Width Quantization Aware FinFET Circuit Design

Jie Gu, John Keane, Sachin Sapatnekar, and Chris Kim

University of Minnesota, Minneapolis

***Abstract*- This paper presents a statistical leakage estimation method for FinFET devices considering the unique width quantization property. Monte Carlo simulations show that the conventional approach underestimates the average leakage current of FinFET devices by as much as 43% while the proposed approach gives a precise estimation with an error less than 5%. Design example on dynamic logic circuits shows the effectiveness of the proposed method.**

I. INTRODUCTION

Double-gate FinFET transistors are recognized as one of the most promising successors of traditional planar bulk devices in the sub-25nm regime due to the significantly reduced leakage current, excellent short channel behavior, and compatible fabrication process with existing SOI or bulk technology [1]. FinFETs utilize a fin-shaped body perpendicular to the wafer surface to carry the current. The fin is sandwiched between the front gate and back gate, and is made very thin to geometrically suppress the short channel effect. Much of the previous work on FinFET devices has been done at the device and process level [1]. At the CAD and circuit level, only few researchers have looked into the FinFET design issues. For example, Joshi et al. compared FinFET-based SRAM cells with planar PD-SOI cells showing reduced delay, less standby power, and acceptable change in read stability due to width quantization [2].

One of the major differences between a FinFET device and a planar device is the fact that the FinFET device consists of multiple small unit fins. This unique width quantization property in FinFETs comes from the constant fin height constraint [3]. In this work, we reveal that width quantization of FinFETs has a large impact on the device leakage distribution under random V_T variations. It is shown that conventional leakage estimation methods significantly underestimate the FinFET leakage current which can potentially lead to chip failure due to insufficient noise margin, inaccurate full chip power estimates, and improper guidelines for leakage-sensitive circuits. This paper proposes a sophisticated mathematical derivation for accurately predicting the leakage current distribution of width-quantized FinFET devices. A real design example is provided to show the importance of using a correct leakage model when designing FinFET circuits. To the best of our knowledge, this paper is the first to analyze, model, and utilize the leakage distribution of FinFET devices considering the width quantization property.

II. FINFET DEVICE MODEL

Fig. 1(a) shows a 3-dimensional FinFET device with a thin fin structure as the body. Double-gate FinFETs have a front and back inversion channel and therefore the effective transistor width of a single fin is twice the fin height; i.e. $W_{fin}=2H$. The body thickness T_{Si} is made extremely thin so that short channel effect is suppressed and the subthreshold leakage is reduced via the improved subthreshold swing [4]. Fig. 1(c) shows the cross section of the FinFET model designed using the Taurus device simulator [5]. The FinFET model has a symmetrical front and back metal gate which are tied together. Table 1 lists the device parameters of the designed FinFET model used throughout this paper.

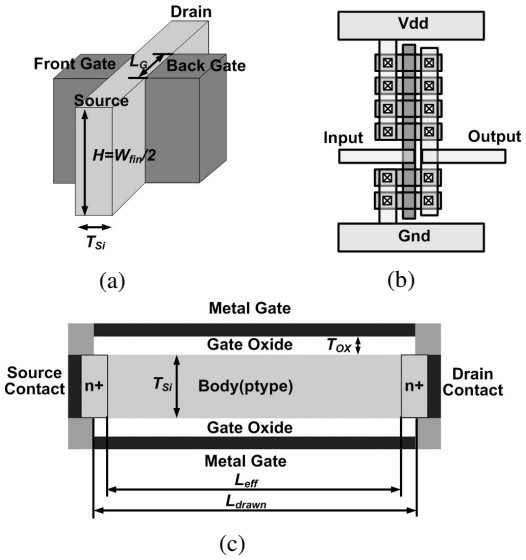


Fig. 1. (a) 3D structure of a FinFET device (b) layout example of a width-quantized FinFET inverter (c) cross section of a 21nm FinFET model designed in Taurus.

