0) - \mu}{\sqrt{2}\sigma}\right)\right] \cdot \sqrt{2\pi} \cdot \sigma} \cdot \dfrac{1}{x} \cdot \exp\left(-\dfrac{\log(x) - \mu}{\sqrt{2\sigma}}\right), 0 < x \le x_0 \\ (1 - r) \cdot \dfrac{1}{\tau} \cdot \left[1 + \dfrac{\xi}{\tau} \cdot (x - x_0)\right]^{-\left(\frac{1}{\xi} + 1\right)}, x > x_0 \end{cases}$$

with $f_2(x)$ its PDF, r the function ratio, μ the location parameter, σ the scale parameter, $x_0$ the truncation point, ξ the shape parameter and τ the second scale parameter. And the good fitting result is demonstrated in Q-Q plot as Fig. 4.23(b). The same as the previous procedure in section 4.2.1, the PDF plot is derived as Fig. 4.23(c) after obtaining all fitting parameters.

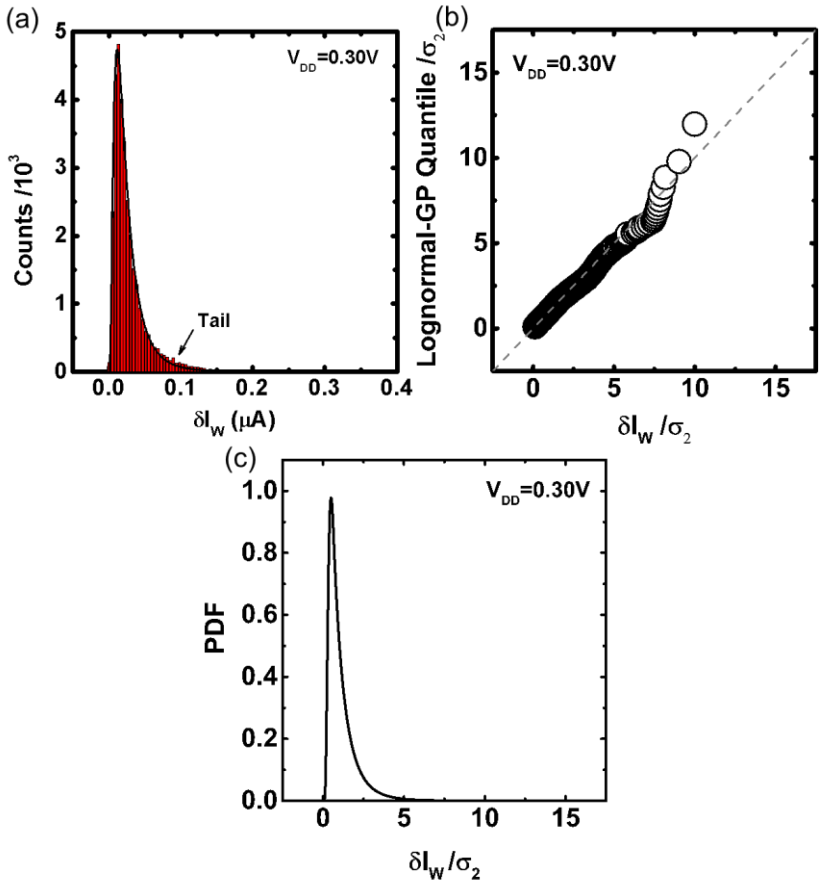


Fig. 4.23. Histogram plot of $\delta I_W$ in 32 kb SOTB SRAM cells at $V_{DD}$ = 0.30 V. The log-normal distribution fit (black curve) shows a large deviation in tail region. (b) Quantile-quantile plot of $\delta I_W$ with lognormal-generalized pareto distribution fit (grey dashed line). $\delta I_W$ is normalized by its standard deviation $\sigma_2$. (c) Calculated PDF plot of $\delta I_W$.

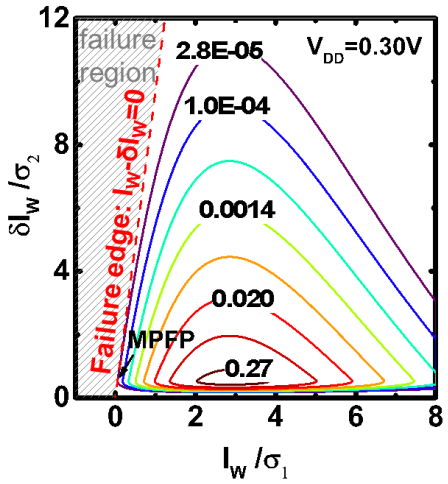


Fig. 4.24. Joint PDF plot of $\delta I_W$ versus $I_W$ at $V_{DD}$ = 0.30 V. Both $I_W$ and $\delta I_W$ are normalized by their standard deviations. The red dashed line and grey patterned area denote failure edge and failure region, respectively. MPFP indicates where the cell fails in the largest probability.

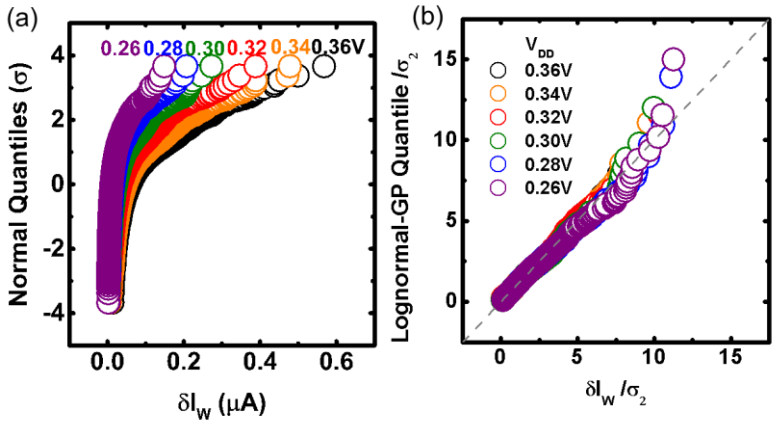


Fig. 4.25. (a) Cumulative plots of measured $\delta I_W$ in 4 kb SOTB SRAM cells at different $V_{DD}$ from 0.36 down to 0.26V. (b) Quantile-quantile plots of $\delta I_W$ with lognormal-generalized pareto distribution fit (grey dashed line). $\delta I_W$ is normalized by its standard deviation $\sigma_2$.

