A Scaling Trend of Variation-Tolerant SRAM Circuit Design in Deeper Nanometer Era

Hiroyuki Yamauchi

Abstract—**Evaluation results about area scaling capabilities of various SRAM margin-assist techniques for random V_T variability issues are described. Various efforts to address these issues by not only the cell topology changes from 6T to 8T and 10T but also incorporating multiple voltage-supply for the cell terminal biasing and timing sequence controls of read and write are comprehensively compared in light of an impact on the required area overhead for each design solution given by ever increasing V_T variation (σ_{VT}). Two different scenarios which hinge upon the EOT (Effective Oxide Thickness) scaling trend of being pessimistic and optimistic, are assumed to compare the area scaling trends among various SRAM solutions for 32 nm process node and beyond. As a result, it has been shown that 6T SRAM will be allowed long reign even in 15 nm node if σ_{VT} can be suppressed to < 70 mV thanks to EOT scaling for LSTP (Low Standby Power) process.**

Index Terms—**SRAM scaling, SRAM design solution, SRAM scaling trend, SRAM margin assist, deeper nanometer**

I. INTRODUCTION

SRAM designers have recognized that increasing of random V_T variation (σ_{VT}) heavily modulates the minimum SRAM operating voltage V_{DD_MIN}, as shown in Fig. 1, thus SRAM scaling often hinges upon the future circuit and device technologies direction. Many techniques to address this issue have attracted much attention in many conferences and papers so far [1-11] but no comprehensive evaluation reports about their scaling capabilities for 22 nm node and beyond have been presented. Thus, in this paper, various SRAM solutions proposed so far are analyzed in light of the impact on area scaling given by 1) overall device technology scaling, 2) cell topologies including 6T[1-5], 8T [6-10], and 10T [11], 3) multiple cell-terminal-biasing for read and write margin assists [12-15], and 4) read modify write sequence control for 8T that can avoid the half-select issue [8].

An introducing high-κ metal-gate (MG) is the best example [7] for the device technology scaling solution that can scale the effective oxide thickness (EOT) while eliminating the depletion layer in the gate, resulting in suppressing the σ_{VT} of random variation. Conventionally, it was believed that the EOT scaling had reached the limit due to the increasing of gate-leakage and the σ_{VT} could not be easily scaled down.

However, such kind of high-κ MG device innovation enables to extend the scaling limit of the MOSFET gate channel size defined by (Lg×Wg).

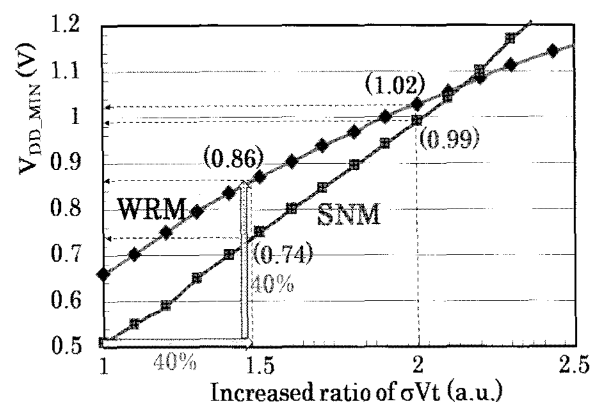


Fig. 1. Increasing V_{DD_MIN} with increasing σ_{VT}.

Manuscript received Mar. 5, 2009; revised Mar. 13, 2009.
Fukuoka Institute of Technology, Fukuoka, Japan, 811-0295
Tel: 81-92-606-4920
E-mail : yamauchi@fit.ac.jp

Meanwhile, cell topology change from 6T to 8T was proposed to alleviate the impact of σ_{VT} increasing on the read and write stability issues. However, it has come to be well known that proposed 8T can't solve the column interleaving issue for write operation [8,16-18]. To address this issue, incorporating a special timing sequence of read modify write with 8T cell was proposed [8], which needs an extra time to write access but can alleviate column interleaving issue. 10T SRAM cell was also proposed to address this kind of column interleaving issue without any extra access time penalty at the expense cost of cell area due to the use of two more transistors, which are used to enable a cross-point access of selected row and column lines [11].

On the other hand, an industry's real trend for this issue looks like there are two trends for 65 nm and 45 nm generations and the both are similarly using 6T SRAM cell, as shown in Fig. 2 (a). One trend is using an SRAM supply voltage regulator like voltage down converter (VDC) to keep the SRAM supply voltage (V_{DD_SRAM}) higher than the minimum (V_{DD_MIN}) and lower than the maximum voltages (V_{DD_MAX}) [20].

The other is incorporating the offset biasing schemes for the read and write design solutions for 6T cell [4,13,15,19] without using VDC. Based on such trend, as shown in Fig. 2 (a), the 6T cell can be expected to be survived even in 32 nm and beyond depending on: 1) amount of σ_{VT}, 2) whether VDC can be incorporated, and 3) whether upsizing 6T cell to suppress σ_{VT} is still smaller than that for 8T or 10T. However, if σ_{VT} would become larger than a certain amount, 8T cell with read modify write sequence control could be better than 6T in terms of overall SRAM core area, as discussed in [8,16,17].

Fig. 2 (a) conceptually shows the possible options of using a different memory cell and its control scheme for 32 nm and beyond: 1) upsizing 6T with smaller σ_{VT} for each transistor, 2) 8T with read-modify-write (time-multiplexing) scheme [8], and 3) 10T with cross-point access [11]. However, if σ_{VT} would become larger than a certain amount and the required minimum operating voltage V_{DD_MIN} becomes lower like [11], the 8T cell with read modify write sequence control or the 10T cell could be better than the upsizing 6T cell in terms of overall SRAM core area, as conceptually sketched in Fig. 2 (b).

When σ_{VT} is low enough to keep a sufficient read and write stability-margins for 6T SRAM even if not using

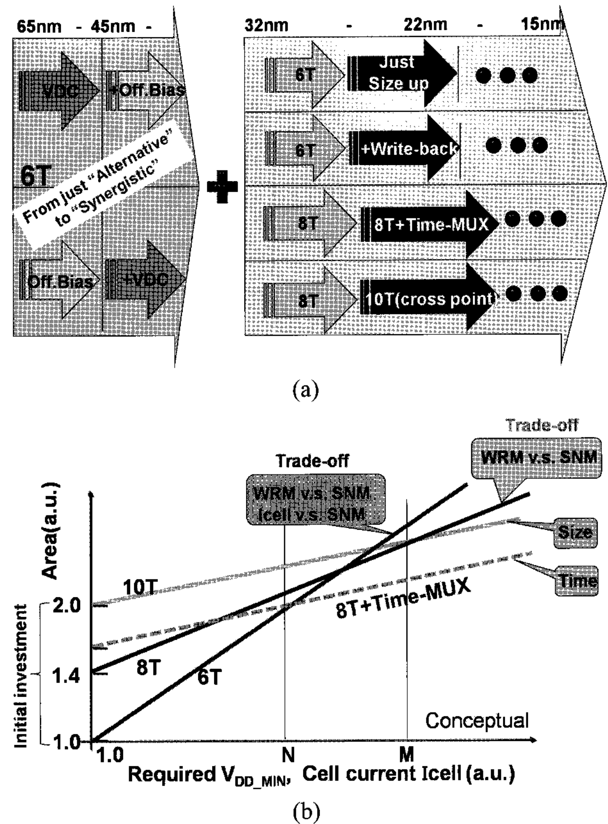


Fig. 2. (a) Various cell options for 32 nm and beyond and (b) Conceptual trend chart of cell area overhead.

any design solutions including the upsizing 6T, 8T and 10T become larger than 6T by 40% and 100%, respecttively, as sketched out in Fig. 2 (b). Such kind of relationships hinge upon the amount of σ_{VT}, the required level of V_{DD_MIN}, and the cell current (access speed). Thus, they are used for X-axis of Fig. 2 (b).

It is crucial for the SRAM designer to predict when such kind of crossover happens. Because an impact on SRAM margins given by the increasing σ_{VT} has become increasingly severe since 65 nm node, various SRAM design solutions have been proposed but no paper has compared their scaling limitations and discussed these kind of trends for 32 nm and beyond so far. Therefore, the purpose of this paper is for the first time to clarify which design solutions will be survived until when, if σ_{VT} will be ever increased. This paper puts primarily emphasis on discussions about 1) read and write stability design solution and its limitation in section II, 2) scaling impact of incorporating of new cell topologies and timing sequence of read modify write on scaling limit extend-ability in section III and IV, respectively, 3) an impact on margin assist capability given by ECC in section V, and 4) area

scaling trend comparisons among the design solutions in section VI, respectively, followed by conclusion in section VII.

