Study of Gate Line Edge Roughness Effects in 50 nm Bulk MOSFET Devices

Shiying Xiong, Jeffrey Bokor
Dept. of Electrical Engineering and Computer Sciences, Univ. of California at Berkeley, CA 94720

Qi Xiang, Philip Fisher, Ian Dudley and Paula Rao
Advanced Micro Devices, P.O. Box 3453, MS 143 Sunnyvale, CA 94088-3453

ABSTRACT

We studied gate line edge roughness (LER) and its effect on electrical characteristics of 50nm bulk MOSFETs. Using simulation, we studied the underlying mechanism of three significant LER effects on the electrical performance of advanced 50 nm gate length bulk devices. First, we found that off-state leakage current is much more sensitive than the on-state drive current to gate LER. Second, we found that high frequency LER can lead to a decrease in effective channel length by enhanced lateral diffusion of the self-aligned source/drain extension. Third, low frequency LER causes local CD variation simply due to the statistical variation of average CD in a finite width sample. We also show how device design parameters, such as halo implant dose, can be used to tradeoff LER sensitivity and device performance.

Keywords: Line edge roughness, MOSFET, simulation, leakage

1. INTRODUCTION

Gate line edge roughness (LER) is the random deviation of gate line edges from ideal definition. There are many coupled factors contributing to it. Some important examples are: photo-resist (PR) line edge roughness, which depends on PR type, thickness, substrate reflectivity, image contrast as well as process conditions; gate polysilicon (poly) etching condition; poly grain size and doping. LER does not scale down with line width. Recently, as the industry is pushing the device research frontier toward 30 nm and even 15nm physical gate lengths, considerable effort has been devoted to the reduction of poly LER in lithography and etching steps.

Attention has also been given to investigating the effects of LER on device electrical parameters\textsuperscript{1, 2}, and the LER tolerance of devices with a specified gate length\textsuperscript{3}. Linton\textsuperscript{1} reported the first 3D simulation of the impact of LER on device characteristics using a simple square wave model for roughness and showed increased off-state leakage with LER. Oldiges\textsuperscript{2} showed a way to calculate the effect with a quasi-2D method, and Diaz\textsuperscript{3} proposed an analytical model following Oldiges’ approach. However, these studies gave little attention to the role of the lateral scale length (or spatial frequency spectrum) of the roughness.

In this work, we describe our study of LER effects on 50 nm bulk MOSFET devices using device simulations coupled with experimental data measured on gate edges. The accuracy of using 2D vs. 3D simulations is compared as the spatial frequency spectrum of the roughness is varied. We use the width of the gate edge autocorrelation function (correlation length) as the parameter that describes the inverse of the width of the spatial frequency spectrum. We also find a previously unrecognized physical effect that becomes important for high spatial frequency roughness. Namely, the lateral diffusion of the source/drain junctions under the gate is enhanced when the roughness correlation length is
comparable to or smaller than the average lateral diffusion length. We also consider how low spatial frequency LER becomes equivalent to statistical variation of the average CD for a given finite gate width. To see how the device design might be re-optimized to account for a given level of LER, we chose the halo implant dose as an example and simulated the dependence of the LER effects as halo dose is varied. Finally, we show experimental results on the physical characterization of etched polysilicon gates with varying LER as a first step toward obtaining electrical characteristics of devices with varying LER magnitude.

2. EFFECT OF GATE LER ON MOSFET $I_{\text{OFF}}$ AND $I_{\text{ON}}$

One of our goals is to establish an efficient method of calculating the effects of gate LER, such that the specifications for the allowable LER in any given process can be established based on electrical performance criteria. It is an inherently three-dimensional effect, but 3D device simulation requires considerable expenditure of effort. In Oldiges et al.\(^2\), a 2D approach was introduced in which the current of small, 2D device slices is summed and they showed that these results agreed well with full 3D simulation. Their model for the LER had ~6nm RMS variation of an 80 nm nominal gate length with 7nm width for the individual slices. The physical meaning of this approach is that it requires that the source-drain current of rough-gate devices have no component perpendicular to the 2D device cross-sectional plane. We have done some further investigation into how well this approach works as the correlation length of the roughness is varied. Our comparison was done on 100nm devices. Using ISE TCAD software\(^4\) we directly sliced device structures after 3D simulation to do 2D simulation and thus exactly keep the doping profile and grid mesh the same. We found that the 2D method can significantly underestimate the current for gate line edge roughness with correlation length smaller than 15nm.