Since the distribution functions of both $I_W$ and $\delta I_W$ are obtained, the joint PDF plot can be generated as Fig. 4.24. Each value is calculated by multiplying the PDF of $I_W$ and that of $\delta I_W$. Failure edge as $I_W - \delta I_W = 0$ is indicated by red dashed line and failure region is highlighted in grey dashed pattern. Along the failure edge, one point named most probable failure point (MPFP) is indicated by black arrow. Above the failure edge, the failure probability can be calculated by integrating PDF over the failure region.

In the same way, $\delta I_W$ shows a long-tailed distribution at different $V_{DD}$ in Fig. 4.25(a) and good fitting results with adjusted parameters are demonstrated in Fig. 4.25(b). On the basis of the distribution functions of $I_W$ and $\delta I_W$, joint PDF plots at different $V_{DD}$ are generated in Fig. 4.26. All curves show a right-skewed pattern and the skewness increases with $V_{DD}$ scaling, which is ascribed to the distribution of $I_W$ at different $V_{DD}$. In addition, these plots can be separated into two groups. One is in sub-0.3 V regime, in which MPFP occurs around 0.6 $\sigma_2$ of $\delta I_W$'s distribution. The other one is in sub-0.4 V regime, such as 0.36 V and 0.34 V, MPFP occurs around 7 $\sigma_2$ of $\delta I_W$ distribution. That means RTN with large amplitude at the tail of $\delta I_W$ distribution dominates the failure, which is different from high-$V_{DD}$ operating bulk SRAM

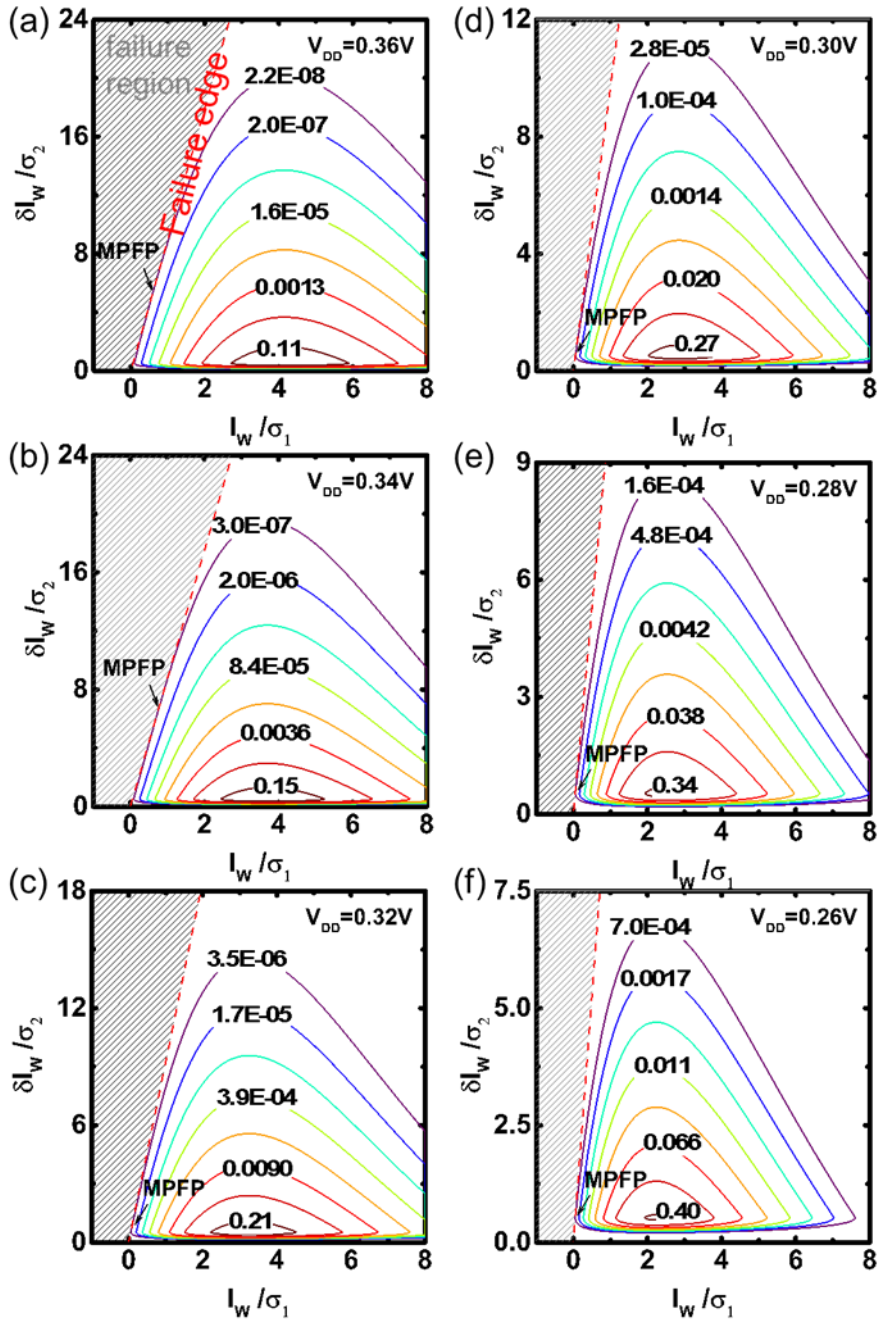


Fig. 4.26. Joint PDF plots of $\delta I_W$ versus $I_W$ at $V_{DD}$ = (a) 0.36 V, (b) 0.34 V, (c) 0.32 V, (d) 0.30 V, (e) 0.28 V, and (f) 0.26 V.

where RTN in the tail does not contribute to the failure [80] and also strengthen the importance of RTN in low-$V_{DD}$ regime.

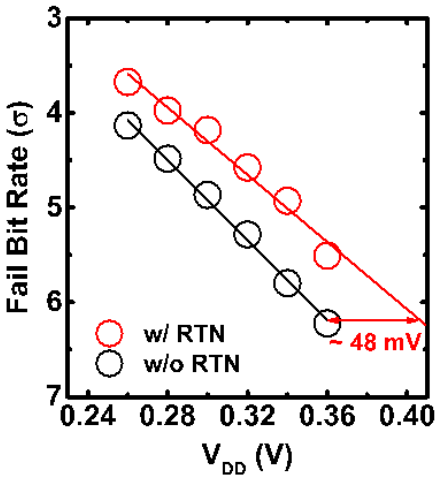


Fig. 4.27. FBR of SRAM at different $V_{DD}$. Linear $V_{DD}$ dependence is proposed both with and without RTN.

