サブ 1V の SRAM におけるリークを 2 桁以上削減する新手法 RRDSV

Fayez ROBERT SALIBA[†]　　Kyeong-Sik MIN[† ‡]

川口 博[†]　　神田 浩一[†※]　　桜井 貴康[†]

E-mail:　[†]{fayez, kawapy, tsakurai}@iis.u-tokyo.ac.jp,

[‡]mks@kookmin.ac.kr,　[※]kouichi.kanda@jp.fujitsu.com

Row-by-Row Dynamic Source-Line Voltage Control (RRDSV) Scheme for Two orders of Magnitude Leakage Current Reduction of Sub-1-V-VDD SRAM's

Fayez ROBERT SALIBA[†]　　Kyeong-Sik MIN[† ‡]　　Hiroshi KAWAGUCHI[†]

Kouichi KANDA[†※]　　and　Takayasu SAKURAI[†]

[†] Institute of Industrial Science, University of Tokyo　4-6-1 Komaba, Meguro-ku, Tokyo 153-8505 Japan

[‡] School of Electrical Engineering, Kookmin University　　Seoul, Korea

[※] System LSI Development Laboratories, Fujitsu Laboratories LTD.　　Kawasaki, Japan

E-mail:　[†]{fayez, kawapy, tsakurai}@iis.u-tokyo.ac.jp,

[‡]mks@kookmin.ac.kr,　[※]kouichi.kanda@jp.fujitsu.com

Abstract A new Row-by-Row Dynamic Source-line Voltage control (RRDSV) scheme is proposed to reduce the active leakage as well as the stand-by leakage in SRAM. By dynamically controlling the source-line voltage of cells row by row, the cell leakage through inactive cells can be reduced by two orders of magnitude. Moreover, the bit-line leakage through pass transistors can be completely cut off. This leakage reduction is caused from the cooperation of reverse body-to-source biasing and Drain Induced Barrier Lowering (DIBL) effects. A test chip has been fabricated using 0.18-μm triple-well CMOS technology to verify the data retention capability of this RRDSV scheme. The minimum retention voltage in the RRDSV is measured to be reduced by more than 60mV, when shielding metal is inserted to protect the memory cell nodes from bit-line coupling noise. It can reduce the leakage by another 50% in addition to the reduction by two orders of magnitude.

Keyword　Low-voltage SRAM, low-power SRAM, row-by-row, low-leakage, leakage reduction technique, subthreshold current

1. Introduction

As supply voltage (V_{DD}) decreases, threshold voltage (V_{TH}) should also be lowered not to degrade the operating speed. Decreasing V_{TH} by as small as 0.1V increases the subthreshold leakage by one order of magnitude. For example, in 70-nm technology, its subthreshold leakage is expected to reach 40 nA/μm at 25 °C [1]. Usually SRAM occupies a largest area among many internal blocks in a chip. Moreover, the area portion occupied by SRAM is more and more increasing as the device scaling goes further down [2]. Hence, if this large leakage is not suppressed, this should be a serious energy sink at strand-by mode. Moreover, this leakage can also be a serious problem even during the active time. At that time, while only one row operates dynamically row by row, the other many inactive rows consume the leakage power [3]. For a SRAM with one million cells, the leakage power consumed in inactive rows is expected to be about 10 times larger than the dynamic power consumed in one active row in future 70-nm technology [4].

Many techniques have been proposed to suppress the leakage in SRAM design [3-7]. Among them, the one technique is dynamically driving bulk voltage of inactive cell to control its V_{TH} [5]. When a row is turned on, its cells are activated by driving V_{DD} and V_{SS} to their n-well and p-well, respectively. When a row becomes inactive, its n-well and p-well voltages are driven to $2V_{DD}$ and $-V_{DD}$ to increase $|V_{TH}|$, respectively. Though this can reduce the leakage of inactive cells during the active time, it needs a large area penalty since wells having different potential voltages should

be separated by at least several microns. In addition, changing n-well and p-well voltages can cause another issue of speed degradation due to large well capacitance.

Another technique of Row-by-Row Dynamic V_{DD} (RRDV) scheme has been proposed in [3], where V_{DD} for cells are reduced to exploit DIBL effect to reduce the cell leakage. When this scheme is applied, however, the negative word line should be used to avoid a faulty read function due to leak pass transistors. If so, bit line should be precharged by V_{DD}-V_{TH} not to impose the high voltage stress on the inactive cell, which in turn reduces the stability of the cell and the write error may occur.

