Electrical impact of line-edge roughness on sub-45-nm node standard cells

Yongchan Ban
The University of Texas at Austin
Department of Electrical and Computer Engineering
2400 Speedway
Austin, Texas 78712
E-mail: ycban@cerc.utexas.edu

Savithri Sundareswaran
Freescale Semiconductor
7700 West Parmer Lane
Austin, Texas 78729

David Z. Pan
The University of Texas at Austin
Department of Electrical and Computer Engineering
2400 Speedway
Austin, Texas 78712

Abstract. Since line-end roughness (LER) has been reported to be of the order of several nanometers and to not decrease as the device shrinks, it has evolved as a critical problem in sub-45-nm devices and may lead to serious device parameter fluctuations and performance limitations for future very large scale integration (VLSI) circuit applications. We present a new cell characterization methodology that uses the nonrectangular gate print images generated by lithography and etch simulations with the random LER variation. We systematically analyze the random LER by taking the impact on circuit performance due to LER variation into consideration. We observed that the saturation current, delay, and leakage current are highly affected by LER as the gate length becomes thinner. Results show that when the root mean square value of LER is 6 nm from its nominal line edge, the worst case saturation current, delay, and leakage current degradation are as much as 10.3% decrease, 12.4% increase, and 7 × increase at a 45-nm-node standard cell. Meanwhile the current, delay, and leakage current degradation at a 32-nm-node cell are up to 19.0% decrease, 21.8% increase, and 4600 × increase, respectively.

1 Introduction
As semiconductor device nodes continue to shrink to 45 nm and below, the complexity of designs is significantly increasing due to process variation. Among multiple variation issues, lithographic printability variation is one of the most fundamental challenges because it directly impacts on yield and performance. Despite advances in resolution enhancement techniques (RETs) such as optical proximity correction (OPC), phase-shifting masks (PSMs), off-axis illumination (OAI), etc., lithographic variation continues to be a challenge.1 There are two types of lithography variations that cause an undesirable performance mismatch in an identically designed transistor: (1) systematic lithography variation and (2) random variation.

The systematic lithography variation is introduced due to deterministic pattern proximity by the limitation of the lithography equipment because each device has different neighboring geometries such as neighboring gates, the convex and concave corner, the jog and line-end overhang, the active shapes, the distance of poly-to-contact landings, etc. To address the problem of systematic lithography variation, several authors have proposed a lithography-aware characterization method.2–5 In Ref. 2, the authors proposed gate slicing and effective gate length (EGL) methods to calculate the impact of nonrectangular gate shapes. Another paper7 proposed a modeling card to combine different EGLs from look-up tables of driving current and leakage current.

The second type of lithography variation is caused by random uncertainties in the fabrication process such as line-edge roughness (LER), the random defects due to missing and/or extra material, etc. At the same time, many nonlithographic sources of variation such as dopant variation6–8 and gate dielectric thickness (Tox) variation9, 10 also result in aggressive scaling. Among them, LER was regarded as a small fraction of the statistical variability in the past since the critical dimensions (CDs) of MOSFETs were orders of magnitude larger than the roughness. However, as the aggressive scaling continues into the nanometer regime, LER does not scale accordingly and becomes an increasingly larger fraction of the gate length.11, 12 As shown in Fig. 1, for channel lengths above 30 nm the random dopants are the dominant source of fluctuations, but below this channel length the LER takes over and becomes the dominant fluctuation source.13

Since LER is mainly caused by erosion of polymer aggregates at the edge of the photoresist (PR) during development and fully depends on some complex chemical formulas, it is difficult to generate the LER image in print-images of layouts. Even though LER is a kind of random variation, it is undesirable and must be analyzed because it highly degrades device performance. LER is of the order of several nanometers, and can be one of the performance-limiting components for 45-nm and below technologies.14–17

To address LER modeling and the impact of LER, much research has been proposed.18–20 Even though many works on LER modeling have been performed, these works focused at the process-level and unit-device-level simulation. Therefore, there is great demand to consider the impact of LER on standard cells and analyze timing impact, in particular delay and leakage current.

In this paper, we propose a comprehensive standard cell characterization method that accounts for random LER variation. Specific contributions in this paper are the following:
We derive an analytical LER variation model that can generically handle any root mean square (rms) amplitude and spatial frequency of LER. Then we integrate the LER variation into our print-image and layout extraction flow so that it can characterize the random LER mismatch variation.