In order to quantitatively evaluate the impact of RTN, Fig. 4.27 shows the calculated FBR at different $V_{DD}$ by taking RTN into consideration. A simple extrapolation shows that Vmin is degraded due to RTN-induced failures by 48 mV at $V_{DD}$ = 0.36 V at the capacity of approximately 6.3 sigma. FBR degradation due to RTN reduces as $V_{DD}$ decreases down to sub-0.3 V regime, which is consistent with the MPFP analysis. This is different from Ref. [80], in which FBR degradation due to RTN is concluded less important in larger SRAM arrays. Thus, we cannot easily extrapolate the high-$V_{DD}$ analysis to low $V_{DD}$, again emphasizing the low-$V_{DD}$ analysis.

## 4.3 Summary

This chapter proposes a statistical model to evaluate the impact of RTN on write stability in SRAM based on RTN measurement in SOTB SRAM cells at low $V_{DD}$ in sub-0.4 V regime. It starts from write stability characterization under time-zero variability using $I_W$ from write N-curve as the write stability metric. Due to the existence of two modes in write N-curves, its

write noise margin $I_W$ deviates from normal distribution at low $V_{DD}$. Thus, a generalized normal distribution function is used to fit $I_W$ distribution followed by FBR calculation.

Besides, RTN-induced $\delta I_W$ in SRAM cells as well as the current fluctuation in cell transistors are measured. A writeable SRAM cell requires a strong write current through access transistor, therefore making current fluctuation in access transistor also dominate $\delta I_W$. $\delta I_W$ is shown to have a long-tailed distribution. However, the commonly used lognormal or exponential distribution from previous publications cannot fit well. Thus, a composite function called lognormal-generalized pareto distribution function is proposed and good fitting results are obtained.

Lastly, the developed statistical model is based on the experimental fact that $I_W$ and $\delta I_W$ have no correlation and can be treated independently. Since the distribution function of each component is derived, the joint PDF plots can be generated by multiplying PDFs of these two components and is followed by FBR calculation. By comparing FBR with and without RTN, it is found that RTN degrades $V_{min}$ up to 48 mV at the capacity ~ 6.3 sigma, which occupies over 10 % of $V_{DD}$. On the other hand, RTN becomes less important in sub-0.3 V regime, since SRAM cells are unstable even without RTN. Therefore, special care must be taken for RTN in SRAM design in sub-0.4 V regime.

# Chapter 5
# Conclusion

## 5.1 Key Parts

SRAM arrays in a larger capacity at a cost-effective price calls for the minimum-size cell transistors. However, variability consisting of time-zero and time-dependent variability increases with size scaling. It limits the design margin for voltage scaling, which is especially important to low-power SRAM since active energy and leakage power both benefit at lower $V_{DD}$. Thus, a comprehensive variability analysis in a large capacity of SRAM arrays is of great importance to both a good understanding of variability from different sources and an accurate estimation of design margin.

Compared to bulk technology in which RDF dominates random variation, SOTB technology utilizes intrinsic channel and facilitates further voltage scaling in SRAM cells. For example, the active operation down to sub-0.4 V regime has been demonstrated in 2 Mb SOTB SRAM cells [32]. But previous works all focused on variability analysis at high $V_{DD}$ whether in bulk or SOTB SRAM arrays. Therefore, this work encompasses three key parts to facilitate the variability-aware design of embedded low-power SRAM:

1) Several commonly used write stability metrics are compared in SRAM at low $V_{DD}$ and the good ones for yield estimation are selected.

Write stability characterization is performed in a large capacity of SOTB SRAM at low

$V_{DD}$ using several write stability metrics, including WSNM from write butterfly curve, WTI/ $I_W$ from read/write N-curve, BLM from bit-line method and CWLM from word-line method. Firstly, WTI is excluded since it cannot detect real write failure in SRAM cells. Among the other four metrics, BLM and CWLM follow good normal distributions. As well, these two metrics are perfectly correlated with each other. Thus, these two metrics are concluded as good candidates for yield estimation even at low $V_{DD}$. On the other hand, WSNM and $I_W$ deviate from normal distributions at low $V_{DD}$. And the origin of non-normality is ascribed to two modes in write butterfly curve or write N-curve when cell transistors operate down to sub-$V_{th}$ region at low $V_{DD}$. This limits the applicability of these two metrics for fast write yield estimation in SRAM arrays.

Considering the origin of non-normality in write butterfly curve at low $V_{DD}$ and the fact that failure mode can be used for write yield estimation, a new extended write butterfly curve is proposed by extending the voltage sweeping range in conventional one. Compared to conventional one, the failure mode emerges clearer in extended write butterfly curve. In addition, good correlation between E-WSNMand $V_{th}$ of write performance-dominant cell transistors as well as CWLM from word-line method is demonstrated. Thus, extended write butterfly curve can be used to extend the applicability of conventional one for write yield estimation at low $V_{DD}$.

In addition, HSPICE simulations are performed to help extend our conclusions up to ±6 sigma. What is more, to demonstrate the universality of our conclusions, similar results have been demonstrated in bulk SRAM cells at low $V_{DD}$ (~ 0.6 V). Our conclusions emphasize that sub-$V_{th}$ operation of cell transistors should be taken into consideration in low-power SRAM design.

2) The impact of time-dependent RTN on write stability in SRAM is evaluated at low

$V_{DD}$.

A statistical model is developed to evaluate the impact of RTN on write stability in SOTB SRAM cells at low $V_{DD}$. $I_W$ from write N-curve is selected as the write stability metric considering its being current-based metric which is good for RTN measurement. But its write noise margin $I_W$ deviates from a normal distribution at low $V_{DD}$. Then generalized normal distribution function is used to fit $I_W$ distribution to define failure region ($I_W < 0$) for FBR calculation.

In addition, RTN-induced fluctuation ($\delta I_W$) in write N-curve is monitored. Different from random variation, $\delta I_W$ shows a long-tailed distribution. To do the same thing, a composite function that is lognormal-generalized pareto distribution function is proposed to fit $\delta I_W$ distribution to define failure region ($I_W - \delta I_W < 0$) for FBR calculation. For comparison, RTN degrades $V_{min}$ up to ~ 48 mV (over 10 %) at capacity of approximately 6.3 sigma. Thus, special care is suggested to be taken for RTN in SRAM design in sub-0.4 V regime.

3) A universal conclusion and implications to low-power SRAM design.

In this work, we specifically selected SOTB SRAM mainly due to its being able to operate at low $V_{DD}$ with elimination of RDF-induced fluctuation in conventional bulk technology. But our conclusions are not uniquely meant for SOTB but applicable to other technologies, such as low-power FinFET SRAM.

## 5.2 Future Work

Future work will be on discussion in dynamic regime. Since write operation in SRAM only takes around the order of picoseconds at high $V_{DD}$ or nanoseconds at low $V_{DD}$, it cannot be accurately discussed in static regime. And the discussion will encompass two parts.