In this paper, a new Row-by-Row Dynamic Source-line Voltage control (RRDSV) technique is proposed, where the source-line of inactive cells is driven by a voltage higher than V_{SS} to reduce the leakage through inactive cells. This increased source-line voltage can enhance V_{TH} by cooperation of Drain-Induced Barrier Lowering (DIBL) and reverse body-t0-source biasing effects. It should be noted here that this RRDSV scheme is completely different with some other dynamic source-line control schemes. See, for example, ref. [6], where cells are made with high V_{TH} to reduce the leakage, even if V_{DD} is below 1V. To compensate the speed degradation in [6], the source line is driven negative when its cells become active, while imposing a higher voltage stress than V_{DD} on the gate oxide. Another technique called by gated-ground scheme has been proposed in [4], where NMOS is inserted between V_{SS} and source line. Here the NMOS cut-off switch in [4] is turned off by V_{SS}, when its cells become inactive, reducing the leakage by 50% roughly [4]. This amount of leakage reduction is not enough to be used in future CMOS technology with very large leakage power consumption. The leakage power of future 70-nm technology is expected to be ~10 times larger than its dynamic power [4]. To eliminate this large leakage power consumption as it is done now, the leakage should be reduced by more than two orders of magnitude.

Finally, it should be noted that a metal shield is inserted to protect memory cell nodes from bit-line coupling noise in this RRDSV scheme. A test chip has been fabricated using 0.18-μm triple-well CMOS technology to verify the leakage reduction by this RRDSV scheme. And, the effectiveness of this scheme is proved in this paper.

2. RRDSV Scheme for leakage reduction

Fig. 1 shows a schematic diagram of the proposed RRDSV scheme. Here, V_{SL} represents the source-line voltage. MN1 and MP1 represent the source-line driver for controlling V_{SL} according to input signal of 'CSL'. V_{SSH} is a voltage lower than V_{DD} and can be generated by a high-efficiency DC-DC converter. In a real SRAM design, only several cells can share one V_{SL} line and its source-line driver. Only 4 cells are connected to one V_{SL} line in this work. If many cells, for example, 8 or 16 cells, share one V_{SL} line and source-line driver, the resulting area penalty decreases. Simultaneously, however, the current through the shared source line will increase significantly. This can cause reliability problems such

as an electromigration and large I-R drop. And, the resulting read access delay may increase. Global V_{SS} lines should be perpendicular to word lines and horizontal to bit lines. The V_{SL} line is connected to the global V_{SS} line through its source-line driver, as shown in Fig. 1. Hence, each V_{SS} line carries the read current of only 4 cells. This current is small enough not to raise the issue of reliability and not to severely degrade the read access delay.

Assume that nodes 'Q' and 'QB' in Fig. 1 are high and low, respectively. If so, MN2, MP3, and MN5 are off in Fig. 1, where off MOSFET's are depicted with dotted line. When all rows are inactive, all source lines are connected to V_{SSH} via MP1. When a row becomes active, its source-line voltage is driven to V_{SS} via MN1, while the other source-lines remain connected to V_{SSH} via MP1, suppressing the leakage through them. Let us see why the leakage can be suppressed by connecting V_{SL} to V_{SSH}. As stated earlier, there are 3 off MOSFET's. For the MN2, V_{GS}=0V, V_{BS}=0V, and V_{DS}=V_{DD}, when V_{SL}=V_{SS}. If V_{SL} is set to V_{SSH}, V_{DS} and V_{BS} of MN2 change to V_{DD}-V_{SSH} and −V_{SSH}, respectively, enhancing V_{TH} by both DIBL and reverse-body-biasing effects. For MN5, V_{GS} changes to −V_{SSH} if V_{SL}=V_{SSH}. It causes the leakage through MN5 to be completely cut off. This bit-line leakage suppression through the MN5 is especially important in sub-1-V operation of SRAM. In this low V_{DD} regime, V_{TH} should also be low which causes the bit-line leakage through the pass transistors to increase significantly. And when the bit line leakage dominates over the cell current, a faulty read operation may occur [8]. As for MP3, its |V_{DS}| is decreased from V_{DD} to V_{DD}-V_{SSH}, and then due to DIBL effect, its V_{TH} increases.

