Design for Suppression of Gate-Induced Drain Leakage in LDD MOSFET’s Using a Quasi-Two-Dimensional Analytical Model

Stephen A. Parke, Student Member, IEEE, James E. Moon, Member, IEEE, Hsing-jen C. Wann, Student Member, IEEE, Ping K. Ko, Member, IEEE, and Chenming Hu, Fellow, IEEE

Abstract—A systematic study of gate-induced drain leakage (GIDL) in single-diffusion drain (SD), lightly doped drain (LDD), and fully gate-overlapped LDD (GOLD) NMOSFET’s is described. Design curves quantifying the GIDL dependence on gate oxide thickness, phosphorus dose, and spacer length are presented. In addition, a new, quasi-2D analytical model is developed for the electric field in the gate-to-drain overlap region. This model successfully explains the observed GIDL dependence on the lateral doping profile of the drain. Also, a technique is proposed for extracting this lateral doping profile using the measured dependence of GIDL current on the applied substrate bias. Finally, the GIDL current is found to be much smaller in lightly doped LDD devices than in SD or fully overlapped LDD devices, due to smaller vertical and lateral electric fields. However, as the phosphorus dose approaches $10^{14}$/cm$^2$, the LDD and fully overlapped LDD devices exhibit similar GIDL current.

I. INTRODUCTION

MINIMIZATION of off-state leakage current in MOSFET’s has become an important issue with the advent of very-low-power, battery-based applications. DRAM cells require a corresponding reduction in leakage currents as the amount of stored charge decreases and the refresh interval increases with scaling. 256-Mb DRAM’s will require a total cell leakage on the order of 100 fA, of which the cell access transistor is only one component. Off-state MOSFET leakage typically consists of subthreshold channel current and surface/bulk thermal generation of carriers in the junction space-charge regions. The power supply voltage, subthreshold slope, junction temperature range, and junction area must all be reduced in order to suppress these leakage mechanisms. In addition, the gate oxide thickness must be scaled to the minimum value permitted by oxide reliability constraints. This minimizes the subthreshold slope and short-channel effects, while maximizing the saturation current and transconductance. However, the oxide and silicon fields increase correspondingly, especially in the gate-to-drain overlap region. In this region, the gate workfunction and high drain concentration serve to enhance the field strength. This large field depletes the surface of the heavily doped drain, giving rise to transport-limited thermal generation. This leakage is exponentially dependent on the drain-to-gate voltage, unlike conventional thermal leakage [1]. At somewhat higher fields, band-to-band tunneling occurs in the gate-overlapped, deep-depleted drain region. This has been proven to be the generation mechanism for Gate-Induced Drain Leakage (GIDL) [2]-[4]. At even higher fields, gate-to-drain leakage occurs, through a combination of Fowler-Nordheim tunneling of electrons from the gate to the drain and the injection of GIDL-generated hot holes into the gate [5], [6]. This off-state gate-to-drain current causes significant hole trapping in the oxide, leading to drain current 'walk-out' and degradation of the gate oxide [7], [8].

These leakage currents are very sensitive to the oxide thickness, drain concentration, and lateral drain doping gradient, as well as the applied drain-to-gate voltage. Various lightly doped drain (LDD) and graded gate oxide (GGO) structures have been studied as a solution for this off-state leakage problem [4], [9], [10]. Nearly complete suppression of the GIDL current has been demonstrated for low-concentration LDD and GGO devices. However, these devices exhibit poor hot-electron, reliability due to reduced gate control of the lightly doped drain region [11]. Fully gate-overlapped LDD devices have been proposed as a reliability solution. These devices have recently been investigated for their GIDL behavior as well [12], [13].

A one-dimensional, vertical field assumption was originally used to model the band-to-band tunneling current in the gate-to-drain overlap region [2]-[4]. However, the field in this region is inherently 2D. Therefore, 2D and
3D numerical simulations have also been used to model the GIDL current in LDD devices [9], [14], [15]. However, these simulations fail to give the intuitive understanding necessary for designers to optimize the LDD drain structure, and therefore a simple, analytical 2D model is needed.

This paper describes a systematic study of the dependence of GIDL on the LDD and fully gate-overlapped LDD design parameters. These parameters are gate oxide thickness, phosphorus dose, spacer length, and the lateral drain doping gradient. Single-diffusion (SD) devices were also studied as experimental controls. In addition, a simple analytical quasi-2D field model is developed for the off-state MOSFET. This model has successfully explained the observed experimental results for all devices studied [16]. In Section II, a description of the process variables and critical device fabrication steps is given. Section III presents the GIDL measurement results for each drain structure. In Section IV, the analytical, quasi-2D GIDL model is described and demonstrated. Section V describes a technique for extracting the lateral doping profile, using the substrate bias dependence of GIDL. Section VI shows the sensitivity of GIDL to the various LDD design parameters. Finally, a summary is given in Section VII.

