Design of FinFET SRAM Cells using a Statistical Compact Model

Darsen D. Lu[1], Chung-Hsun Lin[1], Shijing Yao[1], Weize Xiong[2], Florian Bauer[3],
Cloves R. Cleavelin[2], Ali M. Niknejad[1] and Chenming Hu[1]
[1]Department of Electrical Engineering and Computer Sciences, University of California, Berkeley, CA 94720 USA
[2]Texas Instruments Inc., Dallas, TX USA
[3]Infineon Technologies, Munich, Germany, and Technical University Munich, Germany
Tel: +1-510-642-1010, E-mail: darsen@eecs.berkeley.edu , Addr: 205 Cory Hall #1722, Berkeley, CA 94720

***Abstract*— A study of designing FinFET-based SRAM cells using a compact model is reported. Parameters for a multi-gate FET compact model, BSIM-MG are extracted from fabricated n-type and p-type SOI FinFETs. Local mismatch in gate length and fin width is calibrated to electrical measurements of 378 FinFET SRAM cells. The cell design is re-optimized through Monte Carlo statistical simulations. Variation in readability, writability and static leakage of the cell are studied.**

I. Introduction

Variability has become increasingly troubling as the SRAM cell size is reduced. In the FinFET [1], short channel effects are suppressed by a thin body instead of channel dopants. With a lightly-doped fin as the channel, V_{th} variation due to random dopants is minimized. Reduced variability of FinFET SRAM cells has been demonstrated [2]. FinFET SRAM cell design has been studied with mixed mode TCAD simulations [3]. However, such simulation is very time consuming. A compact model is more suitable when a large number of simulations must be performed (e.g. Monte Carlo simulations).

In this work, a multi-gate compact model BSIM-MG [4] with calibrated variability is used to design a 6T FinFET SRAM cell. Model parameters covering all gate length devices are extracted from measured FinFET I-V. Transistor mismatch within the cell is determined by comparing Monte Carlo simulation with measurement distributions. The model is used to perform optimization for the number of fins in the pull-down devices (NF_{PD}) and gate lengths (L_G) of each transistor. Monte Carlo simulations are performed to study the contribution of each transistor to cell variation.

II. Compact Model Review

BSIM-MG [4] is a physics-based compact model for multi-gate MOSFETs. The core drain current expression is based on the analytical solution of an ideal double-gate MOSFET. Real device effects such as short channel effects, quantum mechanical effects, and the effect of the top gate in the tri-gate FinFET (Fig. 1(b)) are considered. The model is implemented in Verilog-A and can be easily simulated in SPICE simulators.

III. Hardward Calibration

A. Device Fabrication

FinFETs with 60nm fin height (H_{FIN}), 30nm fin width (T_{FIN}), SiON gate dielectric with 1.9nm equivalent oxide thickness (*EOT*), 10nm TiN gate and lightly-doped channels are fabricated on SOI wafers [4][5]. Both stand-alone FinFETs and 6T FinFET SRAM cells are fabricated. The stand-alone devices have 20 fins in parallel; The SRAM cells use single-fin or double-fin devices. An SEM of the SRAM cells is shown in Fig. 2.

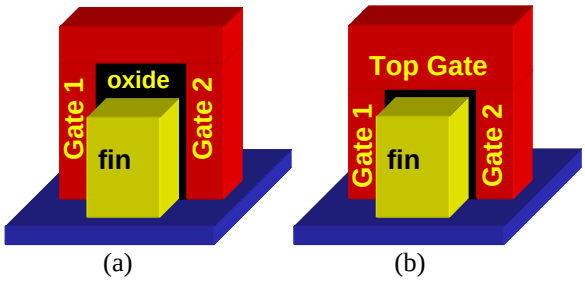


Fig. 1 Illustration of a (a) double-gate FinFET with a oxide hard mask on top of the fin and a (b) tri-gate FinFET without a hard mask.