II. CELL STABILITY IMPROVEMENT AND ITS LIMITATION

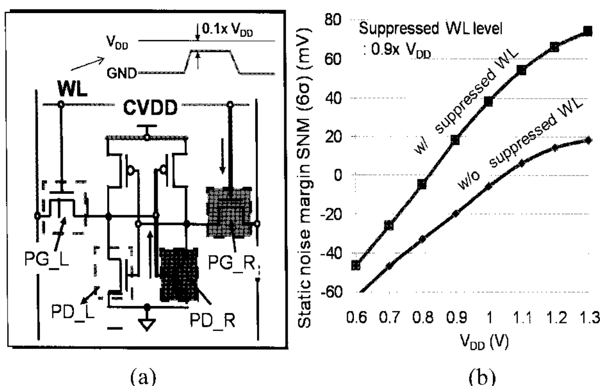
The cell stability depends on the amount of V_T mismatch and the operation voltage V_{DD} as well as the cell ratios of γ-ratio for write and β-ratio for read. The primary reasons for reducing the read and write margins are due to the lost of balance of the cell transistor strengths which are referred to as γ-ratio and β-ratio caused by V_T random variations. Each of the γ-ratio and the β-ratio has to be considered as statistical distribution with a larger variation.

1. Read Margin Assist and Its Limitation

To increase read margin, the potential of the gate electrode of the pull-down NFET (PD_R) can be made higher than that of the pass-gate NFET (PG_R) so that β-ratio becomes larger, as shown in Fig. 3 (a). In order to implement such relationship, there are two techniques: 1) suppress the WL level [4,13] and 2) boost the CVDD level [14,23]. Fig. 3 (b) shows SNM improvements when using the suppressed WL by 10% of V_{DD} as a function of V_{DD}. It can be seen that SNM is increased by about 45 mV for 100 mV-suppressed WL at V_{DD}=1.0 V.

Instead of suppressing WL, boosting CVDD also enables such potential relationship between WL and CVDD.

As described above, the intentional offset biasing between WL and CVDD can compensate the lost read margin due to V_T mismatch. However, the design solu



(a) (b)

Fig. 3. (a) SNM assist basics for 6T SRAM and (b) SNM improvement by suppressed WL as a function of VDD.

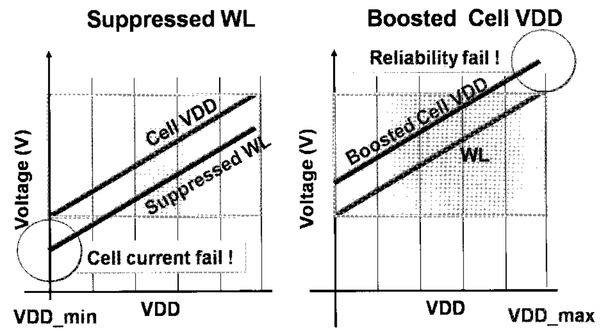


Fig. 4. V_{DD} window for suppressed WL and boosted Cell V_{DD} scheme.

tions for read margin improvement have the limitation of the operating voltage window as shown in Fig. 4.

In case of using suppressed WL scheme and boosted CVDD scheme, it can be seen that the cell current becomes too low to read at the lower V_{DD} boundary due to the suppressed WL level, as shown in Fig. 4.

Higher power supply for the array can also become too excessive to ensure device reliability at higher V_{DD} boundary due to the boosted CVDD level.

2. Write Margin Assist and Its Limitation

To increase write margin WRM, the potential of the gate electrode of the pull-up PFET (PU_L) can be made lower than that of the pass-gate NFET (PG_L) so that γ-ratio becomes smaller as shown in Fig. 5 (a) [12]. In order to implement such relationship, there are two techniques: 1) pull-down of CVDD [12] and 2) negative BL over-driving [15]. WRM improvements when using the pull-down of CVDD is proportional to the amount of CVDD pull-down voltage, as shown in Fig. 5 (b). For example, WRM is increased by 130 mV for the pull-down of CVDD by 10% of V_{DD} at V_{DD}=1.1 V [12].

To ensure the potential of the connected node of PU and PG to be lowered than the trip point of the inverter, the negative BL over-driving scheme has been proposed [15].

Fig. 5 (c) shows that WRM improvements when using the negative BL scheme as a function of V_{DD}. It can be found that WRM is increased by about 280 mV at V_{DD}=1.0 V. The WRM can be increased by the amount of negatively BL over-driving (-280 mV at V_{DD}=1.0 V).

In the case of using pull-down CVDD scheme and negative BL scheme, as shown in Fig. 6, it can be seen that the cell data retention voltage becomes too low to retain the data at the lower V_{DD} boundary due to the pull-

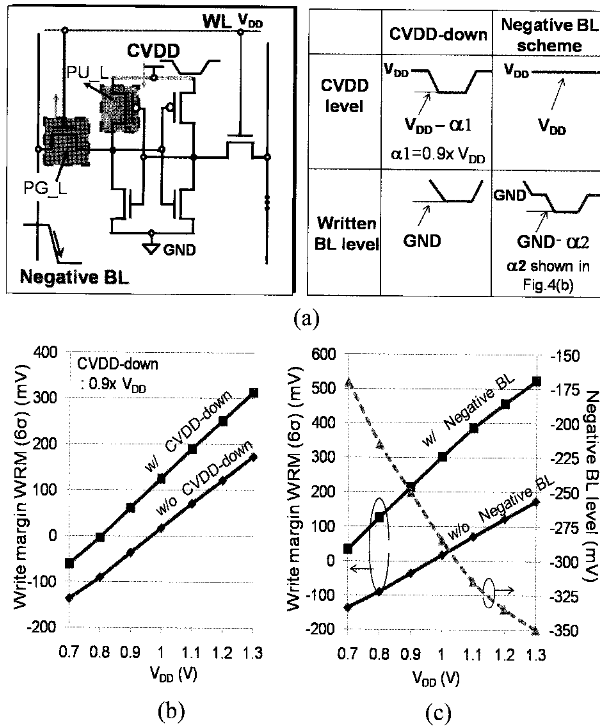


Fig. 5. (a) WRM assist basics, WRM improvements by (b) CVDD-down scheme and (c) negative BL scheme as a function of V_{DD}.

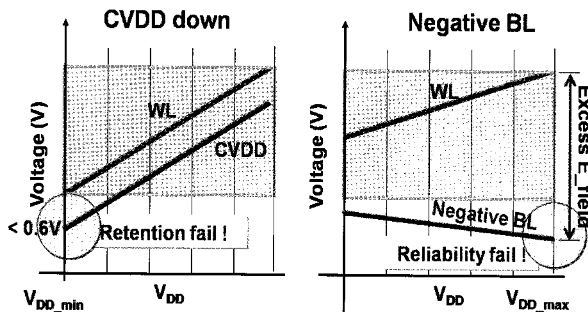


Fig. 6. V_{DD} window for CVDD-down and negative BL schemes.

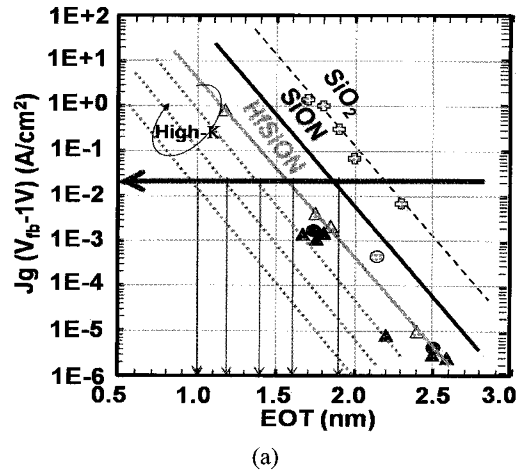
down CVDD level and the applied the electric field becomes too excessive to ensure device reliability at higher V_{DD} boundary due to the negative BL level, respectively [15].

3. V_T Random Variation Trend

The σ_{VT} of MOSFET V_T variation is proportional to Avt / $\sqrt{Lg \times Wg}$, where Avt is Pelgrom coefficient [21] and Lg and Wg are gate channel length and width, respectively [21,27,28]. Since Avt is proportional to the effective oxide thickness EOT [21,27,28], it can be assumed that σ_{VT} is proportional to EOT/$\sqrt{Lg \times Wg}$ as well.

