

Variation Aware Performance Analysis of Gain Cell Embedded DRAMs

Wei Zhang
Department of ECE
University of Minnesota
Minneapolis, MN
zhang758@umn.edu

Ki Chul Chun
Department of ECE
University of Minnesota
Minneapolis, MN
chunx041@umn.edu

Chris H. Kim
Department of ECE
University of Minnesota
Minneapolis, MN
chriskim@umn.edu

ABSTRACT

Gain cell embedded DRAMs are twice as dense as 6T SRAMs, are logic compatible, have decoupled read and write paths providing good low voltage margin, and can drive long bitlines with gain. In this work, we present a variation study of gain cell eDRAM performance using an industrial 1.2V, 65nm low power CMOS process. Two methods are proposed to analyze eDRAM performance which can be used for designing variation tolerant eDRAM circuits, developing redundancy techniques, and guiding the device optimization procedure.

Categories and Subject Descriptors

B.7.2 [Integrated Circuits]: Design Aids - Simulation

General Terms

Design, Performance

Keywords

Embedded DRAM, Gain cell, Process variation, Bitline delay, Monte Carlo simulation

1. INTRODUCTION

Embedded DRAMs (eDRAMs) have been drawing interest in recent years as a potential alternative for mainstream 6T SRAMs. 1T1C type eDRAMs have been adopted for lower level caches in several high performance server chips delivering 5X+ higher bit-cell densities and a random cycle time of 2.0nsec [1-3]. Process and circuit issues such as the additional process steps for the cell capacitor and the destructive read problem make 1T1C cells less attractive with device scaling. Gain cells which are typically 2X denser than 6T SRAM cells are logic compatible and have good low voltage margin thanks to the decoupled read and write paths. They also have a read port capable of driving long bitlines, making them competitive at low voltages. Attaining practical retention times of 100's of μ secs and improving the random access speed remain as formidable challenges for future gain cell eDRAM designs. Fig. 1 compares the circuit parameters of interest for the three types of embedded memory.

A number of gain cell eDRAM designs have been successfully demonstrated in recent literatures [4-6]. However, there has not

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

ISLPED'10, August 18–20, 2010, Austin, Texas, USA.

Copyright 2010 ACM 978-1-4503-0146-6/10/08...\$10.00.

been any in-depth analysis on gain cell eDRAM performance in the presence of variation. The objective of this paper is to investigate this issue in detail and provide insight to gain cell operation especially for those who are not familiar with this type of circuit. The remainder of this paper is organized as follows. Section 2 gives a brief introduction on the circuit operation of a conventional 3T gain cell. In section 3, we examine the impact of TOX and VTH variation on gain cell performance. Section 4 describes practical issues related to running Monte Carlo simulations using standard circuit simulators. Section 5 shows two methods, namely the comprehensive method and the contour plot method, to analyze eDRAM performance variation. Finally, we conclude the paper in section 6.

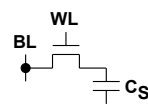
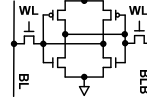
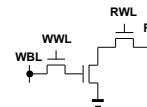
	1T1C eDRAM	6T SRAM	Gain cell eDRAM
Cell schematic			
Low-Vdd margin	Poor (destructive read)	Poor (ratioed)	Good (non ratioed, gain function)
Cell cap	<10 fF	Irrelevant	~ 1 fF
Cell size	15F ²	130F ²	65F ²
Process	Trench cap + thick TOX access TR	Logic compatible	Logic compatible
Access time	2 ns	< 1ns	2 ns

Figure 1. Comparison of various embedded memory options.

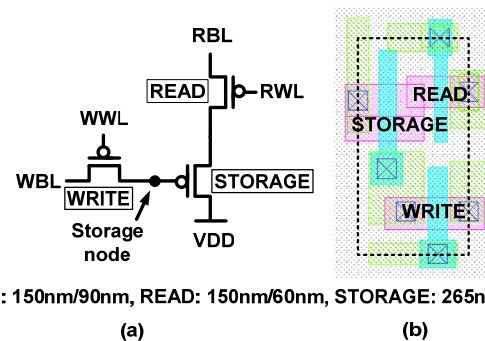


Figure 2. (a) Schematic and (b) layout of a 3T PMOS gain cell. A PMOS cell has roughly an order of magnitude lower gate tunneling leakage than its NMOS counterpart and hence significantly improves the data retention time.

