The Impact of Bit-Line Coupling and Ground Bounce on CMOS SRAM Performance

Li Ding and Pinaki Mazumder

Department of Electrical Engineering and Computer Science
The University of Michigan, Ann Arbor, MI 48109, USA
E-mail: {lding,mazum}@eecs.umich.edu

Abstract

In this paper, we provide an analytical framework to study the inter-cell and intra-cell bit-line coupling when it is superimposed with the ground bounce effect and show how those noises impair the performance of SRAM. The impact of noises is expressed in term of a coupling noise degradation factor and a ground bounce degradation factor. We have used analytical techniques to reduce the governing nonlinear ordinary differential equations to some manageable form and have derived very simple formulas for those degradation factors. Experiments have shown that the results obtained using the derived simple formulas are in good agreement with HSPICE simulation.

1. Introduction

Along with integration level and performance improvement, semiconductor technology scaling has made it a necessity to model inherent 'secondary' physical effects, which, in the past, had little impact on integrated circuit performance and were often neglected. In digital CMOS circuits, such secondary physical effects, including capacitive and inductive coupling among interconnects, power network IR voltage drop, simultaneous switching noise, and so on, are often referred to as digital noises [1] and are receiving increasingly wider attention in recent years.

In contrast to noise in digital CMOS systems, noise in well designed memories typically does not cause hard system failure. Instead, they manifest themselves mainly in the form of degrading system performance. In modern high-performance semiconductor memories, however, every millivolt counts. It is therefore pivotal to study the sources of inherent noises and their impact on memory system performance. In this paper, we study the combined effect of two dominant noise types in static random access memories (SRAM), bit-line coupling and ground bounce, on SRAM performance.

There are two types of capacitive bit-line couplings. Intra-bit-line coupling is caused by the capacitance between the true and the complementary bit-lines of a single memory cell. Due to the differential operation nature of the sense amplifier, intra-bit-line coupling has twice the effect on bit-line delay. Inter-bit-line coupling is due to the coupling between bit-lines of adjacent memory cells. It leads to memory READ time variation depending on the data stored in the memory, as illustrated in Figure 1. In semiconductor memories, the bit-line coupling effect is first seen in dynamic random access memory (DRAM) [2], [3]. Techniques like twisted data-lines are extensively used to cancel out the coupling noise [4], [5]. Bit-line coupling is a lesser problem with SRAM because SRAM cells are significantly larger than DRAM cells resulting to larger separation of the bit-lines. However, aggressive technology scaling in the lateral dimensions while leaving the vertical dimension relatively unchanged has significantly increased the ratio of the bit-line coupling capacitance to total capacitance. This makes the coupling among SRAM bit-lines no longer negligible [6].

Ground bounce, also known as simultaneous switching noise (SSN), is caused by the large instant current, due to the switching of multiple devices, through the parasitic inductance at the ground node. SSN at chip output drivers have been studied extensively in the literature [7] and many driver design techniques considering SSN were also reported [8], [9]. Ground bounce is especially a serious problem in semiconductor memories because the simultaneous switching of a large number of memory cells and sense amplifiers. As the size of memories quadruples every three years or so, the number of columns is slowly increasing, which results to an increase in ground bounce. In this paper, we consider the ground bounce in the SRAM cell array caused by the simultaneous switching of memory cells.

It is the intention of this research to analytically study the combined effects of bit-line coupling and ground bounce on SRAM bit-line delay. In Section II, the problem is formu-

Proceedings of the 16th International Conference on VLSI Design (VLSI'03)
1063-9667/03 $17.00 © 2003 IEEE

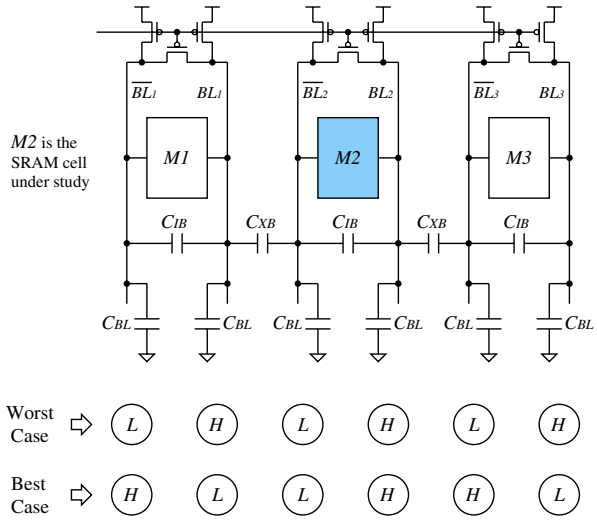

Figure 1. Variation of SRAM access time due to inter-bit-line coupling.