### II. Experiment Design

Three types of experimental drain structures were fabricated, as shown in Fig. 1. Non-LDD, SD devices were fabricated using a $3 \times 10^{15}$ cm$^{-2}$, 80-keV arsenic drain implant. The final junction depth was about 0.2 μm. LDD devices were fabricated using phosphorus implants of 0.5, 1.5, 6, and 12 $\times 10^{15}$ cm$^{-2}$ at 60 keV. CVD oxide sidewall spacers were formed, with deposited thicknesses of 150, 200, and 250 nm, followed by a $3 \times 10^{15}$ cm$^{-2}$, 100-keV arsenic n$^-$ implant. The fully gate-overlapped LDD device is named Total-Overlap Poly Spacer (TOPS), and has been described previously [13]. The key process steps are shown in Fig. 2. A gate stack consisting of the gate oxide, a thin 50-nm polysilicon layer, a 10-nm CVD oxide etch stop layer, and a thick 270-nm poly-gate was deposited. The gate length was then defined by a photoresist ashing procedure. Channel lengths less than 0.2 μm have been achieved by this method. Following the thick poly etch, a 90-keV phosphorus implant was performed through the thin poly layer. Doses of 0.5, 1.5, and 6 $\times 10^{15}$ cm$^{-2}$ were used. The oxide etch stop was stripped, and a polysilicon sidewall spacer layer was deposited. Spacer thicknesses of 150, 200, and 250 nm were fabricated. An anisotropic etch was used to cut through both poly layers, forming the poly sidewall spacers which electrically connect the top and bottom poly layers. Finally, a $3 \times 10^{15}$ cm$^{-2}$, 80-keV arsenic n$^+$ implant was performed. Gate oxide thicknesses of 8.5, 11, and 15 nm were fabricated. The experimental devices are summarized in Table I. They were all fabricated in the UC Berkeley Microfabrication Laboratory.
TABLE I