2. 3T GAIN CELL EDRAM BASICS

Fig. 2(a) shows a conventional gain cell eDRAM implemented with 3 transistors (i.e. WRITE, READ, and STORAGE devices). Note that a PMOS cell has roughly an order of magnitude lower gate leakage than its NMOS counterpart and hence improves the data retention time by a factor of 3~4. Fig. 2(b) shows the layout of a single cell in an industrial 1.2V, 65nm low power CMOS logic process. The Write WordLine (WWL) of the PMOS gain cell is either driven to a negative voltage (typically around -0.5V) during write access to eliminate the threshold voltage drop, or held at a boosted voltage (typically $V_{DD}+0.5V$) during hold mode to cut off the sub-threshold leakage. Due to the positive and negative out-of-rail voltages, the operating voltage of an eDRAM cell must be carefully determined based on the gate dielectric reliability. The W/L of each device is optimized for minimum cell area, longer retention time, and tolerance against process variation. In this work, we use a sizing of 150nm/60nm, 265nm/80nm and 150nm/90nm, for the READ, the STORAGE, and the WRITE devices, respectively.

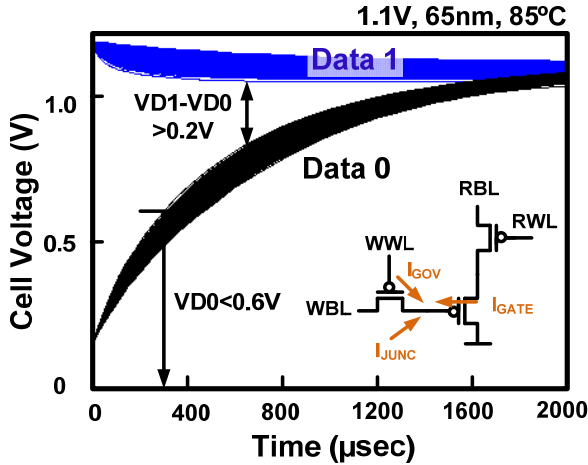


Figure 3. Cell retention characteristics using Monte Carlo simulations. The two criteria for determining the maximum cell retention time are (1) $VD_0 < 0.6V$ for meeting a target read speed and (2) $VD_1 - VD_0 > 0.2V$ to ensure sufficient read current difference.

During hold mode, the leakage currents surrounding the storage node lead to a loss in the cell voltage as depicted in Fig. 3. For a 3T PMOS cell, the steady state voltage is close to V_{DD} as most of the leakage currents in a PMOS device are inherently in the pull up direction. Note that D0 is more critical than D1 as the former becomes the gate overdrive for the read path device, which determines the bitline pull up speed. Degradation in the D0 voltage due to the surrounding leakages affects the bitline delay so a periodic refresh operation restores the cell voltage. In general, the following two criteria determine the maximum cell retention time of a gain cell: (1) the cell node voltage for D0 (or D1 for an NMOS cell) should be lower than a specified value so that a target read speed can be met and (2) the cell voltage difference between D0 and D1 should be greater than a certain value for sufficient sensing margin. In many gain cell designs, the first criterion is more stringent and hence determines the overall retention time but this could change depending on the target operating frequency. Fig. 3 shows a maximum retention time of around 300µsec based

on a criterion of $VD_0 < 0.6$ and $VD_1 - VD_0 > 0.2V$ as annotated in the figure.

Fig. 4 displays the read-writeback timing of a 256 WL x 128 BL 3T gain cell eDRAM array. The bitline voltages are pre-discharged and the RWL is asserted. Depending on the cell data, the bitline voltage either rises (D0) or remains low (D1). Unlike SRAMs, DRAMs do not have a complementary bitline and therefore require a reference bitline voltage for the sense amplifier. A dummy cell produces a reference bitline level that is in between the D1 and D0 bitline levels. A small bitline voltage difference is amplified to a full swing signal by a sense amplifier. This is followed by a write-back operation to restore the cell data. The full read-writeback cycle consists of the precharge delay, the bitline delay, the sense-amplifier delay, and the write-back delay. Among these delay components, the precharge delay, the sense-amplifier delay and the write-back delay are relatively immune to process variations and can be optimized separately. Bitline delay on the other hand, is much more sensitive to process variations as is the case in most memory designs. For this reason, this study will focus on the bitline delay and not the full access cycle. Note that the dummy bitline circuit can be designed to be tolerant to variation through optimal sizing or post-silicon tuning. This is possible due to their limited numbers. Hence, we only consider variation in the accessed cell for our simulations.