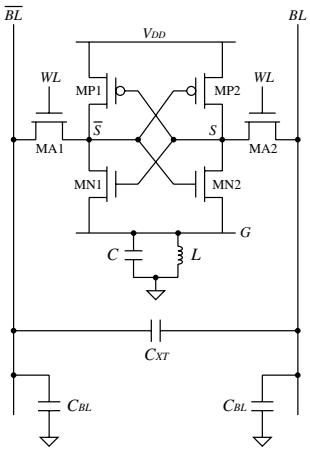

Figure 2. Simplified circuit for noise analysis.

lated into a set of nonlinear ordinary differential equations. In Section III, those equations are first simplified based on observations of the SRAM operation. The simplified equations are then solved analytically. In Section IV, we compare the results predicted by the derived formulation with those obtained using HSPICE simulation.

2. Formulation

Let us study memory cell *M2* in Figure 1. In the figure, C_{BL} is the bit-line capacitance comprised of drain capacitances of the SRAM cells, bit-line wire capacitance, sense amplifier input capacitance, and the capacitances of the precharge and balance transistors. Intra-bit-line capacitance C_{IB} and inter-bit-line capacitance C_{XB} are due to the coupling between the bit-lines.

The coupling noise induced on bit-lines depends on the data pattern stored in the memory cells. The worst case scenario is that all the memory cells store 1 when the coupling from both neighboring cells tends to slow down the development of the differential voltage across $\overline{BL_2}$ and BL_2. In this case, BL_1, BL_2 and BL_3 ($\overline{BL_1}$, $\overline{BL_2}$ and $\overline{BL_3}$) are electrically identical due to the symmetry of the system. Therefore, the coupling capacitance between BL_1 and $\overline{BL_2}$ and that between BL_2 and $\overline{BL_3}$ can be replaced by a single capacitance of value C_{XB} connected between $\overline{BL_2}$ and BL_2. In all, when all memory cells store 1, we have a simplified circuit as shown in Figure 2. It can be shown that the same simplified circuit can be derived for the best case scenario. The only difference between the two cases is the definition of the effective coupling capacitance

$$C_{XT} = \begin{cases} C_{IB}, & \text{best case} \\ C_{IB} + C_{XB}, & \text{worst case.} \end{cases}$$

Denote the parasitic ground inductance, which includes the package inductance and the ground network wiring inductance, as L_0 and the parasitic ground capacitance, which includes the capacitance of the ground network and the package pin, as well as the effective decoupling capacitance, as C_0. Assume there are N columns in the SRAM array, it can be shown that the ground bounce effect caused by N memory cells in a row is identical to that caused by a single memory cell when the following transformation is used

$$C = C_0/N, \quad L = NL_0.$$

In this paper, we assume that the memory bit-lines are precharged to V_{DD}. The analysis is similar when the bit-lines are precharged to other fixed voltages. We also assume the SRAM cell operates under normal condition, i.e., it does not flip state during a READ operation. In this case, node S stays very close to V_{DD}. The access transistor MA2 and the pull-up transistor MP1 remain in the off-state. The voltage at the storage node $\overline{S}$ will increase and then stays at a level such that the current through the access transistor MA1 is approximately the same as that through the drive transistor MN1. For an appropriately sized SRAM cell, this voltage will not reach the logic threshold voltage of the inverter formed by the transistor pair MP2 and MN2.