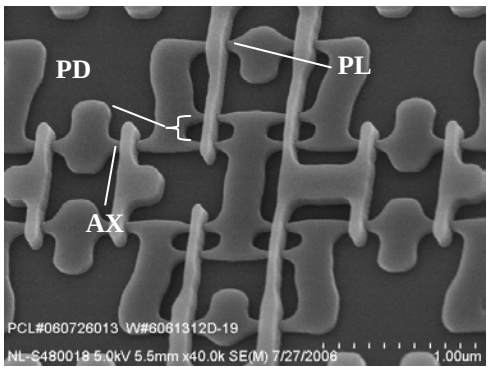
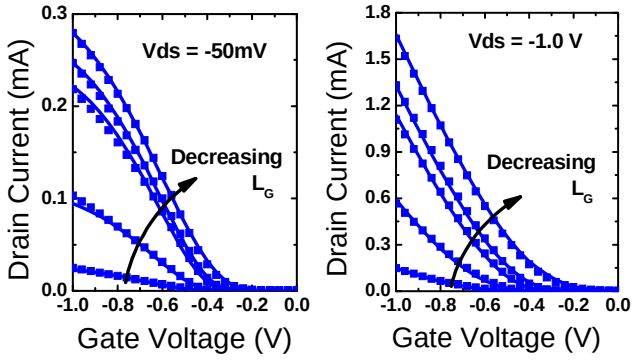


Fig. 2 SEM of a 6T FinFET SRAM (NF_{PD}=2) [5].

B. Nominal Parameter Extraction

The nominal parameters of BSIM-MG are extracted from I-V measurements of stand-alone FinFETs. One set of parameters is extracted from devices with L_G ranging from 75nm to 1μm. Although the binning methodology [6] is offered in BSIM-MG, it is not used in this study. As shown in Fig. 3, the drain current (I_d) versus gate voltage (V_{gs}) for p-type FinFETs in both linear (Fig. 3(a)) and saturation (Fig. 3(b)) modes are well-captured over the entire range of L_G.

This work was supported in part by the Semiconductor Research Corporation under task ID 1451.001.



(a) (b)

Fig. 3 I_d-V_{gs} of p-type FinFET devices at (a) V_{ds}=-50mV and (b) V_{ds}=-1.0V. L_G=75nm, 85nm, 95nm, 235nm and 1µm. Model (lines) and measured data (symbols) agree well.

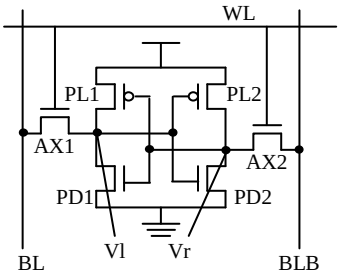


Fig. 4 Circuit schematic of a 6T FinFET SRAM cell.

TABLE I. GATE LENGTH CORRECTION

NF_{PD}	L_{AX} (nm)	L_{PD} (nm)	L_{PL} (nm)
1	+10	+15	+20
2	+0	+20	+20

C. *Adjustment for SRAM FETs*

FinFETs in SRAM cells and stand-alone FinFETs do not have identical physical dimensions and electrical characteristics due to the influence of neighboring patterns. To account for this, we simulate butterfly curves of the SRAM cell, compare it with measured ones, and adjust L_G to account for lithography variation. Fig. 4 illustrates the circuit schematic of the 6T FinFET SRAM cell. Fig. 5 shows the simulated and several measured butterfly curves on the same graph. The discrepancy of the uncorrected model may be caused by a modeled V_{th} value of the pull-down nFETs (PD1, PD2) that is too low. This difference is resolved through the correction of L_G (Table I, Fig. 5). Since only half-cell measurement is available in this study, the butterfly curves are obtained by measuring one curve and mirroring.

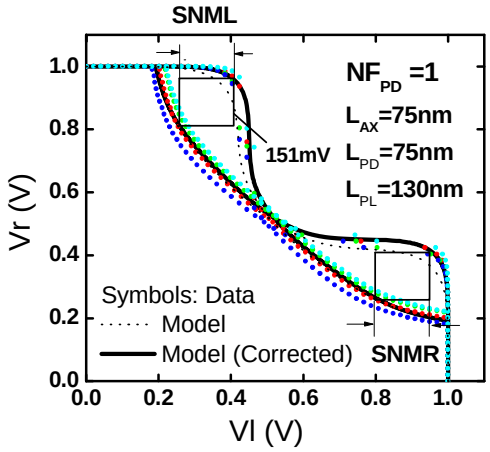
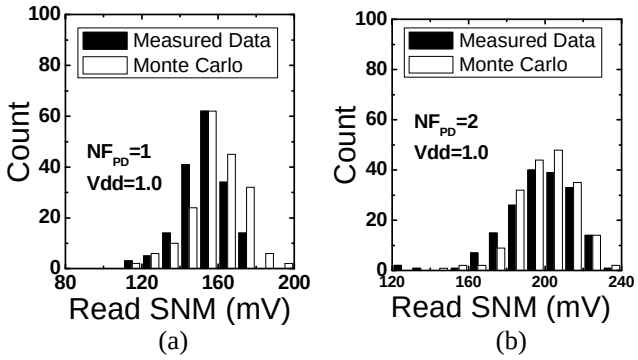