2. The accuracy of our LER-aware layout extraction is validated from the physics-based TCAD simulation, introducing the strain of silicon used in standard cells.

3. We present a method to consider the LER variation in both statistical and deterministic analysis flows, and propose a LER tolerance for 45- and 32-nm standard cells.

The rest of the paper is organized as follows. Section 2 presents the impact of gate length variation on delay and leakage current. Section 3 describes the comprehensive characterization flow. This section presents an effective gate length extraction method and sensitivity characterization method. Experimental results are discussed in Sec. 4, followed by conclusions in Sec. 5.

2 Impact of Gate Length Variation
The most direct impact of systematic gate length variation is the resulting variation of CMOS gate delay and leakage. Figure 2 shows the percent delay variation [Fig. 2(a)] and the percent leakage current variation [Fig. 2(b)] according to the gate length variation in the 45-nm node CMOS inverter. In our experiments on the 45-nm patterning of a silicon wafer, the gate length variation due to systematic and random LER variation was more than 10% of the nominal gate length, which results in pull-up timing transition delay of an unskewed PMOS ($p$-channel metal-oxide semiconductor) to decrease over 25%, as shown in Fig. 2(a). The leakage current variation due to the gate length variation is much longer than that of the saturation current or the delay variation. The 10% gate length decrease causes a more than 10-fold leakage current, as shown in Fig. 2(b).

In a sub-45-nm node standard cell, the gate length variation due to LER is still huge for semiconductor manufacturing in spite of applying a strong RET technique such as OPC, immersion lithography, or an off-axis illumination process. This illustrates that the impact of LER on delay and leakage current should be analyzed in sub-45-nm node devices. In particular, since the standard cells are basic circuit blocks, the characterization of standard cells with regards to LER could be necessary for design and manufacturing cooptimization.
Therefore, the autocorrelation function of \( z(x) \) \( R(\tau) \), is formulated as
\[
R(\tau) = \mathbb{E}^{-1}[S(f)] = \lim_{L \to \infty} \frac{1}{L} \int_{-L/2}^{L/2} z^*(x)z(x+\tau) \, dx. \tag{2}
\]

The rms roughness \( \sigma \) is often defined in terms of \( z(x) \) as
\[
\sigma^2 = \lim_{L \to \infty} \frac{1}{L} \int_{-\infty}^{\infty} |z(x)|^2 \, dx = 2 \int_0^\infty S(f) \, df. \tag{3}
\]

Thus, the autocorrelation function \( R(\tau) \) follows an exponential function by the distance \( r \) for the line edge as
\[
R(r) = \sigma^2 \exp \left[ -\left( \frac{r}{L_c} \right)^{2\alpha} \right]. \tag{4}
\]
where \( L_c \) is the correlation length, \( \sigma \) is the standard deviation of line edge, and \( \alpha \) is related to the fractal dimension \( D (\alpha = 2 \text{ to } D) \). Therefore, PSD is approximated as
\[
S(k) = \frac{2\sigma^2 L_c}{(1 + k^2 L_c^2)^{0.5+\alpha}}, \tag{5}
\]
where \( k = 2\pi f, f = i/(VN\Delta x) \), \( 0 \leq i \leq N/2 \), and \( N \) is the number of points along the line. Hence, the LER for a large number of resists can be characterized by just three numbers, \( \sigma, L_c, \text{ and } \alpha \).

With the magnitude information provided by \( S(k) \), we can reconstruct random line edges by applying a random phase to each frequency component of the PSD to form a unique signal in the frequency domain. A line edge with roughness can be simulated by doing an inverse Fourier transform of this signal. Random lines are distinguished through applied random phases.

Figure 4(a) shows results of line edge roughness from Eq. (5) with \( L_c = 25 \text{ nm}, 3\sigma = 4 \text{ nm}, \) and \( \Delta x = 1 \text{ nm} \) at three different values for \( \alpha \) of 0.2, 0.5, and 0.8. We can see that smaller \( \alpha \) (larger fractal dimension) leads to more local roughness. Meanwhile \( \sigma \), rms amplitude, is the most important parameter for LER. Figure 4(b) shows two simulated roughness profiles with different values of \( \sigma \); \( \sigma \) corresponds to the transversal magnitude to the line, and the larger \( \alpha \) shows greater roughness of the line. On the other hand, the correlation length \( L_c \) shows a longitudinal magnitude along the line. As shown in Fig. 4(c), the larger \( L_c \) depicts the longitudinally wider period of LER. Among the three parameters of LER, \( L_c \) and \( \alpha \) are highly dependent on the photoresist type and relatively less critical than \( \sigma \) (Ref. 24). Thus, in this paper we focus on presenting LER with regard to \( \sigma \).