There are four nodes in the system under study: two bit-lines BL and $\overline{BL}$, the storage node $\overline{S}$, and the ground node G. Each of the first three nodes is governed by a first-order differential equation while the ground node has to be modeled by a second-order differential equation because there

are both capacitance and inductance in the ground node. Those equations can be written as follows.

$$(C_{BL}+C_{XT})\frac{dV_{\overline{BL}}(t)}{dt} - C_{XT}\frac{dV_{BL}(t)}{dt} = -I_{ac}(t) \quad (1)$$

$$(C_{BL}+C_{XT})\frac{dV_{BL}(t)}{dt} - C_{XT}\frac{dV_{\overline{BL}}(t)}{dt} = 0 \quad (2)$$

$$C_{\overline{S}}\frac{dV_{\overline{S}}(t)}{dt} = I_{ac}(t) - I_{dr}(t) \quad (3)$$

$$C\frac{d^2V_G(t)}{dt^2} + \frac{V_G(t)}{L} = \frac{dI_{dr}(t)}{dt} \quad (4)$$

where $I_{ac}(t)$ and $I_{dr}(t)$ are the time-variant current through the access transistor MA1 and the drive transistor MN1, respectively.

For simplicity of analysis, we have assumed the word line switches from zero to the supply voltage immediately. Therefore, the gate voltage of the access transistor is a constant for $t > 0$, and the current through the access transistor is a function of the voltages at the remaining three nodes

$$I_{ac}(t) = f\big(V_{\overline{S}}(t), V_G(t), V_{\overline{BL}}(t)\big). \quad (5)$$

The gate voltage of the drive transistor MN1 approximately stays at V_{DD} and both source and bulk nodes are tied to the ground node. Hence we can write

$$I_{dr}(t) = g\big(V_{\overline{S}}(t), V_G(t)\big). \quad (6)$$

In all, our task is to solve Equations (1)-(6) under the initial condition that $V_{\overline{BL}}(0) = V_{BL}(0) = V_{DD}$, $V_{\overline{S}}(t) = 0$, and $V_G(0) = V'_G(0) = 0$, where the prime mark refers to time direvative.

3. Analysis

Analytical study of Equations (1)-(6) in general form is virtually intractable. In this paper, we focus on the special case that is applicable to our application. We have observed that for typical SRAM READ operations, 1) the bit-line voltage swing is very limited; 2) the access transistor MA1 stays in the saturation region, and 3) the storage node $\overline{S}$ quickly settles to a quasi-static state because of its small nodal capacitance. Those observations will be used to simplify the problem.

First, the set of equations formulated in the previous section need an explicit expression for $I_{ac}(t)$ and $I_{dr}(t)$. A widely used MOSFET I-V model is the α-power law model [10]. However, if we use the α-power law model, the resulting equations are nonlinear and there is no closed-form solution.

A typical approach used in literature on facing similar scenario is to approximate the formulation of the α-power law model such that the resulting equation has a simpler form, e.g., a linear dependency on the node voltages. This approach has two sources of errors, the device modeling error and the error occurred during the approximation. Here we take a shortcut and use a linear model for the transistors directly. We assume that the current equations can be written in the following form

$$I_{ac}(t) = a_0 - a_1 V_{\overline{S}}(t) + a_2 V_G(t) + a_3 V_{\overline{BL}}(t), \quad (7)$$

$$I_{dr}(t) = b_1\big(V_{\overline{S}}(t) - V_G(t)\big). \quad (8)$$

The set of equations (1)-(4) and (7)-(8) are linear ordinary differential equations with no closed-form solution. Based on our previous observation that the bit-line voltage swing is small and that the access transistor is in saturation region, we assume that the bit-line voltage have no direct impact on the current through the access transistor MA1. In the next section, we will show that this is a very good assumption through simulation. Under this assumption, the access transistor current formula is simplified to

$$I_{ac}(t) = a_0 - a_1 V_{\overline{S}}(t) + a_2 V_G(t). \quad (9)$$

Now (3)-(4) are separated from (1)-(2). Therefore, we will first derive the voltage waveforms at the ground node G and the storage node $\overline{S}$. The set of equations is

$$C\frac{d^2V_G(t)}{dt^2} + \frac{V_G(t)}{L} = b_1\big(\frac{dV_{\overline{S}}(t)}{dt} - \frac{dV_G(t)}{dt}\big), \quad (10)$$

$$C_S\frac{dV_{\overline{S}}(t)}{dt} = a_0 - (a_1+b_1)V_{\overline{S}}(t) + (a_2+b_1)V_G(t), \quad (11)$$

which is equivalent to a third-order differential equation and, literally, can be solved analytically. However, we have found the resulting expression is still overly complicated.