1) Using dynamic write noise margin as the metric

The four write stability metrics compared in the main text have been widely used,

however, it was realized that these metrics are known to be optimistic in write stability estimation [89]. Several simulation works [90-91] have been performed to give both definition of dynamic write noise margin and statistical analysis in SRAM. But as processes become increasingly complex and harder to control, designers can no longer rely on model accuracy to fully capture the random effects in large capacity of SRAM cells. On the other hand, Toh et al. [92] proposed a characterization architecture for measuring dynamic SRAM stability through pulsed word-lines calibrated up to 10 ps accuracy. But the measured number of SRAM cells is limited for real yield estimation. Thus, an experimental statistical analysis of dynamic write noise margin in a large capacity of SRAM cells is of necessity.

2) Impact of RTN on dynamic write stability in SRAM

In dynamic regime, the trap with short time constant, which is comparable to word-line pulses, can contribute to degradation of write stability in SRAM. Also, the slow trap can degrade SRAM performance due to hysteretic effects [93]. Therefore, it is important to cover sufficiently wide range of time constants of traps in cell transistors, which adds to the difficulty in RTN measurement. In addition, the fluctuation of write stability in SRAM is a combination of RTN-induced current fluctuations in six cell transistors. It makes gauging the worst write case the key to giving write failure point in SRAM.

Again, considering the great significance of $V_{DD}$ scaling to low-power SRAM, it is necessary to give statistical analysis in dynamic regime at low $V_{DD}$.

# Bibliography

[1] International Technology Roadmap for Semiconductors, Interconnect, 2003 Edition, Semiconductors Industry Assoc. and SEMATECH: http://www.itrs2.net/itrs-reports.html.

[2] A. Pavlov and M. Sachdev, CMOS SRAM Circuit Design and Parametric Test in Nano-Scaled Technologies: Process-Aware SRAM Design and Test (Frontiers in Electronic Testing), New York: Springer, 2008.

[3] K. Zhang, Embedded Memories for Nano-Scale VLSIs, New York: Springer, 2009, pp. 8-9.

[4] Source: Semico Research Corp: http://www.semico.com/.

[5] K. J. Kuhn, "Considerations for ultimate CMOS scaling," IEEE Trans. Electron Devices, vol. 59, no. 7, pp. 1813-1828, Jul. 2012.

[6] A. P. Chandrakasan and R. W. Brodersen, "Minimizing power consumption in digital CMOS circuits," Proc. IEEE, vol. 83, pp. 498–523, Apr. 1995.

[7] H. Soeleman and K. Roy, "Ultra-low power digital subthreshold logic circuits," in Proc. IEEE. Int. Symp. Low Power Electron. Design (ISLPED), 1999, pp. 94-96.

[8] C. Chen and M. Sarrafzadeh, "Simultaneous voltage scaling and gate sizing for low-power design," IEEE Trans. Circuits Syst. II, vol. 49, no. 6, pp. 400- 408, Jun. 2002.

[9] S. K. Gupta, A. Raychowdhury and K. Roy, "Digital computation in subthreshold region for ultralow-power operation: A device-circuit-architecture-codesign perspective", Proc. IEEE, vol. 98, pp. 160-190, Feb. 2010.

[10] T. Sakurai, "Designing ultra-low voltage logic," in Proc. IEEE. Int. Symp. Low Power Electron. Design (ISLPED), 2011, pp. 57-58.

[11] K. Zhang, Embedded Memories for Nano-Scale VLSIs, New York: Springer, 2009, p. 91-93.

[12] N. Verma and A. Chandrakasan, "A 256 kb 65 nm 8T subthreshold SRAM employing sense-amplifier redundancy," IEEE J. Solid-State Circuits, vol. 43, no. 1, pp. 141-149, Jan. 2008.

[13] A. B. Kahng and Y. C. Pati, "Subwavelength lithography and its potential impact on design and EDA," in Proc. IEEE. Design Autom. Conf. (DAC), 1999, pp. 799-804.

[14] L.-T. Pang, “Measurement and analysis of variability in CMOS circuits,” Ph.D. dissertation, University of California, Berkeley, Berkeley, CA, 2008.

[15] A. Asenov, "Random dopant induced threshold voltage lowering and fluctuations in sub-0.1 μm MOSFET's: A 3-D “atomistic” simulation study," IEEE Trans. Electron Devices, vol. 45, no. 12, pp. 2505-2513, Dec. 1998.

[16] K. Takeuchi, T. Fukai, T. Tsunomura, A. T. Putra, A. Nishida, S. Kamohara, and T. Hiramoto, "Understanding Random Threshold Voltage Fluctuation by Comparing Multiple Fabs and Technologies," in IEDM Tech. Dig., 2007, pp. 467-470.

[17] A. Asenov, A. R. Brown, J. H. Davies, S. Kaya, and G. Slavcheva, "Simulation of intrinsic parameter fluctuations in decananometer and nanometer-scale MOSFETs," IEEE Trans. Electron Devices, vol. 50, no. 9, pp. 1837–1852, Sep. 2003.

[18] H. F. Dadgour, K. Endo, V. K. De, and K. Banerjee, "Grain-Orientation Induced Work Function Variation in Nanoscale Metal-Gate Transistors—Part I: Modeling, Analysis, and Experimental Validation," IEEE Trans. Electron Devices, vol. 57, no. 10, pp. 2504-2514, Sep. 2010.

[19] H. F. Dadgour, K. Endo, V. K. De, and K. Banerjee, "Grain-Orientation Induced Work Function Variation in Nanoscale Metal-Gate Transistors—Part II: Implications for Process, Device, and Circuit Design," IEEE Trans. Electron Devices, vol. 57, no. 10, pp. 2515-2525, Sep. 2010.

[20] J. P. Campbell, J. Qin, K. P. Cheung, L. C. Yu, J. S. Suehle, A. Oates and K. Sheng, "Random telegraph noise in highly scales nMOSFETs," in Proc. IEEE. Int. Rel. Phys. Symp. Tech. Dig. (IRPS), 2009, pp. 382-388.

[21] M. T.-Luque, "Correlation of single trapping and detrapping effects in drain and gate currents of nanoscaled nFETsand pFETs," in Proc. IEEE. Int. Rel. Phys. Symp. Tech. Dig. (IRPS), 2012, pp. XT.5.1-XT.5.6.

[22] B. Kaczer, T. Grasser, P. J. Roussel, J. Franco, R. Degraeve, L. A. Ragnarsson, E. Simoen, G. Groeseneken and H. Reisinger, "Origin of NBTI variability in deeply scaled PFETs," in Proc. IEEE. Int. Rel. Phys. Symp. Tech. Dig. (IRPS), 2010, pp. 26-32.