Fig. 5 Simulated and several measured butterfly curves of SRAM cells. (NF_{PD}=1)



(a) (b)

Fig. 6 Read SNM distribution of SRAM cells with pull-down nFETs containing (a) 1 fin (b) 2 fins. (V_{dd}=1.0V)

D. *Calibration of Variation*

To model variation in SRAM cells, we consider physical parameters such as L_G, H_{FIN}, T_{FIN} and *EOT*. Each is assumed to follow a Gaussian distribution. Fluctuation of channel doping is not considered since the fin is lightly-doped. Global variation is assumed to be 3σ=10% of the nominal value for each parameter. σ is the standard deviation. Local mismatch is determined by matching the Monte Carlo simulated read static noise margin (SNM) distribution to measurements. To introduce global variation in Monte Carlo, the same random component is shared by all transistors within the cell. (For example, H_{FIN} of all the transistors vary at the same time.) To model local mismatch, each transistor is assigned an independent and identically distributed random component. To properly describe the statistics, multi-fin transistors are modeled by separate devices connected in parallel with independent local mismatch components.

Butterfly curves of 378 SRAM cells are measured. 189 of the SRAM cells have NF_{PD}=1; the other 189 have NF_{PD}=2. The read SNM of the cells are extracted from the butterfly curves using the conventional method. The 3σ value of local mismatch is found to be 3.1nm for T_{FIN} and 12.6nm for L_G. We neglect the local variation in H_{FIN} and *EOT*, whose values are

not determined by lithography conditions. Fig. 6 shows the good agreement of Monte Carlo simulated read SNM distributions with measurements. A few SRAM cells show read SNM much lower than others (4 cells have 0V SNM).

IV. FinFET SRAM Cell Design

In this section, a variation-aware procedure of designing a FinFET SRAM cell is presented. L_G of each transistor and NF_{PD} are optimized under the constraint that both the read margin and the write margin must satisfy

$$p_{fail} < \Phi(-5.5) \approx 1.9\times10^{-8} \quad (1)$$

where p_{fail} is the read (write) failure probability of a given cell. $\Phi(x)$ is the cumulative distribution function of a standard Gaussian distribution.

A. Design criterion for read and write operation

The word line sweeping write margin (WLWM) follows a Gaussian distribution [7]. Therefore (1) translates to the widely-used criterion for the mean (μ) and σ of WLWM:

$$\mu_{WLWM} - 5.5\sigma_{WLWM} > 0. \quad (2)$$

(2) is adopted as the design criterion for write operation. For read operation, both SNML and SNMR (defined in Fig. 5) are Gaussian but

$$read\ SNM = min(SNML, SNMR)$$

is not [7]. Therefore (2) can not be directly applied to the read SNM. However, we observe that

$$p_{fail} < \Phi(-5.5) = 1-[1-\Phi(-z)]^2. \quad (3)$$

where $\Phi(-z)$ is the probability that SNML is less than zero. Solving (3), we obtain z=5.62. Therefore

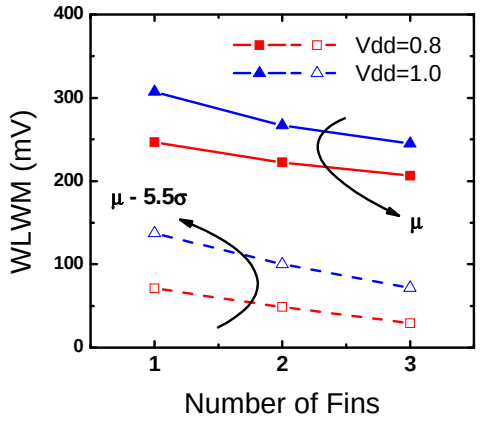
$$\mu_{SNML} - 5.62\sigma_{SNML} > 0 \quad (4)$$

is used as the design criterion for read operation.

