

# On the Efficacy of Write-Assist Techniques in Low Voltage Nanoscale SRAMs

Vikas Chandra  
ARM R&D  
San Jose, CA  
vikas.chandra@arm.com

Cezary Pietrzyk  
ARM R&D  
San Jose, CA  
cezary.pietrzyk@arm.com

Robert Aitken  
ARM R&D  
San Jose, CA  
rob.aitken@arm.com

**Abstract**—Read and write assist techniques are now commonly used to lower the minimum operating voltage ( $V_{min}$ ) of an SRAM. In this paper, we review the efficacy of four leading write-assist (WA) techniques and their behavior at lower supply voltages in commercial SRAMs from 65nm, 45nm and 32nm low power technology nodes. In particular, the word-line boosting and negative bit-line WA techniques seem most promising at lower voltages. These two techniques help reduce the value of  $WL_{crit}$  by a factor of  $\sim 2.5X$  at 0.7V and also decrease the  $3\sigma$  spread by  $\sim 3.3X$ , thus significantly reducing the impact of process variations. These write-assist techniques also impact the dynamic read noise margin (DRNM) of half-selected cells during the write operation. The negative bit-line WA technique has virtually no impact on the DRNM but all other WA techniques degrade the DRNM by 10-15%. In conjunction with the benefit (decrease in  $WL_{crit}$ ) and the negative impact (decrease in DRNM), overhead of implementation in terms of area and performance must be analyzed to choose the best write-assist technique for lowering the SRAM  $V_{min}$ .

## I. INTRODUCTION

Static random access memory (SRAM) is a critical part of most VLSI system-on-chip (SoC) applications. The SRAM bit cell design has to cope with stringent requirement on the cell area leading to minimum (or close to minimum) sized transistors. Due to this scaling trend, device variations and leakage are increasing sharply with each shrinking technology node [1], [14]. Further, the supply voltage is scaled down to reduce dynamic and leakage power consumption. The operation of the SRAM at lower supply voltage becomes even more challenging. The predominant yield loss from increased device variability occurs at minimum operating voltage, a term defined as  $V_{min}$ . The failures at  $V_{min}$  can be due to write failure, read disturb failure, access failure or retention failure [6].

Usually the SRAM  $V_{min}$  is limited by write failure or read disturb failure. It is, however, difficult to predict *a priori* as to which of the two failure mode dominates because it is dependent upon many factors including the bit cell architecture, technology node etc. The read disturb problem can be mitigated by adding a dedicated read port where the bit cell nodes are isolated from the bit-lines [2]. These 8T bit cells are larger in area but they eliminate the read disturb issue. Solving the issue of write failure, however, is more challenging. Write failure is defined as the failure to intentionally flip the value of the bit cell during the write operation. Various write-assist (WA) schemes have been proposed in literature to help the bit cell to flip during the write cycle [9], [10], [13], [15]. For

the SRAMs in which the  $V_{min}$  is dictated by write failures, the WA techniques push the  $V_{min}$  lower and make it limited by other failure modes, say read disturb failures. In this work we analyze the efficacy of the leading WA techniques in low power SRAMs in aggressive nanometer technology nodes. Further, we analyze the impact of supply voltage scaling on the efficacy of the WA techniques. We also investigate the impact of these WA techniques on the dynamic read noise margins of the half-selected cells (half-selected cells exist due to column multiplexing and for these cells the word-line is asserted but the bit-lines are not pulled low). Understanding the impact of technology and voltage scaling on the efficacy of write-assist schemes is crucial in designing low power SRAMs.

The rest of the paper is organized as follows. Section II explains the SRAM write-ability metric and the concept of dynamic write margin. Section III discusses the various write-assist schemes and analyzes their impact on the dynamic write margin. Section IV discusses the impact of technology scaling on the efficacy of the write-assist techniques. Section V analyzes the impact of process variability on the statistical spread of the dynamic write margin. Section VI describes the impact of these write-assist techniques on the dynamic read noise margins of half-selected cells. Section VII summarizes the work and concludes.