[23] S. Zafar, A. Callegari, E. Gusev, and M. V. Fischetti," Charge trapping in high-k gate dielectric stacks," in IEDM Tech. Dig., 2012, pp. 517-520.

[24] C. Hu, S. C. Tam, F. C. Hsu, P. K. KO, T. Y. Chan, and K. W. Terrill, “Hot-electron-induced MOSFET degradation-Model, monitor, and improvement,” IEEE Trans. Electron Devices, vol. ED-32, no. 2, pp. 375-385, Feb. 1985.

[25] N.N. Mahatme, N.J. Gaspard, T.Assis, S. Jagganathan, I. Chatterjee, T.D. Loveless, B.L. Bhuva, L.W. Massengill, S.J. Wen, and R. Wong, “Impact of technology scaling on the combinational logic soft error rate,” in Proc. IEEE. Int. Rel. Phys. Symp. Tech. Dig. (IRPS), 2014, pp. 5F.2.1-5F.2.6.

[26] K. Aadithya, S. Venogopalan, A. Demir and J. Roychowdhury, "MUSTARD: A coupled, stochastic/deterministic, discrete/continuous technique for predicting the impact of random

telegraph noise on SRAMs and DRAMs," in Proc. IEEE. Design Autom. Conf. (DAC), 2011, pp. 292-297.

[27] Source: EE times: http://www.eetimes.com/.

[28] Source: Taiwan Semiconductor Manufacturing Company (TSMC): http://www.tsmc.com/english/default.htm.

[29] Source: SOI Industry Consortium: http://www.soiconsortium.org/.

[30] Source: GlobalFoundries: http://www.globalfoundries.com/.

[31] R. Tsuchiya, M. Horiuchi, S. Kimura, M. Yamaoka, T. Kawahara, S. Maegawa, T. Ipposhi, Y. Ohji and H. Matsuoka, "Silicon on thin BOX: A new paradigm of the CMOSFET for low-power and high-performance application featuring wide-range back-bias control," in IEDM Tech. Dig., 2004, pp. 631-634.

[32] Y. Yamamoto, H. Makiyama, H. Shinohara, T. Iwamatsu, H. Oda, S. Kamohara, N. Sugii, Y. Yamaguchi, T. Mizutani and T. Hiramoto, “Ultralow-voltage operation of silicon-on-thin-box(SOTB) 2 Mbit SRAM down to 0.37 V utilizing adaptive back bias, ” in VLSI Tech. Dig., 2013, pp. T212-T213.

[33] S. Markov, B. Cheng and A. Asenov, "Statistical variability in fully depleted SOI MOSFETs due to random dopant fluctuations in the source and drain extensions," IEEE Electron Device Lett., vol. 3, no. 3, pp. 315-317, Mar. 2012.

[34] T. Hiramoto, A. Kumar, T. Mizutani, J. Nishimura, and T. Saraya, "Statistical advantages of intrinsic channel fully depleted SOI MOSFETs over bulk MOSFETs," in IEEE CICC Tech. Dig., 2011, pp. 5.2-5.5.

[35] P. Bai, C. Auth, S. Balakrishnan, M. Bost, R. Brain, V. Chikarmane, R. Heussner, M. Hussein, J. Hwang, D. Ingerly, R. James, J. Jeong, C. Kenyon, E. Lee, S-H. Lee, N. Lindert, M. Liu, Z. Ma, T. Marieb, A. Murthy, R. Nagisetty, S. Natarajan, J. Neirynck, A. Ott, C.

Parker, J. Sebastian, R. Shaheed, S. Sivakumar, J. Steigerwald, S. Tyagi, C. Weber, B. Woolery, A. Yeoh, K. Zhang, and M. Bohr, "A 65 nm logic technology featuring 35 nm gate lengths, enhanced channel strain, 8 Cu interconnect layers, low-k ILD and 0.57 $um^2$ SRAM Cell," in IEDM Tech. Dig., 2004, pp. 657-660.

[36] F. Boeuf, F. Arnaud, C. Boccaccio, F. Salvetti, J. Todeschini, L. Pain, M. Jurdit, S. Manakli, B. Icard, N. Planes, N. Gierczynski, S. Denorme, B. Borot, C. Ortolland, B. Duriez, B. Tavel, P. Gouraud, M. Broekaart, V. Dejonghe, P. Brun, F. Guyader, P. Morini, C. Reddy, M. Aminpur, C. Laviron, S. Smith, J. P. Jacquemin, M. Mellier, F. Andre, N. Bicais-Lepinay, S. Jullian, J. Bustos and T. Skotnicki, "0.248$um^2$ and 0.334$um^2$ conventional bulk 6T-SRAM bit-cells for 45 nm node low cost—General purpose applications," in VLSI Tech. Dig., 2005, pp. 130-131.

[37] A. Bhavnagarwala, S. Kosonocky, C. Radens, K. Stawiasz, R. Mann, Q. Ye and K. Chin, "Fluctuation limits & scaling opportunities for CMOS SRAM cells," in IEDM Tech. Dig., 2005, pp. 675-678.

[38] Z. Guo, A. Carlson, L.-T. Pang, K. T. Duong, T.-J. K. Liu and B. Nikolić, "Large-scale SRAM variability characterization in 45 nm CMOS," IEEE J. Solid-State Circuits, vol. 44, no. 11, pp. 3174-3192, Nov. 2009.

[39] E. Seevinck, F. List, and J. Lohstroh, "Static-noise margin analysis of MOS SRAM cells," IEEE J. Solid-State Circuits, vol. SC-22, pp. 748-754, Oct. 1987.

[40] E. Grossar, M. Stucchi, K. Maex and W. Dehaene, "Read stability and write-ability analysis of SRAM cells for nanometer technologies," IEEE J. Solid-State Circuits, vol. 41, no. 11, pp. 2577-2588, Nov. 2006.

[41] H. Makino, S. Nakata, H. Suzuki, S. Mutoh, M. Miyama, T. Yoshimura, S. Iwade and Y. Matsuda, "Reexamination of SRAM cell write margin definitions in viewof predicting the

distribution," IEEE Trans.Circuits Syst. II, Exp. Briefs, vol. 58, no. 4, pp. 230-234, Apr. 2011.

[42] H. Qiu, T. Mizutani, Y. Yamamoto, H. Makiyama, T. Yamashita, H. Oda, S. Kamohara, N. Sugii, T. Saraya, M. Kobayashi, and T. Hiramoto, "Statistical Analysis of Four Write Stability Metrics in Fully Depleted Silicon-on-thin-BOX (SOTB) and Bulk SRAM Cells at Low Supply Voltage," in Proc. IEEE Int. Conf. Solid-State Inter. Circuit Technol. (ICSICT), 2014, pp. 987-989.