To obtain an approximate analytical solution to this third order differential equation, we study two extreme cases. First when t is very small, the ground bounce voltage can be neglected. Equation (11) is reduced to

$$C_S\frac{dV_{\overline{S}}(t)}{dt} = a_0 - (a_1+b_1)V_{\overline{S}}(t),$$

which can be analytically solved and the waveform for $V_{\overline{S}}$ is written as

$$V_{\overline{S}}(t) = \frac{a_0}{a_1+b_1}(1 - e^{-\lambda_0 t}), \quad (12)$$

where

$$\lambda_0 = (a_1+b_1)/C_S.$$

Second, when t is large, the left side of the equation can be neglected because C_S is typically 2-3 orders of magnitude smaller than bit-line capacitances. And we can obtain a quasi-steady state equation as

$$V_{\overline{S}}(t) = \frac{a_0}{a_1+b_1} + \frac{a_2+b_1}{a_1+b_1}V_G(t). \quad (13)$$

Therefore, we postulate that the voltage waveform at node $\overline{S}$ has the following form

$$V_{\overline{S}}(t) = \frac{a_0}{a_1+b_1}\left(1-e^{-\lambda_0 t}\right) + \frac{a_2+b_1}{a_1+b_1}V_G(t). \quad (14)$$

It can be easily verified that the above equation reduces to (12) when t is small and to (13) when t is large, respectively.

By inserting the approximate voltage waveform $V_{\overline{S}}(t)$ into (10), we have constructed an approximated equation for the ground node voltage

$$C\frac{d^2V_G(t)}{dt^2} + b_1\frac{a_1-a_2}{a_1+b_1}\frac{dV_G(t)}{dt} + \frac{V_G(t)}{L} = \frac{a_0b_1}{a_1+b_1}\lambda_0 e^{-\lambda_0 t},$$

which can be analytically solved using the initial condition that $V_G(0)=0$ and $V'_G(0)=0$. The voltage waveform at the ground network is derived as

$$V_G(t) = A\left(e^{-\lambda_0 t} + \left(\frac{\lambda_0-\lambda}{\omega}\sin\omega t - \cos\omega t\right)e^{-\lambda t}\right), \quad (15)$$

where

$$\begin{aligned} A &= \frac{a_0b_1}{a_1+b_1}\cdot\frac{\lambda_0}{C\lambda_0^2 - b_1\lambda_0 + 1/L}, \\ \lambda &= \frac{(a_1-a_2)b_1}{2(a_1+b_1)}\cdot\frac{1}{C}, \\ \omega &= \sqrt{\frac{1}{LC}-\lambda^2}. \end{aligned}$$

For typical embedded SRAM, $\lambda_0 \gg \omega$ and $\omega \gg \lambda$. At any reasonably large time t ($t \gg 1/\lambda_0$), the ground node voltage waveform can be further simplified to

$$V_G(t) = \frac{a_0b_1}{a_1+b_1}\sqrt{\frac{L}{C}}e^{-\lambda t}\sin\omega_0 t, \quad (16)$$

where

$$\omega_0 = 1/\sqrt{LC}.$$

Neglecting the effect of the short period before the storage node $\overline{S}$ settles, we have the following approximated expression

$$V_{\overline{S}}(t) = \frac{a_0}{a_1+b_1} + \frac{a_0b_1(a_2+b_1)}{(a_1+b_1)^2}\sqrt{\frac{L}{C}}e^{-\lambda t}\sin\omega_0 t. \quad (17)$$

The time-variant current through the access transistor is hence derived as

$$I_{ac}(t) = K_1 - K_2 e^{-\lambda t}\sqrt{\frac{L}{C}}\sin\omega_0 t, \quad (18)$$

where

$$K_1 = \frac{a_0b_1}{a_1+b_1}, \quad K_2 = \frac{a_0(a_1-a_2)b_1^2}{(a_1+b_1)^2}.$$