## II. SRAM WRITE-ABILITY METRIC

The effectiveness of an SRAM write operation is typically quantified by the minimum (or critical) width of word-line pulse (defined as  $WL_{crit}$ ) during which the bit cell changes state [11]. The definition is important because it captures the dynamic write margin which is more accurate. The static approaches to measure the write margin assume that the word-line pulse width is infinite which could lead to erroneous conclusions. Figure 1 shows how the internal nodes ( $n1$  and  $n2$ ) of a bit cell change when word-line of different pulse widths are applied during the write operation. In Figure 1(a), the pulse width of  $WL$  is smaller than the  $WL_{crit}$  whereas in Figure 1(b), the pulse width of  $WL$  is equal to or greater than  $WL_{crit}$ . In the case of Figure 1(a), the nodes  $n1$  and  $n2$  tend to move towards the other state but they return back to the original state. In Figure 1(b), however, the nodes  $n1$  and  $n2$  are able to move to other stable state. In other words, to have a successful write,  $WL \geq WL_{crit}$ .

Write operation in a bi-stable cell (like a bit cell) can be defined as a transition from one equilibrium state to the other.

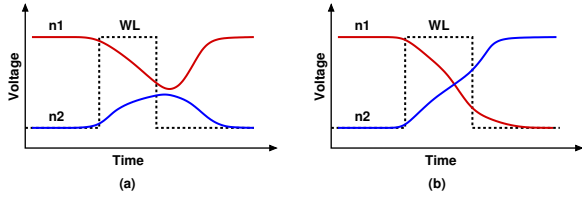


Fig. 1. Internal nodes of a bit cell during write operation (a)  $WL < WL_{crit}$  (b)  $WL \geq WL_{crit}$

The separatrix is defined as the boundary which separates the two stability regions [4]. The importance of the separatrix lies in that a state flip will be generated if and only if the write operation causes the state to be pushed away from the initial stable equilibrium to cross the separatrix. Figure 2 shows the state space trajectory during the write operation with two different word-line pulse widths. The separatrix in this case is a straight line since the cell is symmetrical. The curve in solid is for the case when the  $WL$  pulse width equals  $(WL_{crit} - 1ps)$ . In this case the trajectory starts at the stable point of  $((n1, n2) = (1, 0))$  and ends back at the same stable point. The dotted line represents the case where the  $WL$  pulse width is equal to  $WL_{crit}$ . In this case the trajectory starts at  $((n1, n2) = (1, 0))$ , crosses the separatrix and then ends up at  $((n1, n2) = (0, 1))$ , thus successfully completing the write operation.

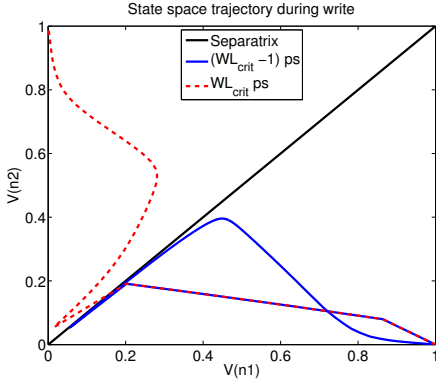


Fig. 2. Write state space trajectory:  $WL < WL_{crit}$  and  $WL = WL_{crit}$

With technology scaling, it is becoming difficult to write to SRAMs even at nominal supply voltage and the challenges become more apparent at lower supply voltages. Figure 3 shows the trend of  $WL_{crit}$  with respect to supply voltage for a 32nm SRAM (the analyses in this paper are based on commercial SRAMs from 65nm, 45nm and 32nm low power technology nodes). The increase in  $WL_{crit}$  is  $\sim 10X$  as the voltage scales down to 0.7V from 1V. This trend is very alarming since supply voltage is frequently dynamically scaled in System-on-Chip designs to reduce power consumption. The substantial increase in  $WL_{crit}$  will increase the  $V_{min}$  of the SRAM and limit its applicability in low power designs. Hence, there is a need to improve the write performance at low supply voltages. The techniques which aid the bit cell in changing the state during write operation are called *write-assist* techniques and now they are widely used in most low power SRAMs. Section III discusses various write-assist schemes in detail and their impact on  $WL_{crit}$ .

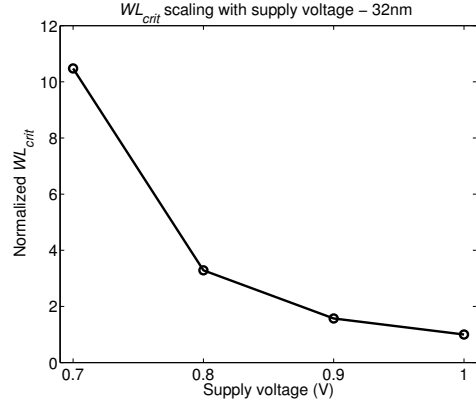


Fig. 3. Change in  $WL_{crit}$  with voltage scaling at 32nm

### III. WRITE-ASSIST SCHEMES

As mentioned in Section II, write-assist (WA) techniques are crucial in reducing the  $V_{min}$  of SRAMs. In essence, WA techniques aid the bit cell to flip state during the write operation. As can be noted from Figure 3, the increase in  $WL_{crit}$  is exponential as the supply voltage scales down. To improve the scaling of  $WL_{crit}$ , several WA techniques ([9], [10], [13], [15]) have been proposed in literature:

