Statistical Leakage Estimation of Double Gate FinFET Devices Considering the Width Quantization Property

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

***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 subthreshold circuits shows the effectiveness of the proposed method.**

***Index Terms*— Leakage currents, IC design, Circuit modeling**

I. INTRODUCTION

As device feature sizes enter the nanometer regime, leakage power consumption in VLSI systems has become one of the main barriers to technology scaling. Precise modeling of leakage current under process variations is crucial for the proper estimation of leakage power consumption and the optimal design of leakage-sensitive circuits such as dynamic circuits, SRAM bitlines and subthreshold circuit [1, 2]. A significant amount of work has been published on the prediction of leakage power under process variation using statistical methods. For example, Chang developed models which consider the spatial correlations of inter-die and intra-die process parameters [3].

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 a fabrication process which is compatible with existing SOI or bulk technology processes [4]. 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. Conventional leakage estimation approach does not consider the width quantization property of FinFET. For example, Ananthan proposed a compact physical model to obtain the leakage distribution of FinFET devices due to gate length and body thickness variation [5]. Rao derived a mathematical equation to predict the mean and variance of full-chip leakage [6]. However, both of them extended the variation for an individual device into multiple devices by simple scaling which ignored the local mismatch and can lead to errors for multi-fin devices such as FinFETs. Srivastawa proposed a full-chip leakage estimation method using principle components approach and accounted for the local mismatch between devices [7]. However, the work requires very involved mathematic computation and is not specific for FinFET device where individual fin possesses identical statistical characteristics. This paper, instead, used an effective V_T to model the V_T change in a multiple-fin device and offered a much simpler solution for leakage prediction of FinFET device. Monte Carlo simulations show that the proposed approach significantly improved the accuracy of leakage estimation for FinFET device.

All simulations in this paper used a FinFET model developed by the Taurus device simulator [8]. Fig. 1 shows the cross section of the FinFET model and Table 1 summarized the device parameters used in this paper.

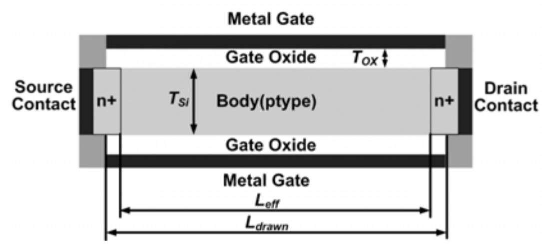

Fig. 1. Cross section of a 21nm FinFET model designed in Taurus.

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
V_{dd}	0.8V
V_T	0.22V

II. IMPACT OF WIDTH QUANTIZATION ON FINFET LEAKAGE ESTIMATION

Width quantization is a unique property of FinFET devices: a large single device always consists of multiple small unit fins. Given the mean μ and standard deviation σ of the single fin V_T, conventional approaches [9] estimate the V_T and leakage distribution of a multi-fin device assuming the same 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:

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

Here, $\sigma_{V_{Tx}}$ is the standard deviation of a single fin V_T.

Manuscript received September 4, 2006; Revised May 2, 2007. This work was supported in part by SRC under Grant xxxxx.

The authors (corresponding email: jiegu@ece.umn.edu) are with the Department of Electrical and Computer Engineering, University of Minnesota, Minneapolis, MN, U.S.A.

Fig. 2 shows the difference between conventional leakage estimation method and the Monte Carlo simulation which serves as the golden result in this example. The conventional approach shows a large error in leakage estimation and more importantly it underestimates the leakage value leading to the potential failure of meeting design targets, such as power budget and noise margin requirements. This error happens because the conventional approach does not capture the exponential relationship between leakage and V_T. As a solution, this paper proposes a precise model for leakage estimation where both μ and σ of the "effective" V_T are functions of the number of fins. The analytical derivation and the experimental results will be shown in the next few sections.