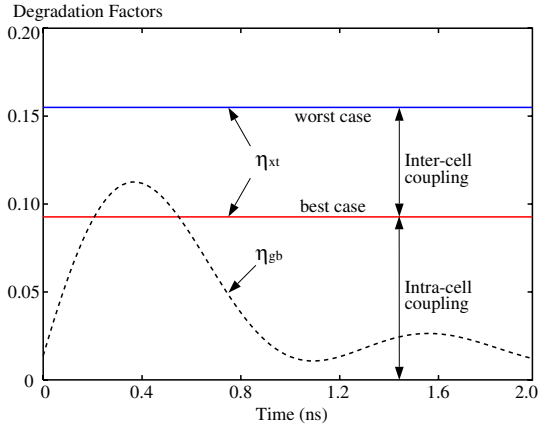

Figure 3. Typical degradation factors.

The voltage across the sense amplifier V_{SA} is derived as

$$V_{SA}(t) \simeq \frac{1}{C_{BL}+2C_{XT}}\left(K_1 t - K_2 L(1-e^{-\lambda t}\cos\omega_0 t)\right). \quad (19)$$

The above equation can be rewritten as

$$V_{SA}(t) \simeq V_{SA,0}(t)\cdot(1-\eta_{xt})\cdot\left(1-\eta_{gb}(t)\right). \quad (20)$$

where $V_{SA,0}(t)$ is the voltage across the sense amplifier assuming there is neither bit-line coupling nor ground bounce, η_{xt} is the bit-line coupling degradation factor, and $\eta_{gb}(t)$ is the ground bounce degradation factor. They are defined as

$$\begin{aligned} V_{SA,0}(t) &= K_1 t/C_{BL}, & (21)\\ \eta_{xt} &= 2C_{XT}/C_{BL}, & (22)\\ \eta_{gb}(t) &= \frac{K_2}{K_1 t}L(1-e^{-\lambda t}\cos\omega_0 t). & (23) \end{aligned}$$

As expected, η_{xt} is a positive number and $\eta_{gb}(t)$ is also positive for any $t>0$. Therefore, capacitive coupling among bit-lines and ground bounce will always deteriorate the performance of the SRAM READ operation. It is also observed that the coupling degradation factor is a constant while the ground bounce degradation factor $\eta_{gb}(t)$ varies with respect to time. The dependency of the degradation factors on time are illustrated in Figure 3 for a typical SRAM.

4. Experiments

In this section, we verify the correctness of the derived formula through HSPICE simulation. We have used a 0.18 micron process with a supply voltage of 1.8 V in this study. The extracted model parameters for the process are shown in Table 1. They are obtained by curve fitting the MOSFET I-V characteristic generated using HSPICE Level 49 model

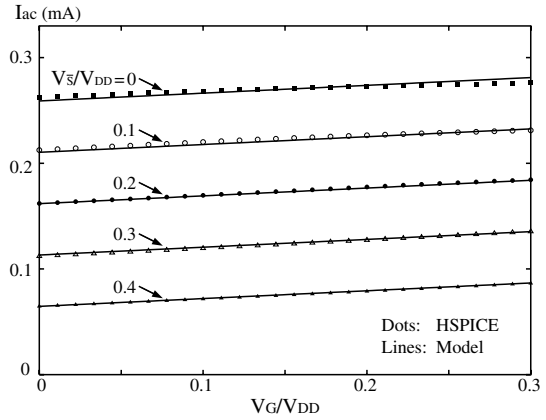

Figure 4. Access transistor current modeling.

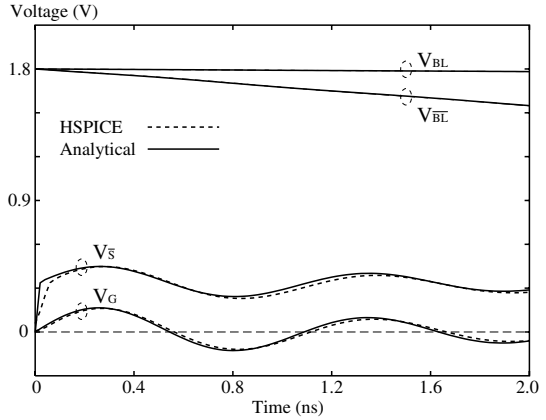

Figure 5. Comparison of simulated and analytical transient waveform.