LER is a random fluctuation in the gate length along the complete width of the device and has influence on both edges of the gate.

LER is often expressed by the power spectral density (PSD), which is theoretically the Fourier transform of the autocorrelation function. Let us define \( z(x) \) as a 1-D distribution of edge locations. The PSD \( S(f) \) is mathematically defined as
\[
S(f) = \lim_{L \to \infty} \frac{1}{L} \left[ \int_{-L/2}^{L/2} \exp(2\pi ifx) \, dx \right]^2 \tag{1}
\]
Therefore, the autocorrelation function of \( z(x) \) \( R(\tau) \), is formulated as
\[
R(\tau) = \mathbb{E}^{-1}[S(f)] = \lim_{L \to \infty} \frac{1}{L} \int_{-L/2}^{L/2} z^*(x)z(x+\tau) \, dx. \tag{2}
\]

The rms roughness \( \sigma \) is often defined in terms of \( z(x) \) as
\[
\sigma^2 = \lim_{L \to \infty} \frac{1}{L} \int_{-\infty}^{\infty} |z(x)|^2 \, dx = 2 \int_0^\infty S(f) \, df. \tag{3}
\]

Thus, the autocorrelation function \( R(\tau) \) follows an exponential function by the distance \( r \) for the line edge as
\[
R(r) = \sigma^2 \exp \left[ -\left( \frac{r}{L_c} \right)^{2\alpha} \right]. \tag{4}
\]
where \( L_c \) is the correlation length, \( \sigma \) is the standard deviation of line edge, and \( \alpha \) is related to the fractal dimension \( D (\alpha = 2 \text{ to } D) \). Therefore, PSD is approximated as
\[
S(k) = \frac{2\sigma^2 L_c}{(1 + k^2 L_c^2)^{0.5+\alpha}}, \tag{5}
\]
where \( k = 2\pi f, f = i/(VN\Delta x) \), \( 0 \leq i \leq N/2 \), and \( N \) is the number of points along the line. Hence, the LER for a large number of resists can be characterized by just three numbers, \( \sigma, L_c, \text{ and } \alpha \).

With the magnitude information provided by \( S(k) \), we can reconstruct random line edges by applying a random phase to each frequency component of the PSD to form a unique signal in the frequency domain. A line edge with roughness can be simulated by doing an inverse Fourier transform of this signal. Random lines are distinguished through applied random phases.

Figure 4(a) shows results of line edge roughness from Eq. (5) with \( L_c = 25 \text{ nm}, 3\sigma = 4 \text{ nm}, \) and \( \Delta x = 1 \text{ nm} \) at three different values for \( \alpha \) of 0.2, 0.5, and 0.8. We can see that smaller \( \alpha \) (larger fractal dimension) leads to more local roughness. Meanwhile \( \sigma \), rms amplitude, is the most important parameter for LER. Figure 4(b) shows two simulated roughness profiles with different values of \( \sigma \); \( \sigma \) corresponds to the transversal magnitude to the line, and the larger \( \alpha \) shows greater roughness of the line. On the other hand, the correlation length \( L_c \) shows a longitudinal magnitude along the line. As shown in Fig. 4(c), the larger \( L_c \) depicts the longitudinally wider period of LER. Among the three parameters of LER, \( L_c \) and \( \alpha \) are highly dependent on the photoresist type and relatively less critical than \( \sigma \) (Ref. 24). Thus, in this paper we focus on presenting LER with regard to \( \sigma \).