[43] T. Hiramoto, M. Suzuki, X. Song, K. Shimizu, T. Saraya, A. Nishida, T. Tsunomura, S. Kamohara, K. Takeuchi, and T. Mogami, "Direct Measurement of Correlation Between SRAM Noise Margin and Individual Cell Transistor Variability by Using Device Matrix Array," IEEE Trans. Electron Devices, vol. 58, no. 8, pp. 2249-2256, Aug. 2011.

[44] K. Takeuchi, T. Mizutani, H. Shinohara, T. Saraya, M. Kobayashi, and T. Hiramoto, "Measurement of SRAM Power-On State for PUF Applications Using an Addressable SRAM Cell Array Test Structure," in IEEE Int. Conf. on Microelectronic Test Structures (ICMTS), 2016, pp. 130-134.

[45] T. B. Hook, "Fully depleted devices for designers: FDSOI and FinFETs," in Proc. CICC, 2012, pp. 1-7.

[46] H. Qiu, T. Mizutani, T. Saraya, and T. Hiramoto, "Comparison and Statistical Analysis of Four Write Stability Metrics in Bulk CMOS SRAM Cells," in Int. Conf Solid State Dev. Mater. (SSDM), 2014, pp. 848-849.

[47] H. Qiu, T. Mizutani, T. Saraya, and T. Hiramoto, “Comparison and statistical analysis of four write stability metrics in bulk CMOS static random access memory cells,” Jpn. J. Appl. Phys., vol. 54, pp. 04DC09(01)-04DC09(04), Feb. 2015.

[48] H. Qiu, K. Takeuchi, T. Mizutani, T. Saraya, M. Kobayashi, and T. Hiramoto, "A New Write Stability Metric Using Extended Write Butterfly Curve for Yield Estimation in SRAM

Cells at Low Supply Voltage," in IEEE Int. Conf. on Microelectronic Test Structures (ICMTS), 2016, pp. 126-129.

[49] K. Takeda, H. Ikeda, Y. Hagihara, M. Nomura and H. Kobatake, "Redefinition of write margin for next-generation SRAM and write-margin monitoring circuit," in ISSCC Dig. Tech., 2006, pp. 630-631.

[50] K. S. Ralls, W. J. Skocpol, L. D. Jackel, R. E. Howard, L. A. Fetter, R. W. Epworth, and D. M. Tennant, "Discrete Resistance Switching in Submicrometer Silicon Inversion Layers: Individual Interface Traps and Low-Frequency (1/f?) Noise," Phys. Rev. Lett., vol. 52, no. 3, pp. 228-231, Jan. 1987.

[51] M. J. Kirton, M. J. Uren, S. Collins, M. Schulz, A. Karmann, and K. Scheffer, "Individual defects at the Si: $SiO_2$ interface," Semicond. Sci. Technol., vol. 4, pp. 1116-1126, Aug. 1989.

[52] T. Nagumo, K. Takeuchi, S. Yokogawa, K. Imai and Y. Hayashi, "New analysis methods for comprehensive understanding of random telegraph noise," in IEDM Tech. Dig., 2009, pp. 759-762.

[53] K. Abe, T. Fujisawa, A. Teramoto, S. Watabe, S. Sugawa and T. Ohmi, "Anomalous random telegraph signal extractions from a very large number of n-metal oxide semiconductor field-effect transistors using test element groups with 0.47 Hz–3.0 MHz sampling frequency," Japanese J. Appl. Phys., vol. 48, pp. 04C044(1)-04C044(5), Apr. 2009.

[54] M. J. Kirton and M. J. Uren, "Noise in solid-state microstructures: A new perspective on individual defects, interface states and low-frequency ($1/f$) noise," Adv. Phys., vol. 38, no. 4, pp. 367-468, Jul. 1989. [55] N. Tega, H. Miki, M. Yamaoka, H. Kume, T. Mine, T. Ishida, Y. Mori, R. Yamada and K. Torii, "Impact of threshold voltage fluctuation due to random telegraph noise on scaled-down SRAM," in Proc. IEEE. Int. Rel. Phys. Symp. Tech. Dig.

(IRPS), 2008, pp. 541-546.

[56] A. L. McWhorter, “1/f Noise and Germanium Surface Properties,” In: Semiconductor Surface Physics, Philadelphia: Univ. of Pennsylvania Press, 1957.

[57] F. Berz, “Theory of low frequency noise in Si MOST’S,” Solid-State Electron., vol. 13, p. 631-647, May. 1970.

[58] S. T. Hsu, "Surface state related 1/*f* noise in MOS transistors," Solid-State Electron., vol. 13, pp. 1451-1459, Nov. 1970.

[59] H.-S. Fu and C.-T. Sah, "Theory and experiments on surface 1/f noise," IEEE Trans. Electron Devices, vol. 19, no. 2, pp. 273-285, Feb. 1972.

[60] Z. Celik and T. Y. Hsiang, “Study of 1/ f noise in N-MOSFET’s: Linear region,” IEEE Trans. Electron Devices, vol. 32, no. 12, pp. 2797-2802, Dec. 1985.

[61] F. N. Hooge, "1/*f* noise is no surface effect," Phys. Lett. A, vol. 29A, no. 3, pp. 139-140, Apr. 1969.

[62] F. N. Hooge, "1/f noise sources," IEEE Trans. Electron Devices, vol. 41, no. 11, pp. 1926-1935, Nov. 1994.

[63] J. Chang, A. A. Abidi, and C. R. Viswanathan, "Flicker noise in CMOS transistors from subthreshold to strong inversion at various temperatures," IEEE Trans. Electron Devices, vol. 41, pp. 1965- 1971, Nov. 1994.

[64] K. K. Hung, P. K. Ko, C. Hu, and Y. C. Cheng, "A unified model for the flicker noise in metal-oxide-semiconductor field-effect transistors," IEEE Trans. Electron Devices, vol. 37, no. 3, pp. 654-665, Mar. 1990.

[65] K. K. Hung, P. K. Ko, C. Hu, and Y. C. Cheng, "A physics-based MOSFET noise model for circuit simulators", IEEE Trans. Electron Devices, vol. 37, no. 5, pp.1323-1333, May. 1990.

[66] T. Tsunomura, A. Nishida, F. Yano, A. T. Putra, K. Takeuchi, S. Inaba, S. Kamohara, K. Terada, T. Hiramoto and T. Mogami, "Analyses of 5 sigma $V_{th}$ fluctuation in 65 nm MOSFETs using Takeuchi plot," in VLSI Tech. Dig., 2008, pp. 156-157.