<table>
<thead>
<tr>
<th>Wafer</th>
<th>Oxide Thickness (nm)</th>
<th>Phosphorus Dose ($\times 10^{12}$/cm$^2$)</th>
<th>Spacer Length (nm)</th>
<th>$G_{sd}$ ($\times 10^{3}$/cm)</th>
<th>$V_{MAX}$ (V)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SD25</td>
<td>8.5</td>
<td>—</td>
<td>—</td>
<td>2.66</td>
<td>3.60</td>
</tr>
<tr>
<td>SD201</td>
<td>15.0</td>
<td>—</td>
<td>—</td>
<td>2.01</td>
<td>4.30</td>
</tr>
<tr>
<td>LDD09</td>
<td>8.5</td>
<td>6</td>
<td>150</td>
<td>6.80</td>
<td></td>
</tr>
<tr>
<td>LDD10</td>
<td>8.5</td>
<td>12</td>
<td>150</td>
<td>5.80</td>
<td></td>
</tr>
<tr>
<td>LDD11</td>
<td>8.5</td>
<td>6</td>
<td>250</td>
<td>7.40</td>
<td></td>
</tr>
<tr>
<td>LDD12</td>
<td>8.5</td>
<td>12</td>
<td>250</td>
<td>7.30</td>
<td></td>
</tr>
<tr>
<td>LDD13</td>
<td>11.0</td>
<td>6</td>
<td>150</td>
<td>8.10</td>
<td></td>
</tr>
<tr>
<td>LDD14</td>
<td>11.0</td>
<td>12</td>
<td>150</td>
<td>7.30</td>
<td></td>
</tr>
<tr>
<td>LDD15</td>
<td>11.0</td>
<td>6</td>
<td>250</td>
<td>8.50</td>
<td></td>
</tr>
<tr>
<td>LDD16</td>
<td>11.0</td>
<td>12</td>
<td>250</td>
<td>7.30</td>
<td></td>
</tr>
<tr>
<td>LDD202</td>
<td>15.0</td>
<td>0.5</td>
<td>200</td>
<td>11.70</td>
<td></td>
</tr>
<tr>
<td>LDD203</td>
<td>15.0</td>
<td>1.5</td>
<td>200</td>
<td>11.00</td>
<td></td>
</tr>
<tr>
<td>TOPS01</td>
<td>8.5</td>
<td>3</td>
<td>150</td>
<td>1.58</td>
<td>4.50</td>
</tr>
<tr>
<td>TOPS02</td>
<td>8.5</td>
<td>6</td>
<td>150</td>
<td>1.26</td>
<td>4.95</td>
</tr>
<tr>
<td>TOPS03</td>
<td>8.5</td>
<td>3</td>
<td>200</td>
<td>1.67</td>
<td>4.45</td>
</tr>
<tr>
<td>TOPS04</td>
<td>8.5</td>
<td>6</td>
<td>200</td>
<td>1.31</td>
<td>4.90</td>
</tr>
<tr>
<td>TOPS05</td>
<td>8.5</td>
<td>3</td>
<td>250</td>
<td>1.41</td>
<td>4.65</td>
</tr>
<tr>
<td>TOPS06</td>
<td>8.5</td>
<td>6</td>
<td>250</td>
<td>1.00</td>
<td>5.15</td>
</tr>
<tr>
<td>TOPS13</td>
<td>11.0</td>
<td>3</td>
<td>150</td>
<td>1.18</td>
<td>5.65</td>
</tr>
<tr>
<td>TOPS14</td>
<td>11.0</td>
<td>6</td>
<td>150</td>
<td>1.03</td>
<td>6.25</td>
</tr>
<tr>
<td>TOPS15</td>
<td>11.0</td>
<td>3</td>
<td>200</td>
<td>1.04</td>
<td>5.80</td>
</tr>
<tr>
<td>TOPS16</td>
<td>11.0</td>
<td>6</td>
<td>200</td>
<td>0.97</td>
<td>5.95</td>
</tr>
<tr>
<td>TOPS17</td>
<td>11.0</td>
<td>3</td>
<td>250</td>
<td>0.66</td>
<td>6.55</td>
</tr>
<tr>
<td>TOPS18</td>
<td>11.0</td>
<td>6</td>
<td>250</td>
<td>1.94</td>
<td>4.55</td>
</tr>
<tr>
<td>TOPS206</td>
<td>15.0</td>
<td>0.5</td>
<td>200</td>
<td>1.64</td>
<td>5.10</td>
</tr>
<tr>
<td>TOPS208</td>
<td>15.0</td>
<td>1.0</td>
<td>200</td>
<td>1.31</td>
<td>6.40</td>
</tr>
<tr>
<td>TOPS210</td>
<td>15.0</td>
<td>3</td>
<td>200</td>
<td>1.31</td>
<td></td>
</tr>
</tbody>
</table>

III. EXPERIMENTAL RESULTS

Typical subthreshold characteristics for the three device types are shown in Fig. 3 for $V_{G}$ = 5.0 V and $t_{ox}$ = 8.5 nm. Since the drain leakage is not a function of channel length, all measurements were performed on $W = 50 \mu m$, $L = 10 \mu m$ NMOSFET's. Large GIDL currents are observed for negative $V_{G}$, while the normal subthreshold behavior is observed for positive $V_{G}$. The grounded-gate configuration ($V_{G} = 0$) is the condition of interest. The oxide-spaced LDD exhibits very low GIDL, while the SD device exhibits severe GIDL. The gate-overlapped TOPS device behavior is in-between. Data from the same three devices is shown in Fig. 4, where the log of the leakage current per micrometer of device width is plotted as a function of the applied drain voltage. The gate and bulk were grounded, while the source was left floating to insure that no subthreshold current was measured. This is the typical measurement for observing GIDL. The results of the new model, to be described in Section IV, are plotted as solid curves in Fig. 4. An excellent fit is achieved over most of the voltage range. At low drain voltages, the leakage current is dominated by transport-limited thermal generation [1], which is near or below the measurement noise level for a $W = 50 \mu m$ device at room temperature. At high drain voltages, hot-hole trapping in the gate oxide occurs, which reduces the silicon field and resulting GIDL current below the model. At even higher voltages (not shown in the figure), avalanche multiplication of the combined thermal leakage and GIDL leads to a sharp rise in $I_{d}$. $I_{d}$ is eventually limited by the series resistance of the device.