LER is a random fluctuation in the gate length along the complete width of the device and has influence on both edges of the gate. To implement the LER effect in the print-images, we convert these two line edge fluctuations to a single fluctuation with an equivalent variation given by
\[
\sigma_{\text{ler}}^2 = \sigma_l^2 + \sigma_r^2 - 2\rho_1\sigma_l\sigma_r, \tag{6}
\]
where, the \( \sigma_{\text{ler}} \) is the line-width roughness, \( \rho_1 \) is the correlation coefficient between the left and the right edges of a line, which means that the \( \rho_1 \) is 0 for no correlation and 1 for perfect correlation. When we calculate the effective gate length in a cell, we add the gate length variation due to LER to the top of the systematic component. The gate length

**3 Model Formulation and Simulation**

**3.1 Random LER Modeling**

LER, one of the dominant random variations, is caused by the interaction of light and thermal bombardment with the molecular nature of photoresist materials in the acid generation, the acid diffusion, and the development process in chemically amplified resists (CARs), as shown in Fig. 3(a). As shown in Fig. 3(b), the severe CD variation evolves at the line edge, despite patterning a straight-line structure. LER is a random fluctuation in the gate length along the complete width of the device and has influence on both edges of the gate.

LER is often expressed by the power spectral density (PSD), which is theoretically the Fourier transform of the autocorrelation function. Let us define \( z(x) \) as a 1-D distribution of edge locations. The PSD \( S(f) \) is mathematically defined as
\[
S(f) = \lim_{L \to \infty} \frac{1}{L} \left[ \int_{-L/2}^{L/2} \exp(2\pi ifx) \, dx \right]^2 \tag{1}
\]
Fig. 4 Demonstration of LER simulation on a gate edge with (a) \( \alpha \), (b) \( \sigma \), and (c) \( L_c \).

variation is calculated as a function of \( \sigma_{\text{LWR}} \). We assume the rms amplitude of the left LER (\( \sigma_l \)) equals the rms amplitude of the right LER (\( \sigma_r \)) and the correlation coefficient \( \rho \) is randomly determined when the LER is generated.

In a conventional Spice circuit simulation, one particular gate length is used for each transistor. Due to LER, the gate printed images show a nonrectangular transistor shape. Therefore, we should get an effective gate length by a nonrectangular layout extraction as in the following section.

Algorithm 1 LER aware nonrectangular gate length calculation

1: \textbf{Require:} A set of lookup table, print-images \( I \)
2: \text{Table gate} = \text{poly} \cap \text{active}
3: \text{nmos} = \text{gate} \cap \text{nwell}, \text{pmos} = \text{gate} – \text{nmos}
4: \text{for each cell} \( C \in I \) \text{ do}
5: \text{for each \text{nmos} } N \in C \text{ do}
6: \text{Find intersection points between poly \& active}
7: \text{Set } W_{\text{eff}} \text{ \& diffusion rounding}
8: \text{for each slice } = \Delta z S \in N \text{ do}
9: \text{Reconstruct segmented polygon}
10: \text{calculate } I_{\text{seg}} \text{ from } I_m \text{ \& } I_{\text{off}} \text{ lookup tables}
11: \text{sum} + = \omega I_{\text{seg}} \text{; where } \omega \text{ is the weighting factor of narrow width effect.}
12: \text{end for}
13: \text{Update sum from Eq. (7) and (8)}
14: \text{Calculate } L_{\text{eff}}
15: \text{end for}
16: \text{for each \text{pmos} } P \in C \text{ do}
17: \text{Same sequence as \text{nmos}}
18: \text{end for}
19: \text{end for}

3.2 LER-Aware Nonrectangular Gate Extraction

In this step, we extract the effective gate length for postlithography print-images using a gate-segmentation technique. Lithography variations result in a nonrectangular shapes for both poly and diffusion layers. For a standard cell, the area of the diffusion region defines the drive-strength of the cell. The active diffusion rounding has a nontrivial impact on the nonrectangular gate because the contours in this layer show rounding patterns connecting to power rails which causes much variation of the effective gate length and width.

In our experiments, the area difference of gates between drawn diffusion and printed diffusion is over 6%; the effective gate length when considering diffusion rounding is up to 7% different from that due to no diffusion rounding. The difference in drawn and printed diffusion layer dimensions results in the drive strength difference to be about 8% in our 45-nm standard cell.

The proposed algorithm is illustrated in Algorithm 1 for random LER-aware extraction. To extract the print-image, we first construct four look-up tables for the on current \( I_{\text{on}} \) and the off current \( I_{\text{off}} \) of the NMOS (N-type metal-oxide semiconductor) and PMOS devices using commercial simulation tool.\(^{25}\) We then find the four intersection points using poly and diffusion print-images. These points represent the gate/channel region. From these points, we identify the effective gate width (\( W_{\text{seg}} \)) and rounded diffusion area, as shown in Fig. 5.

