

Impact of Gate-to-Contact Spacing on ESD Performance of Salicided Deep Submicron NMOS Transistors

Kwang-Hoon Oh, *Student Member, IEEE*, Charvaka Duvvury, *Senior Member, IEEE*, Kaustav Banerjee, *Member, IEEE*, and Robert W. Dutton, *Fellow, IEEE*

Abstract—Electrostatic discharge (ESD) failure threshold of NMOS transistors is known to degrade with the use of silicided diffusions owing to insufficient ballast resistance, making them susceptible to current localization, which leads to early ESD failure. In general, the gate-to-contact spacing of salicided devices is known to have little impact on their ESD strength. However, experimental results presented in this paper show that the ESD strength depends on the gate-to-contact spacing independent of the silicided process. Subsequently, a detailed investigation of the influence of gate-to-source and gate-to-drain contact spacings is carried out for a salicided 0.13- μm technology which provides new insight into the behavior of deep submicron ESD protection devices. It is shown that the reduction in current localization and increase in the power dissipating volume with increase in the gate-to-contact spacings are the primary causes of this improvement, which implies that even for silicided processes, the gate-to-contact spacing should be carefully engineered for efficient and robust ESD protection designs.

Index Terms—Ballast resistance, CMOS technology, electrostatic discharge, gate-to-contact spacing, NMOS transistor, n-p-n transistor, silicides, substrate bias, thermal capacity.

I. INTRODUCTION

NMOS transistors are widely used as protection devices against electrostatic discharge (ESD), which is a major reliability concern for all categories of integrated circuits [1]. It is well known that for nonsilicided or silicide-blocked NMOS protection transistors, the second breakdown triggering current (It_2), which is widely used to monitor the ESD strength, can be increased with larger drain contact spacing because of more uniform triggering of the lateral n-p-n structure obtained with ballast resistance¹ effects [1]. In addition, it is also well established that effectiveness against ESD is reduced in the case

of devices with silicided diffusions [1], since the ballast resistance is negligible. In silicided CMOS processes, the primary cause of the degradation of ESD failure threshold is known to be nonuniform lateral bipolar conduction, which is attributed to insufficient ballasting resistance in the fully silicided source/drain structures [2]. This decrease in ESD strength imposes severe restrictions on the efficient design of ESD protection. Therefore, to avoid localized current conduction and improve It_2 , device structures with sufficient ballasting resistance are realized by introducing the silicide blocking option, or by implementing well resistors on the drain side, or by inserting local interconnect layers [3]–[5]. However, these options either require an extra mask or more process complexity and result in increased process cost and chip area. Hence, use of salicided² devices is often preferred for cost effectiveness in providing advanced ESD protections. The It_2 of the silicided devices is generally believed to be independent of the gate-to-contact spacing parameter. Hence, the impact of the gate-to-contact spacing of a salicided NMOS transistor on the ESD failure strength has not been fully explored.

However, contrary to conventional understanding, for advanced deep submicron salicided technologies with shallow trench isolation (STI) structures, we have recently reported that the gate-to-drain contact spacing has an impact on It_2 [6]. Moreover, the gate-to-source contact spacing has also been observed to affect It_2 [6]. This work investigates the above new phenomenon in advanced salicided transistors and describes the different mechanisms that are observed at the source and drain sides, respectively. The physical mechanism causing unexpected It_2 improvement in silicided devices with increased gate-to-contact spacing is identified. Furthermore, it is shown that the ESD strength of the protection device becomes independent of the gate-to-contact spacing when adequate substrate bias is applied. These observations have significant implications for ESD performance improvement simply through optimization of the device layout, even without introducing expensive process options.

II. EXPERIMENTS

A. It_2 Dependence on Contact Spacing

In this work, the dependence of second breakdown triggering current It_2 [mA/ μm] is investigated as a function of the gate-to-source/drain salicided (CoSi₂) contact spacing

²A self-aligned silicide process.

Manuscript received February 25, 2002; revised July 8, 2002. This work was supported by Texas Instruments, Inc., Dallas, TX. The review of this paper was arranged by Editor C.-Y. Lu.

K.-H. Oh and R. W. Dutton are with the Center for Integrated Systems, Stanford University, Stanford, CA 94305 USA (e-mail: okhoon@gloworm.stanford.edu; dutton@ee.stanford.edu).

C. Duvvury is with Silicon Technology Development, Texas Instruments, Dallas, TX 75243 USA (e-mail: c-duvvury@ti.com).

K. Banerjee is with the Department of Electrical and Computer Engineering, University of California, Santa Barbara, CA 93106 USA (e-mail: kaustav@ece.ucsb.edu).

Digital Object Identifier 10.1109/TED.2002.803627

¹Ballast resistance effect makes the ESD currents flow more uniformly in the drain diffusion region.