One more issue to be considered here is that the source-line driver is controlled by a negative voltage of V_N when cells that belong to the source-line driver are inactive. V_N should be low enough to completely cut off the leakage through MN1. If not, the leakage can still flow through MN1. V_N can be easily generated by using the conventional charge pumps, since usually V_N is slightly below V_{SS}. One possible issue of high-voltage stress can occur when V_N is below V_{SS}. This problem, however, can easily be solved by setting |V_N| to a value lower than V_{DD}-V_{SSH}.

Next, the role of MP1 should be discussed. By turning on the MP1 by V_N, the source line is not floating or in high impedance state but is connected to V_{SSH} via MP1, when its cells are inactive. A high-impedance state is very susceptive to external noise such as bit-line coupling noise and soft-error related noise. Hence, driving V_{SL} to V_{SSH} is very helpful in improving the data retention capability when the data is retained in inactive cells. Fig. 2 shows how V_{SL} varies with V_N in both cases; when MP1 is used in the RRDSV scheme and when it is not. With the MP1, V_{SL} is connected to V_{SSH} via MP1. In this case, the retention voltage in Fig. 1 becomes V_{DD}-V_{SSH} in inactive cell. If this retention voltage is larger than V_{TH}, the data can be retained when its row is inactive. And, the data can be later restored to V_{DD} and V_{SS} when its row becomes active. If the MP1 is not used, V_{SL} is strongly dependent on V_N. With small variation of V_N by 0.1V, V_{SL} can change by as much as 0.4V. Increased V_{SL} reduces the retention voltage so much

that it may destroy the stored data at the end. To avoid this kind of data destruction, the source line should be connected to V_{SSH} via the MP1 when its cells are inactive.

Fig. 3 shows the simulated subthreshold characteristics of future 70-nm NMOSFET, which is from Berkeley Predictive Technology Model (BPTM) [9]. If V_{SL} is increased from $V_{SS}=0V$ to $V_{SSH}=0.7V$ when $V_{DD}=0.9V$, the leakage can be reduced by two orders of magnitude, as indicated in Fig. 3.

3. Design of SRAM with RRDSV scheme

Fig. 4 shows a level shifter whose output is connected to the source-line driver. This level shifter converts voltage swing between V_{SS} and V_{DD} to a swing between V_N and V_{DD} without causing a high voltage stress that exceeds V_{DD}. Here G1 represents the decoder that selects a word line among many word lines according to input address. An output of G1 is represented by 'A' in Fig. 4. When a word line is selected, 'A' changes from V_{SS} to V_{DD} and 'B' of G2 goes from V_{SSH} to V_{SS}. It should be noted here that the MN1 might be leaky since its V_{GS} is larger than V_{SS} when turned off. This leakage current, however, can be neglected because only one row is active while the other rows are inactive. Using MN1 with high V_{TH} in Fig. 4 can eliminate this leaky turn-off problem of MN1. In this case, the speed degradation due to high-V_{TH} MN1 will occur not at the inactive-to-active transition but at the active-to-inactive transition. The active-to-inactive transition does not have an effect on the data access time, even though it is longer.

Fig. 5 indicates the normalized read-out delay with varying MN1 size of source-line driver, which is shown in Fig. 1. Here, N presents the number of cells that share the same source-line driver. W_{MN1} and W_{MN2} is the width of MN1 of the source-line driver and the width of MN2 in Fig. 1, respectively. Compared with the read-out delay when the source line is directly connected to V_{SS} without the source-line driver, the delay with the source-line driver is about 6% longer. This delay penalty includes the time for driving the source-line driver and the degraded bit-line delay due to the additional MN1 of the source-line driver. The bit-line delay can be defined by the time in which the potential difference between two bit lines (ΔV_{BL}) reaches 200 mV. As you can see in Fig. 5, the delay penalty begins to increase sharply when the ratio of W_{MN1} to $N*W_{MN2}$ becomes smaller than 0.7.

Fig. 6 shows the array architecture, where neighboring 4 cells share one source-line driver. The area penalty due to this added source-line driver is estimated very roughly to be 25%, if the peripheral area outside the cell array is neglected. This penalty becomes smaller as the number of cells sharing one source line increases. Simultaneously, the current flowing through this shared source-line, however, increases with the number of cells increasing. This may introduce the reliability problem such as an electromigration and high I-R drop, and longer access delay during the read. To avoid these problems, the number of cells sharing one source-line driver should be limited. In this work, as shown in Fig. 6, only 4 cells are connected to one source-line driver. Each source-line driver is connected to global V_{SS} line, which is perpendicular to word-line. Since the word line is activated row by row, only one source-line driver draws the read current into the global V_{SS} line.