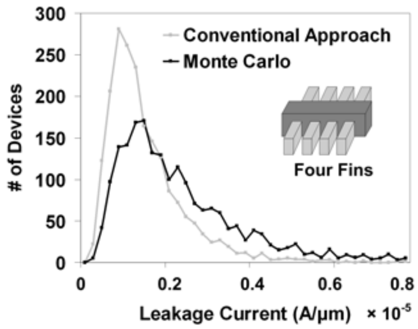

Fig. 2. Leakage distribution of a 4-fin device from the conventional estimation approach and Monte Carlo simulation (golden) showing a large discrepancy.

III. STATISTICAL LEAKAGE ESTIMATION UNDER WIDTH QUANTIZATION

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),

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

where W is the total width of the FinFET device, T is the temperature, m is the body effect coefficient, q is electron charge, k is Boltzmann's constant, and C is a technology parameter. q/mkT is referred as constant B for simplicity. The threshold voltage ($V_{Ti,}$) changes due to factors such as channel length variation and random dopant fluctuation (RDF). The threshold voltage of each fin can be modeled using correlated Gaussian random variables because:

(1) RDF introduces uncorrelated V_T variations because the device dopant concentration, which significantly influences the V_T value, can be random even for devices within a small area. Xiong et.al, showed that as device dimensions scale below 25nm, the number of dopant atoms per device becomes less than 100, and thus V_T can vary significantly due to the fluctuation in the number and placement of dopants [10]. Recently, Chiang shows that although undoped silicon is likely the material of choice for FinFET devices, even a single impurity atom randomly deposited in the channel region can lead to significant fluctuation in threshold voltage because of the ultra-thin body [11]. As a result, RDF will still remain as one of the major sources of variation in FinFETs.

(2) Process parameters such as channel length and fin height show a strong spatial correlation. Our analysis also considers the spatial correlation between the fins in order to develop a general leakage estimation framework that can be applicable for different variation sources.

As shown in equation (2), 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 provides a simple approximation for modeling the sum of lognormals [12]. In Wilkinson's approach, a sum of lognormals $\left(\sum_{i=1}^{n}\frac{W}{n}e^{x_i}\right)$ can be approximated as another lognormal $\left(We^{y}\right)$ 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:

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

$$\begin{aligned} u_2 = E(S^2) &= \frac{1}{n^2}\Big(\sum_{i=1}^{n} e^{2m_{x_i}+2\sigma_{x_i}^2} \\ &\quad + 2\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} e^{m_{x_i}+m_{x_j}} e^{(\sigma_{x_i}^2+\sigma_{x_j}^2+2r_{ij}\sigma_{x_i}\sigma_{x_j})/2}\Big) \\ &= e^{2m_y+2\sigma_y^2} \end{aligned} \tag{3}$$

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. Therefore, by solving equation (3), the mean and standard deviation of the new Gaussian variable in the lognormal equation is found as follows:

$$\begin{aligned} m_y &= m_x + \frac{1}{2}\Delta \\ \sigma_y^2 &= \sigma_x^2 - \Delta \end{aligned} \tag{4}$$

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

Reader can easily prove that Δ is a non-negative number.

Finally, the average and standard deviation of the new equivalent V_T can be derived from (4) by including the constant B defined in the subthreshold current equation (2).

$$\boxed{\begin{aligned} \mu_{V_{Ty}} &= \mu_{V_{Tx}} - \frac{1}{2}\Delta / B \\ \sigma_{V_{Ty}}^2 &= \sigma_{V_{Tx}}^2 - \Delta / B^2 \qquad (\Delta \geq 0) \end{aligned}} \tag{5}$$

where $\Delta = B^2\sigma_{V_{Tx}}^2 - \ln\left(\frac{e^{B^2\sigma_{V_{Tx}}^2} + (n-1)\cdot e^{rB^2\sigma_{V_{Tx}}^2}}{n}\right)$.