[67] K. Abe, S. Sugawa, S. Watabe, N. Miyamoto, A. Teramoto, Y. Kamata, K. Shibusawa, M. Toita and T. Ohmi, "Random telegraph signal statistical analysis using a very large-scale array TEG with 1 m MOSFETs," in VLSI Tech. Dig., 2007, pp. 201-211.

[68] S. Realov and K. L. Shepard, "Analysis of random telegraph noise in 45-nm CMOS using on-chip characterization system," IEEE Trans. Electron Devices, vol. 60, no. 5, pp. 1716-1722, May. 2013.

[69] N. Tega, H. Miki, Z. Ren, C. P. D'Emic, Y. Zhu, D. J. Frank, M. A. Guillorn, D.-G. Park, W. Haensch, and K. Torii, "Impact of HK / MG stacks and Future Device Scaling on RTN," in Proc. IEEE. Int. Rel. Phys. Symp. Tech. Dig. (IRPS), 2011, pp. 6A.5.1-6A.5.6.

[70] N. Tega, H. Miki, F. Pagette, D. J. Frank, A. Ray, M. J. Rooks, W. Haensch, and K. Torii, "Increasing threshold voltage variation due to random telegraph noise in FETs as gate lengths scale to 20 nm," in VLSI Tech. Dig., 2009, pp. 50-51.

[71] A. Ghetti, C. M. Compagnoni, F. Biancardi, A. Lacaita, S. Beltrami, L. Chiavarone, A. Spinelli and A. Visconti, "Scaling trends for random telegraph noise in deca-nanometer Flash memories," in IEDM Tech. Dig., 2008, pp. 835-838.

[72] M. J. M. Pelgrom, A. C. J. Duinmaijer and A. P. G. Welbers, "Matching properties of MOS transistors,", IEEE J. Solid-State Circuits, vol. 24, no. 5, pp. 1433-1439, Oct. 1989.

[73] M. Pelgrom, H. Tuinhout, and M. Vertregt, "Transistor matching in analog CMOS applications," in IEDM Tech. Dig., 1998, pp. 915-918.

[74] K. Takeuchi, T. Nagumo, S. Yokogawa, K. Imai and Y. Hayashi, "Single-charge-based modeling of transistor characteristics fluctuations based on statistical measurement of RTN

amplitude," in VLSI Tech. Dig., 2009, pp. 54-55.

[75] A. Asenov, R. Balasubramaniam, A. R. Brown, and J. H. Davies, "RTS amplitudes in decananometer MOSFETs: a 3D simulation study," IEEE Trans Electron Devices, vol. 50, no. 3, pp.839-845, Mar. 2003.

[76] M. Agostinelli, J. Hicks, J. Xu, B. Woolery, K. Mistry, K. Zhang, S. Jacobs, J. Jopling, W. Yang, B. Lee, T. Raz, M. Mehalel, P. Kolar, Y. Wang, J. Sandford, D. Pivin, C. Peterson, M. DiBattista, S. Pae, M. Jones, S. Johnson and G. Subramanian, "Erratic fluctuations of SRAM cache $V_{min}$ at the 90 nm process technology node," in IEDM Tech. Dig., 2005, pp. 655-658.

[77] M. Tanizawa, S. Ohbayashi, T. Okagaki, K. Sonoda, K. Eikyu, Y. Hirano, K. Ishikawa, O. Tsuchiya and Y. Inoue, "Application of a statistical compact model for random telegraph noise to scaled-SRAM $V_{min}$ analysis," in VLSI Tech. Dig., 2010, pp. 95-96.

[78] K. Takeuchi, T. Nagumo, K. Takeda, S. Asayama, S. Yokogawa, K. Imai and Y. Hayashi, "Direct observation of RTN-induced SRAM failure by accelerated testing and its application to product reliability assessment," in VLSI Tech. Dig., 2010, pp. 189-190.

[79] M. Yamaoka, H. Miki, A. Bansal, S. Wu, D. J. Frank, E. Leobandung, and K. Torii, "Evaluation methodology for random telegraph noise effects in SRAM arrays," in IEDM Tech. Dig., 2011, pp. 745-748.

[80] S. O. Toh, Y. Tsukamoto, Z. Guo, L. Jones, T.-J. K. Liu, and B. Nikolić, "Impact of Random Telegraph Signals on $V_{min}$ in 45nm SRAM," in IEDM Tech. Dig., 2009, pp. 767-770.

[81] B. Zimmer, O. Thomas, S. O. Toh, T. Vincent, K. Asanović, and B. Nikolić, “Joint impact of random variations and RTN on dynamic writeability in 28nm bulk and FDSOI SRAM,” in Proc. IEEE Eur. Solid-State Device Res. Conf. (ESSDERC), 2014, pp. 98-101.

[82] H. Qiu, T. Mizutani, Y. Yamamoto, H. Makiyama, T. Yamashita, H. Oda, S. Kamohara,

N. Sugii, T. Saraya, M. Kobayashi, and T. Hiramoto, “Impact of random telegraph noise on write stability in silicon-on-Thin-BOX (SOTB) SRAM cells at low supply voltage in sub-0.4V regime,” in VLSI Tech. Dig., 2015, pp. T38-39.

[83] M. B. Wilk and R. Gnanadesikan, "Probability Plotting Methods for the Analysis of Data," Biometrika, vol. 55, no. 1, pp. 1-17, Mar. 1968.

[84] S. Realov and K. Shepard, "Random telegraph noise in 45-nm CMOS: Analysis using an on-chip test and measurement system," in IEDM Tech. Dig., 2010, pp. 624-627.

[85] H. Miki, N. Tega, M. Yamaoka, D. J. Frank, A. Bansal, M. Kobayashi, K. Cheng, C. P. D’Emic, Z. Ren, S. Wu, J-B. Yau, Y. Zhu, M. A. Guillorn, D.-G. Park, W. Haensch, E. Leobandung, and K. Torii, “Statistical measurement of random telegraph noise and its impact in scaled-down high-/metal-gate MOSFETs,” in IEDM Tech. Dig., 2012, pp. 450-453.

[86] R. Gusmeroli, C. M. Compagnoni, A. Riva, A. S. Spinelli, A. L. Lacaita, M. Bonanomi and A. Visconti, "Defects spectroscopy in $SiO_2$ by statistical random telegraph noise analysis," in IEDM Tech. Dig., 2006, pp. 483-486.