![Fig. 3. Subthreshold characteristics for the SD, TOPS, and LDD devices at $V_{d} = 5$ V and $t_{ox} = 8.5$ nm. GIDL current is observed below threshold. The LDD device exhibits substantially less GIDL than the SD device.](image)

![Fig. 4. GIDL characteristics for the devices from Fig. 3. The new, quasi-2D model fits the experimental data very well, up to voltages where hot-hole trapping in the oxide causes "walk-out." The criterion for $V_{MAX}$ is defined as $I_{d} = 0.1 \mu A/\mu m$.](image)

The maximum drain voltage that can be sustained for a 0.1-\mu A/\mu m leakage criterion is defined as the figure of merit for comparing the various device designs. Note that this criterion may need to be even lower for 256-Mb DRAM's or low-power applications. As seen in Fig. 4, for an oxide thickness of 8.5 nm, the SD device can sustain only 3.6 V while the TOPS and LDD devices can sustain 5.1 and 6.2 V, respectively. The maximum sustainable voltage ($V_{MAX}$) for this leakage criterion is shown in Fig. 5, as a function of gate oxide thickness. For a given drain concentration, $V_{MAX}$ varies linearly with $t_{ox}$ if only the vertical field is considered. $V_{MAX}$ decreases as the drain concentration ($N_{D}$) increases, since the vertical field is proportional to $\sqrt{N_{D}}$. This vertical field model predicts that a critical oxide field of only 1.9 MV/cm will result in 0.1 \mu A/\mu m of leakage [3], while the new quasi-2D model to be described predicts this critical field to be 5.6 MV/cm, in the absence of any lateral field. When the lateral field is included, the behavior becomes nonlinear and $V_{MAX}$ drops below the 5.6 MV/cm line.
Fig. 5. The maximum sustainable voltage as a function of $t_{OX}$. The previous 1D model predicted $E_{crit}$ of 1.9 MV/cm, while the new model predicts $E_{crit}$ = 5.6 MV/cm, for zero lateral field. As the lateral field $G$ increases, $V_{MAX}$ decreases and becomes a non-linear function of $t_{OX}$. As $t_{OX}$ becomes small, $V_{MAX}$ approaches $E_B + V_{FB}$ and becomes less sensitive to the lateral field.

Fig. 6. $V_{MAX}$ as a function of the drain concentration, as predicted by the 1D model with $E_{crit}$ = 5.6 MV/cm. The LDD data match the $\sqrt{N_D}$ dependence. However, the TOPS and SD data are scattered and fall below the model, indicating the presence of a contributing lateral field in these devices. The lightly doped LDD samples exhibit avalanche current at 12 V before reaching $V_{MAX}$.

For a large enough drain concentration, the band bending at the gate edge is no longer sufficient to permit tunneling ($\psi_S < 1.2$). This is seen in Fig. 6, where $E_{crit} = 5.6$ MV/cm is assumed. This concentration is defined as $N_{crit}$, and is found to be $9 \times 10^{18}$/cm$^3$. As $N_D$ is increased above $N_{crit}$, the maximum tunneling point (at which $\psi_S = 1.2$) moves away from the gate edge, toward the metallurgical junction, and $V_{MAX}$ becomes insensitive to $N_D$. This is seen on the right side of Fig. 6.

Since both the SD and TOPS devices have a region where the gate overlaps the $N_D = 10^{19}$ arsenic junction, the 1D model predicts that their $V_{MAX}$ data points should fall on the 5.6 MV/cm line in Fig. 5 and on the straight-line portions of Fig. 6. However, the measured data points lie significantly below these lines. As mentioned in [3], some variation in $V_{MAX}$ is expected between different processes due to varying amounts of oxide "birdsbeak" at the gate edge. However, these devices were all fabricated with the same post-gate process, thus requiring another explanation for the variations. We also observed that the TOPS device $V_{MAX}$ actually increases with increasing phosphorus dose, opposite to the LDD behavior. Also, both the LDD and TOPS devices exhibit increasing $V_{MAX}$ as the spacer length is increased. The existing 1D, vertical field models, summarized in Fig. 7(a), are unable to explain these results.

IV. QUASI-2D MODEL

These newly observed results have been successfully modeled by 1) adding a lateral electric field term, and 2) using an indirect band-to-band tunneling formulation. The new, quasi-2D model is summarized in Fig. 7(b). For devices where an abrupt, $n^+$ arsenic junction is overlapped by the gate (SD, TOPS), a substantial, built-in lateral field exists and must be added to the vertical field. This increases the GIDL current and lowers $V_{MAX}$. In fact, this built-in lateral field can exceed 1 MV/cm for shallow arsenic junctions with $x_j < 0.2$ μm, and is comparable to the vertical field strength. The vertical field is given by