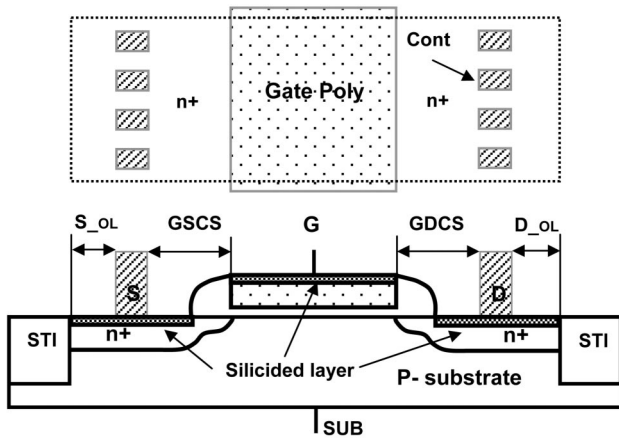


Fig. 1. Schematic of a silicided NMOS transistor indicating the gate to source/drain contact spacing (GSCS/GDCS) and the n^+ overlap of the source/drain contact (S_{OL}/D_{OL}). For the 1.5-V NMOS, S/D contact opening (S/D) = $0.15\ \mu\text{m}$ and $S_{OL} = D_{OL} = 0.1\ \mu\text{m}$. For the 3.3-V NMOS, the S/D contact opening (S/D) = $0.15\ \mu\text{m}$ and $S_{OL} = D_{OL} = 0.125\ \mu\text{m}$. In the test structures, the S/D contact opening and n^+ overlap of S/D contact (S_{OL}/D_{OL}) remain unchanged despite the variations of GSCS/GDCS.

(GSCS/GDCS) for various test structures with grounded gates. Test structures were fabricated using a $0.13\text{-}\mu\text{m}$ mixed-voltage CMOS technology. Two types of transistors are investigated in this study: low (1.5 V) and high (3.3 V) voltage NMOS transistors with different gate oxide thickness and drawn channel lengths. The 1.5-V NMOS transistor has a $27\text{-}\text{\AA}$ -thick gate oxide and $0.175\text{-}\mu\text{m}$ -long gate poly, while the 3.3-V NMOS transistor has a $70\text{-}\text{\AA}$ -gate oxide and $0.5\text{-}\mu\text{m}$ -long gate poly. However, the finger width for both transistors is $20\ \mu\text{m}$. Fig. 1 shows the schematic cross section of the ESD NMOS transistor used in this study. The contact spacing is measured from the gate poly edge to the near edge of contact opening and the contact opening width is fixed at $0.15\ \mu\text{m}$ for all the test structures. Since the test structure uses a shallow trench isolation (STI) with constant n^+ overlap with the source and drain contact (S_{OL}/D_{OL}) for a given device rating, the total size of the source and drain structures are changed proportional to the variation of the GSCS and GDSCS. The second breakdown triggering current (It_2) was measured using the transmission line pulsing (TLP) method for a voltage pulse width of 200 ns. As expected, since the resistance of silicided region is relatively small compared to other parasitic resistances in the source/drain structure, a change in resistance proportional to the GDSCS/GSCS is not apparent from the dc current-voltage (I - V) measurements. The drive current (I_{drive}) of the 1.5-V NMOS transistor was tested with GDSCS/GSCS variations. As shown in Fig. 2, the measured drive currents show that the difference in the resistance due to increased gate-to-contact spacing with the silicided diffusion is hardly apparent. In addition, the inset in Fig. 2 shows a sample of TLP curves for the 1.5-V transistor with two different GDSCS and GSCS. It can be clearly seen that the It_2 value is about doubled with an increase in GDSCS/GSCS from $0.1\ \mu\text{m}$ to $0.5\ \mu\text{m}$. However, the slope of the high current regions is almost identical for the two

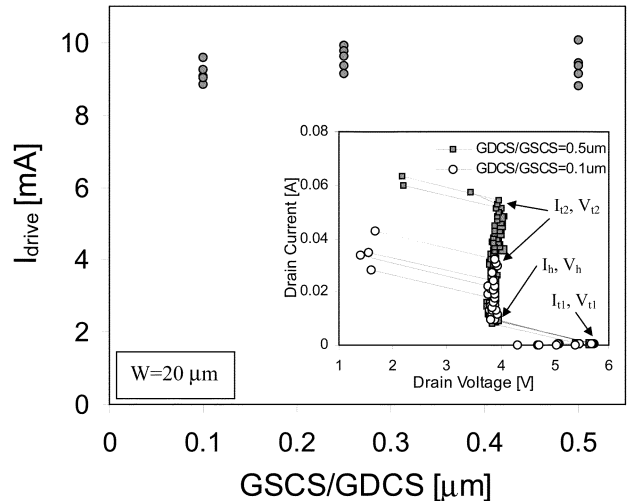


Fig. 2. Measured drive currents (I_{drive}) for the 1.5-V NMOS transistors with different gate-to-contact spacings show no apparent differences, where $V_{gs} = V_{ds} = 1.5\ \text{V}$. The inset shows high-current TLP curves for the 1.5-V NMOS transistor, which clearly show the impact of the gate-to-contact spacings on It_2 despite the silicided diffusion.