Next, we segment the gate region by a set of equal width rectangular polygons. Each segment then has a width \( W_{\text{seg}} \). The current for each segment \( I_{\text{seg}} \) is computed using the nominal current from the rectangular device. For more accurate extraction, we can consider the gate narrow-width effect\(^{1,26}\) by multiplying the current weighting factor \( \omega \) by the current of each segment. The equivalent or total current for the gate region is computed by summing all these segment currents.
Prior to obtaining $L_{\text{eff}}$ for each device, we update the equivalent current with that due to the rounded diffusion area. We use the formulation in Ref. 27 to compute the equivalent currents due to diffusion rounding. The device currents $I_{\text{on}}$ and $I_{\text{off}}$ are updated using following formulations:

\begin{equation}
I_{\text{on}} = I_{\text{on-nom}} \left[ 1 + 0.5 \frac{(W_{\text{top}} + W_{\text{btm}})}{W_{\text{nom}}} \right],
\end{equation}

\begin{equation}
I_{\text{off}} = I_{\text{off-nom}} \exp \left( \frac{L_{\text{nom}}}{L'} \right),
\end{equation}

where $I_{\text{on-nom}}, I_{\text{off-nom}}, L_{\text{nom}},$ and $W_{\text{nom}}$ are the on current, the off current, the gate length, and the gate width of the nominal rectangular device, respectively; $W_{\text{top}}$ and $W_{\text{btm}}$ are the top height and the bottom height of the rounded diffusion area, respectively, as shown in Fig. 5; $C$ is a fitting parameter; and $L'$ is the effective channel length at the edge of rounded diffusion. The final effective channel lengths ($L_{\text{eff}}$) for on and off-states are directly calculated from the total $I_{\text{on}}$ and $I_{\text{off}}$ current, which are described in look-up tables.

### 3.3 TCAD Simulation and Validation

Using our LER-aware nonrectangular gate extraction model, we characterize 45- and 32-nm standard cells in terms of delay and leakage current. Prior to introducing our characterization approach, we validate our LER model with a rigorous TCAD simulation result. The gate shape due to LER is actually subject to change in accordance with relevant process (lithography, etch, etc.) and environmental condition. Even

in silicon experimental data, the polysilicon width ($\approx$ gate length) in a gate could be different in every measured point and shows a form of a distribution. For this reason, we did thousands of LER simulation for a certain LER value and used the average CD in measuring a trend of device performance due to LER.

To verify the proposed LER model on device performance in terms of the driving current, we employ a TCAD simulator\textsuperscript{28} with the strained silicon in which tensile strain is introduced in the NMOS channels by using a postsalicide silicon-nitride capping layer. To save a simulation time and memory usage, we use a quasi-3-D simulation, as shown in Fig. 6(a), in which the LER implemented print-image is considered in the TCAD simulation, then a set of 2-D simulations are carried out. The left and right edges of a gate have a same amount of rms edge roughness, yet the correlation coefficient $\rho$ is randomly chosen. Some of the most important parameters of the device are the range of gate lengths caused by LER is from 25 to 60 nm (the nominal gate length is 40 nm), oxide thickness is 1.2 nm, and the capping layer thickness is 75 nm.

We compare the result in terms of the amount of LER between the rigorous TCAD simulation and the circuit simulation used for LER characterization. To compensate for the
internal difference between the TCAD simulator and circuit simulator, we normalize the current value to the current of a device without LER. Figure 6(b) shows the result for validation of our proposed LER model. The result reports the percent variation of the saturation current with the amount of LER and shows the great agreement. The maximum error between TCAD simulation and our proposed result is within 5.3%, and the average error is about 1.2%, when comparing the current variation due to LER.

\[ \Delta_i = \frac{I_i - I_0}{I_0} \]

where \( I_0 \) has a systematic component \( L_{sys} \) and a component due to the random LER variations \( \Delta L_{LER} \). This can be represented as

\[ L_{pi} = L_{sys} + \Delta L_{LER}. \]

To characterize for the effect of systematic lithography variations, the standard characterization procedure is used. The characterization is carried out by annotating \( L_{sys} \) for each device in the cell. The \( L_{sys} \) is a deterministic value and a standard delay/leakage characterization by setting each device to the new effective gate length/width due to systematic variations is performed. To characterize for random LER variations, the standard cell is characterized for sensitivity to \( \Delta L_{LER} \). During sensitivity characterization, the variations in each device need to be accounted. Let \( p \) be number of devices in a cell. Let the random LER variation for each device \( k \) be \( \Delta L_k \). Since these random variations is much smaller than the nominal \( L_{pi} \), performance characteristics of the cells are almost linear functions within the range of the variations \( \Delta L_i \).