Fig. 7 (a) shows the gate leakage scaling for different EOT and gate materials (SiO_2, SiON, HfSiON, and other High-κ) [24-26]. The international technology roadmap for semiconductors 2007 (ITRS2007) [30] shows that an upper limit of the gate leakage density Jg is about 1E-1 A/cm^2 for low standby power (LSTP) process. The required EOT to meet this target depends on the gate material. It can be found that SiON for 1.9 nm, HfSiON for 1.6 nm, and the new High-k for < 1.4 nm are needed to suppress the maximum gate leakage Jg (V_{fb}-1 V) to < 2E-2 A/cm^2 (5-times margin for the upper limit of Jg for LSTP process) [27,30].

Fig. 7 (b) shows the trend of EOT and σ_{VT} from 65 nm to 15 nm process node. The σ_{VT} could be suppressed to < 75 mV even when 15 nm process node if EOT could be successfully scaled down, otherwise, it could be increased up to 135 mV, as shown in Fig. 7 (b).



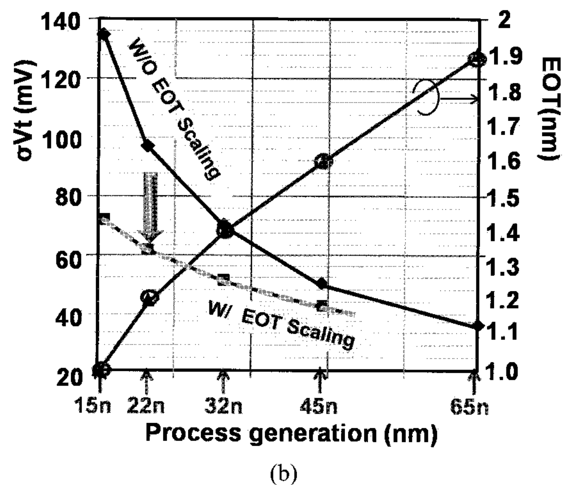


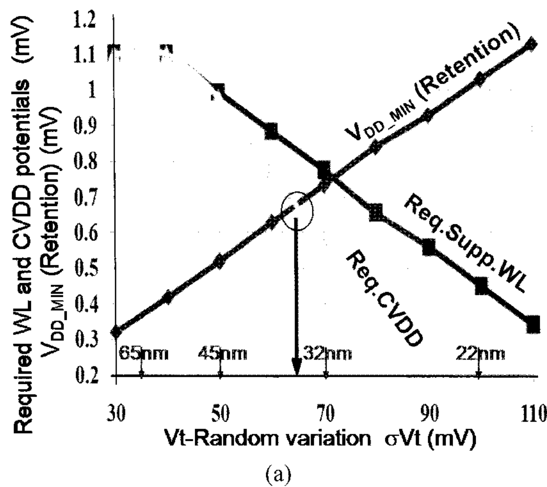
Fig. 7. (a) Gate leakage scaling with different dielectrics and EOT and (b) σ_{VT} scaling trend w/ and w/o EOT scaling.

4. Various Assumptions to Study the Impact on Limitation of Assist Circuits Given by σ_{VT} Increasing

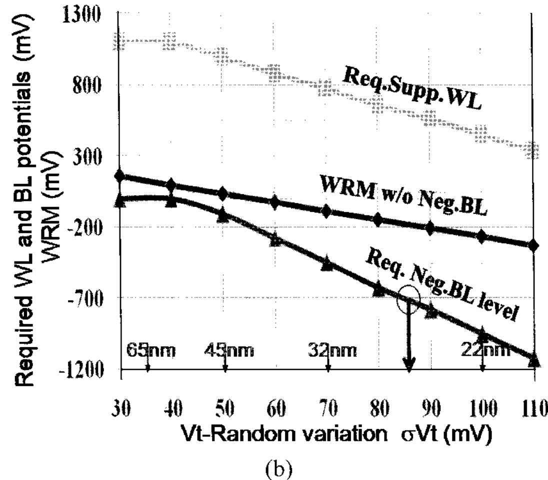
a) V_T variations are assumed as follows:

1) σ_{VT} of random V_T variation for PMOS minimum device in SRAM is assumed to be 0.76-times smaller than that for NMOS minimum device based on my experience in the industry. This trend is not so different from the other references [28,29], although it depends on the process technologies and its ratio of device sizes of NMOS and PMOS [28].

2) To show a trend of limitation of assist circuits caused by process scaling, random σ_{VT} is used as an X-axis instead of using a "process node" like 65 nm, 45 nm, and so on, as shown in Fig. 8 and 9. This is because a different type of process like high performance (HP) process and LSTP process bring about a different σ_{VT} [27,28, 33-36] even if assuming the same process node. This is because the σ_{VT} itself depends on EOT, V_T, channel doping amount Na, gate work function and so on, besides gate channel dimensions of Lg and Wg[27,28,33-36].



(a)



(b)

Fig. 8. Limitations of control levels for a) CVDD-down and b) negative BL.

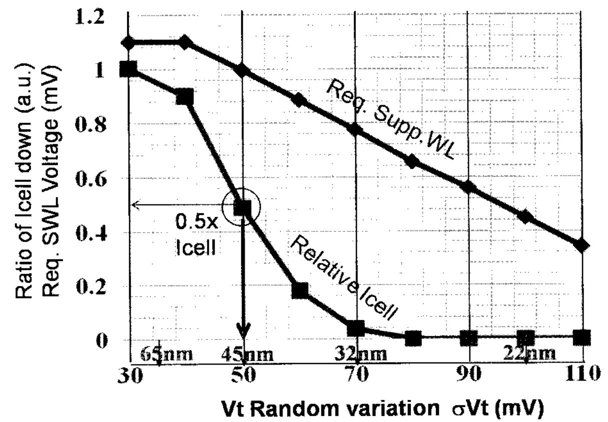


Fig. 9. Icell down impacts given by suppressed WL.

Hence, the σ_{VT} versus area impact is more universal.

3) Increasing pace of random σ_{VT} every generation is assumed as the two cases with and without EOT scaling, respectively, as shown in Fig. 7 (b).

Indicators for process node being used in Fig. 9 and others, excluding Fig. 20 (b), are assumed that an increasing pace of σ_{VT} is only one case without EOT scaling for simplicity.

For example, σ_{VT} of 40 mV and 50 mV for 45 nm process node with the Hf-doped silicate gate [33] and without using High-κ gate are referred to [33], respectively.

b) Test bench setup items used for simulation are as follows:

1) Static noise margin SNM and write margin WRM are measured by DC analysis.

2) Importance sampling technique for Monte Carlo (MC) simulation is used to accurately and efficiently take into account the distribution of SNM and WRM from μ-6σ to μ+6σ [37,38].

5. σ_{VT} Increasing Impact on Limitation of Assist Circuits

If σ_{VT} for NMOS minimum device of SRAM is increased from 30 mV to 110 mV, the minimum data retention V_{DD} (V_{DD_MIN} (Retention)) is increased from 0.32 V to 1.14 V at Z=4 point (Z: number of σ unit) based on the analysis by using a Monte Carlo simulation, as shown in

Fig. 8 (a).

In the similar way, it is found that the required pull-down cell-V_{DD} (CVDD) is lowered with increasing of σ_{VT}, as shown in Fig. 8 (a).

CVDD-down assist scheme will come to end when the required CVDD level to make a WRM (6σ) larger than zero at V_{DD}=1.1 V crosses over that for data retention V_{DD_MIN}, as marked by circle in Fig. 8 (a). This is because further column-based CVDD-down causes data retention failure in the selected column. Fig. 8 (a) also shows the required suppressed-WL level to make a SNM (6σ) larger than zero at V_{DD}=1.1 V as a function of σ_{VT}.

It can be seen that if σ_{VT} is increased to > 70 mV, the required suppressed WL level has to be < 0.8 V, resulting in a significant reduction of cell current (Icell).

Meanwhile, the required negative BL level to make a WRM (6σ) larger than zero at V_{DD}=1.1 V depends on the suppressed WL level, as shown in Fig. 8 (b). As a result, it can be seen that if σ_{VT} is increased to > 85 mV, the required negative BL level has to be shallower than forward diode turn-on voltage (-0.7 V).

Moreover, the deeper suppressed WL level causes larger decrease of Icell, as shown in Fig. 9. It can be seen that if σ_{VT} is increased to > 50 mV, Icell becomes less than half of that for without using suppressed WL.

6. Limit of Design Solutions with Increasing σ_{VT}

As explained above, the increase of σ_{VT} has led to the degradation in cell stability and write margins. Fig. 10 illustrates the design limits. It can be seen that the su ppressed WL scheme for SNM assist could become non-practical if σ_{VT} becomes > 50 mV due to the significant Icell reduction by > 50%.