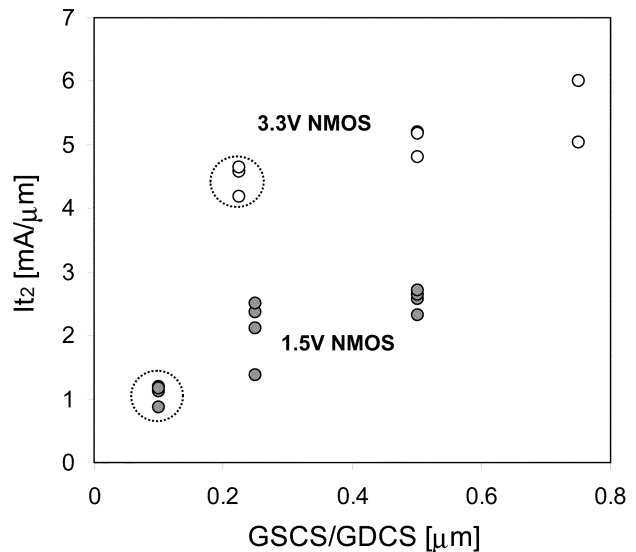


Fig. 3. Second breakdown triggering current (It_2) with the gate-to-source/drain contact spacing for two different silicided NMOS transistors. The two dotted circles indicate It_2 for each device with minimum contact spacing: $0.1\ \mu\text{m}$ and $0.225\ \mu\text{m}$ for the 1.5-V NMOS ($W/L_{poly} = 20/0.175\ \mu\text{m}$) and 3.3-V NMOS ($W/L_{poly} = 20/0.5\ \mu\text{m}$), respectively.

different test structures, which implies that the on-resistance of the test structures is nearly the same despite the different gate-to-contact spacings.

For the TLP measurements of the 1.5-V and 3.3-V transistors as shown in Fig. 3, It_2 of both the devices surprisingly improves with the GSCS/GDCS. Compared with It_2 values of the transistor with the minimum gate-to-contact spacing ($0.1\ \mu\text{m}$ for 1.5-V NMOS and $0.225\ \mu\text{m}$ for 3.3-V NMOS), the improvement of It_2 is approximately 100% and 40% for the low-voltage and high-voltage transistors, respectively. This implies that the

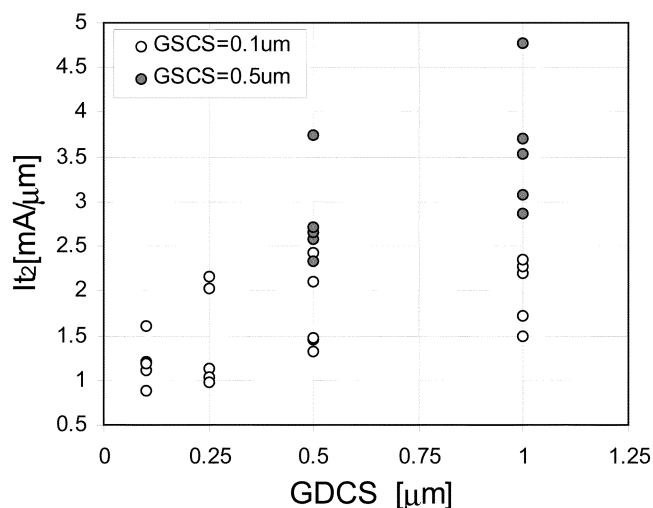


Fig. 4. It_2 values for the 1.5-V NMOS transistors with various GDCS and GSCS. It_2 depends on both GDCS and GSCS within the scatter of data. $W = 20 \mu\text{m}$.

gate-to-contact spacing is an important design parameter determining ESD strength for the gate grounded ESD protection devices. It also suggests the possibility of achieving increased ESD robustness through optimizing the layout of the silicided protection devices without any extra processing steps or structure options. However, the primary cause of this improvement of It_2 has not been explored and needs comprehensive modeling and analysis in order to improve understanding of the device physics involved in this effect. This will also enable the establishment of robust ESD protection design approaches through proper design of the devices.