For delay characterization, the delay of a timing arc \( D \) can be represented as

\[ D = D_0 + \sum_{k=1}^{p} d_k \Delta L_k, \]

where \( D_0 \) is the nominal delay value and is characterized by extracting \( L_{eff} \) and \( L_{sys} \) due to printed contours in poly and diffusion layers. Each device LER, \( \Delta L_k \) is modeled as a distribution \( \mathcal{N}(0, \sigma) \). The quantities \( d_k \) are direct sensitivities of cell delay with respect to the LER variations \( \Delta L_k \).

Thus, each cell in the library is characterized for a nominal delay \( D_0 \) by setting all devices to their corresponding contour-based effective gate lengths and zero LER. Additionally, the cells are characterized for sensitivity to LER on each device by setting a separate random variable, \( \Delta L_k \) and the corresponding delay variation is computed as in Fig 7. Assuming delay variation due to each device is statistically independent, the cell’s delay sensitivity can then be obtained using the following relation:

\[ d_{eq} = \left( \sum_{i} \frac{d_i^2}{\sigma_i^2} \right)^{1/2}. \] (11)

### 4 Experimental Results

We implemented gate LER using Tcl and Perl script language and tested with Nangate 45- and 32-nm open cell library. The nominal drawn gate CDs of 45- and 32-nm cells are 40 and 30 nm, respectively. We used Calibre-WB from Mentor Graphics for model-based OPC and printed images. The timing analysis and characterization were done by H-Spice circuit simulator from Synopsys. We directly implemented LER on the poly layer, where we applied the LER just on the gate region (≈ poly on active). This assumption is reasonable due to the following two reasons: one objective is to save the simulation time, and the other reason is due to that the poly layout besides gate regions does not affect on the effective gate length of a gate. In a sub-45-nm node design, the gate region (not poly layout) is usually drawn with 1-D type structure due to restricted design rule (RDR).

We generated more than 1000 LER patterns for a particular rms value of LER so that the results are shown as a distribution similar to a normal distribution. For 45- and 32-nm circuit simulations, we used Predictive Technology Model (PTM). We swept the LER variation from zero to 10 nm of the nominal gate length. Figure 8 shows that more LER causes a more severe pattern distortion and gate length variation.

Figure 9 illustrates the overall flow of our model-based geometrical and electrical analysis. The flow is divided into three main steps:

1. **Printed image simulation**: This step involves simulating the lithography models and generating of non-rectangular contours/shapes due to the printed image. We get printimages of the nominal condition and the process corners. After finishing lithography print-image simulation, we apply etch simulation which is done by a rule-based correction in which the rule table is defined from the empirical experimental data. Once we get the final print-images, we can also simulate the impact of LER. Input LER conditions are first requested, then the LER variations are added on the edge of the final print-images.

2. **Layout extraction with printed image**: This step extracts device dimensions considering the nonrectangular shape in the poly and diffusion layers due to print image. The basic idea is to convert a nonrectangular transistor into several slices such that
the nonrectangular gate shape is modeled as a single equivalent rectangular transistor with an effective gate length.

3. **Characterization for several corners**: This step characterizes the cells for delay and leakage information using the extracted parameters from previous step. We measure delay and power of a cell for each process corners.

We first investigated the device saturation current variation and the leakage current variation with the amount of LER in a 45-nm inverter cell (Fig. 10) and a 32-nm inverter standard cell (Fig. 11). Figure 10(a) shows the impact of LER on the saturation current of a conventional 45-nm NMOS device. The black circled dot represents the average of the variation, and the small bars show the upper and lower bounds of the variation. The upper and lower bounds are equivalent to $+3\sigma$ and $-3\sigma$ from the nominal value. As shown in the results, the deviation between the upper bound and the lower bound is highly increased as LER increases, while the average values are slightly increased.