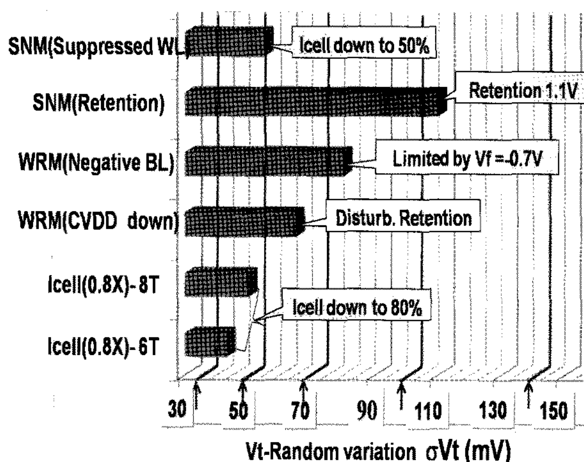


Fig. 10. Figure of merit for assist schemes.

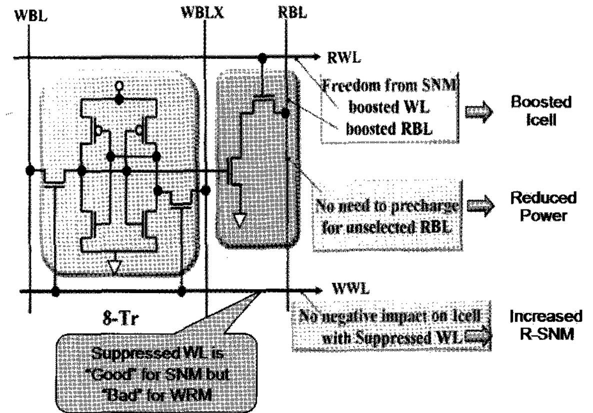


Fig. 11. Basics of margin trade-off in 8T SRAM.

If σ_{VT} is larger than 105 mV, the minimum data retention voltage would reach 1.1 V. The negative BL scheme and the CVDD-down scheme for WRM assists becomes unable to be adopted any more if σ_{VT} will become > 75 mV and > 60 mV, respectively.

The most sensitive parameter to σ_{VT} is the cell current, Icell. If 20% reduction (0.8x Icell) is the limit, often true, allowable σ_{VT} can only be < 40 mV and < 50 mV for 6T-SRAM and 8T-SRAM, respectively.

III. New Cell Topology and its Impact

1. 8T-SRAM

In order to eliminate the conflicting requirements between the cell current and the read stability in 6T SRAM cell design, two NFETs along with extra read-word-line (RWL) and read-bit-line (RBL) are added to provide designated read port SRAM, as shown in Fig. 11, which is often referred to as 8T-SRAM [6-10]. 8T-SRAM can improve the following aspects of the SRAM cell: 1) the cell current Icell, 2) the dynamic and leakage power, and 3) read stability. However, 8T-SRAM can not address the half-select issue [11,16,17] in write operation when requiring the column-interleaving operation. The trade-off between WRM and SNM has to be taken in account for the cell design, as shown in Fig. 11.

2. 10T-SRAM

In order to avoid the half-selected access issue, two more transistors can be added to enable a cross-point

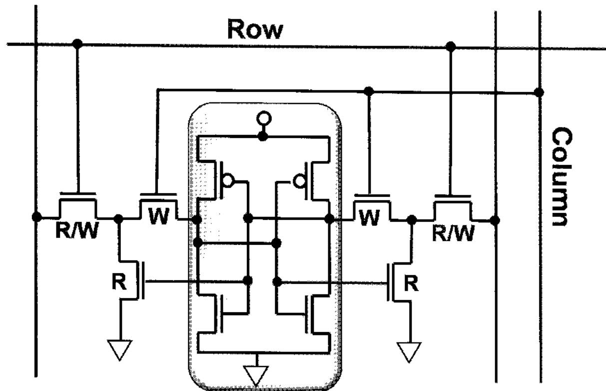


Fig. 12. 10T-SRAM.

access of selected row and column [11], as shown in Fig. 12. The pass-gate for write operation consists of the two NFETs connected in series between BL and the storage nodes, which are selected by row and column decoding, respectively, as shown in Fig. 12. Meanwhile, the pass-gate for read operation bypasses the column-decoded NFET which is connected to the storage node so that the read disturbance can be eliminated. As a result, the performances of read stability and the cell current drivability becomes the same as the read port of 8T-SRAM cell.

On the other hand, since the pass-gate for write operation consists of the two NFETs connected in series, the equivalent γ-ratio might be reduced compared with 8T. However, since the drivability of each of pass-gate NFETs can be increased without any concern for the β-ratio, the γ-ratio can be improved independently. This 10T-SRAM cell can completely eliminate the trade-off between the SNM, WRM, and the cell current at the expense cost of cell area due to the use of more transistors.

IV. Read and Write Multiplexing and its Impact

As explained above, to have the designated ports for read and write by adding two or four NFETs is one of the effective ways to avoid the conflicting requirements from read and write stability issues. In this section, another alternative way of using read and write multiplexing is described.

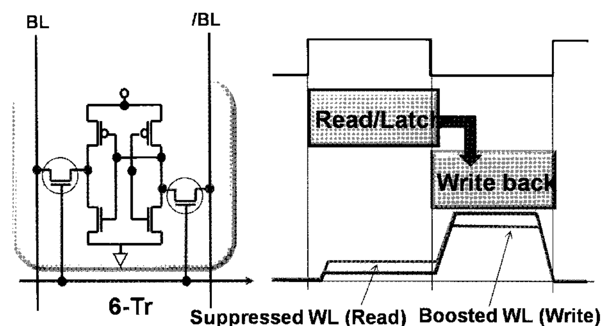
1. Time Division for Read with Suppressed WL and Write Operation

The limitations in using suppressed WL scheme stem from the difficulty in finding a good balance between read and write requirements. In order to relax such constrain, the potential level of WL can also be modulated so that sufficient read stability can be maintained without any difficulties in writing.

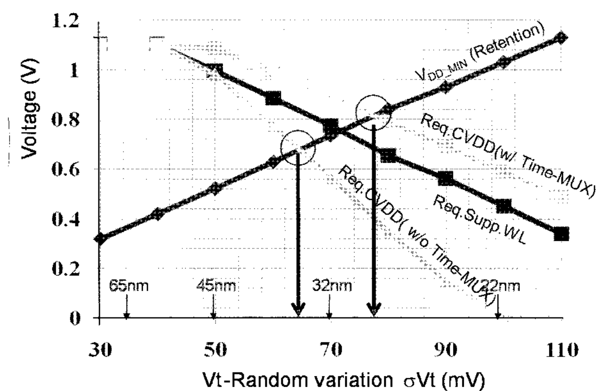
The potential levels of WL-on can be optimized separately for sequential read and write WL accesses so that its level becomes lower and higher for read and write operations, respectively, as shown in Fig. 13 (a). This scheme can shallow the required level of CVDD pull-down because of no need for the suppressed WL in write back period as shown in Figs. 13 (a) and (b). As a result, it can be seen that the cross-points of the required CVdd pull-down levels with the V_{DD_MIN} for data retention level can be extended from σ_{VT}=65 mV to 78mV by using the read and write time multiplexing, as shown in Fig. 13 (b).

In the same way, the required negative BL levels also can be relaxed from σ_{VT}=85 mV to >110 mV because of no need for the suppressed WL in write back period, as shown in Fig. 14 (a).

However, it can be found that the cell current improve



(a)



(b)

Fig. 13. (a) Time-division for read and write multiplexing for 6T-SRAM, (b) CVDD-down limit-point shifting impacts given by time multiplexing of read and write operation.

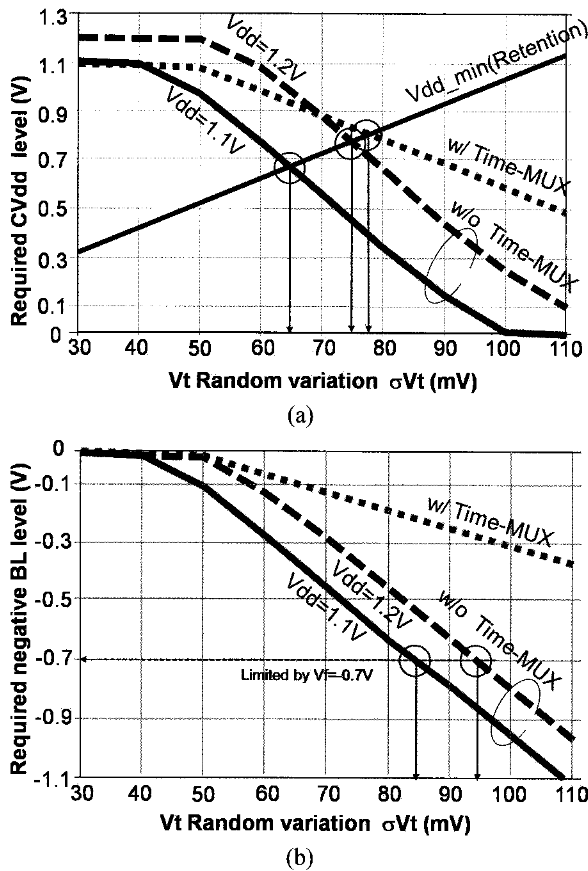


Fig. 14. Time multiplexing and raising Vdd impact on limit-points for (a) CVdd-down and (b) negative BL.

ment can't be expected at all even if using this kind of sequential control, as shown in Fig. 14 (b). It can be seen that the time-multiplexing scheme can shift the limit point for the CVDD pull-down and negative BL schemes more than that for increasing of V_{DD} from 1.1 V to 1.2 V, as shown in Fig. 14 (b).