Table 1 Device parameters of the Taurus FinFET model

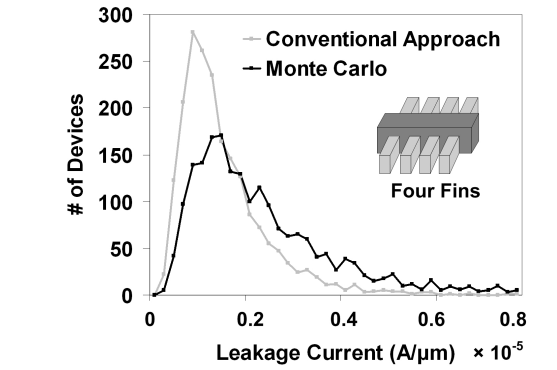
Device Parameters	Values
Drawn Channel Length L_{drawn}	25nm
Effective Channel Length L_{eff}	21nm
Oxide Thickness T_{ox}	14Å
Body Thickness T_{Si}	5nm
Device Height H	30nm
Vdd	0.8V
V_T	0.22V
Temperature	110°C
Subthreshold Swing S at 110°C	83mV/dec

III. IMPACT OF WIDTH QUANTIZATION ON FINFET LEAKAGE DISTRIBUTION

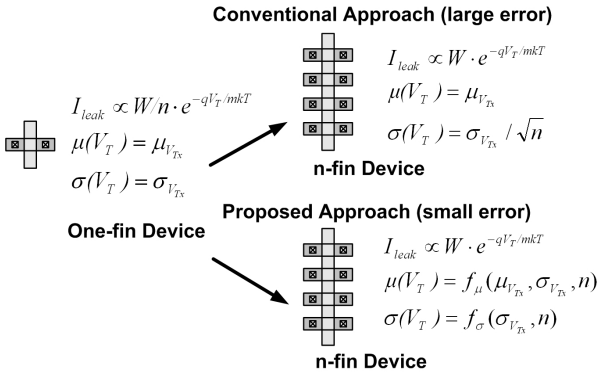
Width quantization of FinFET occurs from the fact that every fin has to have an equal height (H) due to process restrictions [3]. As a result, a FinFET device with a large width has to be discretized into multiple minimum unit fins. Fig. 1(b) shows a layout example of a FinFET inverter whose pull-up and pull-down are both quantized into smaller unit fins. Given the mean μ and standard deviation σ of the single fin V_T, conventional approaches [6] estimate the V_T and leakage distribution of a multi fin device assuming a fixed mean V_T value and a σ which is inversely proportional to the square root of the device area (or the number of fins in FinFET) as:

$$\sigma_{V_T} \propto \frac{1}{\sqrt{WL_{eff}}} = \frac{\sigma_{V_{Tx}}}{\sqrt{n}} \quad (1)$$

Here, $\sigma_{V_{Tx}}$ is the standard deviation of a single fin V_T. The conventional approach fails to accurately estimate the statistical



(a)



(b)

Fig. 2. (a) Leakage distribution of a 4-fin device based on conventional approach and Monte Carlo simulation showing large error exists in conventional leakage estimation approach. (b) Equations used in conventional and proposed leakage estimation approach for n-fin devices.

leakage distribution since they assume that the average of the multi fin device leakage is simply a multiple of the single fin leakage by the number of fins [6].

Fig. 2(a) shows the difference between conventional leakage estimation and the Monte Carlo simulation which serves as the golden result in this example. Conventional approach shows large error in leakage estimation and more importantly it underestimates the leakage value leading to the potential failure of design target, such as power budget and noise margin. As will be shown later, this error becomes even more pronounced in larger devices consisting of more number of fins and with larger V_T variation. Essentially, this error happens because the conventional approach does not capture the exponential relationship between leakage and V_T; i.e. a leaky fin overweighs an unleaky fin in determining the overall device leakage current. This paper proposes a precise model for leakage estimation where both μ and σ of the "effective" V_T value are functions of the number of fins. Our model shows a much smaller estimation error and can help optimize the design of width quantized FinFET circuits. Fig. 2(b) summarizes the key modeling difference between the conventional and proposed approach.