The impact of LER on the gate leakage current is much more critical than that of the saturation current, as shown.
In a similar way, the device saturation current variation and the leakage current variation of a 32-nm NMOS device have a similar trend with those of the 45-nm device. Even its similar impact of LER on the saturation current, as shown in Fig. 11(a), the leakage current variation is much higher than in the 45-nm device. The results show that since LER does not shrink, the gate leakage is highly increased as device shrinks.

Table 1 shows why the average current due to LER is slightly increased for $I_{\text{on}}$ and exponentially escalated for $I_{\text{off}}$. As LER increases, the nominal effective gate length $L_{\text{eff}}$ for the driving current becomes smaller, which causes the nominal $I_{\text{on}}$ to slightly increase. Meanwhile, the $L_{\text{eff}}$ for the leakage current is decreased more, and the $L_{\text{eff}}$ deviation for leakage current is also much wider than the $L_{\text{eff}}$ of the saturation current. For example, when the rms $3\sigma$ LER is 7 nm, the standard deviation $\sigma$ of on-current $L_{\text{eff}}$ is 1.45 nm, while the $\sigma$ of off-current $L_{\text{eff}}$ is 2.07 nm.

As mentioned, the gate length variation due to LER follows a distribution which has the upper bound corner and the lower bound corner. The variation due to LER is defined for three different conditions: (1) a typical condition, (2) $+3\sigma$, and (3) $-3\sigma$ variations. The $\pm 3\sigma$ variations result in the lower (~thinner line) and upper (~thicker line) bounds. A gate layer is simulated with three different conditions in a circuit simulation: the best condition for delay occurs when the gate length due to LER has a minimum value, which causes the delay of a cell to decrease; meanwhile the worst condition represents the maximal gate length for delay. When the gate length due to LER has the minimum value, the worst corner for the leakage current $I_{\text{off}}$ occurs.

![Fig. 12](image1.png)

We analyzed delay and leakage variation with LER in a 45-nm inverter cell (Fig. 12) and a 32-nm inverter cell (Fig. 13). The results indicate that the delay variation is trivial at a small amount of LER (less than 3 nm of $3\sigma$ LER). However, we found that the delay difference between the best and the worst corner is steep when the roughness of LER increases. This is because the gate length variation is

![Fig. 11](image2.png)

![Fig. 13](image3.png)
caused due to LER, which causes the saturation current to increase. The gate leakage current variation is much higher than the delay, as mentioned for Fig. 10(b). The $Y$ axis is a log scale, thus the results show that the leakage current increases more than 7 times compared to the typical value when the rms roughness of LER is 6 nm, and the difference between the best and the worst corner is dramatically increased as LER increases. As a result, if designers want to keep the worst case delay variation within 10% from the typical value and to keep the worst case leakage variation under 5 times from the typical leakage, more than 5 nm LER should be avoided for a 45-nm standard cell library.

In a similar way, we analyzed delay and leakage current variation in a 32-nm inverter cell, as shown in Fig. 13. The percent delay variation of a 45-nm cell is similar to that of a 32-nm cell. However, as mentioned for Fig. 11(b), the leakage variation is dramatically increased. Compared with Fig. 12(b), when the $3\sigma$ LER is 6 nm, the leakage variation between the typical and the worst corner is $7.1 \times$ at the 45-nm inverter cell, whereas the leakage variation for a 32-nm inverter cell is as much as 4677$\times$ above the typical leakage value. This means that the leakage current is highly dependent on LER as the transistor shrinks even though the impact on delay variation is not severe. Therefore, the leakage minimization for LER is the most important issue in sub-45-nm node device. For instance, to decrease the worst case leakage within $5 \times$ from the typical leakage, the process engineer should control LER under 5 nm in the 45-nm node design, meanwhile the LER for a 32-nm node devices should be under 2 nm, which makes the need of process and design cooptimization more important in semiconductor manufacturing.

We also computed the effective delay sensitivity using the formulations in Sec. 3.4 for the device LER variations at various lithography corners. The results for few cells from the 45-nm bulk technology libraries are presented in Table 2.