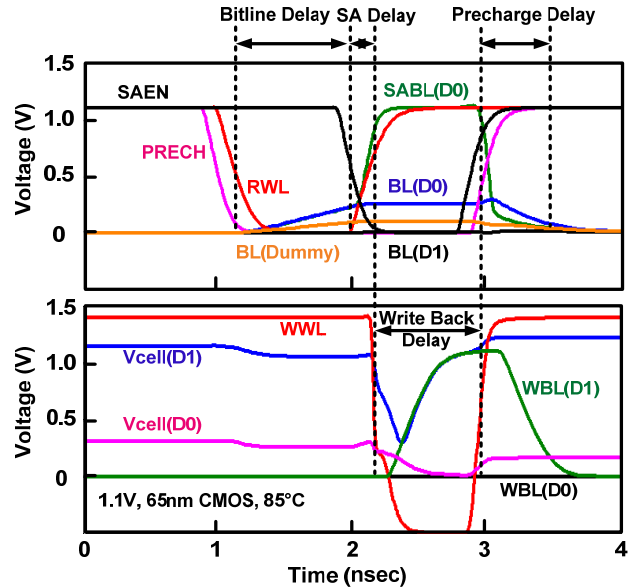


Figure 4. Read-writeback timing of a conventional 256 WL x 128 BL 3T gain cell eDRAM array.

3. BITLINE DELAY VARIATION

3.1. Corner Simulation Pitfalls

Designers use corner parameters such as SS (slow NMOS, slow PMOS) to estimate worst case circuit performance. To examine whether standard process corners capture the worst case performance for gain cells, we simulated the bitline delay while varying the TOX and V_{TH} in the same direction for all devices in a 3T gain cell. The $\sigma(V_{TH})$ values were calculated according to the gate areas of each individual device. We used 3σ values of 0.11V, 0.07V and 0.09V, for the V_{TH} s of the READ, STORAGE, and WRITE devices, respectively. 3σ value of TOX was 0.45nm for all three devices. Since we are considering a PMOS cell, we

are interested in the read access time for D0 which determines the PMOS drive current and is more vulnerable to process variations due to the leakage in the pull up direction.

Fig. 5(a) shows the bitline delay at a hold time of 0 μ sec, which is equivalent to the cell being read immediately after being written. In this case, the SS and FF corner delays are close to the worst and best case delays. However, after a 300 μ sec hold time, neither process corners captures the worst nor best cases as shown in Fig. 5(b). This is due to the fact that a fast corner device has both larger leakage and higher read current, and therefore, even though the read device is stronger, the larger leakage current makes the cell voltage degrade faster resulting in an overall increase in read delay. Similarly, for the slow corner, the weak read device is “compensated” by the smaller loss in the cell voltage due to the lower leakage current which leads to a non-worst case delay. Fig. 5(b) also implies that after a long hold period, a thin TOX high VTH combination will result in the worst case, which is not a standard process corner included in device models. For this reason, it is necessary to examine the impact of TOX and VTH separately.

