



HAL
open science

A Novel Dummy Bitline Driver for Read Margin Improvement in an eSRAM

Michael Yap San Min, Philippe Maurine, Magali Bastian Hage-Hassan, Michel Robert

► **To cite this version:**

Michael Yap San Min, Philippe Maurine, Magali Bastian Hage-Hassan, Michel Robert. A Novel Dummy Bitline Driver for Read Margin Improvement in an eSRAM. DELTA 2008 - 4th IEEE International Symposium on Electronic Design, Test and Applications, Jan 2008, Hong Kong, China. pp.107-110, 10.1109/DELTA.2008.72 . lirmm-00243966

HAL Id: lirmm-00243966

<https://hal-lirmm.ccsd.cnrs.fr/lirmm-00243966v1>

Submitted on 27 Jun 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

A Novel Dummy Bitline Driver for Read Margin Improvement in an eSRAM

M. Yap San Min^{1,2}, P. Maurine¹, M. Bastian² and M. Robert¹

¹LIRMM, Montpellier, France

²INFINEON TECHNOLOGIES, Sophia Antipolis, France

{michael.yapsanmin, pmaurine, michel.robert}@lirmm.fr, {michael.yap, magali.bastian}@infineon.com

Abstract

Aggressive scaling of transistors is often accompanied by an increase in variability of its intrinsic parameters. In this paper, we point out the importance of considering sensitivity performances due to process variations during SRAM design. We propose a novel dummy bitline driver, an essential component in a self-timed memory, which is less sensitive to process variations. A statistical sizing method of this dummy bitline driver is introduced so as to improve the read timing margin, while ensuring a high timing yield. The memory considered is a 256kb SRAM design in 90nm technology node.

Keywords: dummy bitline driver, low power, self-timed memory, SRAM, statistical design

1 Introduction

Technology fabrications have led to the realization of system on chip whereby functional blocks coexist, like embedded memories which can occupy up to 80% of the chip's area. Hence, the overall performances and the fabrication yield of the chips rely heavily on memory's yield. Simultaneously with the rapid increase of memory blocks within the chips, technology evolution is accompanied by an increase of variability effects owing to process variations, which appear during the manufacturing steps.

Generally, process variability can be classified into 2 distinct groups of manufacturing processes namely: global and local variations. Global variations originate from numerous factors: non uniform chemical mechanical polishing [1], lens aberrations [2] and non-uniformity of temperature [1], whereas local variations stem from a variety of factors like random dopant fluctuations [3] and line edge roughness [4]. In fact transistor scaling has exacerbated the impact of local and global variations, affecting performances of integrated circuits like maximum operation frequency and static power consumption.

To handle the impact of process variations in circuit design, corner based methodology is performed by characterizing the circuit across process corners. However, the increase of variability in manufacturing process results in an underestimation of performances in the operating frequency of an integrated circuit. This can therefore impact on the convergence of the design flow. In this paper, we highlight the importance of considering process variations in the design of an SRAM. We propose a novel dummy bitline driver which tracks the discharge time of the bitline in a read operation and triggers the sense amplifier at the right time. This structure is less

sensitive to process variation. A statistical sizing method of the driver is also introduced to improve the read timing margin while guaranteeing a high timing yield.

The paper is organized as follows. Section 2 introduces a simple way of computing the required read timing margin without being too pessimistic and of calculating the probability of fulfilling this constraint. Section 3 presents a new structure, dubbed Dummy Bitline Driver (DBD), having its timing performances less sensitive to manufacturing processes compared to a more classic DBD [5]. In section 4, we will introduce a statistical sizing method of the DBD which is independent of the process corner. Section 5 compares the results obtained between the proposed and the classic structures.

2 Modelling Approach

Conventionally, the characterization of a circuit involves performing several simulations across best and worst case corners to verify whether its performances and timing constraints are met under all conditions. For example, the worst case delay is defined by considering that principal parameters p_i of transistors have their values at $\pm m_i \cdot \sigma_{p_i}$ ($m_i \in \mathbb{N}$) around their mean values μ_{p_i} (σ_{p_i} represents the standard deviation of the statistical distribution of parameter p_i). The set up of such a simple approach, through a proper choice of m_i values, allows the worst case to be defined at $n \cdot \sigma_D$, with σ_D being the standard deviation of the delay distribution.