IV. STATISTICAL LEAKAGE VARIATION UNDER WIDTH QUANTIZATION FOR FINFET DEVICES

In a width-quantized FinFET device, the total leakage of an *n*-fin device is the sum of the leakage currents of each unit fin. Hence it can be expressed as the sum of lognormal terms as shown in (2),

$$I_{leak} = \sum_{i=1}^{n} C \frac{W}{n} e^{-\frac{qV_{Ti}}{mkT}} = \sum_{i=1}^{n} C \frac{W}{n} e^{-BV_{Ti}} \quad (2)$$

where W is the total width of the FinFET device, T is the temperature, m is the body effect coefficient, and C is a technology parameter. q/mkT is referred to as constant B for simplicity. In equation (2), the threshold voltage V_{Ti} changes due to factors such as channel length variation and random dopant fluctuation. V_{Ti} for each unit fin can be considered as a Gaussian random variable with a certain correlation between each other. Equation (2) indicates that the leakage of a large FinFET device can be expressed as a sum of lognormals. Although a closed form expression for a sum of lognormals does not exist, Wilkinson's method [7] provides a simple expression for modeling the sum of lognormals. In Wilkinson's approach, a sum of lognormals $\sum_{i=1}^{n} \frac{W}{n} e^{x_i}$ can be approximated as another lognormal We^y where y is a new Gaussian variable with a calculable mean and standard deviation. This approximation is completed by matching the first and second moment of both equations. Let (m_{x_i}, σ_{x_i}) and (m_{y_i}, σ_{y_i}) be the mean and standard deviation of the original Gaussian variables x_i and the new Gaussian variable y of the lognormal functions, respectively.

Let r_{ij} be the correlation coefficient of each random variable and n be the number of fins in a device. By equating the first two moments of the original lognormal equation and the new lognormal equation, we get:

$$u_1 = E(S) = \sum_{i=1}^{n} \frac{1}{n} e^{m_{x_i} + \sigma_{x_i}^2/2} = e^{m_y + \sigma_y^2/2} \quad (3)$$

$$u_2 = E(S^2) = \frac{1}{n^2} \left(\sum_{i=1}^{n} e^{2m_{x_i} + 2\sigma_{x_i}^2} + 2 \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} e^{m_{x_i} + m_{x_j}} e^{(\sigma_{x_i}^2 + \sigma_{x_j}^2 + 2r_{ij}\sigma_{x_i}\sigma_{x_j})/2} \right)$$

$$= e^{2m_y + 2\sigma_y^2}$$

Solving (3), we obtain the mean and standard deviation of the new Gaussian variable y as:

$$m_y = 2\ln u_1 - 1/2 \ln u_2$$
$$\sigma_y^2 = \ln u_2 - 2\ln u_1 \quad (4)$$

In a FinFET device, it is fair to assume every fin has the same mean and variance of V_T, and the same correlation between each other. The mean and standard deviation of the first two moments u_1 and u_2 then become:

$$u_1 = e^{m_x + \sigma_x^2/2}$$
$$u_2 = \frac{1}{n} e^{2m_x + \sigma_x^2} \left(e^{\sigma_x^2} + (n-1) \cdot e^{r\sigma_x^2} \right) \quad (5)$$

Substituting equation (5) into (4) gives the mean and standard deviation of the new Gaussian variable in the lognormal equation as:

$$m_y = m_x + \frac{1}{2}\Delta$$
$$\sigma_y^2 = \sigma_x^2 - \Delta \quad (6)$$

where $\Delta = \sigma_x^2 - \ln\left(\frac{e^{\sigma_x^2} + (n-1) \cdot e^{r\sigma_x^2}}{n}\right)$ and r is the correlation coefficient of V_T between fins. The change in the average and standard deviation of y is expressed using a single parameter Δ. It can be easily shown that Δ is a non-negative number.

$$\Delta = \sigma_x^2 - \ln(\frac{e^{\sigma_x^2} + (n-1)\cdot e^{r\sigma_x^2}}{n}) = \ln e^{\sigma_x^2} - \ln(\frac{e^{\sigma_x^2} + (n-1)\cdot e^{r\sigma_x^2}}{n})$$

$$= \ln(\frac{ne^{\sigma_x^2}}{e^{\sigma_x^2} + (n-1)\cdot e^{\sigma_x^2}}) = \ln(\frac{e^{\sigma_x^2} + (n-1)e^{\sigma_x^2}}{e^{\sigma_x^2} + (n-1)e^{r\sigma_x^2}}) \geq 0$$

$$(for\ \ n > 1,\ 0 \leq r \leq 1)$$

Finally, the average and standard deviation of the new equivalent V_T can be derived from (6) by including the constant B defined in the subthreshold current equation (2).