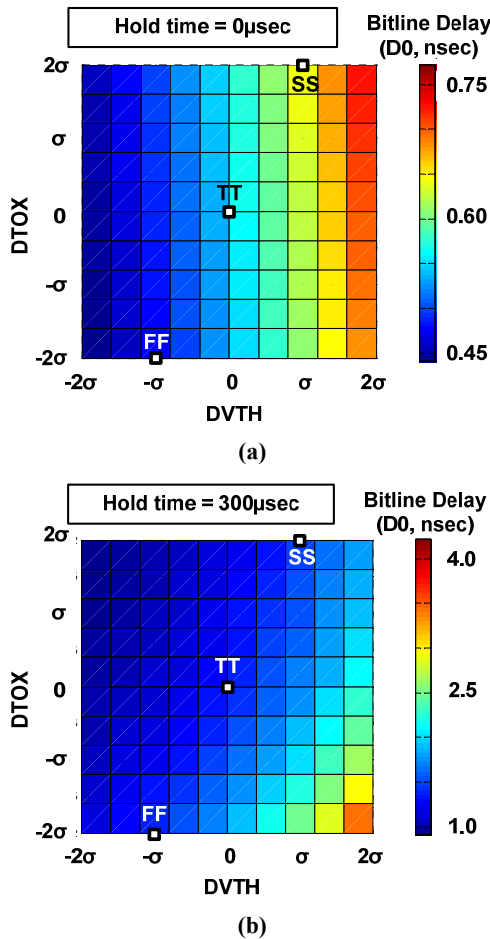


Figure 5. Bitline delay dependency on TOX and VTH after a hold period of (a) 0 μ sec (b) 300 μ sec. Standard process corners such as SS, TT, and FF do not represent the worst or best case conditions at longer hold periods.

3.2. Cell Node Voltage and Read Path Strength

Bitline delay of a 3T PMOS cell is determined by two circuit parameters: (1) the D0 cell node voltage at the time the cell is accessed, and (2) the read path drive strength. The former is determined by the various leakage components surrounding the storage node, namely the gate tunneling leakage of the STORAGE and WRITE transistors, and the junction leakage of the WRITE transistor. The latter factor (i.e. the read path drive strength) simply depends on the VTH of the STORAGE and READ devices that comprise the read path. Note that the sub-threshold leakage of the WRITE device with a boosted WWL must be negligible compared to the other leakage components. This is imperative for any type of DRAM cell that has to maximize the cell retention time. Fig. 6 shows the VD0 (D0 cell node voltage) after a 100 μ sec hold period. It indicates that TOX of the STORAGE and WRITE devices have the strongest impact on VD0. Other parameters such as the TOX of the READ device or the VTH of any device have little impact on VD0. On the other hand, the bitline delay at a fixed VD0 voltage in Fig. 7 indicates that the VTH of the STORAGE and READ determine the read path strength. Note that TOX variation in the STORAGE and READ devices does affect the read path drive current through the change in COX but this effect is negligible compared to the VTH effect as shown in Fig. 7.

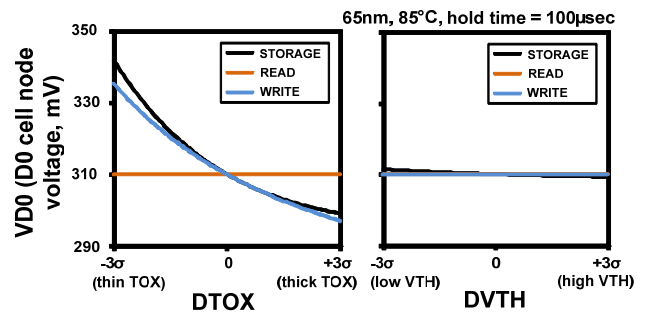


Figure 6. VD0 after 100 μ secs of hold mode for different (a) TOX and (b) VTH values. The VD0 voltage primarily depends on the TOX of the STORAGE and WRITE devices.

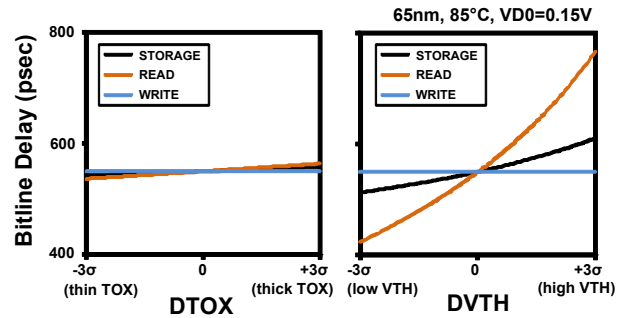


Figure 7. Bitline delay at VD0=0.15V for different (a) TOX and (b) VTH values. Bitline delay at a fixed VD0 depends on the VTH of the STORAGE and READ devices.

Fig. 8 summarizes the findings from Figs. 6 and 7. TOX affects only the cell node voltage and not the read path strength while VTH affects only the read path strength. This explains why standard corner simulations cannot capture the worst case cells. Since both TOX and VTH are moving in the same direction for standard corners (e.g. FF corner has a thin TOX and low VTH, SS corner has a thick TOX and high VTH, and so on), the TOX and

VTH effects compensate each other resulting in a non-worst case delay. In reality however, a device with a thin TOX and high VTH combination gives the worst case delay due to the detrimental effect on both the cell node voltage and the read path strength.