- Vdd lowering WA
- Vss raising WA
- Word-line boosting WA
- Negative bit-line WA

To make sure that the comparison is fair, we impose the same percentage of increase or decrease in voltage levels for all the WA techniques. For the Vdd lowering WA scheme, the core voltage is lowered by 10%. For the Vss raising WA scheme, the core ground is raised by 10%. For the word-line boosting WA scheme, the word-line is boosted by 10% and for the negative bit-line scheme, the bit-line is lowered by 10% below the core ground. The following sections describe each of the WA technique and its impact on  $WL_{crit}$ .

#### A. Vdd lowering

$WL_{crit}$  can be reduced by weakening the pull-up device with respect to the pass-gate device. Once the pull-up device is weakened, it is easier to write a new data to the bit cell. Figure 4 shows the timing relationships using the Vdd lowering WA scheme. This WA scheme is implemented using a second

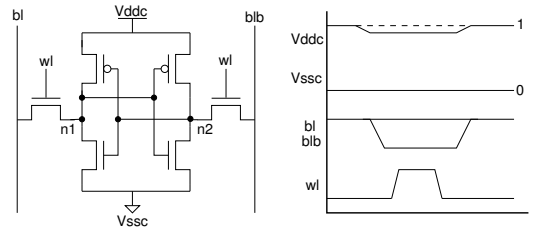


Fig. 4. Write-assist based on lowering of core Vdd

external lower supply which is connected via a multiplexer to the write-selected columns [15]. On-chip regulators can also be used to generate the lower supply voltage during write [8].

Some other techniques to lower the core Vdd voltage include floating the selected write columns [12]; charge sharing between the selected column and an appropriately sized pre-discharged dummy capacitance to create the lower voltage level [7] etc. The main challenge with this technique is to make sure that the lowered column voltage is still higher than the retention voltage of the unselected bit cells in the same column. To simplify the implementation, sometime the voltage supply of the whole array is lowered during the write operation. However, this decreases the dynamic read noise margin (DRNM) of the half-selected bit cells.

Figure 5 shows the impact of the Vdd lowering based WA on the  $WL_{crit}$  for a 32nm bit cell. Although the WA based

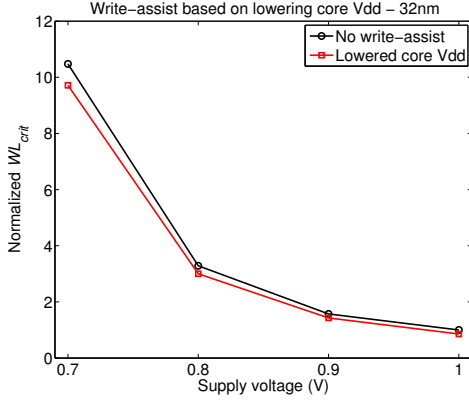


Fig. 5. Normalized  $WL_{crit}$  for write-assist based on lowering of core Vdd

scheme consistently performs better than the nominal case, the gain is minuscule. This is due to the fact that the pull-up PMOS is already very weak in current generation bit cells and hence making it further weak does not help much.

### B. Vss raising

A raised ground scheme is another way to aid the write operation thus decreasing the value of  $WL_{crit}$ . This WA technique reduces the risk of data retention failure [13]. The idea is still to weaken the pull-up PMOS but in this scheme it is done by weakening the PMOS gate voltage instead of the source voltage. The core ground (shown as Vssc in Figure 6) is raised during the write operation. This extra ground voltage can be routed as a separate ground or can be generated internally using a regulator. This WA technique also impacts the DRNM of half-selected bit cells, if implemented globally for the whole array.

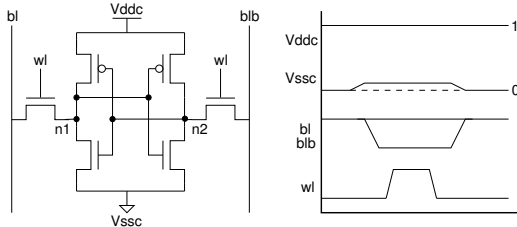


Fig. 6. Write-assist based on raising of core Vss

Figure 7 shows the impact of the Vss raising based WA on the  $WL_{crit}$ . The  $WL_{crit}$  for the WA case is better than

the non WA case but the gain is very small. The reason for this marginal increase is similar to that of the Vdd lowering WA case which is that the PMOS is already weak and hence weakening the gate drive does not further weaken it.