(BSIM3). In Figure 4, we compare the access transistor current calculated by the linear model with that obtained using HSPICE level 49 model. The comparison is for the variable ranges that are interested to us: the values of V_G are limited to $0.3V_{DD}$ and the values of $V_{\overline{S}}$ are limited to $0.4V_{DD}$. As shown in the figure, the results match very well.

Table 1. Extracted model parameters.

Parameter	Value (mA/V)
a_0	0.259
a_1	0.270
a_2	0.041
b_1	0.510

The parameters used in our simulations are as follows. Bit-line capacitance, intra-cell coupling capacitance, and inter-cell coupling capacitance are 1.2 fF, 0.056 fF, and 0.038 fF, respectively, for each cell. A lumped capacitance of 15 fF is added to the bit-line capacitance representing the sense amplifier input capacitance and the capacitance of the precharge and balance transistors. The storage node capacitance C_S is 4.1 fF. We also use typical values for parasitic ground inductance and capacitance: $L_0 = 3\,nH$ and $C_0 = 10\,pF$.

Transient waveforms obtained using HSPICE simulation as well as that derived based on the proposed equations are compared in Figure 5 for a typical SRAM (1024 rows, 64 columns). The derived voltages at the ground node and the storage node $\overline{S}$ match very well with simulation results. This shows that the postulation (14) we made is valid. The difference in voltage waveforms is indistinguishable for the bit-line voltages at the resolution of the figure. The voltage across the sense amplifier is plotted in Figure 6(a) for

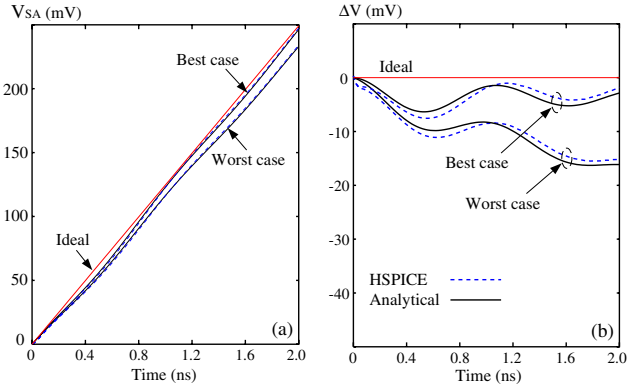

Figure 6. (a) Voltage across sense amplifier. (b) Voltage difference from ideal value.

the same SRAM configuration. The curve marked as ideal is obtained assuming, 1) it is under best inter-cell coupling scenario, and 2) there is no ground bounce by setting the ground inductance to zero. Again, the proposed formula matches the experimental results extremely well. The difference is not distinguishable at the resolution of this figure. So in Figure 6(b), we use $\Delta V(t) = V_{real} - V_{ideal}(t)$ as the vertical axis of the plot to amplify the difference between formula and experiment. The error of the formula is observed to be less than 2 mV.

In Table 2, we compare the time for the voltage across the sense amplifier to reach 100 mV. Delay variation due to inter bit-line coupling is observed to be around 5 % for our technology. Typical variation is 14 mV for 256-row SRAM array and 40 mV for 1024-row SRAM array. Analytical results on delay variation match simulation results very well.

Table 2. Variation of voltage across sense amplifier.

Number of rows	Number of cols	Best case (ns)		Worst case (ns)		Variation (ns)	
		Simulation	Analysis	Simulation	Analysis	Simulation	Analysis
256	32	0.226	0.218	0.238	0.230	0.012 (5.3%)	0.012 (5.5%)
	64	0.237	0.230	0.250	0.244	0.013 (5.5%)	0.014 (5.7%)
	128	0.260	0.263	0.276	0.278	0.016 (6.2%)	0.015 (5.7%)
512	32	0.438	0.431	0.462	0.455	0.024 (5.5%)	0.024 (5.7%)
	64	0.464	0.455	0.488	0.479	0.024 (5.2%)	0.024 (5.3%)
	128	0.509	0.498	0.534	0.521	0.025 (4.9%)	0.023 (4.6%)
1024	32	0.820	0.817	0.862	0.860	0.042 (5.1%)	0.043 (5.3%)
	64	0.838	0.830	0.878	0.872	0.040 (4.8%)	0.042 (5.1%)
	128	0.872	0.855	0.911	0.896	0.039 (4.5%)	0.041 (4.8%)