Failing to account for local variations across process corners is not a serious problem as far as simple data paths are considered. However, this issue is far more complex for data path with racing conditions. This approach incurs optimistic and pessimistic estimations of worst and best case methods.

node Z to 0 V at the beginning of a read cycle operation. When the internal signal WLSUM is at '1' during a read mode, inputs pins Iadj_i (i=1 to 4) are activated by hardcoding them individually at V_{dd}. Hence, they can be used to adjust the discharge current of the DBD with respect to the actual supplied voltage of the memory. In doing so, the read timing margin can be adapted to the supply voltage applied. It should be noted that the reference DBD has also the same functionalities as the proposed DBD. The main difference lies in the use of stacked transistors for representing pass gate and pull down transistors of the SRAM cell. This condition causes the sensitivity of the read current flowing through PG_i and PD_i (i=1..4) to be less representative of the read current flowing through PG_{cc1} and PD_{cc1} in the 6T SRAM cell.

4 Statistical Sizing Method

In order to perform comparisons between the reference and proposed DBDs under constant timing yield, we have developed a sizing methodology.

Step 1 (identification of most critical condition): Starting from an initial solution, the first step involves identifying the voltage and temperature (V, T)_{Crit} conditions having the poorest timing yield. To identify the critical condition, transient simulations of the timing performances of critical paths A and B in the memory are done under different temperature and voltage conditions covering this whole range to obtain μ_A and μ_B . The critical condition corresponds to the highest numerical value of the following expression:

$$\frac{\mu_A \cdot \mu_B}{(\mu_B - \mu_A)^2} \quad (4)$$

Step 2 (variability estimation): The second step requires the estimation of the variability of paths A and B involved in the signal races. To do so, Monte Carlo simulations of the critical path are performed at the critical conditions (V, T)_{Crit} found in step 1. Once these statistical simulations are performed and the values of μ_A , μ_B , σ_A , σ_B and ρ are obtained, the value of the required timing margin μ_D^{Yield} corresponding to a timing yield is computed using (3).

Step 3 (sizing for a given timing yield): The third step consists in sizing the DBD at a typical process and under (V, T)_{Crit} to obtain the computed μ_D^{Yield} .

Step 4 (first verification step of the timing yield): Once the above sizing procedure is over, the first verification step consists in performing Monte Carlo simulations on the critical path at (V,T)_{Crit} to obtain μ_A , μ_B , σ_A , σ_B and ρ values. The constraint of the timing yield is then evaluated using (2). If the computed value fulfills the predefined constraint, we proceed with the second verification step. Otherwise, we reiterate step 3 with the new values of μ_A , μ_B , σ_A , σ_B and ρ .

Step 5 (second verification step of the timing yield): It implies verifying that the constraint of the timing yield satisfies all temperature and supply voltage conditions. This is done through Monte Carlo simulations in order to estimate the values of μ_A , μ_B , σ_A , σ_B and ρ for different values of V and T. Once the statistical simulation has been done, the timing yield is processed. If the values obtained for the various (V, T) couples are greater than the predefined constraint at (V, T)_{Crit}, the verification step is over. However, if the constraint is not satisfied, step 1 should be repeated with the new sizing obtained.

5 Performance Comparisons

To perform performance comparisons, both reference and proposed DBDs have been placed in the critical path of a 256kb SRAM memory. The model card, used in Hspice simulations, is the bsim4.3.0 which takes into account local and global variations. The sizing methodology developed in section four has been applied to pass gate transistors PG1 to PG4 and pull down transistors PD, PD1 to PD4 (Fig. 3a and 3b) at four operating voltages considered i.e. 1.0V, 1.08V, 1.2V and 1.32V. The timing yield had been set at 99.87% i.e. n=3 and the correlation value ρ considered was equal to 0.9. At each operating voltage, the appropriate adjustments of pins Iadj_i were performed.