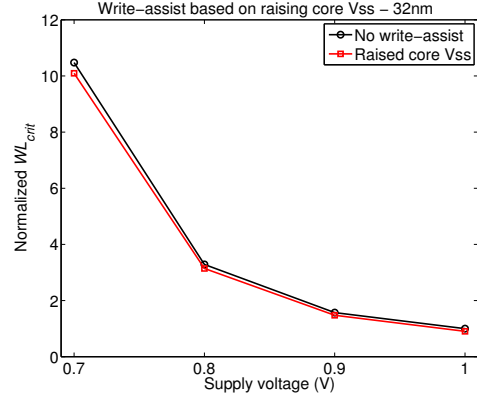


Fig. 7. Normalized  $WL_{crit}$  for write-assist based on raising of core Vss

### C. Word-line boosting

Another technique which assists the bit cell to flip during a write event is boosting the word-line higher than the supply voltage (Figure 8). The boosting increases the  $V_{gs}$  of the access transistor and hence increases its drive strength. The increased drive strength of the access transistor aids significantly in flipping the bit cell. The boost voltage can be routed as a

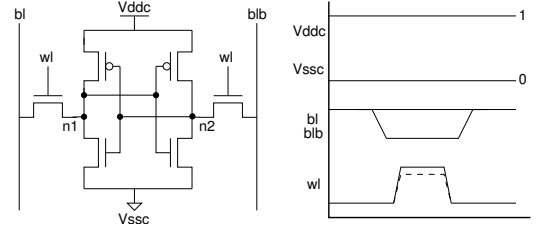


Fig. 8. Write-assist based on word-line boosting

separate power supply or it can be generated internally by a charge pump [10] or by capacitive coupling [5]. Unlike the techniques mentioned in Sections III-A and III-B which work on a column, the word-line boosting technique works on a row. Hence all the half-selected cells in a row are more prone to an upset due to reduction in their dynamic read noise margins.

Figure 9 shows the impact of word-line boosting based WA on the  $WL_{crit}$ . The  $WL_{crit}$  in this case is substantially better than the nominal case with no WA. As shown in Figure 9, the benefits of this WA scheme increases significantly as the supply voltage is scaled down.

### D. Negative bit-line

To create a larger  $V_{gs}$  for the NMOS pass transistor, either the gate voltage needs to be increased (Section III-C) or the source voltage needs to be decreased. The approach of negative bit-line based WA swings the bit-line voltage below "0" during the write operation (Figure 10). The increase in  $V_{gs}$  causes the access transistor to become stronger and hence can flip the bit

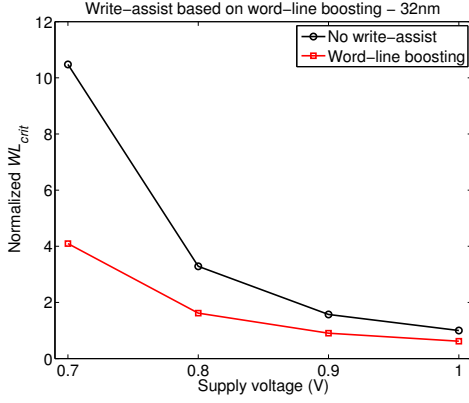


Fig. 9. Normalized  $WL_{crit}$  for write-assist based on word-line boosting

cell easily [9]. Similar to the word-line boosting, the negative bit-line voltage can be generated internally using a charge pump technique or using a capacitive coupling technique. This

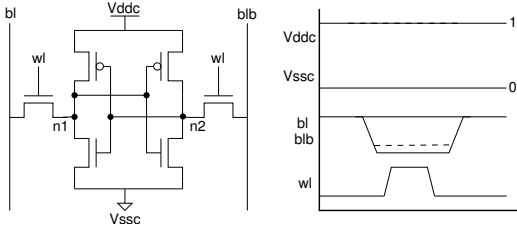


Fig. 10. Write-assist based on negative bit-line

WA technique works on a column, hence all the unaccessed bit cells in the column see an increase in  $V_{gs}$  on the pass transistor. Since the word-lines are not asserted for those cells, the increase in the pass gate  $V_{gs}$  is well lower than the  $V_t$  of the access transistor, and hence their DRNMs are not affected.