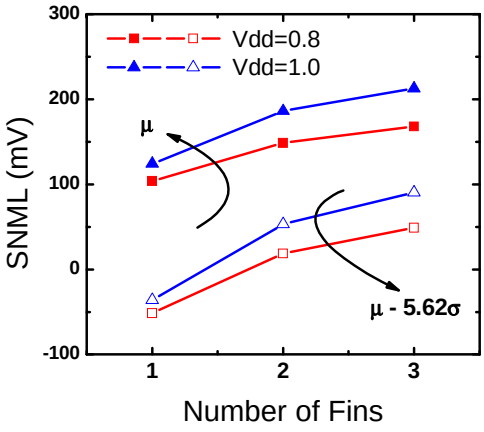
B. Cell Optimization

We first study the effect of changing NF_{PD} and V_{dd}. NF_{PD} is varied from 1 to 3 at V_{dd}=0.8 and V_{dd}=1.0. 1000 Monte Carlo circuit simulations are performed for each combination of NF_{PD} and V_{dd} (Fig. 7). The strength of the pull-down nFET increases with NF_{PD}. Therefore with increasing NF_{PD}, SNML is improved and WLWM is slightly degraded. At NF_{PD}=1, SNML does not satisfy the design criterion (4). However, this will be overcome through further optimization of L_G. V_{dd}=0.8 is chosen for low power operation.

Next L_G of the access transistor and the pFET load are optimized for the two cases: NF_{PD}=1 and NF_{PD}=2. NF_{PD}=3 is not considered since both SNML and WLWM constraints are satisfied at NF_{PD}=2 with a smaller cell area. For NF_{PD}=2, L_G of the access transistor is chosen to be the minimum value (75nm) since the design constraints are already satisfied (Fig. 7). For NF_{PD}=1, we vary L_G of the access transistor from 75nm to 105nm and perform Monte Carlo simulations (Fig. 8). The



(a)



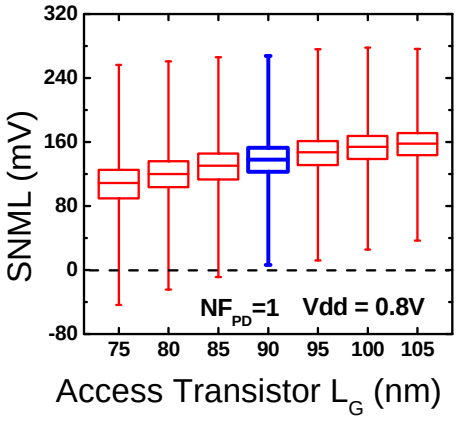
(b)

Fig. 7 (a) Word line sweep write margin (WLWM) and (b) SNML (defined in Fig. 5) versus fin number of pull-down nFETs for V_{dd}=0.8 and V_{dd}=1.0. (L_{NA}=75nm, L_{PD}=75nm, L_{PL}=130nm)

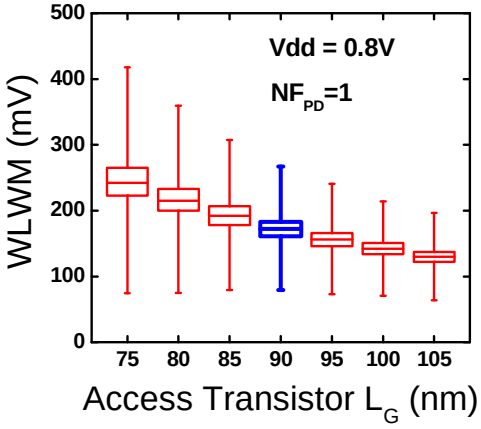
minimum access transistor L_G that satisfy the SNML constraint is 90nm (Fig. 8(a)). At 90nm the WLWM constraint is also satisfied (Fig. 8(b)). Similar optimization is performed for L_G of the pFET load for both NF_{PD}=1 and NF_{PD}=2. Table II summarizes the optimization results. The cell area is estimated according to the 65nm design rule [5]. When NF_{PD}=1 the cell area is about 30% smaller due to the smaller number of fins.