$$\mu_{V_{Ty}} = \mu_{V_{Tx}} - \frac{1}{2}\Delta / B \qquad (7)$$
$$\sigma^2_{V_{Ty}} = \sigma^2_{V_{Tx}} - \Delta / B^2 \qquad (\Delta \geq 0)$$

where $\Delta = B^2\sigma^2_{V_{Tx}} - \ln(\frac{e^{B^2\sigma^2_{V_{Tx}}} + (n-1)\cdot e^{rB^2\sigma^2_{V_{Tx}}}}{n}) \geq 0$.

Here, V_{Ty} denotes the threshold voltage of an effective large single-fin device and V_{Tx} denotes the threshold voltage of the original single fin. This can be understood from the following relationship.

$$I_{leak} = CWe^{-\frac{qV_{Ty}}{mkT}} = C\sum_{i=1}^{n}\frac{W}{n}e^{-\frac{qV_{Tx_i}}{mkT}} \qquad (8)$$

In the rest of the paper, we will refer to V_{Ty} as the "effective threshold voltage". By introducing the effective threshold voltage concept, one can efficiently find the leakage distribution of a width-quantized FinFET without running Monte Carlo simulations for n number of random variables.

The expression for the effective V_T in (7) reveals that the average of V_T is reduced compared to that of a single fin. The amount of change in the average is determined by a single non-negative parameter Δ. The standard deviation $\sigma_{V_{Ty}}$ also decreases with larger number of fins due to the Δ parameter in equation (7). The change of the standard deviation value agrees with the prediction from conventional approach based on Fig. 2(b) while the proposed approach shows a more accurate value as will be shown later. Equation (7) indicates that conventional leakage estimation approach that simply treats FinFETs as one single device without sophisticatedly calculating $\mu_{V_{Ty}}$ and $\sigma_{V_{Ty}}$ as shown in the equation will underestimate the leakage of width-quantized FinFET devices.

Experiments were carried out for cases with and without spatial correlation between the fins to show the accuracy of the new leakage estimation method. The leakage distribution is compared for the following three cases: (i) Monte Carlo simulation assuming random variation for each fin which offers the most realistic "golden" leakage distribution; (ii) conventional leakage estimation which is based on the equations in Fig. 2(b); and (iii) the new leakage estimation method using (7).

A. *Leakage distribution with no correlation between fins*

Fig. 3 shows the leakage distribution with correlation coefficient r=0 for different number of fins and different $\sigma_{V_{Tx}}$ values. Here $\sigma_{V_{Tx}}$ refers to the standard deviation of a single fin as defined earlier. The conventional method significantly underestimates the FinFET device leakage for all different cases. Table 2 summarizes the accuracy of the different leakage estimation approaches. For 4 fins and 20% V_{Tx} standard deviation ($\sigma_{V_{Tx}} / \mu_{V_{Tx}}$ =20%), the estimation error of average leakage value is 42.3% using the conventional approach. On the other hand, the proposed method based on equation (7) gives a precise estimation of the total leakage with an error of only 4.1%. The table also indicates that leakage estimation error from the conventional approach increases significantly as the number of fins and standard deviation increase while the error from proposed approach is below 5% under all conditions.

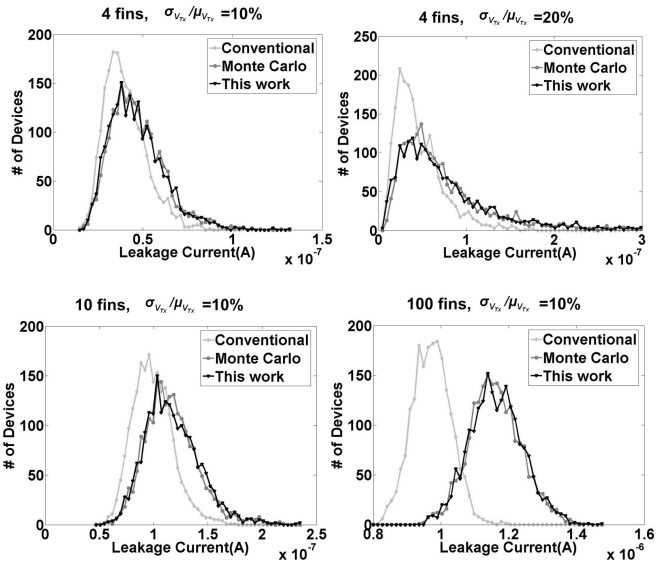


Fig. 3. Leakage distribution for various numbers of fins and different standard deviation of V_{Tx}.