Figure 11 shows the impact of negative bit-line based WA on the  $WL_{crit}$ . The benefits due to the negative bit-line WA are very similar to that of the word-line boosting WA scheme. The reason being that both techniques ultimately increase the  $V_{gs}$  of the access transistor. The benefits of this WA scheme increases as well when the supply voltage is scaled down.

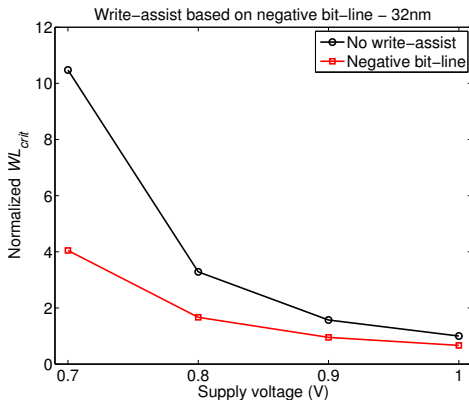


Fig. 11. Normalized  $WL_{crit}$  for write-assist based on negative bit-line

#### IV. TECHNOLOGY SCALING TRENDS

Figure 12 shows the impact of technology scaling on  $WL_{crit}$  for SRAMs with no write-assist. Two observations can be made from Figure 12. First, for all the three technology nodes, the  $WL_{crit}$  goes up sharply with reduction in the supply voltage. Second, at each supply voltage shown in Figure 12, the  $WL_{crit}$  scales nicely with technology scaling. Even though the absolute value of  $WL_{crit}$  decreases with technology scaling, writing to the SRAMs still gets challenging due to significant reduction in the clock period.

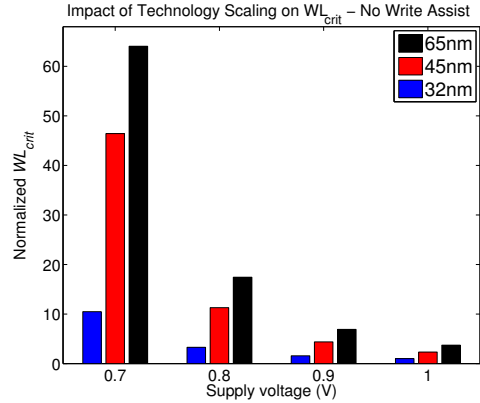


Fig. 12. Impact of technology scaling on  $WL_{crit}$  with no write-assist

Figure 13 shows the impact of technology scaling on the efficacy of various WA techniques described in the sections above. Figures 13(a), 13(b), 13(c) and 13(d) show the technology trends for the Vdd lowering, Vss raising, word-line boosting and negative bit-line WA schemes respectively. The y-axis in each case represents  $WL_{crit}$  (normalized against the case with no write-assist). As can be noted from Figure 13, all the WA techniques perform better than the nominal case. However, the gains due to the Vdd lowering and Vss raising WA techniques are marginal and the benefits decrease as the voltage is scaled lower. On the other hand, the word-line boosting and negative bit-line WA techniques perform very well as compared to the nominal case. Further, the benefits increase as the supply voltage is scaled down.

With just the bit cell write-ability as a metric, word-line boosting and negative bit-line WA techniques seem potential solutions to lower the  $V_{min}$  of the SRAM. However, until now we have evaluated the efficacy of the WA techniques at nominal process corner only. In Section V we explore how the benefits of these WA techniques scale in the presence of process variability.

#### V. IMPACT OF PROCESS VARIABILITY

The significance and complexity of process variations is increasing with technology scaling. In sub-45nm technology, these variations can be classified into two groups, based on the mechanism of the variation - systematic and random. These sources of variations severely limit the ability to push the performance of a design. We used Monte-Carlo simulation framework to understand the impact of process variations on the  $WL_{crit}$  for various write-assist methods. The Monte-Carlo