However, since the Icell improvement can't be expected at all by only using the time-multiplexing scheme, raising V_{DD} (keeping Vdd as higher as possible) is also needed to meet the three functional margins for SNM, WRM, and Icell.

2. Time Division for Read with De-coupled Read Port and Write Operation

In order to solve the remaining issues discussed above, decoupled read port from the write port is introduced to eliminate the requirements of suppressed WL for read operation [8] as shown in Fig. 15. As a result, the decoupled read port of 8T-SRAM avoids the slowdown in the read access speed due to the suppressing WL, as

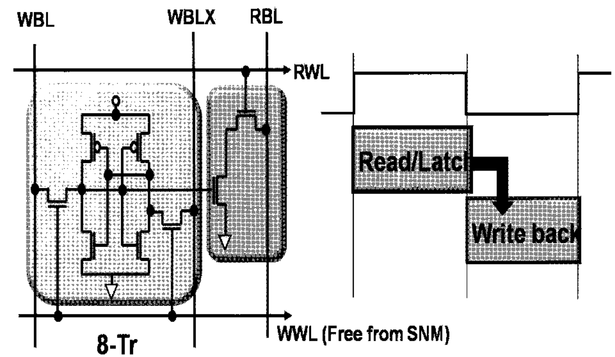


Fig. 15. Time-division for read and write multiplexing for 8T-SRAM.

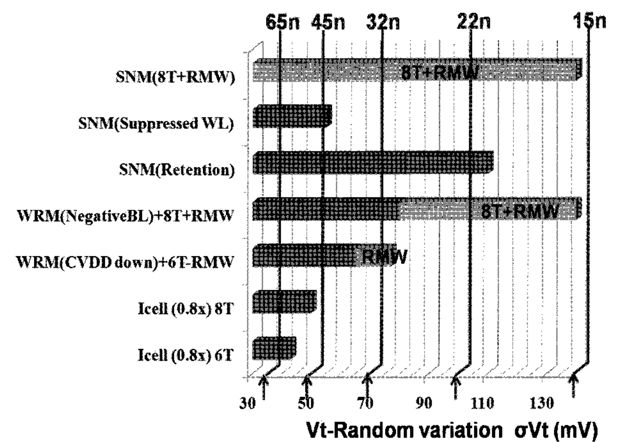


Fig. 16. Impacts on each design solution's availability given by read-modify write with decoupled R/W port.

shown in Fig. 13 (a). This scheme can completely eliminate the trade-off between the SNM, WRM, and the cell current at the expense of two-more NFETs area and sequential timing slot of read-modify-write, as shown in Fig. 15.

Fig. 16 shows how each scheme can play the key role for extending the limit of design solution as a function of σ_{VT}. It can be seen that read modify write operation with decoupled read port referred to as 8T+RMW can extend the timing of assist failure for SNM and WRM so that it can't be limited even if σ_{VT} becomes > 135 mV.

V. ECC Scheme for Redundancy

1. ECC Scheme for Yield Improvement

Error-checking and correcting (ECC) scheme is well known that it can reduce the failure rate, resulting in an increase of yield. Table 1 defines the meanings of several parameters which are used in the expressions of (1), (2)

Table 1. Definition of meaning of several parameters.

Parameter	Definition	Number
N	Total number of bits for each ECC word, $N = N_0 + N_P$	136
N_0	The number of data bits	128
N_P	The number of parity bits	8
r	Bit error rate of each bit	Z=3.4-5.4
W	The number of ECC words excluding redundant words	1024
R	The number of redundant ECC words	R=0~4
q	The number of bits out of ECC-words of W that have uncorrectable error of 2 or more	q =0~W

and (3) for estimating the yield improvement of SRAM when using the ECC words as shown below. ECC word includes N bits consisting of N_0 data bits and N_P parity bits (e.g., N_0=128, N_P=8 and N=136), which can correct only one defect cell in the word [2]. Meanwhile, the ECC word with two or more defective cells can be replaced by adding redundant ECC words of R. The yield improvement of SRAM using both of ECC words of W and ECC redundancy words of R can be estimated as follows.

If the bit-error rate is r, the probability (p) of an ECC word having two or more defective cells is expressed as:

$$p = 1-(1-r)^N - N \times r \times (1-r)^{(N-1)} \qquad (1)$$

The probability that q out of ECC-words of W has uncorrectable error of 2 or more is expressed as:

$$P_q = {}_W C_q \times p^q \times (1-p)^{(W-q)} \qquad (2)$$

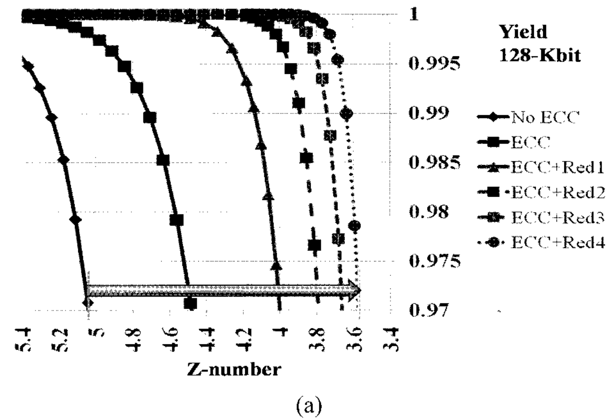
As a result, the yield can be expressed as:

$$Y = \sum_{q=0}^{R} P_q \qquad (3)$$

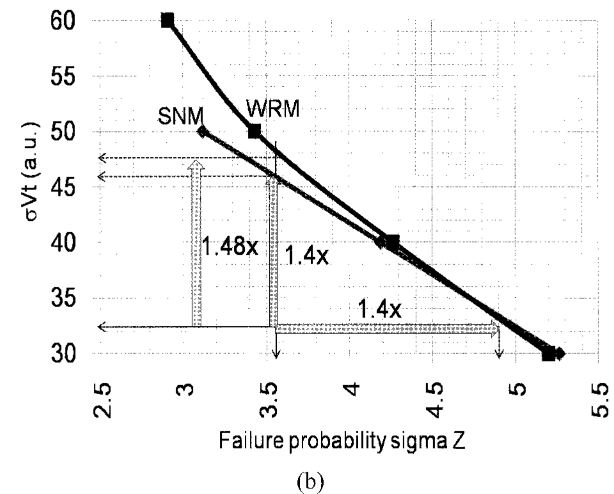
, where the probability for defects of redundant ECC words of R is neglected for simplicity [22].

Based on the above equation, ECC impact on yield improvement as a function of the bit-error rate r (which is represented by the number of Z) is calculated as shown in Fig. 17 (a), where each parameter is assumed as shown in Table-1.

It can be seen that ECC plus its redundancy can significantly improve the bit-error tolerance. If R=4, failure



(a)



(b)

Fig. 17. (a) ECC Impacts on yield improvement and (b) relationship between failure probability sigma Z and σVt.

probability sigma Z can be reduced from 5 to 3.6. This means that even if the σ_{VT} is increased by 1.4-times (=5/3.6) due to device scaling to the half (which corresponds to about one process generation), the similar yield can be maintained in the similar range of Z if the relationship between the failure probability (due to SNM and WRM) and σ_{VT} is a linear relation. Based on the reference [31], the relationships between the failure probability (due to SNM and WRM) and σ_{VT} can be known as a roughly linear in the range of Z (failure probability sigma) from 5 to 3.6, as shown in Fig. 17 (b). This implies that ECC combined with its redundancy can extend the limit of SRAM scaling by roughly one process generation even if σ_{VT} is increased by 1.4-times (=5/3.6) due to device scaling to the half without EOT scaling.