B. *Leakage distribution with correlation between fins*

Fig. 4 shows the leakage distribution for different values of correlation coefficient r for the three test cases. The distribution when r=1 is also shown for comparison. The following conclusions can be drawn from the simulation results.

(i) In the presence of correlation, V_T is a weighted sum of the correlated component V_{Tcor} and the uncorrelated component V_{Tuncor}. As the value of r increases, the correlated component becomes dominant in the total leakage distribution. Therefore, as shown in the figure, when r increases from 0.1 to 0.4, the total distribution is becoming closer to a fully correlated case with r=1.

(ii) Even with correlation, the approach developed in this paper matches very well with the Monte Carlo simulation.

Table 2 Comparison of conventional and new FinFET leakage estimation techniques.

4 fins, $\sigma_{V_{Tx}} / \mu_{V_{Tx}}$ =10%		
	Conventional	This work
Error in leakage μ	13.4%	3.4%
Error in leakage σ	17.9%	3.5%
4 fins, $\sigma_{V_{Tx}} / \mu_{V_{Tx}}$ =20%		
	Conventional	This work
Error in leakage μ	42.3%	1.8%
Error in leakage σ	57.4%	4.1%
10 fins, $\sigma_{V_{Tx}} / \mu_{V_{Tx}}$ =10%		
	Conventional	This work
Error in leakage μ	14.5%	0.3%
Error in leakage σ	21.4%	1.1%
100 fins, $\sigma_{V_{Tx}} / \mu_{V_{Tx}}$ =10%		
	Conventional	This work
Error in leakage μ	16.4%	0.1%
Error in leakage σ	24.5%	0.1%

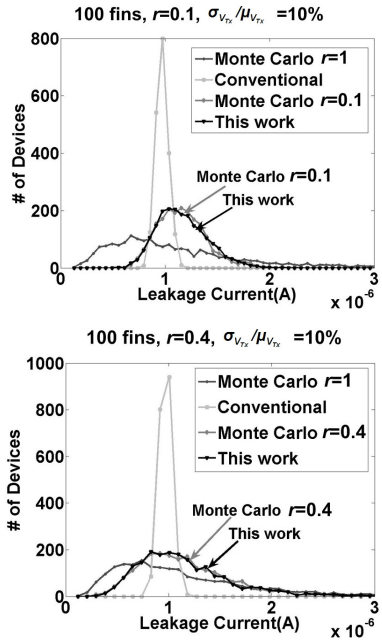


Fig. 4. Leakage distribution with various correlation coefficients.

(iii) In both correlated and uncorrelated cases, the conventional estimation exhibits large errors in leakage prediction because the width quantization effect has not been considered.

V. FINFET DYNAMIC CIRCUIT DESIGN

This section shows a design example of a FinFET dynamic circuit where the developed leakage estimation method can be beneficial. Static noise margin (SNM) is one of the primary design constraints in dynamic circuits in leakage dominant technologies. SNM is defined as the input DC noise level which will produce a certain amount of voltage droop (assumed 10% of Vdd in this paper) on the dynamic node. SNM of a dynamic circuit is directly related to the leakage current of the pull-down circuits. Under the presence of leakage, the dynamic node can collapse and result in a faulty evaluation.

Fig. 5 shows a schematic of a wide-OR domino circuit. To compensate the pull-down leakage current and replenish the charge loss, a static keeper is deployed to hold the voltage of the dynamic node. For optimal keeper sizing, the leakage current of the pull-down circuit has to be accurately modeled [8].