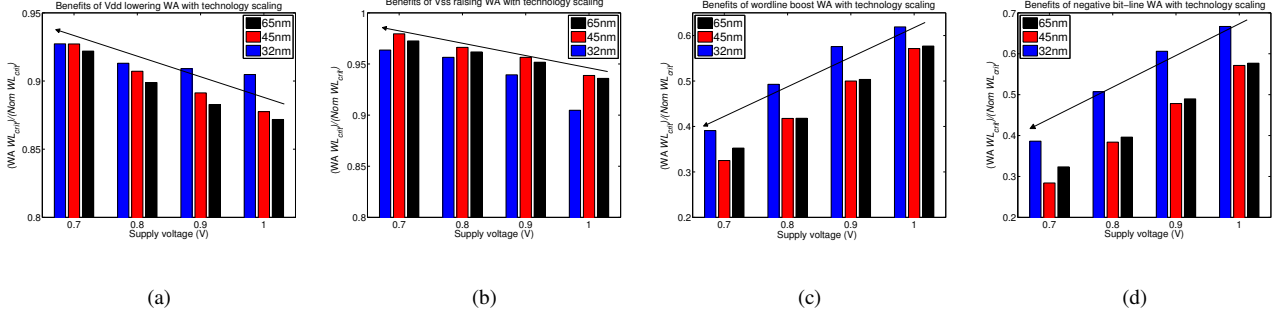


Fig. 13. Efficacy of write-assist techniques with technology scaling (a) Vdd lowering (b) Vss raising (c) Word-line boosting (d) Negative bit-line

simulations were done at a supply voltage of 0.7V, since the impact of write assist on the  $WL_{crit}$  is much higher at lower supply voltages. Figure 14 shows the statistical spread in  $WL_{crit}$  for the non WA and the other four WA cases.

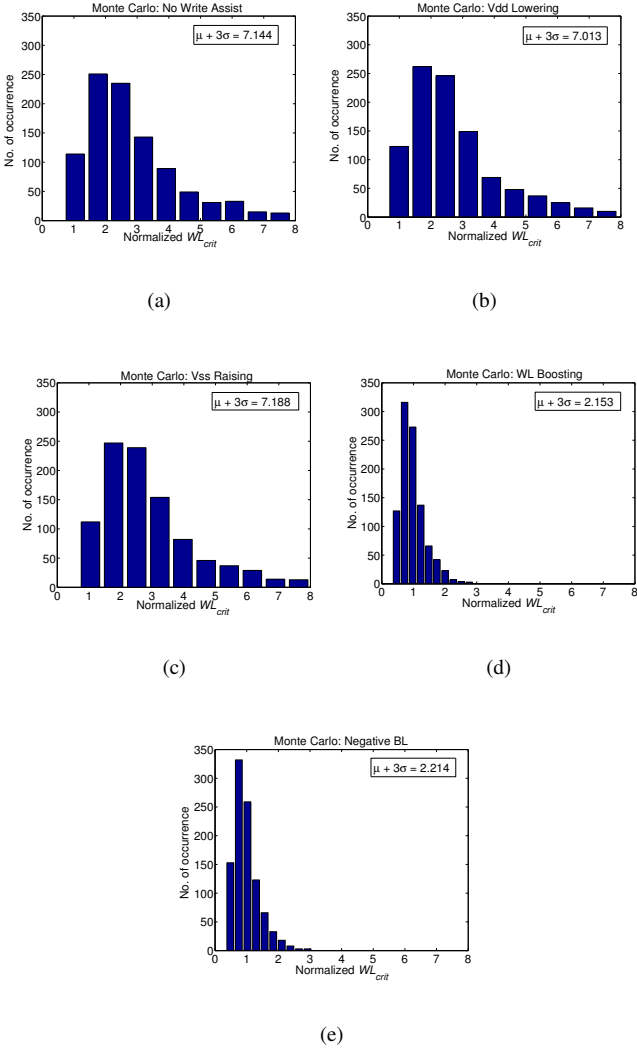


Fig. 14. Statistical spread of  $WL_{crit}$  at 0.7V (a) No write-assist (b) Vdd lowering (c) Vss raising (d) Word-line boosting (e) Negative bit-line

The variability in device parameters makes the  $WL_{crit}$  follow extreme order statistics, specifically, the Gumbel distribution characterized by long tails. It can be observed from Figure 14 that the statistical spread of  $WL_{crit}$  is much higher for the non WA, Vdd lowering and Vss raising WA cases

as compared to the word-line boosting and negative bit-line WA cases. It can be noted that the values of  $\mu$  and  $\sigma$  do not change much for the Vdd lowering and Vss raising cases as compared to the case with no write-assist. However, both  $\mu$  and  $\sigma$  are reduced substantially for the word-line boosting and negative bit-line WA cases. One pertinent metric to consider is the value of  $WL_{crit}$  which is  $3\sigma$  away from the mean. Figure 15 shows the values of  $\mu + 3\sigma$  for various WA schemes. As observed earlier, the spread of  $WL_{crit}$  is much tighter for word-line boosting and negative bit-line WA schemes as compared to the Vdd lowering and negative bit-line WA schemes. This clearly shows that the worst case choice for  $WL_{crit}$  will be much smaller for the word-line boosting and negative bit-line WA schemes.