2. SRAM Scaling Impact Given by ECC

Fig. 18 shows how each of 1) read-modify write with

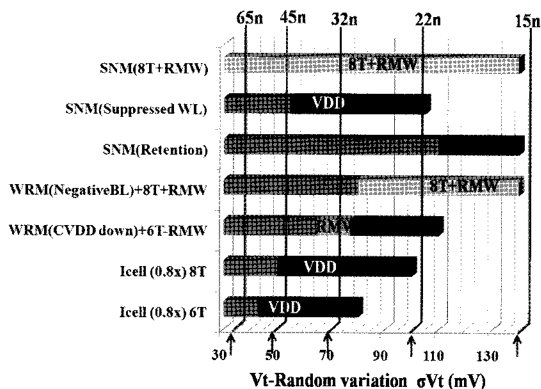


Fig. 18. Comparisons of SRAM scaling impacts given by each solution.

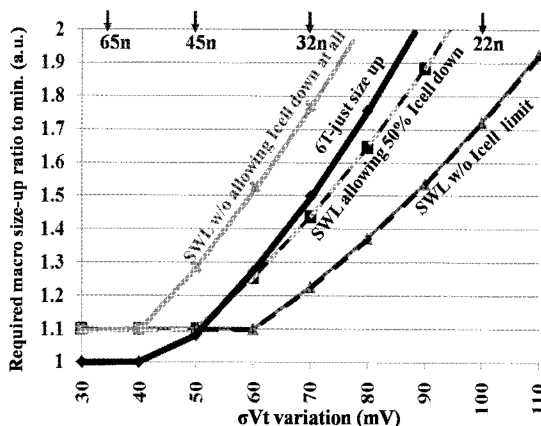


Fig. 19. Required macro up-sizing comparisons between suppressed WL and cell size-up.

8T-SRAM, 2) V_{DD} regularity and 3) ECC can relax the limitations of each design solutions as a function of σ_{VT}.

VI. Area Scaling Trend Comparisons

The design solutions for margin assists including 8T and 10T-SRAM cells are realized at the cost of additional area penalties caused by adding two or four MOSFETs and required associated peripheral circuits. Meanwhile, up-sizing 6T can reduce σ_{VT} and improve the functional margins at the expense of additional cell area due to larger channel width Wg. In that sense, the area comparisons of SRAM macro as a function of σ_{VT} is needed to discuss which design solution including up-sizing 6T becomes better solution for each request.

1. Area Comparisons of SNM and WRM Assists

As one of examples, up-sizing 6T compares with the suppressed WL depending on how much Icell down can be allowed, as shown in Fig. 19, which shows the normalized required up-sizing ratio comparisons as a function of σ_{VT} among 1) up-sizing 6T, 2) suppressed WL (SWL) without care of Icell down, 3) SWL with allowing 50% Icell down, and 4) SWL without allowing Icell down at all. It can be seen that if 50% Icell down could be allowed, suppressed WL would have greater area-advantage than just up-sizing 6T when σ_{VT} becomes > 60 mV. Meanwhile, when σ_{VT} is less than 50 mV, up-sizing 6T is superior in terms of macro area saving than the suppressed WL due to no need of additional area overhead of peripheral circuits to generate suppressed WL level.

2. Area Comparisons of Various Design Solutions

Area benefits from the cells with more transistors depend on the following requirements and conditions: 1) increasing cell current, 2) lowering VDD operation, and 3) increasing σ_{VT}. As a result, there are cross over points if areas of 6T, 8T, 8T with time-multiplexing and 10T are compared as a function of the above conditions, as shown in Fig. 20 (a) and (b).

At the initial point, the minimum macro sizes for 10T and 8T are about 2.0-times and 1.4-times larger than 6T, respectively. However, the required conditions like Icell or σ_{VT} becomes severer, the required area for 6T crosses over 8T first and then 10T, as shown in Fig. 20 (a). It can be seen that 8T-SRAM combined with time-multiplexing can be smallest at the end of the day if the time-multiplexing (read-modify-write operations) can be allowed. This is due to no need of trade-off among SNM, WRM, and Icell in 8T cell.

In order to highlight the trend of area advantage as a function of σ_{VT}, Fig. 20 (a) shows the normalized required up-sizing ratio comparisons as a function of σ_{VT} among 1) up-sizing 6T, 2) 6T combined with both of negative BL and suppressed WL, 3) 8T combined with negative BL and read-modify-write, 4) 8T combined with negative BL and suppressed WL, and 5) cross-point 10T.

It can be seen that simply up-sizing 6T crosses over the two 8Ts combined with negative BL for WRM and read-modify-write and suppressed WL for SNM when σ_{VT} becomes > 70 mV and > 80 mV, respectively.

6T combined with negative BL and suppressed WL crosses over 8T combined with negative BL and supp-

ressed WL when σ_{VT} becomes > 90 mV. 10T crosses over 8T combined with negative BL and suppressed WL when σ_{VT} becomes > 110 mV.

If σ_{VT} could be suppressed > 70 mV thanks to EOT scaling even at a 15 nm CMOS generation, as shown in Fig. 6 (b), 6T cells would be allowed long reign, as shown in Fig. 20 (b).

Fig. 20 (b) shows how EOT-scaling can play a key role for SRAM scaling by comparing with the case of non EOT-scaling (shown in Fig. 20 (a)).

It can be seen that if EOT can be scaled down to 1.9 nm for 65 nm node → 1.6 nm for 45 nm node → 1.4 nm for 32 nm node → 1.2 nm for 25 nm node → 1.0 nm for 15 nm node as shown in Fig. 6 (b), 6T-SRAM can survive with V_{DD}=1.2 V or ECC V_{DD}=1.1 V without using a read modify write operation with 8T-SRAM.

3. Various Design Options

As discussed above, many design options have been

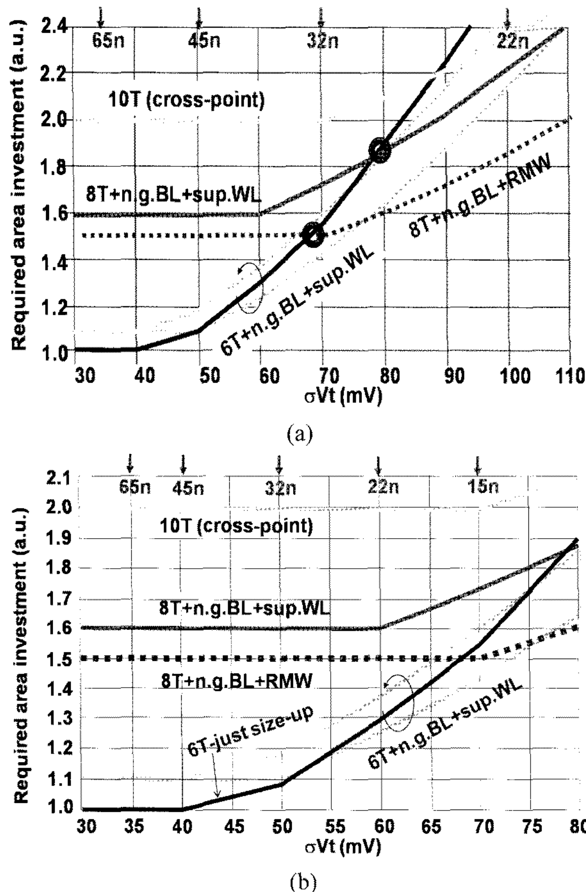


Fig. 20. Area comparisons of various SRAM cell options for the case of (a) optimistic σ_{VT} scaling (b) pessimistic σ_{VT} scaling.

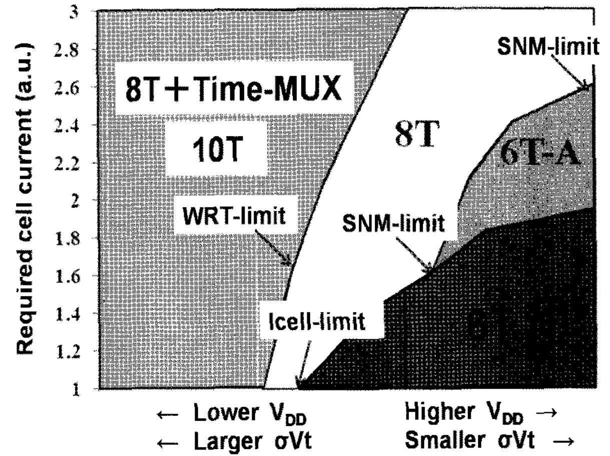


Fig. 21. Sketch for various design options.

proposed. However, it depends on the requirements and conditions that which can provide the best performance. Fig. 21 sketches out the concept for which option can become majority in each application. In Fig. 21, each application zone is roughly partitioned based on the required the cell current Icell as Y-axis and V_{DD} range or the amount of σ_{VT} as X-axis.