Here column I shows the sensitivities due to LER when considering no systematic lithography variations. Columns II, IV, VI are delay sensitivities due to LER when considering systematic lithography variations at the typical, best, and worst corners, respectively. Columns III, V, VII, are the errors in these three corners when compared with that due to no systematic variations. The results indicate that the sensitivities due to LER variations increase at typical and best case corners when compared with that due to no systematic variations. Therefore, the leakage minimization for LER is the most important issue in sub-45-nm node device. For instance, to keep the worst case leakage within $5 \times$ from the typical leakage, the process engineer should control LER under 5 nm in 45-nm and $3\sigma$ 2 nm in 32-nm node processes, which are consistent with ITRS roadmap.31 Our further work will concern the impact of LER on metal lines, and a layout-proximity-aware LER model from neighboring layout interface is being investigated for LER-aware layout optimization.

### Table 2: Delay sensitivity due to LER variations.

<table>
<thead>
<tr>
<th></th>
<th>$d_{\text{eff}}$ for $L_{\text{nom}}$</th>
<th>$d_{\text{eff}}$ at typical</th>
<th>Error (%)</th>
<th>$d_{\text{eff}}$ at best</th>
<th>Error (%)</th>
<th>$d_{\text{eff}}$ at worst</th>
<th>Error (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inv</td>
<td>3.933</td>
<td>4.077</td>
<td>3.7</td>
<td>5.819</td>
<td>42.7</td>
<td>3.349</td>
<td>−17.8</td>
</tr>
<tr>
<td>NOR</td>
<td>3.544</td>
<td>4.070</td>
<td>14.8</td>
<td>5.581</td>
<td>37.1</td>
<td>3.612</td>
<td>−11.2</td>
</tr>
<tr>
<td>NAND</td>
<td>3.189</td>
<td>3.962</td>
<td>24.2</td>
<td>4.955</td>
<td>25.1</td>
<td>2.787</td>
<td>−29.6</td>
</tr>
<tr>
<td>DFF delay</td>
<td>6.947</td>
<td>7.913</td>
<td>13.9</td>
<td>9.803</td>
<td>23.9</td>
<td>8.452</td>
<td>6.8</td>
</tr>
</tbody>
</table>
Acknowledgments

This work is sponsored in part by SRC, NSF CAREER Award, and equipment donations from Intel.

References


Yongchan Ban is pursuing his PhD degree in electrical and computer engineering at the University of Texas at Austin. His research focuses on lithography-driven design automation, DFM (design for manufacturing) and very large scale integration (VLSI) design automation. He was a senior engineer with the Semiconductor R&D Center, Samsung Electronics, from 2002 to 2007. He was with Freescale Semiconductor, Austin, Texas during the summer of 2008, and with Globalfoundries, Sunnyvale, California, during the summer of 2009 as a hardware engineering intern. He worked for Synopsys, Austin, Texas in 2010 as a software engineering intern. He received two Best Paper awards at Symposium on VLSI Design Technical Conference in 2004 and 2006, was awarded an SPIE Scholarship in Optical Science and Engineering in 2010, and has published over 45 papers and 5 patents.

Savithri Sundareswaran received her Bachelors degree in electrical and electronics engineering and her Masters (Honors) degree in physics from Birla Institute of Technology and Sciences in 1993 and 1994, respectively, and her PhD degree in computer engineering from the University of Texas at Austin in 2009. She was a scientist with the Central Electronics Engineering Research Institute in India from 1993 to 1995. Dr. Sundareswaran has been with Motorola and then Freescale since 1995, working in design analysis and optimization including timing, power, and reliability analysis. She is a senior member of the IEEE and has served as member of VLSI’s Signal Integrity and DFM committees. She has served on the technical program committees of several conferences. Her research interests include circuit-level modeling and analysis for manufacturability, statistical timing characterization, and optimizations.

David Z. Pan received his PhD degree in computer science from University of California, Los Angeles, in 2000. From 2000 to 2003, he was a research staff member with the IBM T. J. Watson Research Center, Yorktown Heights, New York. He is currently an associate professor and director of the UT Design Automation (UTDA) Laboratory, Department of Electrical and Computer Engineering, the University of Texas at Austin. He has published over 120 refereed papers in international conferences and journals, and is the holder of 8 U.S. patents. His research interests include nanometer very large scale integration (VLSI) physical design, design for manufacturing, vertical integration of technology, design and architecture, and design/CAD for emerging technologies.