One more issue considered in the array design is noise-related problem. Since the retention voltage between Q and QB in Fig. 1 is reduced almost to a level of V_{TH}, Q and QB can be susceptive to external noise. To avoid an unexpected data flip in the retention time, the metal shield is introduced in Fig. 6. Here, the dotted region in Fig. 6 means that the region is shielded by metal. Hence, the cell under the dotted area of Fig. 6 can be electrically isolated by metal to protect cell nodes from the bit-line coupling noise. The area penalty of this metal shield is negligible but metal layers as many as 4 are required to route the local interconnection, bit-line, word-line, and the shield line, respectively.

4. Simulation and measurement

Fig. 7 shows how much the subthreshold leakage will be reduced by using the RRDSV scheme in future 70-nm and sub-1-V technology, which is from Berkeley Predictive MOSFET Technology Model (BPTM) [9]. From this figure, it is expected that the leakage with the RRDSV scheme becomes 1/100 compared with the conventional SRAM scheme when $V_{DD}=0.9V$ and $V_{SSH}=0.7V$. When $V_{DD}=0.6V$ and $V_{SSH}=0.4V$, the leakage with the RRDSV scheme becomes 1/20. This is because the reverse source-to-body voltage becomes smaller from 0.7V to 0.4V. A drastic leakage reduction can be observed at the bit-line leakage through pass transistor. Since the gate-to-source voltage of pass transistor becomes negative by increasing the source-line voltage, the bit-line leakage can be completely suppressed as can be seen in Fig. 7. When the temperature is 100°C, the leakage is reduced to 1/50 with $V_{DD}=0.9V$ and $V_{SSH}=0.7V$.

The SRAM with RRDSV scheme was fabricated in 0.18-μm CMOS technology. Fig. 8 shows the chip micrograph of the SRAM array with 128x16x8 bits. Fig. 8 shows the measured shmoo plots for the data retention capability. Here, the x-axis and y-axis are V_{DD} and V_{SSH}, respectively. Figs. 8 (a) and (b) are for the RRDSV scheme without the bit-line shield and with the bit-line shield, respectively. From the measurement, the minimum retention voltage is about 0.2V and this value is very comparable to V_{TH}. This minimum retention voltage can be reduced more when shielding metal is inserted between cell nodes and bit lines. Since the retention voltage in the RRDSV scheme is so small that the stored data may be susceptible to the bit-line coupling noise. The bit-line shield can protect the stored data from being flipped by the bit-line coupling noise. The minimum retention voltage is measured to be improved by 60mV by using this bit-line shield. It can cause the leakage to be more suppressed by 50% in addition to the reduction by two orders of magnitude.

5. Conclusion

A new SRAM scheme is proposed, where the active leakage can be suppressed as well as the stand-by leakage. In sub-1-V and 70-nm

era, V_{TH} is lowered to less than 0.2V so that leakage power of memory cells becomes more dominant than the dynamic power. By dynamically controlling the source-line voltage of cells row by row, the leakage through inactive cells can be reduced by two orders of magnitude. It is caused by the cooperation of reverse body, source biasing effect and Drain Induced Barrier Lowering (DIBL) effect. The data retention capability is verified by measurement when retention voltage is as small as V_{TH}. The measured minimum retention voltage is improved by 60mV when shielding metal is inserted to protect the memory cell nodes from bit-line coupling noise. It can cause the leakage to be more suppressed by 50% in addition to the reduction by two orders of magnitude.

6. Acknowledgement

7. References

[1] International Technology Roadmap for Semiconductors (ITRS99)

[2] T. Sakurai, "Perspectives on power-aware electronics," Digest of IEEE International Solid-State Circuits Conference, pp. 26-29, Feb. 2003.

[3] K. Kanda, T. Miyazaki, K. Min, H. Kawaguchi, and T. Sakurai, "Two Orders of Magnitude Reduction of Low Voltage SRAM's by Row-by-Row Dynamic VDD Control (RDDV) Scheme," Proceedings of IEEE International ASIC/SOC Conference, pp. 381-385, Rochester in USA, Sep. 2002.

[4] A. Agarwal, H. Li, and K. Roy, "A single Vt low-leakage gated-ground cache for deep submicron," IEEE Journal of Solid-State Circuits, vol. 38, no. 2, pp. 319-328, 2003.