	TOX Variation	VTH Variation
VD0 @ fixed hold time	WRITE, STORAGE	Negligible impact
Read path strength @ fixed VD0	Negligible impact	READ, STORAGE

Figure 8. Summary of the results from Figs. 6 and 7. TOX affects only the VD0 while VTH affects only the read path strength.

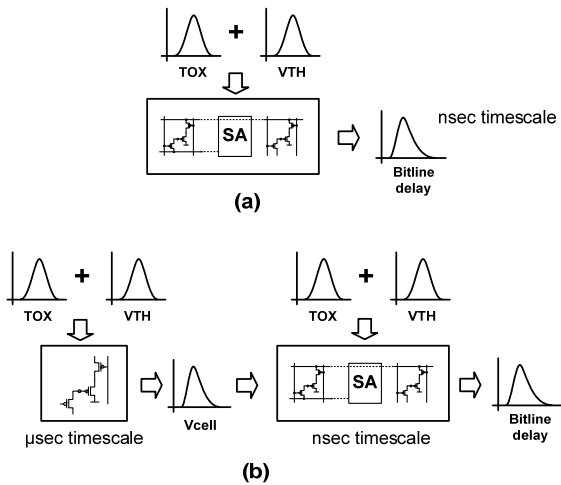


Figure 9. (a) One step and (b) two step Monte Carlo simulation methods for estimating gain cell eDRAM performance using standard circuit simulator features.

4. PRACTICAL ISSUES WITH MONTE CARLO SIMULATIONS

This section discusses the simulation strategies for analyzing gain cell performance under process variation. The following two strategies (see Fig. 9) can be considered when using the standard Monte Carlo features of circuit simulators. Other elaborate simulation strategies can be devised but will not be discussed in this paper as they involve a considerable amount of programming effort.

- *One step Monte Carlo*: This straightforward method simulates the entire hold mode and read access sequence for each Monte Carlo parameter set. This can be considered as the golden method which represents the real hardware. The main drawback of this method is that, depending on the type of circuit simulator, the designer may not be able to view the accurate signal waveforms. Such anomalies can occur when running simulations with a discrepancy in time scale. For example, for our studies, the hold mode simulation requires a time frame of 100's of μ secs while the read access requires a few nsecs.

- *Two step Monte Carlo*: Here, the hold mode and read access Monte Carlo simulations are run separately. That is, the cell node voltage distribution is obtained in the first Monte Carlo run, and

the resultant distribution is fed into the second Monte Carlo in which the read access time is simulated using a fresh set of device parameters. Since the time scale and simulation time steps can be determined separately for the two Monte Carlo runs, the simulator doesn't introduce any anomalies and provides accurate signal waveforms. However, this method uses different parameter sets for the two Monte Carlo runs, which is not realistic. Moreover, the cell voltage distribution from the first Monte Carlo may deviate from a pure Gaussian which could introduce further errors.

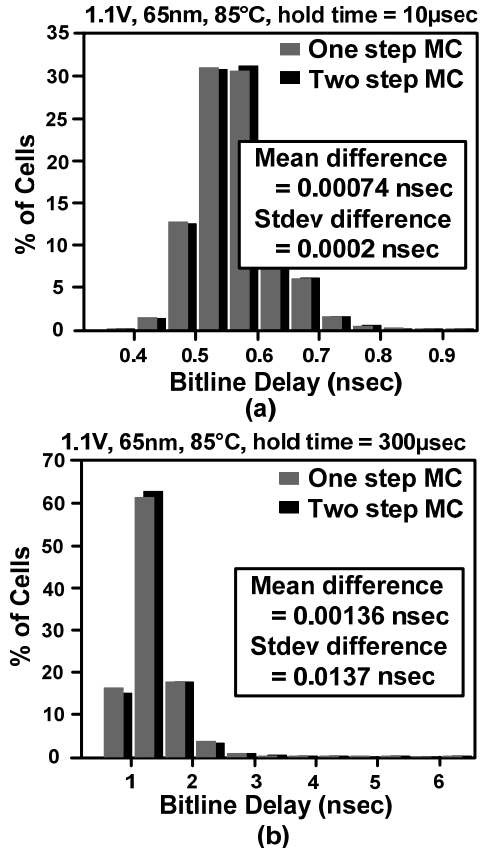


Figure 10. Bitline delay results of the one step and two step Monte Carlo methods at a hold mode time of (a) 10 μ sec and (b) 300 μ sec.