Once the statistical sizing method has been done, we performed 2000 Monte Carlo runs in order to obtain the mean values (μ_A and μ_B) and standard deviation values (σ_A and σ_B) of the characteristic delays of the signal races of paths A and B over the whole voltage and temperature ranges considered. The results obtained were used to compute in table 1 the reduction in the delay variance (ΔV_B) of path B between proposed (prop) and reference (ref) DBDs and in table 2, the probability P^V (2) of meeting the timing constraint, the read timing margin (1) of the reference μ_{dref} and proposed μ_{dprop} DBDs and subsequently the reduction in read timing margin $\Delta\mu_D$ between proposed and reference DBDs.

Table 1 shows the reduction in variability obtained. The first column Iadj_i corresponds to the respective branches of transistors selected with respect to supply voltage. For instance Iadj_i=1, 2 means that branches Iadj1 and Iadj2 are selected at V_{dd}= 1.08V. The reduction in variability ($\Delta V_B/V_{Bref}$) is quite important, lying between 5.8% and 24.7%. This reduction has been achieved by using pass gate transistors PG_i in the proposed DBD which is 2 to 3 times the size of the PG_i used in the reference DBD.

In table 2, we can see that the values of the probability P^V of fulfilling the read timing constraint have been computed. As expected, the values of P^V are very close to the required 99.87% (3 σ) for both the reference and proposed structures. Simultaneously, we observe a reduction in the read

timing margin $\Delta\mu_D/\mu_{Dref}$ lying between 14.5% to 25.2%.

Table 1: Variability reduction

Indji	Vdd (v)	T (c)	μ_A (ps)	V_A (%)	μ_{Bref} (ps)	V_{Bref} (%)	μ_{Bprop} (ps)	V_{Bprop} (%)	$\Delta V_B/V_{Bref}$ (%)
1	1.00	-40	1919	9.0	2412	10.8	2303	9.8	93
		125	2209	7.7	2497	9.1	2453	6.8	24.7
1,2	1.08	-40	1534	7.5	1793	8.5	1734	7.5	11.7
		125	1827	6.7	2029	7.7	1998	6.0	22.6
1,2,3	1.20	-40	1181	6.0	1306	6.9	1277	5.9	13.7
		125	1448	5.7	1585	6.5	1565	5.2	19.8
1,2,3,4	1.32	-40	963	5.2	1043	5.4	1023	5.1	5.8
		125	1200	5.0	1315	5.5	1289	4.8	13.3

Table 2: Reduction of read timing margin

Indji	Vdd (v)	T (c)	μ_A (ps)	μ_{Bref} (ps)	P_{ref} (%)	μ_{Bprop} (ps)	P_{prop} (%)	$\Delta\mu_D/\mu_{Dref}$ (%)
1	1.00	-40	1919	493	99.99	384	100.00	222
		125	2209	288	99.81	244	99.97	15.4
1,2	1.08	-40	1534	259	99.99	201	99.99	225
		125	1827	201	99.86	170	99.95	15.5
1,2,3	1.20	-40	1181	125	99.94	96	99.88	232
		125	1448	137	99.91	117	99.96	14.5
1,2,3,4	1.32	-40	963	80	99.97	60	99.67	252
		125	1200	115	99.99	89	99.97	22.7

6 Conclusion

Due to the pessimism of corner analysis method, we have proposed a simple way of computing the required read timing margin and of calculating the probability of meeting this constraint. A statistical optimization method has also been developed to ensure a predefined timing yield. The developed design approach has been particularly introduced to optimize the critical path of the SRAM memory, in which the dummy bitline driver has been replaced by a more robust structure to manufacturing process variations. Results have demonstrated that the use of the optimization method and the proposed dummy bitline driver improves significantly the reduction in the design timing margins, while ensuring a given timing yield.