V. Discussion

To further analyze the optimized cell, Monte Carlo simulation is performed with local mismatch added to one pair of transistors at a time. Global variation is switched off. Fig. 9(a) shows the contribution of each transistor to read SNM variation. The pull-down nFET has the largest contribution. Therefore, increasing L_G of the pull-down nFET may be another option to reduce variability. Fig. 9(b) shows that the variation of WLWM is primarily due to access transistor variation. This is reflected in Fig. 8(b), where we see a strong L_G dependence of WLWM variation. The static leakage power



(a)



(b)

Fig. 8 (a) SNML (defined in Fig. 5) versus access transistor gate length. Whiskers mark $\mu \pm 5.62\sigma$. (b) word line sweeping write margin (WLWM) versus access transistor gate length. Whiskers mark $\mu \pm 5.5\sigma$. (V_{dd}=0.8, 1 fin pull-down nFET)

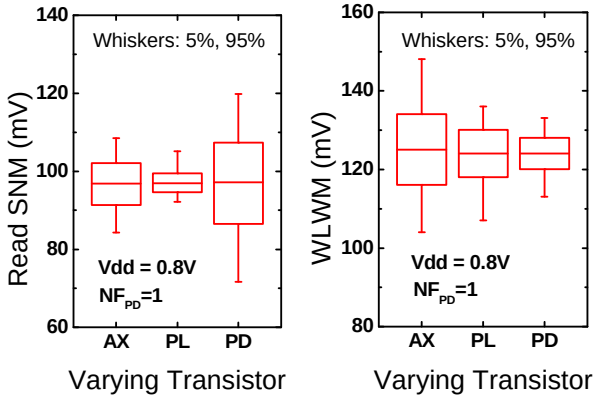
TABLE II. GATE LENGTH OPTIMIZATION RESULT

NF_{PD}	L_{PD}(nm)	L_{AX} (nm)	L_{PL} (nm)	Area (μm^2)
1	75	90	90	0.702
2	75	75	95	1.027

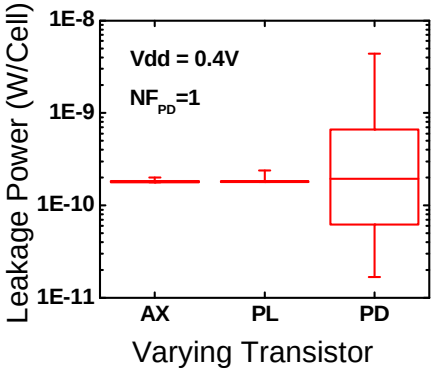
is dominated by the pull down nFET. Therefore it has the largest contribution to leakage variation (Fig. 9(c)).

CONCLUSION

Multi-gate compact model BSIM-MG is calibrated to I-V measurements of stand-alone FinFET devices and 6T FinFET SRAM cells. Variation in the SRAM cell is determined by calibrating to measured read static noise margin data. The cell is re-designed through a variation-aware procedure. Read margin, write margin and static leakage of the cell are studied through Monte Carlo circuit simulations.



(a) (b)



(c)

Fig. 9 Impact of access (AX), pFET load (PL), and pull-down (PD) device variation on (a) read static noise margin (V_{dd}=0.8V), (b) word line sweeping write margin (WLWM) (V_{dd}=0.8V) and (c) static leakage power per cell (V_{dd}=0.4V). (Whiskers mark the 5% and 95% quantiles. L_{AX}=90nm, L_{PL}=90m, L_{PD}=75nm)

ACKNOWLEDGMENT

The authors would like to thank IMEC, Leuven, Belgium and ATDF, Austin, Texas, USA for testchip fabrication.

REFERENCES

[1] X. Huang et al., "Sub 50-nm FinFET: PMOS," *IEDM Tech. Dig.*, pp. 67-70, 1999.

[2] H. Kawasaki et al., "Embedded Bulk FinFET SRAM Cell Technology with Planar FET Peripheral Circuit for hp32 nm node and beyond," *VLSI Symp. Tech. Dig.*, pp. 70-71, 2006.

[3] Z. Guo, S. Balasubramanian, R. Zlatanovici, T.-J. King and B. Nikolic, "FinFET-Based SRAM Design," *ISLPED*, pp. 2-7, 2005.

[4] M. V. Dunga et al, "BSIM-MG: A Versatile Multi-Gate FET Model for Mixed-Signal Design," *VLSI Symp. Tech. Dig.*, pp. 60-61, 2007.

[5] F. Bauer et al., "Layout Options for Stability Tuning of SRAM Cells in Multi-Gate-FET Technologies," *ESSCIRC*, pp. 392-395, 2007.

[6] Y. Cheng and C. Hu, MOSFET Modeling and BSIM3 User's Guide, Springer, 1999.

[7] H. Yamauchi, Variation-Tolerant SRAM Circuit Designs, *ISSCC Tutorial*, 2009.