[5] H. Kawaguchi, Y. Iataka, and T. Sakurai, "Dynamic leakage cut-off scheme for low-voltage SRAM's," Technical digest of Symposium on VLSI circuits, pp. 140-141, June, 1998.

[6] H. Mizuno and T. Nagano, "Driving source-line cell architecture for sub-1-V high-speed low-power applications," IEEE Journal of Solid-State Circuits, vol. 31, no. 4, pp. 552-557, April 1996.

[7] K. Osada, Y. Saitoh, E. Ibe, and K. Ishibashi, "16.7fA/cell tunnel-leakage-suppressed 16Mb SRAM for handling cosmic-ray-induced multi-errors," IEEE International Solid-State Circuits Conference, pp. 302-303, 2003.

[8] K. Agawa, H. Hara, T. Takayanagi, and T. Kuroda, "A bit line leakage compensation scheme for low-voltage SRAM's," IEEE Journal of Solid-State Circuits, vol. 36, no. 5, pp. 726-734, 2001.

[9] Berkeley predictive technology model web site: http://www-device.eecs.berkeley.edu/~ptm
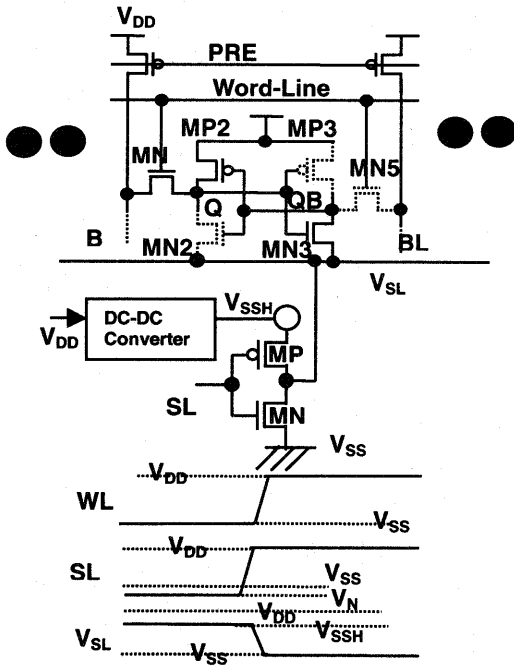
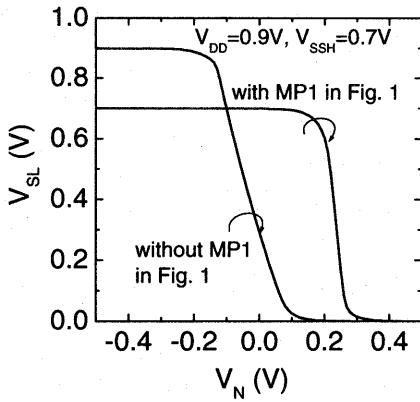Fig. 1 Schematic Diagram of RRDSV scheme with shared source line

Fig. 2 Plot of V_{SL} with varying V_N when MP1 in Fig. 1 is used and when MP1 in Fig. 1 is not used
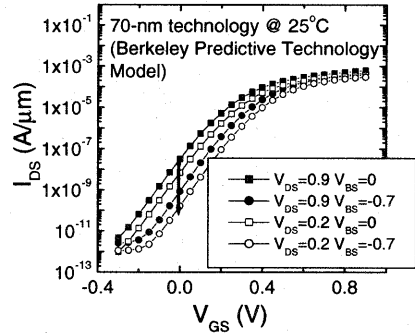

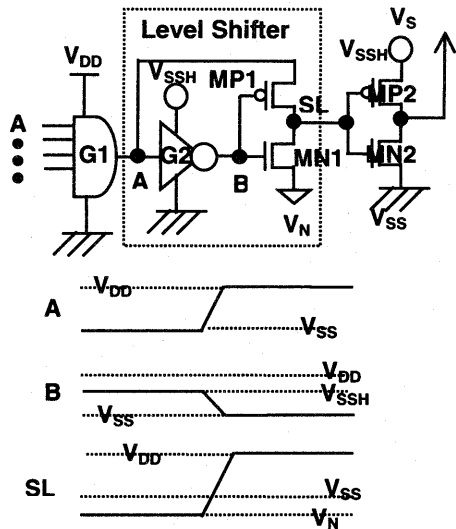Fig. 3 Subthreshold characteristics of future 70-nm technology
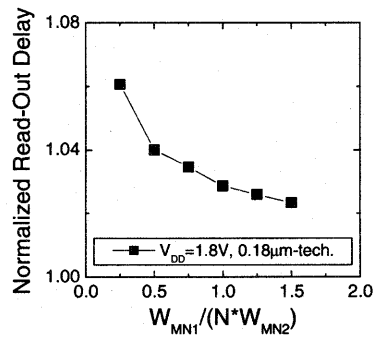