If V_{DD} is excessively lowered and σ_{VT} becomes extremely larger for some applications, cross-point 10T-SRAM or 8T-SRAM combined with read-modify write would be more needed than upsizing 6T-SRAM, as shown in Fig. 21 and they would become majority in such region.

Meanwhile, if the required Icell is extremely larger for higher-speed application for other applications, 8T-SRAM would become majority even if σ_{VT} is not extremely larger and V_{DD} is not excessively lowered, as sketched out in Fig. 21. 6T-A (Asymmetrical 6T-SRAM) also was proposed [12] targeting for around this region, as shown in Fig. 21.

VII. CONCLUSIONS

The comparisons of area scaling trend of various SRAM margin-assist solutions for V_T variability issues have been discussed for the first time. It is found that if the optimistic and pessimistic scenarios for increasing pace of σ_{VT} are assumed that σ_{VT} gets < 70 mV and > 130 mV at 15 nm node, respectively, 6T SRAM would be allowed long reign even in 15 nm if σ_{VT} could be suppressed to < 70 mV thanks to EOT scaling for LSTP process, otherwise 10T and 8T with read modify write would be needed after σ_{VT} becomes > 85 mV and 75 mV, respectively.

ACKNOWLEDGMENTS

This research is partially supported by a grant-in-aid for scientific research from Japan society for the promotion of science and from Information science laboratory in Fukuoka Institute of Technology.

REFERENCES

[1] F. Hamzaoglu, K. Zhang, Y. Wang, H. J. Ahn, U. Bhattacharya, Z. Chen, Y. G. Ng, A. Pavlov, K. Smits, and M. Bohr, "A 153Mb SRAM Design with Dynamic Stability Enhancement and Leakage Reduction in 45 nm Hi-K Metal Gate CMOS Technology," *in Digest of Technical Papers IEEE Solid States Circuits Conference 2008*, pp.376-377, Feb. 2008.

[2] H. Pilo, V. Ramadurai, G. Braceras, J. Gabric, S. Lamphier, and Y. Tan "A 450ps Access-Time SRAM Macro in 45 nm SOI Featuring a Two-Stage Sensing-Scheme and Dynamic Power Management," *in Digest of Technical Papers, IEEE Solid States Circuits Conference 2008*, pp.378-379, Feb. 2008.

[3] J. Pille, C. Adams, T. Christensen, S. Cottier, S. Ehrenreich, T. Kono, D. Nelson, O. Takahashi, S. Tokito, O. Torreiter, O. Wagner, and D. Wendel, "Implementation of the CELL Broadband Engine in a 65 nm SOI Technology Featuring Dual-Supply SRAM Arrays Supporting 6 GHz at 1.3 V," *in Digest of Technical Papers, IEEE Solid States Circuits Conference 2007*, pp.322-323, Feb. 2007.

[4] M. Yabuuchi, K. Nii, Y. Tsukamoto, S. Ohbayashi, S. Imaoka, H. Makino, Y. Yamagami, S. lshikura, T. Terano, T. Oashi, K. Hashimoto, A. Sebe, G. Okazaki, K. Satomi, H. Akamatsu, H. Shinohara, "A 45 nm Low-Standby-Power Embedded SRAM with Improved Immunity Against Process and Temperature Variations," *in Digest of Technical Papers, IEEE Solid States Circuits Conference 2007*, pp.326-327, Feb. 2007.

[5] Y. Wang, H. Ahn, U. Bhattacharya, T. Coan, F. Hamzaoglu, W. Hafez, C. -H. Jan, R. Kolar, S. Kulkarni, J. Lin, Y. Ng, I. Post, L. Wel, Y. Zhang, K. Zhang, M. Bohr, "A 1.1 GHz 12 μA/Mb-Leakage SRAM Design in 65 nm Ultra-Low-Power CMOS with Integrated Leakage Reduction for Mobile Applications," *in Digest of Technical Papers, IEEE Solid States Circuits Conference 2007*, pp.324-325, Feb. 2007.

[6] L. Chang, D. M. Fried, J. Hergenrother, J. W. Sleight, R. H. Dennard, R. K. Montoye, L. Sekaric, S. J. McNab, A. W. Topol, C. D. Adams, K. W. Guarini, and W. Haensch, "Stable SRAM Cell Design for the 32 nm Node and Beyond," *in Digest of Technical Papers, 2005 Symposium on VLSI Technology*, pp.128-129, June 2005.

[7] L. Chang, R. K. Montoye, Y. Nakamura, K. A. Batson, R. J. Eickemeyer, R. H. Dennard, W. Haensch, and D. Jamsek, "An 8T-SRAM for Variability Tolerance and Low-Voltage Operation in High-Performance Caches," *IEEE Journal of Solid-State Circuits*, vol. 43, no. 4, pp.956-963, April 2008.

[8] Y. Morita, H. Fujiwara, H. Noguchi, Y. Iguchi, K. Nii, H. Kawaguchi, and M. Yoshimoto, "An Area-Conscious Low-Voltage-Oriented 8T-SRAM Design under DVS Environment," *in Digest of Technical Papers, 2007 Symposium on VLSI Circuits*, pp. 256-257, June 2007.

[9] L. Chang; Y. Nakamura, R. K. Montoye, J. Sawada, A. K. Martin, K. Kinoshita, F. H. Gebara, K. B. Agarwal, D. J. Acharyya, W. Haensch, K. Hosokawa, and D. Jamsek, "A 5.3 GHz 8T-SRAM with Operation Down to 0.41 V in 65 nm CMOS," *in Digest of Technical Papers, 2007 Symposium on VLSI Circuits*, pp.252-253, June 2007.

[10] V. Naveen and A. P. Chandrakasan, "A 65 nm 8T Sub-Vt SRAM Employing Sense-Amplifier Redundancy," *in Digest of Technical Papers, IEEE Solid States Circuits Conference 2007*, pp.328-329, Feb. 2007.

[11] I-J. Chang, J-J. Kim, S. Park, K. Roy, and West Lafayette, "A 32 kb 10T Subthreshold SRAM Array with Bit-Interleaving and Differential Read Scheme in 90 nm CMOS," *in Digest of Technical Papers, IEEE Solid States Circuits Conference ISSCC 2008*, pp. 388-622, Feb. 2008.

[12] K. Zhang, U. Bhattacharya, Z. Chen, F. Hamzaoglu, D. Murray, N. Vallepalli, Y. Wang, B. Zheng, and M. Bohr, "A 3 GHz 70 Mb SRAM in 65 nm CMOS Technology with Integrated Column-Based Dynamic Power Supply," *IEEE Journal of Solid-State Circuits*, vol.31, no.1, pp.474-611, Jan. 2006.

[13] S. Ohbayashi, M. Yabuuchi, K. Nii, Y. Tsukamoto, S. Imaoka, Y. Oda, M. Igarashi, M. Takeuchi, H.

Kawashima, H. Makino, Y. Yamaguchi, K. Tsukamoto, M. Inuishi, K. Ishibashi, and H. Shinohara, "A 65 nm Soc Embedded 6T-SRAM Design for Manufacturing with Read and Write Cell Stabilizing Circuits," *in Digest of Technical Papers, 2006 Symposium on VLSI Circuits*, pp. 17-18, June 2006.

[14] H. Yamauchi, T. Iwata, H. Akamatsu, T. Fujita "A 0.8 V/100 MHz/sub-5 mW-operated mega-bit SRAM cell architecture with charge-recycle offset-source driving (OSD) scheme," *in Digest of Technical Papers, 1996 Symposium on VLSI Circuits 1996*, pp.126-127, June 1996.

[15] D. P. Wang, H. J. Liao, H. Yamauchi, W. Hwang, Y. L. Lin, Y. H. Chen, and H. C. Chang, "A 45 nm Dual-Port SRAM with Write and Read Capability Enhancement at Low Voltage," *in Digest of Technical Papers, IEEE SOCC 2007*, pp. 211-214, Oct, 2007.

[16] H. Yamauchi, "Embedded SRAM circuit design technologies for a 45 nm and beyond," *in Digest of Technical Papers, ASICON 2007*, pp. 1028-1033, Oct. 2007.

[17] H. Yamauchi, "Embedded SRAM trend in nano-scale CMOS," *in Digest of Technical Papers, IEEE MTDT 2007*, pp. 19-22, Dec. 2007.

[18] V. Ramadurai, R. Joshi, and R. Kanj, "A Disturb Decoupled Column Select 8T SRAM Cell," *in Digest of Technical Papers, 2007 Custom Integrated Circuits Conference, CICC 2007*, pp.25-28. Sept. 2007.