5. Conclusion

In this paper, we have analytically studied the issues of bit-line coupling and ground bounce in SRAM. The impact of those two sources of noise on SRAM performance is formulated using two degradation factors. The bit-line coupling degradation factor is a constant with respect to time while the ground bounce degradation factor varies with time. It is observed that both noise types may cause over 10 % increase in bit-line delay and that the variation of bit-line delay due to different data pattern in the SRAM is about 5% in our technology. We have analytically derived simple formulas for the degradation factors using an application-specific linear model for the MOSFET transistors. Experiments have shown that those simple formulas are in good agreement with HSPICE simulation results.

Acknowledgments

This work is supported in part by the AFOSR MURI grant and in part by an NSF grant. The authors would like to thank Gaurav Mittal for having several stimulating discussions on noise issues in SRAM.

References

[1] K. L. Shepard and V. Narayanan, "Noise in Deep Submicron Digital Design," in *Proc. of International Conference on Computer Aided Design*, pp. 524-531, Nov. 1996.

[2] Y. Nakagome, M. Aoki, S. Ikenaga, *et al*., "The Impact of Data-Line Interference Noise on DRAM Scaling," *IEEE Journal of Solid-State Circuits*, vol. 23, no. 5, pp. 1120-1127, Oct. 1988.

[3] Y. Konishi, M. Kumanoya, H. Yamasaki, *et al*., "Analysis of Coupling Noise Between Adjacent Bit Lines in Megabit DRAM's," *IEEE Journal of Solid-State Circuits*, vol. 24, no. 1, pp. 35-42, Feb. 1989.

[4] M. Aoki, Y. Nakagome, M. Horiguchi, *et al*., "A 60-ns 16-Mbit CMOS DRAM with a Transposed Data-Line Structure," *IEEE Journal of Solid-State Circuits*, vol. 23, no. 5, pp. 1113-1119, Nov. 1988.

[5] H. Hidaka, K. Fujishima, Y. Matsuda, *et al*., "Twisted Bit-Line Architectures for Multi-Megabit DRAM's," *IEEE Journal of Solid-State Circuits*, vol. 24, no. 1, pp. 21-27, Feb. 1989.

[6] A. Kinoshita, S. Murakami, Y. Nishimura, *et al*., "A Study of Delay Time on Bit Lines in Megabit SRAM's," *IEICE Transactions on Electrons*, vol. E75-C, pp. 1383-1386, Nov. 1992.

[7] R. Senthinathan and J. L. Prince, "Simultaneous Switching Ground Noise Calculation for Packaged CMOS Devices," *IEEE Journal of Solid-State Circuits*, vol. 26, no. 11, pp. 1724-1728, Nov. 1991.

[8] H. I. Hanafi, R. H. Dennard, C.-L. Chen, R. J. Weiss, and D. S. Zicherman, "Design and Characterization of a CMOS Off-Chip Driver/Receiver with Reduced Power-Supply Disturbance," *IEEE Journal of Solid-State Circuits*, vol. 27, no. 5, pp. 783-791, May 1992.

[9] R. Senthinathan and J. L. Prince, "Application Specific CMOS Output Driver Circuit Design Techniques to Reduce Simultaneous Switching Noise," *IEEE Journal of Solid-State Circuits*, vol. 28, no. 12, pp. 1383-1388, Dec. 1993.

[10] T. Sakurai and A. R. Newton, "Alpha-Power Law MOSFET Model and Its Application to CMOS Inverter Delay and Other Formulas," *IEEE Journal of Solid-State Circuits*, vol. 25, no. 2, pp. 584-594, Apr. 1990.

[11] T. Wada, S. Rajan, and S. A. Przybylski, "An Analytical Access Time Model for On-Chip Cache Memories," *IEEE Journal of Solid-State Circuits*, vol. 27, no. 8, pp. 1147-1156, Aug. 1992.