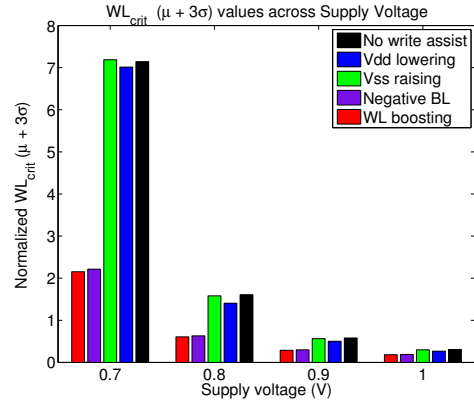


Fig. 15. Statistical deviation ( $\mu + 3\sigma$ ) for  $WL_{crit}$  across supply voltage

While WA techniques reduce the  $WL_{crit}$  for bit cells which are intended to be written, they also impact the stability of half-selected cells. Specifically, the dynamic read noise margins of half-selected cells get affected when the WA techniques are applied. Section VI examines the impact of all the four WA techniques on the stability of half-selected bit cells.

## VI. IMPACT ON HALF-SELECTED BIT CELLS

As described earlier, some cells are half-selected during the write operation due to column multiplexing. For these cells the word-line is asserted but the bit-lines are not pulled low. In a way, these half-selected cells undergo a read operation when the selected columns are written to. To ensure that the half-selected cells do not have read disturb issues, the dynamic read



noise margins (DRNM) need to be evaluated especially when WA techniques are used. DRNM is defined as the minimum voltage difference, over time, between the internal nodes in the bit cell during the read operation [3].

Since the half-selected cells have the word-line enabled, the WA schemes which impact the DRNM need to be row based. Word-line boosting based WA definitely impacts the half-selected bit cells since it raises the word-line voltage. Vdd lowering and Vss raising WA techniques could also impact the half-selected cells if the respective supply voltages are scaled for the whole core array. However, if the supply voltages are scaled only for the selected columns, it would not impact the half-selected cells. Figure 16 shows normalized DRNM values for half-selected cells during the write operation when various

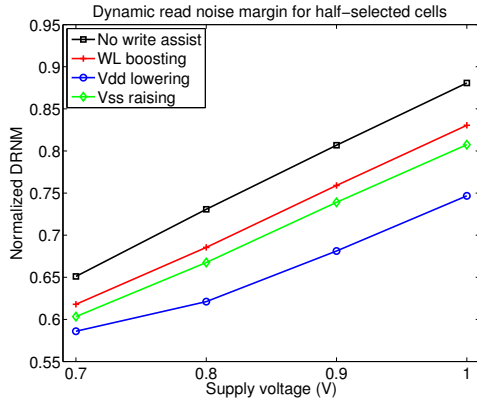


Fig. 16. Normalized DRNM for half-selected cells

WA techniques are used. Negative bit-line WA technique has no impact on the DRNM of half-selected cells since it impacts only the column in which the cells are written, hence it has been omitted from the figure. From Figure 16, it can be observed that the scheme with no WA has the best DRNM across all supply voltages. All the WA techniques decrease the DRNM of the half-selected cells and the degradation is highest for the Vdd lowering WA ( $\sim 15\%$  at 1V supply voltage).

Figure 17 shows the statistical spread of DRNM in the presence of process variability. Similar to  $WL_{crit}$ , the process

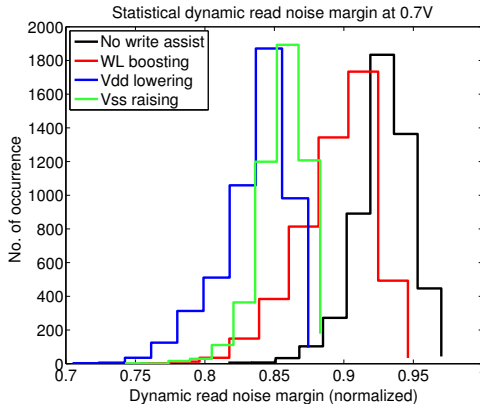


Fig. 17. Statistical spread of DRNM at 0.7V

variability makes the DRNM also follow the extreme order statistics, specifically, the Gumbel distribution. The case with no WA scheme has the best mean DRNM ( $\mu$ ) as well as the

best worst-case DRNM ( $\mu+3\sigma$ ). Similar to the observations from Figure 16, the  $\mu$  and the  $\mu+3\sigma$  DRNM values are worst for the Vdd lowering WA scheme.