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:

$$We^{-\frac{qV_{Ty}}{mkT}} = \sum_{i=1}^{n}\frac{W}{n}e^{-\frac{qV_{Tx_i}}{mkT}} \tag{6}$$

In the rest of the paper, we will refer to V_{Ty} as the "effective threshold voltage". By introducing the effective threshold voltage concept, we can efficiently find the leakage distribution of a width quantized FinFET without having to run Monte Carlo simulations for *n* number of random variables. The expression for the effective V_T in (5) 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 the larger number of fins due to the Δ parameter in equation (5). Simulations were run for the cases with and without spatial correlation between the fins. 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 considers a FinFET device as a large single-fin device; and (iii) the new leakage estimation method using (5).

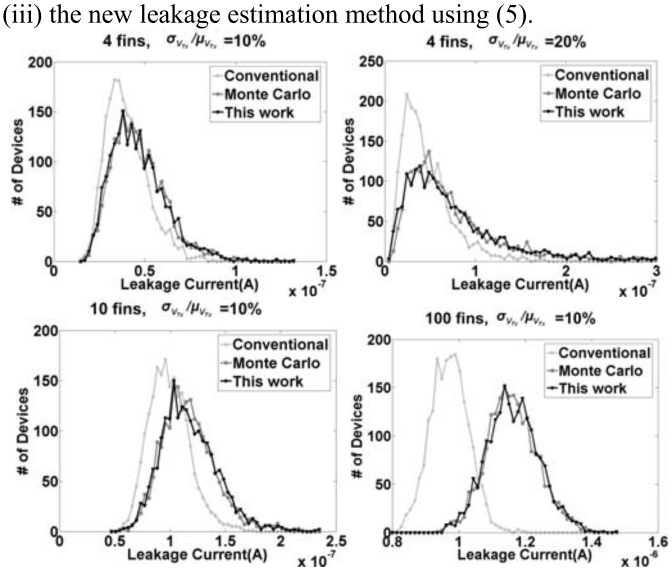

Fig. 3. Leakage distribution for various numbers of fins and different standard deviation of V_{Tx}.

A. Leakage distribution with no correlation between fins

Fig. 3 shows the leakage distribution with r = 0 for different numbers 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 cases. For four fins and 20% V_{Tx} standard deviation (i.e. $\sigma_{V_{Tx}} / \mu_{V_{Tx}}$ =20%), the error of the mean value of the estimated effective V_T using the conventional approach is about 12%. This corresponds to a 42.3% error in terms of the mean leakage current value. On the other hand, the proposed method based on equation (5) gives a precise estimation of the total leakage with an error of less than 5%.

Fig. 4 shows the changes of estimation errors in comparison between the conventional approach and proposed approach as we vary the number of fins and standard deviation of V_{Tx}. The figure shows the improvement of estimation accuracy using the proposed approach. Although the error in the conventional approach does not vary much with the increase in the number of fins, it grows quickly with the standard deviation $\sigma_{V_{Tx}}$. On the other hand, the error using the proposed approach is maintained at a very small value under the different conditions. We also notice that the error in the proposed approach becomes relatively large when $\sigma_{V_{Tx}} / \mu_{V_{Tx}}$ reaches 30% or higher. This is because the Wilkinson's method itself is no longer accurate when $\sigma_{V_{Tx}}$ becomes considerable [12]. In reality, $\sigma_{V_{Tx}} / \mu_{V_{Tx}}$ is less than 30% in most CMOS processes. Note that the leakage of a large number of independent fins will produce a Gaussian-like distribution curve from the central limit theorem [13]. That is consistent with our results shown in Fig. 3 when the fin number increases from 4 to 100. Unlike the central limit theorem, our leakage estimation approach can also handle the correlated situations as well.