$$E_{VERT}(y) = \frac{\epsilon_{OX} (V_{DG} - \psi_S(y))}{\epsilon_S}$$

(1)

where $V_{DG} = V_{DS} - V_{FB}$ and $y$ is the lateral position as measured from the gate edge. The new lateral field term is simply defined as the gradient of the surface potential $\psi_S(y)/dy$. Since the tunneling occurs mainly at the surface, it is assumed that the drain concentration is not a function of depth. Applying the depletion approximation, the surface potential may be expressed as

$$\psi_S(y) = V_{DG} + \psi_0(y) - \sqrt{2V_{DG}\psi_0(y) + \psi_0(y)^2}$$

(2)

where

$$\psi_0(y) = \frac{q_x N_0(y)}{\epsilon_{OX}} \left( \frac{1}{C_{OX}} \right).$$

Differentiating $\psi_S$ with respect to $y$, the lateral field expression simplifies to

$$E_{LAT}(y) = \frac{d\psi_S}{dy} = G(y)\psi_S(y) \frac{V_{DG} - \psi_S(y)}{V_{DG} + \psi_S(y)}$$

(3)

where $G(y)$ is defined as

$$G(y) = \frac{d(ln(N_0(y)))}{dy}.$$
The right-hand fraction in (3) is plotted as a function of $V_{DG}$ in Fig. 8. For large drain voltages, this term approaches unity, but for voltages of interest, it is less than 0.5. Therefore, for a given $V_{DG}$, the lateral field is simply proportional to the product of the lateral doping gradient and the surface potential, at each position $y$ in the overlap region. The total field is then approximated by the vector sum of the vertical and lateral field components

$$E_{TOT}(y) = \sqrt{E_{VERT}(y)^2 + E_{LAT}(y)^2}. \tag{4}$$

This is equivalent to treating the lateral field as a perturbation to the vertical field, and is only strictly valid for $E_{LAT} \ll E_{VERT}$, hence the name “quasi-2D.” However, this simplified model has been found to work well even when $E_{LAT}$ is nearly equal to $E_{VERT}$. The effect of including the lateral field is shown in Fig. 5 for various values of the lateral gradient $G$. The SD devices exhibit an effective $G$ value of $2.5 \times 10^6$/cm, while the TOPS devices $G$ values range between 1.0 and $2.0 \times 10^6$/cm.

Indirect, phonon-assisted tunneling is also assumed in this new formulation. It is well known that indirect tunneling is, in fact, the predominant tunneling mechanism in silicon. Previous GIDL models adopted the direct formulation possibly because the experimentally obtained values for $B$ (ignoring the lateral field) matched the direct tunneling theoretical value of 21.3 MV/cm more closely than the indirect tunneling value of 36.2 MV/cm.

In order to extract an experimental value for $B_I$, we chose the LDD devices with long sidewall spacers as a reference, and assumed that they had negligible lateral field ($G = 0$). This assumption is reasonable since the phosphorus gradient is several times smaller than the $n^+$ arsenic gradient, and no gate overlap of the arsenic exists in these devices. As shown in Fig. 9, a linear fit to the tunneling characteristics for these devices yields an experimental value of $B_I = 45$ MV/cm. This value is much larger than previously reported values due to inclusion of the lateral field in the model. However, it is not unreasonable when compared with the indirect tunneling theoretical value. This value of $B_I$ fits well for all devices, independent of their drain doping profile and oxide thickness. However, the coefficient $A_I$ appears to be proportional to $f_{ox}$, as can be seen from the $y$ intercepts in Fig. 9. This dependence is expected from the equation for $A_I$, given in Fig. 7.

Fig. 10 shows the application of this analytical, quasi-2D model to SD, LDD, and TOPS device. A large lateral field is observed in the SD device, while the LDD device has a negligible lateral field in the region of interest. For the SD and TOPS devices, the electric field and tunneling current peaks are located several nanometers from the gate edge, near the position with $N_D = N_{sat}$. For the LDD device, the tunneling current peak is located exactly at the gate edge, where the vertical field is maximum. The phosphorus and arsenic profiles add in the TOPS device, and this interaction of the two profiles reduces the field contribution from the arsenic gradient in the region of $N_{sat}$. Therefore, the lateral field and GIDL current are reduced and $V_{MAX}$ is increased, with respect to the SD device.