[19] H. Pilo, J. Barwin, G. Braceras, C. Browning, S. Burns, J. Gabric, S. Lamphier, M. Miller, A. Roberts, and F. Towler, "An SRAM Design in 65 nm and 45 nm Technology Nodes Featuring Read and Write-Assist Circuits to Expand Operating Voltage," *in Digest of Technical Papers, 2006 Symposium on VLSI Circuits*, pp.15-16, June 2006.

[20] Y. H. Chen, W. M. Chan, S. Y. Chou, H. J. Liao, H. Y. Pan, J. J. Wu, C. H. Lee, S. M. Yang, Y. C. Liu, and H. Yamauchi, "A 0.6 V 45 nm Adaptive Dual-rail SRAM Compiler Circuit Design for Lower VDD_min VLSIs," *in Digest of Technical Papers, 2008 Symposium on VLSI Circuits*, pp.210-211, June 2008.

[21] M. Pelgrom, A. Duinmaijer, and A. Welbers, "Matching properties of MOS transistors," *IEEE Journal of Solid-State Circuits*, vol.24, Issue 5, pp.1433-1439, Oct. 1989.

[22] K. Itoh, M. Horiguchi, and M. Yamaoka "Low-Voltage Limitations of Memory-Rich Nano-Scale CMOS LSIs," *in Digest of Technical Papers, 33rd European Solid State Circuits Conference, ESSCIRC 2007*, pp.11-13, Sept. 2007.

[23] Y. Chung and S-W. Shim, "An Experimental 0.8 V 256-kbit SRAM Macro with Boosted Cell Array Scheme," *ETRI Journal*, vol. 29, no. 4, August 2007.

[24] M. Suzuki, "Lanthanum Aluminate Gate Dielectric Technology with Direct Interface," *Toshiba review*, vol. 62, no. 2, pp.37-41, 2007.

[25] E. P. Gusev, V. Narayanan, and M. M. Frank, "Advanced high-j dielectric stacks with polySi and metal gates: Recent progress and current challenges," *IBM Journal Res. & Dev*,. vol. 50 no. 4/5, pp.387-410, July/September, 2006.

[26] S. J. Lee, "Characteristics of TaN gate MOSFET with ultrathin hafnium oxide," *in Digest of Technical Papers, IEDM, 2000*, p.39-42, 2000.

[27] K. J. Kuhn, "Reducing Variation in Advanced Logic Technologies: Approaches to Process and Design for Manufacturability of Nanoscale CMOS," *in Digest of Technical Papers, IEDM, 2007*, pp.471-474, 2007

[28] 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 Digest of Technical Papers, IEDM. 2007*, pp. 467-470, 2007.

[29] A. Asenov, A. Cathignol, B. Cheng, K. P. McKenna, A. R. Brown, A. L. Shluger, D. Chanemougame, K. Rochereau, and G. Ghibaudo, "Origin of the Asymmetry in the Magnitude of the Statistical Variability of n- and p-Channel Poly-Si Gate Bulk MOSFETs," *IEEE Electron Device Letters*, vol. 29, no. 8, pp. 913-915, August 2008.

[30] International Technology Roadmap for Semiconductors 2007, http://www.itrs.net/Links/2007ITRS/2007_Chapters/2007_PIDS.pdf

[31] M. Saibal, M-M. Hamid, and R. Kaushik, "Modeling of failure probability and statistical design of SRAM array for yield enhancement in nanoscaled CMOS," *IEEE Trans. on CAD of Integrated Circuits and Systems*, vol. 24, no. 12, pp.1859-1880, 2005.

[32] M. Yamaoka, N. Maeda, Y. Shimazaki, and K. Osada, "65 nm Low-Power High-Density SRAM Operable

at 1.0 V under 3σ Systematic Variation Using Separate Vth Monitoring and Body Bias for NMOS and PMOS," *in Digest of Technical Papers IEEE Solid States Circuits Conference 2008*, pp.384-385, Feb. 2008.

[33] T. Tsunomura, A. Nishida, F. Yano, A. T. Putra, K. Takeuchi, S. Inaba, S. Kamohara, K. Terada, T. Hiramoto, and T. Mogami, "Analyses of 5σ Vth Fluctuation in 65 nm MOSFETs using Takeuchi plot," *in Digest of Technical Papers, 2008 Symposium on VLSI Technology*, pp.156-157, June, 2008.

[34] G. Tsutui, K. Tsunoda, N. Kariya, Y. Akiyama, T. Abe, S. Maruyama, T. Fukase, M. Suzuki, Y. Yamagata, and K. Imai, "Reduction of Vth variation by work fuction optimization for 45-nm node SRAM cell," *in Digest of Technical Papers, 2008 Symposium on VLSI Technology*, pp.158-159, June 2008.

[35] S. Ekbote, K. Benaissa, B. Obradovic, S. Liu, H. Shichijo, F. Hou, T. Blythe, T. W. Houston, S. Martin, R. Taylor, A. Singh, H. Yang, and G. Baldwin, "45 nm low-power CMOS SoC technology with aggressive reduction of random variation for SRAM and analog transistor," *in Digest of Technical Papers, 2008 Symposium on VLSI Technology*, pp.158-159, June 2008.

[36] X. Chen, S. Samavedam, V. Narayanan, K. Stein, C. Hobbs, C. Baiocco, W. Li, D. Jaeger, M. Zaleski, H. S. Yang, N. Kim, Y. Lee, D. Zhang, L. Kang, J. Chen, H. Zhuang, A. Sheikh, J. Wallner, M. Aquilino, J. Han, Z. Jin, J. Li, G. Massey, S. Kalpat, R. Jha, N. Moumen, R. Mo, S. Kirshnan, X. Wang, M. Chudzik, M. Chowdhury, D. Nair, C. Reddy, Y. W. Teh, C. Kothandaraman, D. Coolbaugh, S. Pandey, D. Tekleab, A. Thean, M. Sherony, C. Lage, J. Sudijono, R. Lindsay, J. H. Ku, M. Khare, and A. Steegen, "A Cost Effective 32 nm High-K/ Metal Gate CMOS Technology for Low Power Applications with Single-Metal/Gate-First Process," *in Digest of Technical Papers, 2008 Symposium on VLSI Technology*, pp.88-89, June 2008.

[37] T. S. Doorn, E. J. W. Maten, J. A. Croon, A. D. Bucchianico, and O. Wittich, "Importance sampling Monte Carlo simulations for accurate estimation of SRAM yield," *in Digest of Technical Papers, IEEE 34th European Solid-State Circuits Conference ESSCIRC 2008*, pp. 230-233, Sep. 2008.

[38] R. Kanj, R. Joshi, and S. Nassif, "Mixture importance sampling and its application to the analysis of SRAM designs in the presence of rare failure events," *in Digest of Technical Papers, Design Automation Conference 2006*, pp. 69-72, July 2006.

Hiroyuki Yamauchi (M'95) received the Ph.D. degree in engineering from Kyushu University, Fukuoka, Japan, in 1997. His doctoral dissertation was on "Low Power Technologies for Battery Operated Semiconductor Random Access Memories." In 1985 he joined the Semiconductor Research Center, Panasonic, Osaka, Japan. From 1985 to 1987 he had worked on the research of the submicron MOS FET model-parameter extraction for the circuit simulation and the research of the sensitivity of the scaled sense amplifier for ultrahigh-density. DRAM's which was presented at the 1989 Symposium on VLSI Circuits. From 1988 to 1994, he was engaged in research and development of 16-Mb CMOS DRAM's including the battery-operated high-speed 16 Mbit CMOS DRAM and the ultra low-power, three times longer, self-refresh DRAM which was presented at the 1993 and 1995 ISSCC, respectively. He also presented the charge-recycling bus architecture and low-voltage operated high-speed VLSI's, including 0.5 V/100 MHz-operated SRAM and Gate-Over-Driving CMOS architecture, which was presented at the Symposium on VLSI Circuits in 1994 and 1996, respectively, as well as the 1997 ISSCC. After experienced general manager for development of various embedded memories, eSRAM, eDRAM, eFlash, and eFeRAM for system LSI in Panasonic, he has moved to Fukuoka Institute of Technology and become a professor since 2005. His current interests are focused on study for memory circuit designs for nano-meter era. He holds 87 U.S. Patents and has presented over 40 papers at proceedings and international conferences. He received the 1996 Remarkable Invention Award from Science and Technology Agency of Japanese government and the ISOCC2008 Best Paper Award. He served a program committee of IEEE Symposium on VLSI Circuits from 1998 through 2000 and has come back since 2008. Also he has been serving a program committee of ISSCC since 2002, and ASSCC since 2008.