Table 3 shows the 3σ SNM values of a wide-OR dynamic circuit with a constant keeper size. Results are shown for the three different leakage estimation approaches. No spatial correlation is assumed in this experiment to represent the worst case situation. The

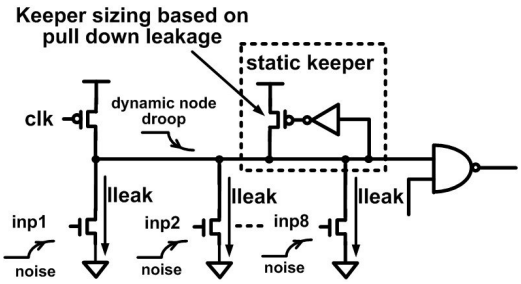


Fig. 5. Schematic of a dynamic circuit with static keeper for sufficient SNM. Accurate estimation of pull-down leakage using proposed method can assist designers in determining proper keeper sizes for FinFET circuits.

Table 3. $SNM_{3\sigma}$ predicted by different leakage estimation approaches. Conventional leakage estimation overestimates the SNM values.

# of fins=400	$SNM_{3\sigma}$(V)			Required keeper size-up
	Monte Carlo	Conv.	This work	
$\sigma_{V_{Tx}} / \mu_{V_{Tx}}$ =5%	0.129	0.132	0.129	5%
$\sigma_{V_{Tx}} / \mu_{V_{Tx}}$ =10%	0.122	0.130	0.122	20%
$\sigma_{V_{Tx}} / \mu_{V_{Tx}}$ =15%	0.112	0.128	0.112	50%

conventional approach overestimates the SNM of FinFET dynamic circuits by 2% to 14% which can lead to serious yield issues. The proposed method accurately predicts the SNM for a wide range of standard deviation. The last column presents how much the keeper size has to be increased to achieve the SNM value set by the conventional approach. The proposed leakage estimation technique predicts a 5% to 50% upsizing of the keeper to meet the target SNM.

VI. CONCLUSION

Double-gate FinFET devices are considered as one of the most promising successors for conventional MOSFET devices. Due to their physical fin structure, the width of a FinFET device is quantized. In this paper, we show that the impact of width quantization on leakage distribution is significant in FinFET devices. We developed a new leakage estimation model which can accurately capture the statistical characteristics of FINFET leakage current under process variation. Monte Carlo simulation has been used to prove the correctness of the proposed method. Simulation results show that the conventional approach for leakage estimation can significantly underestimate the leakage current by as much as 43% while the proposed approach gives an error of less than 5%. Design example on FinFET dynamic circuit shows that the keeper size has to be upsized by 5-50% to suppress the extra leakage due to width quantization.

REFERENCES

[1] J. Kedzierski, M. Ieong; T. Kanarsky, et al., "Fabrication of metal gated FinFETs through complete gate silicidation with Ni," *IEEE Transactions on Electron Devices*, vol. 51, no. 12, pp. 2115-2120, 2004.

[2] R.V. Joshi, R. Williams, E. Nowak, K. Kim, et al., "FinFET SRAM for high-performance low-power applications", *European Solid-State Device Research Conference*, pp. 69-72, 2004.

[3] E. J. Nowak, I. Aller, T. Ludwig, K. Kim, et al., "Turning silicon on its edge [double gate CMOS/FinFET technology]," *IEEE Circuits and Devices Magazine*, vol. 20, no.1, pp. 20-31, 2004.

[4] H. Wong, D. J. Frank, Y. Taur, M.C. Stork, "Design and performance considerations for sub-0.1 μm double-gate SOI MOSFET's," *International Electron Devices Meeting*, pp. 747-750, 1994.

[5] Taurus v. 2004. 12, Synopsys, Inc. , 2004.

[6] M. J. M. Pelgrom, A. C. J. Duinmaijer, and A. P. G. Welbers, "Matching properties of MOS transistors," *IEEE Journal of Solid-State Circuits*, vol. 24, no. 5, pp. 1433-1440, 1989.

[7] A. A. Abu-Dayya and N. C. Beaulieu, "Comparison of Methods of Computing Correlated Lognormal Sum Distributions and Outages for Digital Wireless Applications," *IEEE 44th Vehicular Technology Conference*, vol. 1, pp. 175-179, 1994.

[8] A. Alvandpour, P. Edefors and C. Svensson, "A Leakage Tolerant Multi-Phase Keeper for Wide Domino Circuit," *IEEE Intl. Conf. on Electronics, Circuits, and System*, pp. 209-212, 1999.