V. Substrate Bias Dependence

The quasi-2D model provides a way to extract the lateral doping profile in the gate–drain overlap region, by observing the variation of the GIDL current with respect to the substrate bias. It is known that the GIDL current is affected by the substrate bias through formation of an inversion layer in the gate–drain overlap region [18]. According to the quasi-2D model, the surface potential is still mainly determined by the vertical field. Varying the substrate voltage will not modify the surface potential significantly. This assumption remains valid as long as the region is deep-depleted. Once an inversion layer is formed, band bending larger than the bandgap is no longer possible, since the surface potential is pinned at $2|\psi_0|$. Hence no tunneling can occur in this region. The total GIDL current is then given as

$$I_t = W \times \sum_{j=0}^{y_{sat}+1.2} J(y) \, dy \tag{5}$$
Fig. 10. Demonstration of the quasi-2D model for sample SD, LDD, and TOPS devices. The drain cross section, doping profile, surface potential, electric fields, and tunneling current density are shown for each device, as a function of position from the gate edge. The SD and TOPS devices exhibit sharp lateral field peaks associated with their abrupt arsenic junctions, leading to higher GIDL current.
where the $y$(inv) stands for the position at which the inversion layer is formed. As the substrate voltage is raised, $y$(inv) moves from the metallurgical junction toward the gate edge. The maximum doping concentration which can be inverted with a substrate voltage $V_B$, is given as follows:

$$N_{inv} = \frac{C_{ox}(V_{BG} - 2|\psi_n|)^2}{2\varepsilon_s q(V_{DB} - 2|\psi_n|)}.$$  \hspace{1cm} (6)

The exact position of this concentration depends on the drain profile. Fig. 11 shows the GIDL current as a function of the substrate voltage for the SD, LDD, and TOPS devices with the gate grounded and $V_g = 6.0$ V. The SD and TOPS devices have higher GIDL currents than the LDD device, and this current drops more significantly at high $V_B$. Both of these devices have dominant regions which contribute most of the tunneling current. This "dominant tunneling point" is located where the band bending is just above 1.2 eV and becomes inverted when $V_B$ is large (about 1 V smaller than $V_D$). After this point is inverted, the GIDL current is reduced significantly. For the LDD device, there is no major dominant tunneling point. Thus the tunneling current density decreases gradually from the gate edge. As $V_B$ increases, the inversion layer approaches the gate edge and a gradual reduction of the GIDL current, rather than a more abrupt change (as in the SD and TOPS devices) is observed. Fig. 12 compares the measured SD device behavior with the quasi-2D model. Good agreement is shown. Based on the one-to-one correspondence between the lateral doping profile and the $I_D$-$V_B$ dependence, the lateral doping profile can be extracted from the experimental data. The tunneling current density, evaluated at the boundary of the inversion region, is given by

$$J|_{N_D = N_{inv}} = \frac{G}{G_{N_{inv}}} \frac{dI_D}{dV_B}.$$  \hspace{1cm} (7)

where the right-hand term is simply the incremental slope of Fig. 12. The doping gradient at $N_{inv}$ is given by

$$G_{N_{inv}} = \frac{d\ln(N_{inv})}{dV_B} = \frac{1}{V_{DB} + 2|\psi_n|} + \frac{2}{V_{BG} - 2|\psi_n|}.$$  \hspace{1cm} (8)

By solving for $G$ at each $V_B$, an inverse function of the doping profile is found

$$y(N_D) = \int_0^{V_B} \frac{G_{N_{inv}}}{G} dV_B.$$  \hspace{1cm} (9)

The basic idea behind these calculations is to change variables from $y$ to $V_B$. When $V_B$ is increased by $\Delta V_B$, the experimentally observed reduction of $I_D$ corresponds to $J(y) \Delta V_B G_{N_{inv}} / G$. Therefore, $N_D(y)$ can be reconstructed. However, care must be taken with this technique since the solution will be sensitive to the extracted constants $A$ and $B$ in the tunneling equation.

VI. GIDL SENSITIVITY TO LDD DESIGN PARAMETERS

The sensitivity of GIDL to the gate oxide thickness has already been quantified in Fig. 5. As $t_{OX}$ scales, $V_{MAX}$ approaches 1.2 V for all devices, and the lateral field becomes less important. It should be noted that this is actually the oxide thickness at the drain position corresponding to $N_{CRIT}$, which could be significantly larger than the channel $t_{OX}$ if a large "birdsbeak" exists at the gate edge.

The sensitivity of GIDL to the phosphorus dose is summarized in Fig. 13. LDD devices exhibit decreasing $V_{MAX}$ for doses less than $1 \times 10^{14}$/cm², due to increasing vertical field. On the other hand, TOPS devices exhibit increasing $V_{MAX}$, due to the decreasing lateral gradient caused by the interaction of the phosphorus profile with that of the arsenic. For $N_D > N_{CRIT}$, the TOPS and LDD devices behave similarly.