Simulation results in Fig. 10 indicate that the bitline delay distributions estimated by the two methods are extremely similar with a mean and sigma difference of only 0.00136nsec and 0.0137nsec, respectively (Fig. 10(b)). This can be attributed to the fact that the impacts of TOX and VTH variation on bitline delay are independent, as discussed in section 3. In other words, the TOX variation impacts only the first of the two Monte Carlo runs while VTH affects the second run only so the fact that the two Monte Carlo runs use different variation data does not affect the final results. Approximating the cell voltage distribution to an ideal Gaussian did not introduce a noticeable error. Therefore, it can be concluded that the two step Monte Carlo can prevent simulator anomalies at the cost of a minute error compared to the golden method. Although we chose to use the golden method (i.e. one step Monte Carlo) for this work, one can consider using the two step Monte Carlo if the detailed waveforms are needed for further circuit inspection.

5. STATISTICAL SIMULATION RESULTS

Based on the gain cell variation behavior discussed in section 3, and using the golden simulation methodology described in section 4, this section investigates the statistical performance of gain cell eDRAM. The test vehicle is a 256 WL x 128 BL 3T PMOS gain cell eDRAM array built in a 1.2V, 65nm LP CMOS process but operated at 1.1V to reduce the voltage over stress caused by the boosted WWL.

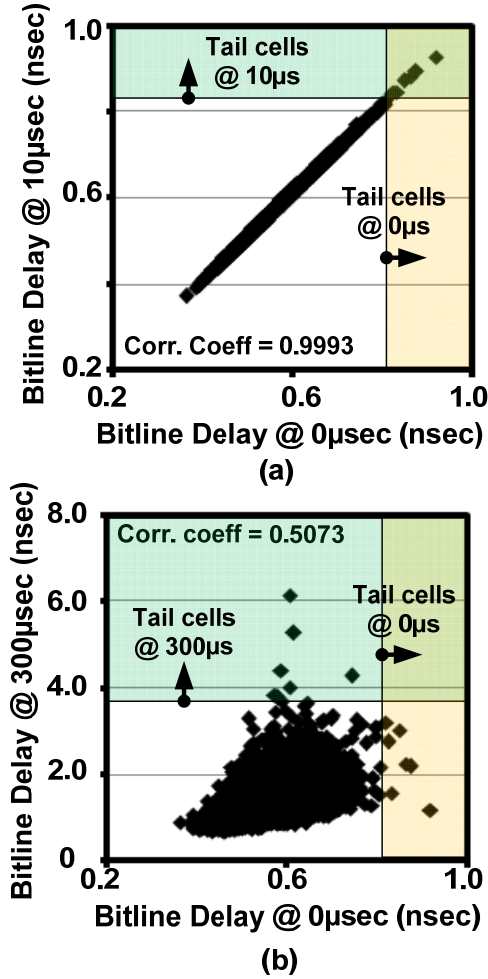


Figure 11. Scatter plot shows the correlation between bitline delays at different hold periods. (a) 10µsec vs. 0µsec and (b) 300µsec vs. 0µsec.

5.1. Comprehensive Analysis

First we show results from a comprehensive Monte Carlo simulation including both global and local TOX and VTH variations for each device. This method can effectively capture the worst case cells and give an accurate estimate on the amount of redundancy needed for a target yield. The simulation setup incorporates both the systematic and random variation components. We chose a reference bitline level that equalizes the 3σ bitline delays for D0 and D1 which in turn maximizes the 3σ point performance. Care has been taken to ensure that the simulation setup and the bitline/peripheral delays are in good agreement with real eDRAM systems [4-5]. Even though the cell retention time according to Fig. 3 is around 300µsec, we evaluate

the gain cell performance at shorter hold periods (e.g. 10µsec) to fully understand the variation effect. Fig. 11 shows the simulation results where the correlation between bitline delays at different hold periods are plotted. The bitline delay at 0µsec is simply the read path strength since the cell voltage has not degraded, so the x-axes in Figs. 11(a) and 11(b) are equivalent to the VTH variation of the read path. Results show that after a 300µsec hold period, the TOX variation starts to influence the bitline performance. The correlation coefficient between the bitline delay at 300µsec and that at 0µsec (i.e. Fig. 11(b)) is only 0.5073 which is substantially reduced from the value at 10µsec (i.e. 0.9993).