[87] K. Fukuda, Y. Shimizu, K. Amemiya, M. Kamoshida and C. Hu, "Random telegraph noise in Flash memories—Model and technology scaling," in IEDM Tech. Dig., 2007, pp. 169-172.

[88] C. M. Compagnoni, R. Gusmeroli, A. Spinelli, A. Lacaita, M. Bonanomi and A. Visconti, “Statistical model for random telegraph noise in Flash memories," IEEE Trans. Electron Devices, vol. 55, no. 1, pp. 388-395, Jan. 2008.

[89] M. Yamaoka, K. Osada, and T. Kawahara, “A cell-activation-time controlled SRAM for low-voltage operation in DVFS SoCs using dynamic stability analysis,” in Proc. IEEE Eur. Solid-State Circuit Res. Conf. (ESSCIRC) Dig., 2008, pp. 286–289.

[90] D. Khalil, M. Khellah, N.-S. Kim, Y. Ismail, T. Karnik, and V. K. De, “Accurate

estimation of SRAM dynamic stability," IEEE Trans. VLSI, vol. 16, no. 12, pp. 1639–1647, Dec. 2008.

[91] J. Wang, S. Nalam, and B. H. Calhoun, "Analyzing static and dynamic write margin for nanometer SRAMs," in Proc. IEEE. Int. Symp. Low Power Electron. Design (ISLPED), 2008, pp. 129-134.

[92] S. Toh, G. Zheng, T.-J. K. Liu, and B. Nikolic, "Characterization of dynamic SRAM stabilityin 45 nm CMOS," IEEE J. Solid-State Circuits, vol. 46, no. 11, pp. 2702-2712, Nov. 2011.

[93] H. Miki, N. Tega, Z. Ren, C. P. D'Emic, Y. Zhu, D. J. Frank, M. A. Guillorn, D.-G. Parks, W. Haensch, and K. Torii, "Hysteretic drain-current behavior due to random telegraph noise in scaled-down FETs with high-k/metal-gate stacks," in IEDM Tech. Dig., 2010, pp. 620-623.

# Publication List

## [Journals]

[1] H. Qiu, T. Mizutani, T. Saraya, and T. Hiramoto, “Comparison and statistical analysis of four write stability metrics in bulk CMOS static random access memory cells,” **Jpn. J. Appl. Phys.**, vol. 54, pp. 04DC09(01)-04DC09(04), Feb. 2015.

[2] H. Qiu, K. Takeuchi, T. Mizutani, Y. Yamamoto, H. Makiyama, T. Yamashita, H. Oda, S. Kamohara, N. Sugii, T. Saraya, M. Kobayashi, and T. Hiramoto, “Statistical Write Stability Characterization in SRAM Cells at Low Supply Voltage”, submitted to **IEEE Trans. Electron Devices**, 2016.

## [International Conferences]

[1] H. Qiu, K. Takeuchi, T. Mizutani, T. Saraya, M. Kobayashi, and T. Hiramoto, "A New Write Stability Metric Using Extended Write Butterfly Curve for Yield Estimation in SRAM Cells at Low Supply Voltage," in IEEE Int. Conf. on Microelectronic Test Structures (**ICMTS**), 2016, pp. 126-129.

[2] H. Qiu, T. Mizutani, Y. Yamamoto, H. Makiyama, T. Yamashita, H. Oda, S. Kamohara, N. Sugii, T. Saraya, M. Kobayashi, and T. Hiramoto, “Impact of random telegraph noise on write stability in silicon-on-Thin-BOX (SOTB) SRAM cells at low supply voltage in sub-0.4V regime,” in **VLSI Tech. Dig.**, 2015, pp. T38-39.

[3] H. Qiu, T. Mizutani, Y. Yamamoto, H. Makiyama, T. Yamashita, H. Oda, S. Kamohara, N. Sugii, T. Saraya, M. Kobayashi, and T. Hiramoto, "Statistical Analysis of Four Write Stability Metrics in Fully Depleted Silicon-on-thin-BOX (SOTB) and Bulk SRAM Cells at Low Supply Voltage," in Proc. IEEE Int. Conf. Solid-State Inter. Circuit Technol. (**ICSICT**), 2014, pp. 987-989. (*This paper received the best paper award.)*

[4] H. Qiu, T. Mizutani, T. Saraya, and T. Hiramoto, "Comparison and Statistical Analysis of Four Write Stability Metrics in Bulk CMOS SRAM Cells," in Int. Conf Solid State Dev. Mater. (**SSDM**), 2014, pp. 848-849.

## [Domestic Conferences]

[1] H. Qiu, K. Takeuchi, T. Mizutani, T. Saraya, M. Kobayashi, and T. Hiramoto, "A New Write Stability Metric for Yield Estimation in SRAM Cells at Low Supply Voltage," **The 77th JSAP Autumn Meeting**, Toki Messe, Niigata, Japan, September 14th, 2016. (to be presented)

[2] H. Qiu, T. Mizutani, Y. Yamamoto, H. Makiyama, T. Yamashita, H. Oda, S. Kamohara, N. Sugii, T. Saraya, M. Kobayashi, and T. Hiramoto, “Impact of random telegraph noise (RTN) on write stability in silicon-on-thin-BOX (SOTB) SRAM cells in sub-0.4V regime,” **The 76th JSAP Autumn Meeting**, Nagoya Congress Center, Nagoya, Japan, September 15, 2015.

[3] 邱浩, 水谷朋子, 山本芳樹, 槇山秀樹, 山下朋弘, 尾田秀一, 蒲原史朗, 杉井信之, 更屋拓哉, 小林正治, 平本俊郎, “完全空乏型 Silicon-on-Thin-BOX SRAM セルの低電源電圧の書き込み安定性における RTN の影響,” **応用物理学会シリコンテクノロジー分科会第 184 回研究集会**, 甲南大学ネットワークキャンパス東京, 東京, 日本, 2015 年 8 月 17 日.

[3] H. Qiu, T. Mizutani, Y. Yamamoto, H. Makiyama, T. Yamashita, H. Oda, S. Kamohara, N. Sugii, T. Saraya, M. Kobayashi, and T. Hiramoto, "Statistical analysis of four write stability metrics in fully depleted silicon-on-Thin-BOX (SOTB) SRAM cells at low supply voltage down to 0.4V," **The 62nd JSAP Spring Meeting**, Tokai University, Kanagawa, Japan, March 12, 2015.

[4] H. Qiu, T. Mizutani, T. Saraya, and T. Hiramoto, "Measurement and statistical comparison of four write stability metrics in bulk CMOS SRAM cells," **The 62nd JSAP Spring Meeting**, Tokai University, Kanagawa, Japan, March 12, 2015.