As seen in Fig. 10, the tunneling current peaks very sharply at a given position in the gate overlap region. As-
Fig. 13. $V_{\text{MAX}}$ as a function of $n^-$ phosphorus dose. As the dose approaches $1 \times 10^{14} \text{cm}^{-2}$, the LDD devices exhibit decreasing $V_{\text{MAX}}$ due to increasing vertical field. For low doses, the TOPS devices behave like SD devices, with low $V_{\text{MAX}}$. However, as the dose increases, the TOPS $V_{\text{MAX}}$ increases due to reduction of the arsenic lateral field by the $n^-$ phosphorus. The quasi-2D model fits both data well.

Fig. 14. The effective lateral doping gradient increases linearly with increasing phosphorus dose for LDD devices. For TOPS devices with long spacers, $G_{\text{eff}}$ decreases from the arsenic value to zero when $N_{\text{D}} = N_{\text{As}}$. This corresponds to a dose of $1 \times 10^{14} \text{cm}^{-2}$. TOPS and LDD devices with very short spacers behave similarly.

associated with this dominant tunneling position is an effective lateral doping gradient $G_{\text{eff}}$. The value of $G_{\text{eff}}$ has been extracted for each device by fitting the model to the experimental data, assuming $B = 45 \text{ MV/cm}$. These values are given in Table 1. In Fig. 14, this effective gradient is plotted as a function of the phosphorus dose. As long as the LDD spacer length is longer than the lateral diffusion of the $n^-$ arsenic junction, no dependence on the arsenic junction or the sidewall spacer length are observed. However, when the spacer is short enough, the arsenic profile encroaches on the phosphorus profile. This causes $G_{\text{eff}}$ to increase toward the arsenic value of $2.5 \times 10^6 \text{cm}^{-1}$, as either the spacer length or phosphorus dose are reduced. When the spacer is very short, the LDD and TOPS devices are essentially identical, as seen in Fig. 14. The TOPS device with a long spacer has a slightly different behavior. In this case, the phosphorus concentration reaches a plateau under the gate. The gradient drops steeply as the phosphorus dose is increased. When the plateau concentration exactly equals $N_{\text{As}}$, the gradient drops to nearly zero. This occurs for a dose of $1 \times 10^{14} \text{cm}^{-2}$. The gradient then rises to the phosphorus value as the dose continues to increase.

The effect of the spacer length on GIDL is shown in Fig. 15. $V_{\text{MAX}}$ is independent of $L_s$ for spacers longer than the arsenic lateral diffusion (150 nm). As $L_s$ approaches zero, both the LDD and TOPS devices approach the behavior of the SD device, with a $V_{\text{MAX}}$ of 3.6 and 4.3 V for $t_{\text{OX}} = 8.5$ and 11 nm, respectively. The spacer “lengths” plotted in Fig. 15 are actually the deposited film thicknesses before performing the anisotropic etch. The final physical spacer lengths are somewhat shorter. For both LDD and TOPS devices, it is clearly desirable to keep the spacer length longer than the arsenic lateral diffusion length, in order to minimize GIDL and maximize $V_{\text{MAX}}$.
VII. CONCLUSIONS
Hot-electron reliability, current drive, short-channel effects, and device parasitics (Rd and Cgd) must be considered along with the GIDL in order to optimize the deep-submicrometer LDD MOSFET. Typically, a larger n+ concentration is used for the LDD device than for the fully overlapped LDD (TOPS) device, in order to minimize LDD series resistance and minimize TOPS overlap capacitance. For n+ concentrations less than 9 x 10^{18}/cm^3, this means that the LDD device will continue to have lower GIDL current than TOPS or conventional SD devices. However, for a larger n+ concentration and/or thinner gate oxide, the differences between these devices diminishes. A new, quasi-2D GIDL model has been successfully applied to a wide variety of drain profiles and oxide thicknesses. The built-in lateral field has been found to play an important role in the dependence of GIDL on the drain doping profile and spacer length.

ACKNOWLEDGMENT
The authors wish to thank J. Chen for his helpful discussions throughout the course of this work.

REFERENCES

Stephen A. Parke (S'80) was born in Evansville, IN, on October 15, 1960. He received the A.A. degree from Olivet Nazarene University, Kankakee, IL, in 1980, and the B.S. and M.S. degrees from Purdue University, West Lafayette, IN, in 1982 and 1984.