Tracking the tail cells (i.e. cells with the longest bitline delays) at different hold periods can give insight into how different process parameters affect the overall chip performance. The data points in the highlighted areas in Fig. 11 represent the tail cells. At a hold time of 0µsec, cells with a higher VTH become tail cells. The tail cells do not change at 10µsec as the hold time is too short to cause any significant degradation in the cell node voltage. However at 300µsec, cells with thinner TOX emerge as new tail cells as they suffer more from the gate tunneling leakage during the long hold period. Fig. 11(b) suggests that the tail cells switch from the ones with a higher VTH at short hold periods to the ones with thinner TOX at longer hold periods. It is worth noting that although the TOX variation effect becomes more prominent at longer hold times, VTH remains equally influential on bitline delay. This can be seen from the correlation coefficient value of 0.5073 at 300µsec.

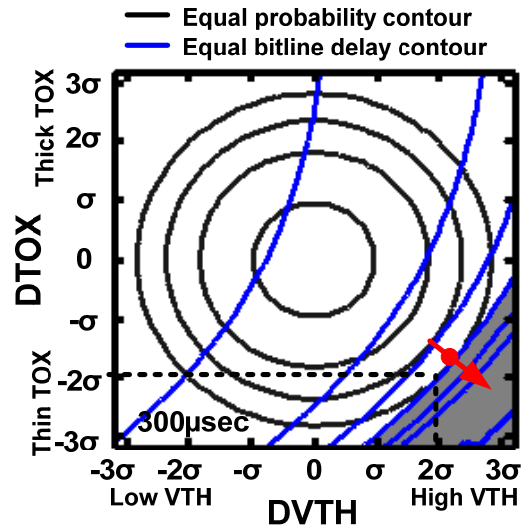


Figure 12. Equal performance contours for a hold period of 300µsec. The cumulative probability of the highlighted area is 0.20%.

5.2. Contour Plot Analysis

Another way to understand how the influence of TOX and VTH shifts with increasing hold periods is through a contour plot such as the one shown in Fig. 12. Equal bitline delay contours (blue curves) and equal probability contours (black concentric circles) are shown in the TOX and VTH space. The figure shows an example for a 300µsec hold time. Since we are concerned of the worst case only, and since we would like to be able to plot the contours in a 2D space, it is assumed that TOX or VTH of all three transistors are moving in the same direction. In case a more

thorough evaluation is needed, for example for studying the random mismatch between the three devices in a gain cell, the same methodology can be extended to an N-dimensional space by independently sweeping the TOX and VTH for each individual device. The blue contours in Fig. 12 have a positive slope indicating that the bitline delay is increasing as we move to the high VTH and thin TOX corner. Suppose we consider the concentric circle which passes the $(TOX, VTH) = (-2\sigma, +2\sigma)$ point. It can be seen that the red dot has the worst delay out of the possible TOX and VTH combinations. The arrow indicates the direction of maximum bitline delay increase. The cumulative probability of the dark grey area wherein the bitline delay is greater than that of blue contour is 0.20%.

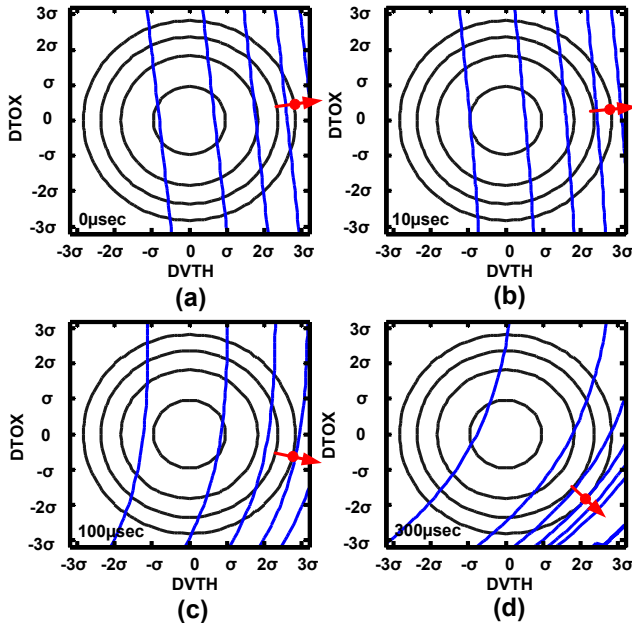


Figure 13. Bitline delay contour plot in the TOX and VTH space for hold times of (a) $0\mu\text{sec}$ (b) $10\mu\text{sec}$ (c) $100\mu\text{sec}$ and (d) $300\mu\text{sec}$.