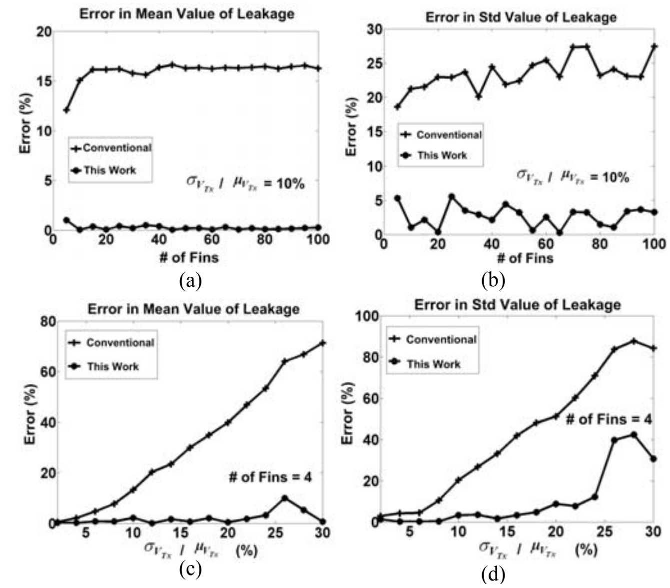

Fig. 4. Leakage estimation error compared between the conventional approach and the proposed approach. (a) Error in mean value of leakage for various numbers of fins, (b) error in std value of leakage for various numbers of fins, (c) error in mean value of leakage for various $\sigma_{V_{Tx}} / \mu_{V_{Tx}}$, (d) error in std value of leakage for various $\sigma_{V_{Tx}} / \mu_{V_{Tx}}$.

B. Leakage distribution with correlation between fins

Fig. 5 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 are drawn from the simulation results.

(i) Even with correlation, the approach developed in this paper matches very well with the Monte Carlo simulation. Note that in a fully correlated case where r=1, Δ in equation (5) becomes zero and the mean and standard deviation of the effective V_T becomes same as those of a unit fin. The leakage distribution is then simply a multiplication of the unit-fin leakage. As shown in equation (5) and (6), our approach is still accurate for this extreme case.

(ii) In presence of correlation, V_T is a weighted sum of the correlated component and the uncorrelated component. As 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 close to a fully correlated case with r=1.

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

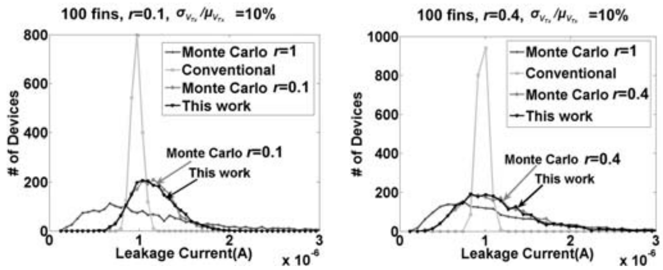

Fig. 5. Leakage distribution with various correlation coefficients.

IV. WIDTH QUANTIZATION AWARE CIRCUIT DESIGN

This section shows an example where the proposed approach can be beneficial for FinFET circuit design.

The supply voltage for minimum energy operation in low-to-medium-performance applications has been shown to lie in the subthreshold regime [14]. FinFET devices could prove to be especially useful in this design regime due to the decreased impact of short-channel effects and the improved subthreshold swing. However, in subthreshold designs, leakage current becomes the drive current, and therefore both current and delay are modeled as lognormal random variables that are exponentially dependent on the threshold voltage.

In order to approximate the delay through a chain of FinFET gates operating in the subthreshold regime, one must first sum the current distributions of the fins that comprise each individual gate. The delay through each gate, which is a function of the leakage computed in that first step, is subsequently summed to produce the final delay distribution. That is, due to width quantization, one must apply equations (5) twice in order to find the delay distribution of FinFET devices. The delay through n gates is computed as [14]:

$$delay_{inv\_chain} = \frac{1}{2}\eta C_S V_{dd} \sum_{j=1}^{n} \frac{1}{I_{on,j}} \quad (7)$$

where η is the delay factor due to a non-ideal input waveform, and C_s is the load capacitance driven by each gate. Fig. 7 shows the delay distribution of a FinFET inverter chain operated in subthreshold region. As seen in figure, the delay distribution predicted under our approach accurately matches the results of the Monte Carlo simulations. Because the overall leakage is increased after considering width quantization, the average delay is reduced by 5% compared to that from the conventional method. Basing design parameters on conventional calculations would lead to an over-design, which is especially detrimental in subthreshold designs, as this would lead to increased power consumption.