## VII. CONCLUSIONS

As the technology scales in deep nanometer era, the challenges in designing an SRAM increase substantially. The challenges arise due to increase in write, read, access and retention failures. It is now becoming common for SRAMs to have read and write assist techniques to enable robust operation at lower supply voltages. In this work, we reviewed the efficacy of four leading write-assist (WA) techniques across the operating range of supply voltages. Our analyses suggest that the word-line boosting and negative bit-line WA techniques seem most promising at lower supply voltages. These two techniques help reduce the value of  $WL_{crit}$  by a factor of  $\sim 2.5X$  at 0.7V and also significantly reduce the impact of process variations. Also, most of the WA techniques degrade the dynamic read noise margin (DRNM) of the half-selected cells by 10-15%. Understanding these trade-offs and the overhead in terms of area and performance are crucial to choose the best write-assist technique in nanoscale SRAMs to enable robust low voltage operation.

## REFERENCES

- [1] A. J. Bhavnagarwala *et al*, "The impact of intrinsic device fluctuations on CMOS SRAM cell stability," *IEEE Journal of Solid-State Circuits*, Vol. 36, pp. 658-665, Apr. 2001.
- [2] L. Chang *et al*, "An 8T-SRAM for Variability Tolerance and Low-Voltage Operation in High-Performance Caches," *IEEE Journal of Solid-State Circuits*, Vol. 43, pp. 956-963, Apr. 2008.
- [3] W. Dehaene *et al*, "Embedded SRAM design in deep deep submicron technologies," *European Solid-State Circuits Conference (ESSCIRC)*, pp. 384-391, 2007.
- [4] W. Dong *et al*, "SRAM Dynamic Stability: Theory, Variability and Analysis," *Intl. Conf. on Computer Aided Design (ICCAD)*, pp. 378-385, 2008.
- [5] M. Iijima *et al*, "Low Power SRAM with Boost Driver Generating Pulsed Word Line Voltage for Sub-1V Operation," *Journals of Computers*, Vol. 3, No. 5, May 2008.
- [6] M. Khellah *et al*, "Read and Write Circuit Assist Techniques for Improving Vccmin of Dense 6T SRAM Cell," *Intl. Conf. on IC Design and Technology*, June 2008.
- [7] K. Nii *et al*, "A 45-nm Bulk CMOS Embedded SRAM with Improved Immunity Against Process and Temperature Variations," *IEEE Journal of Solid-State Circuits*, Vol. 43, Jan 2008.
- [8] H. Pilo *et al*, "An SRAM design in 65nm technology node featuring read and write-assist circuits to expand operating voltage," *IEEE Journal of Solid-State Circuits*, Vol. 42, Apr 2007.
- [9] N. Shibata *et al*, "A 0.5V 25 MHz 1mW 256kb MTCMOS/SOI SRAM for Solar-Power-Operated Portable Personal Digital Equipment - Sure Write Operation by Using Step-Down Negatively Overdriven Bitline Scheme," *IEEE Journal of Solid-State Circuits*, Vol. 41, Mar 2006.
- [10] C. C. Wang *et al*, "A Boosted Wordline Voltage Generator for Low Voltage Memories," *ICECS*, 2003.
- [11] J. Wang *et al*, "Analyzing Static and Dynamic Write Margin for Nanometer SRAMs," *ISLPED*, 2008.
- [12] M. Yamaoka *et al*, "90-nm Process-Variation Adaptive Embedded SRAM Modules with Power-Line-Floating Write Technique," *IEEE Journal of Solid-State Circuits*, Vol. 41, Apr 2006.
- [13] H. S. Yang *et al*, "Scaling of 32nm Low Power SRAM with High-K Metal Gate," *IEEE Intl. Electron Devices Meeting*, 2008.
- [14] K. Zhang *et al*, "Low Power SRAMs in nanoscale CMOS technologies," *IEEE Trans. Electron Devices*, Vol. 55, No. 1, pp. 145-151, Jan. 2008.
- [15] K. Zhang *et al*, "A 3 GHz 70 Mb SRAM in 65nm CMOS technology with integrated column-based dynamic power supply," *IEEE Journal of Solid-State Circuits*, Vol. 41, No. 1, Jan 2006.