Using the contour plots in Fig. 13, we can evaluate how the sensitivity of bitline delay with respect to TOX and VTH changes at different hold times. For example, for a $0\mu\text{sec}$ hold time, the bitline delays are relatively insensitive to the TOX as shown in the vertical contours in Fig. 13(a). As the hold time increases, the blue contours gradually exhibit a positive slope. This becomes most apparent in the lower right corners of Figs. 13 (c) and (d). For hold times longer than $300\mu\text{sec}$ in which the TOX variation effect is even more severe, it can be anticipated that the red dot will traverse towards the bottom of the outer concentric circle. The contour plot method can be useful in guiding the device optimization for improving performance and predicting the amount of redundancy needed to meet a target yield.

6. CONCLUSIONS

Gain cell eDRAMs can be implemented in a generic logic process achieving roughly 2X higher bit cell densities compared to 6T SRAMs. Furthermore, gain cells have a wider read/write margin than 6T SRAMs since there is no contention between the read and write paths. Despite the recent advances in eDRAM circuit technology, there has not been any in-depth study on gain cell retention time or performance considering process variation. In this work, we carry out a simple variation study that can offer a better understanding of gain cell eDRAM operation under process variation. The investigation begins on the premise that traditional corner parameters cannot capture the worst case delay and makes the case for a device level statistical simulation framework. Next, we show that TOX variation affects only the cell node voltage while VTH affects only the read path drive strength. As a result, the performance at short hold times is impacted more by the VTH, while at longer hold times, both TOX and VTH have significant impact on performance. Two Monte Carlo simulation strategies are compared for obtaining the eDRAM delay distribution. Finally, we utilize the correlation between bitline delays at different hold times as well as the delay contour plots to explain how the bitline delay's sensitivity to TOX and VTH varies with respect to hold time. These two methods for analyzing eDRAM performance can help optimize the sizing, build circuits and redundancy techniques for variation tolerance, and guide the device optimization process.

7. REFERENCES

- [1] Klim, P. J., Barth, J., Reohr, W. R., *et al.*, "A 1 MB Cache Subsystem Prototype With 1.8 ns Embedded DRAMs in 45 nm SOI CMOS," *Solid-State Circuits, IEEE Journal of*, vol. 44, no. 4, pp. 1216-1226, 2009.
- [2] Nakayama, M., Sakakibara, H., Kusunoki, M., *et al.*, "A 16 MB cache DRAM LSI with internal 35.8 GB/s memory bandwidth for simultaneous read and write operation," *Solid-State Circuits Conference, 2000. Digest of Technical Papers. ISSCC. 2000 IEEE International*. pp. 398-399, 472-3.
- [3] Romanovsky, S., Katoch, A., Achyuthan, A., *et al.*, "A 500MHz Random-Access Embedded 1Mb DRAM Macro in Bulk CMOS," *Solid-State Circuits Conference, 2008. ISSCC 2008. Digest of Technical Papers. IEEE International*. pp. 270-612.
- [4] Somasekhar, D., Yibin, Y., Aseron, P., *et al.*, "2 GHz 2 Mb 2T Gain Cell Memory Macro With 128 GBytes/sec Bandwidth in a 65 nm Logic Process Technology," *Solid-State Circuits, IEEE Journal of*, vol. 44, no. 1, pp. 174-185, 2009.
- [5] Chun, K. C., Jain, P., Lee, J. H., *et al.*, "A sub-0.9V logic-compatible embedded DRAM with boosted 3T gain cell, regulated bit-line write scheme and PVT-tracking read reference bias," *VLSI Circuits, 2009 Symposium on*. pp. 134-135.
- [6] Luk, W. K., Jin, C., Dennard, R. H., *et al.*, "A 3-Transistor DRAM Cell with Gated Diode for Enhanced Speed and Retention Time," *VLSI Circuits, 2006. Digest of Technical Papers. 2006 Symposium on*. pp. 184-185.