The 2D model can be understood using a simple analysis of the dependence of $I_{\text{OFF}}$ and $I_{\text{ON}}$ on gate length $L$. These can be generally fit by the following functions.

$$I_{\text{Off}} (L) = C_0 e^{\frac{f(V_d)}{kT} \frac{\sigma}{q}}$$

and

$$I_{\text{On}} (L) = C_1 + C_2 \frac{1}{L}$$

where $C_0$, $C_1$, $C_2$, and $f$ are fitting parameters, and $V_d$ is the applied source-drain voltage. In the absence of 3D effects, the net current in a device with gate LER can then be written as:

$$I_{3D} = \frac{W}{N} \sum_{L_{\text{min}}}^{L_{\text{max}}} P(L) I_{2D}(L)$$

where $W$ is the device width, $N$ is the number of length samples in the sum, and $P(L)$ is the probability distribution function for the gate roughness. Eqs. 1 and 3 give us the understanding of increased leakage current with increase of LER. The strongly non-linear increase of leakage current with shrinking gate length causes the leakage of a device with larger LER to be greater than that of a smooth gate device with the same average gate length. The increase of drive current with shorter gate length is much weaker. As a result, on-state current is only elevated by several percent.

Fig. 1 shows the results of 2D process and device simulations of NMOS $I_{\text{off}}$ for a 50 nm minimum gate length logic technology. The squares show the results of the simulations and the solid line shows the best fit, using Eq. 1.

Fig. 1: Super exponential increase of leakage current with shrinking gate length
The fit is excellent, indicating that only a relatively small number of 2D process and device simulations are needed to determine the parameters $C_0$ and $l$, and that the analytical expression (1) can be used in Eq. 3.

In order to illustrate the effect of the correlation length of the roughness, we compared the 2D analysis method described above with full 3D device simulations. In this case, a 100 nm gate length SOI technology design was used. The results are shown in Fig. 2 for two different correlation lengths. For 5 nm correlation length, there is a very significant 3D effect, and the 2D model is not accurate.

![Graph](image)

In this case, we can fit the results to an expression of the form:

$$I_{3D}(A) = \frac{W}{N} f_{3D}(L_0, V_{DS}, V_G, L_C, \sigma) \sum_{L_{min}}^{L_{max}} P(L) I_{2D}(L)$$

$$f_{3D} = 1 + c_1 r + c_2 r^2 + \ldots$$, of $r = \frac{\sigma}{L_C}$

In this equation, $f_{3D}$ is a factor describing the 3D effect of LER on the total current, $r$ is the relative roughness, $\sigma$ is the root mean square value of the LER, $L_0$ is the nominal gate length, and $L_C$ is the correlation length of the gate LER. The coefficients in the expression of $f_{3D}$ are dependent on the process, nominal gate length, gate and drain to source voltage. By using $f_{3D}$ to second order we can gain enough accuracy for high frequency LER with $L_C$ as small as 5nm (See Fig. 2). This model allows for the characterization of LER effect in the regime of small correlation length roughness with a minimal number of 3D simulations.

The 3D effect factor $f_{3D}$ is a function of relative roughness. It is acceptable to use Eq. 4 to its zero order for low frequency LER with large $L_C$ and small $\sigma$. Fig. 2(b) shows that ignoring the 3D effect (let $f_{3D} = 1$) underestimates the leakage current by less than 20% for LER with $L_C$ around 15nm and $\sigma < 5nm$. 

![Graph](image)
3. HIGH FREQUENCY LER ENHANCED LATERAL DIFFUSION

The source/drain junctions of CMOS devices are formed from self-aligned implantation. Gate edges with roughness result in a “rough” under-gate doping profile. Numerical solution of a simple 2D lateral diffusion equation shows that the doping profile with a rough initial boundary diffuses more than a smooth one for the same thermal budget. We refer to this as LER enhanced lateral diffusion (Fig. 3). We define the diffusion length as the distance from the mean location of the gate edge to the mean location of the junction.

More accurate results about the source/drain extension implantation and its diffusion can be found from full 3D process simulation. We used the process simulator Taurus, in which more complete implantation and diffusion models are computed. For the typical RTA condition used to make the device in our process, we find that if the correlation length of gate LER is larger than 30nm, the resulted channel edges approximately follow the roughness of gate edge, but for high frequency line edge roughness (HFLER) with \( L_c \) as small as 5nm, significant smoothing occurs (Fig. 4), which is caused both by scattering in implantation and diffusion in high temperature annealing. Gate LER causes the lateral diffusion of source/drain extensions to be measurably enhanced. The average under gate lateral source/drain extension length of a rough gate device is longer than that of a smooth edge device. Additionally, the lateral extension length of HFLER device is longer than that of a LFLER device if we keep both the average and \( \sigma \) value of gate edges the same (see Fig. 5). The source/drain of devices with gate LER, especially HFLER, encroaches more under gate and causes additional effective channel length shortening. This could bring several nm of effective channel length reduction and increase gate to source/drain overlap capacitance. It also leads to a doping profile that is less abrupt, which results in increased parasitic resistance. This effect depends on the amplitude and frequency of the LER but also on the source/drain

![Fig. 3: Lateral diffusion of a rough doping boundary and comparison with a smooth boundary by solving 2D isotropic diffusion equation](image)

![Fig. 4: Smoothing off of the source-channel junction for HF gate LER (3D Taurus simulation)](image)

![Fig. 5: Average under gate S/D lateral diffusion length for gate LER with different \( L_c \) (3D Taurus simulation)](image)
extension implantation and thermal process conditions.