He was with the IBM Thomas J. Watson Research Center, Yorktown Heights, NY, in 1983, where he worked on an advanced microprocessor design. In 1984, he joined IBM in Essex Junction, VT, where he worked on DRAM cell design and array peripheral circuits for the 4-, 16-, and 64-Mb DRAM designs. In 1989, he was awarded the IBM Resident Study Fellowship and began pursuing the Ph.D. degree at the University of California at Berkeley. He is currently engaged in LDD CMOS and SOI BiCMOS device and technology research at UC Berkeley.

James E. Moon (S'87-M'89-S'89-M'90) received the B.S. degree in chemical engineering in 1976 from Carnegie-Mellon University, Pittsburgh, PA. The M.B.A. degree in 1981 from the University of Rochester, Rochester, NY, with concentration in finance and applied economics, and the M.S. and Ph.D. degrees in electrical engineering from the University of California at Berkeley in 1988 and 1990, respectively.

Since 1976 he has been employed by the Eastman Kodak Company, Rochester, NY, having had design and development responsibilities in a variety of product lines. From 1985 to 1990 he was on assignment at the University of California at Berkeley as the recipient of the Doctoral Award from Kodak. His doctoral research was focused on high-speed deep-submicrometer MOSFET devices and technology. He is currently Director of New Business Opportunities in the Professional Photography Division at Kodak. His responsibilities include definition and development of electronic and hybrid imaging systems.

Hsing-jen C. Wann (S'90) received the B.S. degree from National Taiwan University in 1988. He started his graduate study at University of California, Berkeley in 1990 and expects to receive the M.S. degree in 1992. His current research projects include SOI MOSFET device physics, thin dielectric material, and nonvolatile memories devices. He was a Book Coupon Award recipient at NTU.

Mr. Wann is currently the president of the Chinese Institute of Engineers, student chapter of U.C. Berkeley.
Ping K. Ko (S’78-M’81-S’82-M’82-M’89) received the B.S. degree in physics with special honors from Hong Kong University in 1974, and the M.S. and Ph.D. degrees in electrical engineering from the University of California at Berkeley, in 1978 and 1982, respectively.

During 1982 and 1983, he was a Member of Technical Staff at Bell Laboratories, Holmdel, NJ, and was responsible for developing high-speed MOS technologies for communication circuits. He joined the University of California at Berkeley in 1984, where he is now Professor and Vice Chairman of the Department of Electrical Engineering and Computer Sciences. He is also the director of the Berkeley Microfabrication Laboratory. His present research interests include high-speed VLSI technologies and devices, device modeling for circuit simulation, and electronic neural network. He has authored or co-authored one book and over 100 research papers.

Dr. Ko has served on the program committees of the International VLSI Technology Symposium and the International Electron Device Meeting. He was Associate Editor of IEEE Transactions on Electron Devices from 1988 to 1990.

Chenming Hu (S’71-M’76-SM’83-F’90) received the B.S. degree from the National Taiwan University and the M.S. and Ph.D. degrees in electrical engineering from University of California, Berkeley, in 1970 and 1973, respectively.

From 1973 to 1976 he was an Assistant Professor at Massachusetts Institute of Technology, Cambridge. In 1976 he joined the University of California, Berkeley, as Professor of Electrical Engineering and Computer Sciences. He is Director of Joint Services Electronics Program at Berkeley. While on industrial leave from the University in 1980–1981 he was manager of nonvolatile memory development at National Semiconductor. Since 1973 he has served as a consultant to the electronics industry. He has also been an advisor to many government and educational institutions. His present research areas include VLSI devices including silicon-on-insulator devices, hot-electron effects, thin dielectrics, electromigration, circuit reliability, simulation, and nonvolatile semiconductor memories. He has also conducted research on electrooptics, solar cells, and power electronics. He has been awarded several patents on semiconductor devices and technology. He has authored or co-authored three books and over 250 research papers. He has delivered a dozen keynote addresses and invited papers at scientific conferences, and has received five best paper awards. He is an Honorary Professor of Beijing University and Tsinghua University, China and of the Chinese Academy of Science. He was appointed the first National Science Council Invited Chair Lecturer, Republic of China, in 1987. Dr. Hu was Board Chairman of East San Francisco Bay Chinese School from 1988 to 1991. He received the 1991 Design News Excellence in Design Award for leading the development of an IC Reliability Simulator, BERT.


PARKE et al.: SUPPRESSION OF GATE-INDUCED DRAIN LEAKAGE IN LDD MOSFET'S