Appendix

A.1) Estimation of design margin μ_D at $n\cdot\sigma_D$

Suppose that we want to have a read margin μ_D at $n\cdot\sigma_D$, such that:

$$\mu_D - n\cdot\sigma_D = 0 \quad (\text{A.1})$$

As we have seen previously in section 2, μ_D and σ_D can be defined using (1). Expression (A.1) can therefore be represented by the following equation:

$$(\mu_B - \mu_A) - n\sqrt{\sigma_A^2 + \sigma_B^2 - 2\cdot\sigma_A\cdot\sigma_B\cdot\rho} = 0 \quad (\text{A.2})$$

$$\text{Let } V_A = \frac{\sigma_A}{\mu_A} \text{ and } V_B = \frac{\sigma_B}{\mu_B} \quad (\text{A.3})$$

As the delay of signal B should be greater than that of signal A, delay μ_B in (A.2) becomes:

$$\mu_B = -\frac{a}{b} \left\{ \sqrt{1 - \frac{b\cdot c}{a^2}} + 1 \right\} \quad (\text{A.4})$$

$$a = n^2 \cdot V_B \cdot \sigma_A \cdot \rho - \mu_A \quad b = 1 - n^2 \cdot V_B^2 \quad c = \mu_A^2 - n^2 \cdot \sigma_A^2$$

Once the delay of μ_B is computed, the required design margin μ_D^{Yield} in (1) is given by:

$$\mu_D^{Yield} = -\frac{a}{b} \left\{ \sqrt{1 - \frac{b\cdot c}{a^2}} + 1 \right\} - \mu_A \quad (\text{A.5})$$

A.2) Identification of critical condition (V, T)_{crit}

The probability P^V of fulfilling a timing constraint is given by (2). In fact, since (V, T)_{crit} represents the condition showing the highest probability of the occurrence of a timing constraint violation P^V should be minimum at this condition. Thus, P^V is minimum if

$$\text{expression } \frac{\mu_D^2}{\sigma_D^2} \text{ is also minimum} \quad (\text{A.6})$$

$$\text{Let } \alpha = \frac{\sigma_A}{\mu_A} \approx \frac{\sigma_B}{\mu_B} \quad (\text{A.7})$$

By substituting both σ_D by (1) and σ_A and σ_B by (A.7) in (A.6), expression (A.6) can be represented by:

$$\frac{\mu_D^2}{\alpha \sqrt{1 + \frac{2\mu_A\mu_B}{(\mu_B - \mu_A)^2} (1 - \rho)}} \quad (\text{A.8})$$

It can be clearly seen that expression (A.8) is minimum if expression $\frac{\mu_A\mu_B}{(\mu_B - \mu_A)^2}$ is the largest.

7 References

- [1] O. S. Unsal, J. W. Tschanz, K. Bowman, V. De, X. Vera, A. Gonzalez, and O. Ergin, "Impact of Parameter Variations on Circuits and Microarchitecture," *IEEE Computer Society*, vol.26, no. 6, pp 30-39 (2006).
- [2] M. Orshansky, L. Milor, and C. Hu, "Characterization of Spatial Intrafield gate CD variability, its impact on Circuit performance, and Spatial Mask-level Correction," *IEEE Transactions on Semiconductor Manufacturing*, vol. 17, no. 1, pp 2-11 (2004).
- [3] S. Mukhopadhyay and K. Roy, "Modeling of Failure Probability and Statistical Design of SRAM Array for Yield Enhancement in Nanoscaled CMOS," *IEEE Trans. Comput.-Aided Des. Integr. Circuits Syst.*, vol. 24, no. 12, pp 1859-1880 (2005).
- [4] J. A. Croon, G. Storms, S. Winkelmeier, I. Pollentier, M. Ercken, S. Decoutere, W. Sansen, and H.E. Maes, "Line Edge Roughness: Characterization, Modeling and Impact on Device Behavior," *Proc Electron Devices Meeting*, pp 307-310 (2002).
- [5] B. S. Amrutur, M. A. Horowitz, "A Replica Technique for Wordline and Sense Control in Low Power SRAMs," *IEEE Journal of Solid State Circuits*, vol. 33, no. 8, pp 1208-1219 (1998).