One way to understand why the rough doping profile diffuses farther when the correlation length is short, is to think of the initial doping profile as that of a smooth gate device, but with reduced abruptness. In other words, the gate edge is not sharp; it is “fuzzy”. The junction location occurs where the source/drain doping concentration equals to the lightly doped channel concentration, which is 2-3 orders of magnitude below the peak source/drain doping concentration. Thus even before any diffusion occurs, the junction location is farther from the mean gate edge due to the “fuzziness” of the gate edge. This is clearly shown in Fig. 5.

4. LER BUDGET ESTIMATION

To establish a method to estimate the allowable LER for a design, we should include the coupling effects with other process variations such as across wafer CD non-uniformity. 8nm was used as the $3\sigma_{CD}$ value of across wafer CD variation in the following simulations. The gate length of all MOSFETs with the same coded gate length on a wafer is assumed in the range of $L = L_0 \pm k\sigma_{CD}$, where $k$ is a chosen parameter (e.g.: $k=3$ for 99.7% and 1.5 for 86.6% in normal distribution). We used equation (3) to make a statistical computation of the leakage and on-state current of devices with gate LER, under the assumption that $L_c$ is much larger than 15nm, which is applicable to our process. A Gaussian distribution truncated at $\pm 3\sigma_{CD}$ was assumed for the device gate length statistics. We first computed the $I_{on}$-$I_{off}$ curves of a smooth gate device and devices with increasing gate length RMS values. LER with a gate length RMS value over 4nm makes a significant difference on the universal curves shown in Fig. 6. It is clear that a device with smaller average physical gate length will be more sensitive to LER. Considering this worst case, we obtained the results for the leakage current of 50nm nominal gate length devices with different LER RMS values (shown in Fig. 7). The LER budget estimation can be obtained from this figure. The results are summarized in Table 1.
Table 1: LER budget estimation for devices designed with 50nm nominal gate length

<table>
<thead>
<tr>
<th>Control the leakage current spread of devices with the same nominal gate length</th>
<th>( \frac{I_{\text{off}, 3\sigma_{\text{CD}}}}{I_{\text{off, nom}}} )</th>
<th>( \frac{I_{\text{off}, 3\sigma_{\text{ED}}}}{I_{\text{off, nom}}} )</th>
<th>( \frac{I_{\text{off}, 1.5\sigma_{\text{CD}}}}{I_{\text{off, nom}}} )</th>
<th>( \frac{I_{\text{off}, 1.5\sigma_{\text{ED}}}}{I_{\text{off, nom}}} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Specified value</td>
<td>6</td>
<td>&lt; 4</td>
<td>3</td>
<td>&lt; 1.8</td>
</tr>
<tr>
<td>Line width RMS budget</td>
<td>3.5nm</td>
<td>Impossible because the across wafer CD variation</td>
<td>4.5nm</td>
<td>Impossible because the across wafer CD variation</td>
</tr>
<tr>
<td>Line edge RMS budget</td>
<td>2.5nm</td>
<td></td>
<td>3.2nm</td>
<td></td>
</tr>
</tbody>
</table>

In the above table, the line width RMS is assumed to be \( \sqrt{2} \) times of the edge RMS.

5. HALO IN EFFECTING LER SENSITIVITY

Linton, et al suggested controlling the increased leakage current by making the physical gate length slightly longer\(^1\). This causes some degradation of the drive current. Halo implantation is generally used to reduce short channel effects in short channel MOSFETs; it also helps to reduce the sensitivity of devices to gate LER. For our 50nm bulk device design, we intentionally doubled the halo dose and recalculated the LER effect on the leakage current from the 2D interpolation method. Comparing Fig. 8 with Fig. 7 we can see that the heavy halo dose can effectively reduce the device leakage current and pull it back to the range of control. The leakage sensitivity to LER amplitude is suppressed in a wider RMS value range. On the other hand, drive current dropped by about 12% in this specific case, which is probably acceptable. However, judicious optimization of halo dose, possibly combined with slight increase of minimum gate length can very likely increase the LER budget significantly with little penalty in drive current.