Fig. 4 Level Shifter without high voltage stress

Fig. 5 Normalized read-out delay with varying NMOS size of source-line driver shown in Fig. 1
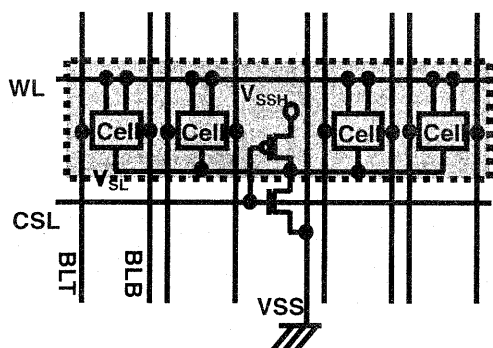
Fig. 6 Array configuration of RRDSV scheme, where the dotted area represents the cells shielded by metal to protect cell nodes with small retention voltage from bit-line coupling noise
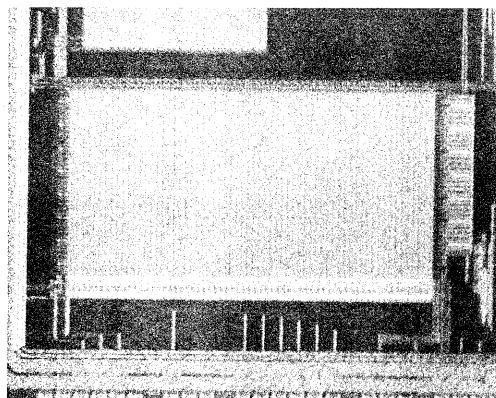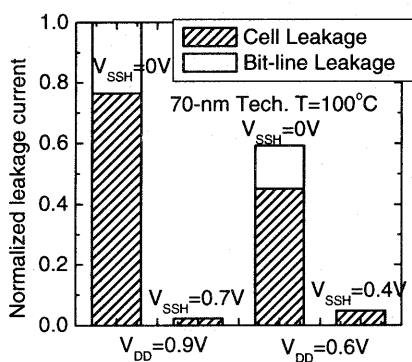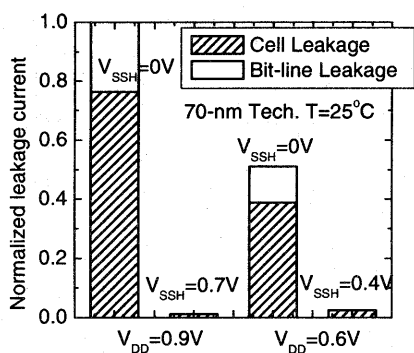

Fig. 8 Chip micrograph of 16-K SRAM

Fig. 7 (a) Leakage comparison between conventional SRAM and RRDSV-SRAM at T=25°C (b) Leakage comparison between conventional SRAM and RRDSV-SRAM at T=100°C
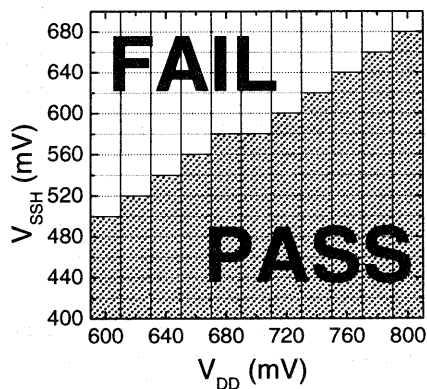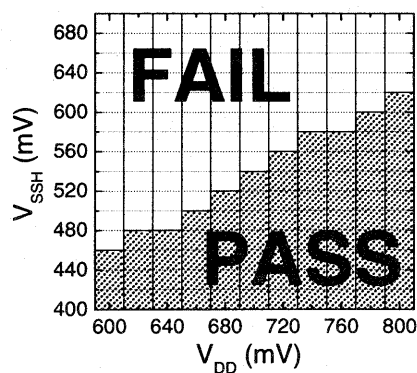

Fig. 9 (a) Measured shmoo plot of SRAM array without bit-line shield (b) Measured shmoo plot of SRAM array with bit-line shield