V. CONCLUSION

Double-gate FinFET devices are considered as one of the most promising successors of conventional MOSFET devices. Due to the physical fin structure, the width of a FinFET device is quantized. In this paper, we show that the impact of width quantization on statistical leakage estimation is significant for FinFET devices. We developed a new leakage estimation method which can accurately capture the statistical characteristics of leakage current under process variation. Monte Carlo simulation has been used to prove the accuracy of the proposed method. Simulation results show that the conventional approach for leakage estimation can significantly underestimate the average leakage current by as much as 43% while the proposed approach gives an error of less than 5%. A design example is provided on how to apply our leakage estimation approach to a subthreshold FinFET circuit. The result shows that the average delay is 5% less than what is expected using the conventional technique because of the width quantization property.

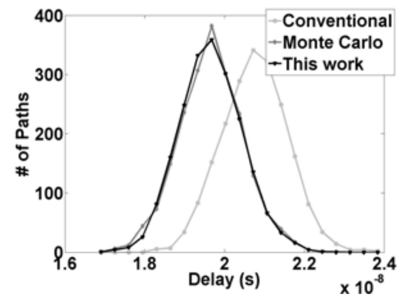

Fig. 7. Delay distribution of an inverter chain operating in the subthreshold region. A total of 80 fins with 10% standard deviation of V_{Tx} are used. (V_{dd}=0.2V)

REFERENCES

[1] C. H. Kim, K. Roy, S. Hsu, et al., "A Process Variation Compensating Technique for Sub-90nm Dynamic Circuits," *Symposium on VLSI Circuits*, pp.205-206, 2003.

[2] S. Borkar, T. Karnik, S. Narendra, et al., “Parameter Variations and Impact on Circuits and Microarchitecture,” *Design Automation Conference*, pp. 338-342, 2003.

[3] H. Chang and S. S. Sapatnekar, “Full-Chip Analysis of Leakage Power under Process Variations, Including Spatial Correlations,” *Design Automation Conference*, pp. 523-529, 2005.

[4] E. J. Nowak, 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.

[5] H. Ananthan and K. Roy, “A Fully Physical Model for Leakage Distribution under Process Variations in Nanoscale Double-gate CMOS,” *Design Automation Conference*, pp. 24-28, 2006.

[6] R. R. Rao, A. Devgan, D. Blaauw, et al., “Parametric Yield Estimation Considering Leakage Variability,” *Design Automation Conference*, pp. 442-447, 2004.

[7] A. Srivastava, S. Shah, K. Agarwal, et al. “Accurate and Efficient Gate-Level Parametric Yield Estimation Considering Correlated Variations in Leakage Power and Performance,” *Design Automation Conference*, pp. 535-540, 2005.

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

[9] 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.

[10] S. Xiong, J. Bokor, “Sensitivity of Double-Gate and FinFET Devices to Process Variations,” *IEEE Transactions on Electron Devices*, vol. 50, no. 11, pp. 2255-2261, 2003.

[11] M. Chiang, J. Lin, K. Kim, et al., “Discrete Dopant Fluctuation in Limited-Width FinFETs for VLSI Circuit Application: A Theoretical Study,” *IEEE International Conference on Integrated Circuit Design and Technology*, pp. 1-4, 2006.

[12] 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.

[13] H. Tijms, “Understanding Probability: Chance Rules in Everyday Life”, Cambridge: Cambridge University Press, 2004.

[14] B. Zhai, S. Hanson, D. Blaauw, et al., “Analysis and Mitigation of Variability in Subthreshold Design”, *International Symposium on Low Power Electronics and Design*, pp. 20-25, 2005.