6. LOW FREQUENCY LER INDUCED INTRA-DIE CD FLUCTUATION

Typically, the largest CD variation occurs across wafer. However, local CD variation is observed for small device width. The local CD variation can be attribute to low frequency line edge roughness. If the width is 200nm, experimentally we observe a 3\( \sigma \) value of local CD variation of 3-4nm. However, if the width is doubled to 400nm, we measured a 3\( \sigma \) CD value of about 2-3nm. Although the intra-die CD variation is small compared with the
across wafer CD variation, since it happens locally it will cause local CD mismatch problems particularly in analog circuits. If we assume the power spectral density (PSD) of line edge roughness is normal, then it is straightforward to show that the variance of the mean CD is proportional to the inverse of the square root of the width that is measured. Fig. 9 illustrates this.

This means if for 200nm device width, the $3\sigma_{\text{CD}}$ value is 3nm, then if we want to get a moderate matching of $\frac{dL}{L} \sim 1\%$ for 50nm nominal gate length devices, the device width used has to be more than 7.2µm.

Fig. 9: Intra-die CD distribution for different device width

7. OBSERVATION AND CHARACTERIZATION OF LER

We developed our own software to extract line edge waveforms by processing SEM current data recorded from scans over poly lines. The SEM tool is OPAL from Applied Materials. Multi-box measurements were made on each poly line. A maximum number of 64 scans with less than 3nm minimum scan spacing along the lines can be obtained in every measuring box (see Fig. 10). The minimum pixel size in the transverse direction is about 0.8nm. Our method requires a large amount of SEM raw current data but does not require special equipment adjustment except for resolution considerations. The LER results are analysis of the poly line edge waveform data. After calibration, the method has good throughput and good CD agreement with the results from other methods. The LER results are repeatable after cleaning the wafer and re-measuring the same poly line. For the specific 50nm process, we studied the LER of poly lines with approximately 50nm CD. We can approximate them by a Gaussian distribution with slight modification. The observed maximum edge deviations from the means are quite consistently around three times the line edge RMS value.

Fig. 10: Extraction of line edge waveforms from SEM current data
Spectral properties of the LER can be obtained from the Fourier Transform of the line edge waveforms. The typical power spectrum and autocorrelation function are shown in Fig. 11. We used the autocorrelation length (Lc) defined in the plot as the characteristic length. Variation with period larger than Lc was observed to contribute over 80% of the total LER power.

8. EXPERIMENTS ON DEVICES WITH DIFFERENT LER

One of our goals is to experimentally study the electrical behavior of devices with different LER. The first step was to be able to produce different line edge roughness in a controlled process. Engineering the photo-resist process gave us pronounced differences in poly LER. Fig. 12 shows a SEM picture of poly lines formed in an optimized process in contrast with poly lines with rough edges. The extracted top line width RMS values are shown in Fig. 13. We obtained up to 4 ~ 5 fold increase of the top line width RMS value. By comparing the power spectra in Fig. 14, we can see more clearly that the power density of line edge variation with period larger than 30nm was effectively increased. After gate formation splits as illustrated in Fig. 12, devices will be fully processed normally. Measurements of the difference in electrical parameters of these devices will be compared with the simulations reported here.

9. SUMMARY

We have investigated the effect of gate LER on 50nm minimum gate length NMOSFETs using device simulations. For gate LER with Lc much larger than 15nm, and RMS<5nm, a 2D statistical model ignoring 3D effects can give satisfactory results. In other cases, the effort of 3D simulation cannot be saved, but we show how it can be minimized. By considering across wafer CD variation, the LER budget can be determined. Depending on the frequency properties
of LER produced in a specific process, LER effects may be a great concern in the sub-50nm technology. High frequency LER results in enhanced lateral diffusion, which adds to effective channel length reduction. Low frequency LER causes local CD fluctuation for devices with small device width, and as the result we have to choose either longer gate length or much wider width for devices in analog matching circuits. Specific device design parameters that control short channel effects such as halo doping, also influence LER sensitivity. A controlled experiment with different poly gate LER is in process to demonstrate the LER effects on the electrical parameters of 50nm bulk MOSFETs, which will be compared to our simulations.

ACKNOWLEDGEMENT

This work was supported by UC-SMART program under contract of 97-01. We also wish to acknowledge the long time support of TCAD vendors ISE and Avanti in providing their most advanced 3D simulation tools to make the 3D LER simulation possible.

REFERENCES

4. ISE TCAD software: A package of tools in lithography, process, device and circuit simulations from Integrated System Engineering (ISE). DESSIS is the tool for multi-dimensional device simulations.
5. Taurus: A multi-dimensional process simulator from Avanti.