As shown in Fig. 3, for the 1.5-V NMOS transistors, the It_2 dependence on the gate-to-contact spacing is more apparent than that for the 3.3-V devices. In general, for advanced devices with silicided diffusion, the improvement of It_2 is not easily achieved due to early failure caused by current localization. In this regard, any amount of improvement in It_2 for silicided technology with no process changes has significant implications. As can be seen in Fig. 4, It_2 values are influenced by both GDCS and GSCS within the scatter of data. Despite the silicided process, the increasing trend of It_2 is obvious as GDCS and GSCS increase. For the minimum gate-to-contact spacing of $0.1 \mu\text{m}$, It_2 ranges from 1 to $1.5 \text{ mA}/\mu\text{m}$. However, with the increased GDCS = $1 \mu\text{m}$ and GSCS = $0.5 \mu\text{m}$, It_2 values are clustered around $3.5 \text{ mA}/\mu\text{m}$. The data clearly show that for a given spacing (GDCS), increase in GSCS strongly affects the ESD hardness. This experimental result implies that the impact of gate-to-source contact spacing is as significant as that of the gate-to-drain contact spacing for a silicided technology.

B. Effect of Substrate Bias

For silicided devices, it has been reported that It_2 improves with a forward substrate bias by enlarging the turn-on finger width of the devices [2]. With sufficient external substrate bias ($V_{\text{sub}} > 0.7 \text{ V}$), the It_2 dependence on the gate-to-contact spacing disappears as the emitter-base junction of the lateral bipolar transistor fully turns on (Fig. 5). As shown in [2], the phenomenon of It_2 improvement is associated with the extent

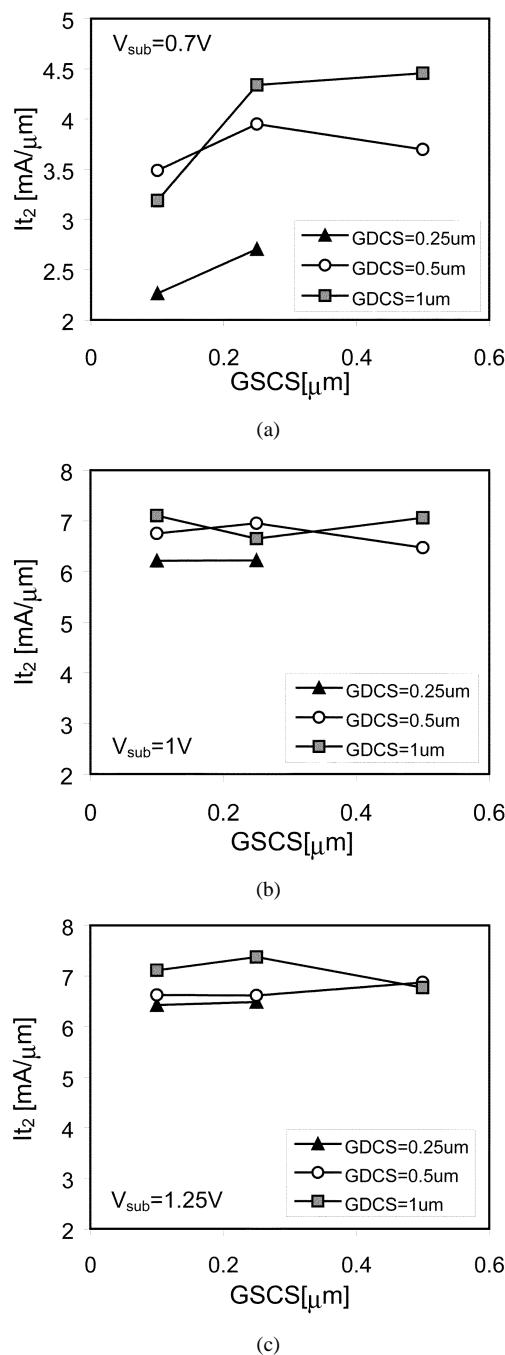


Fig. 5. (a) It_2 of the 1.5-V NMOS transistors is dependent on GSCS and GDCS with an external substrate bias of 0.7 V , and is nearly independent of GSCS and GDCS with sufficient external substrate bias of (b) 1 V and (c) 1.25 V .

of uniformity of the lateral current distribution, since the ESD current becomes more uniform along the channel width with increased external substrate bias. In addition, it should be noted that the It_2 dependence on the gate-to-contact spacing for the 1.5-V transistor is stronger, since the extent of uniformity of the lateral current distribution of the two devices varies (Fig. 6). It is believed that the devices with shorter channel length and shallower junction depth experience nonuniform bipolar conduction more strongly since the relative sensitivity to the statistical random variation is higher for given process conditions. In order to identify the underlying physical mechanisms for the

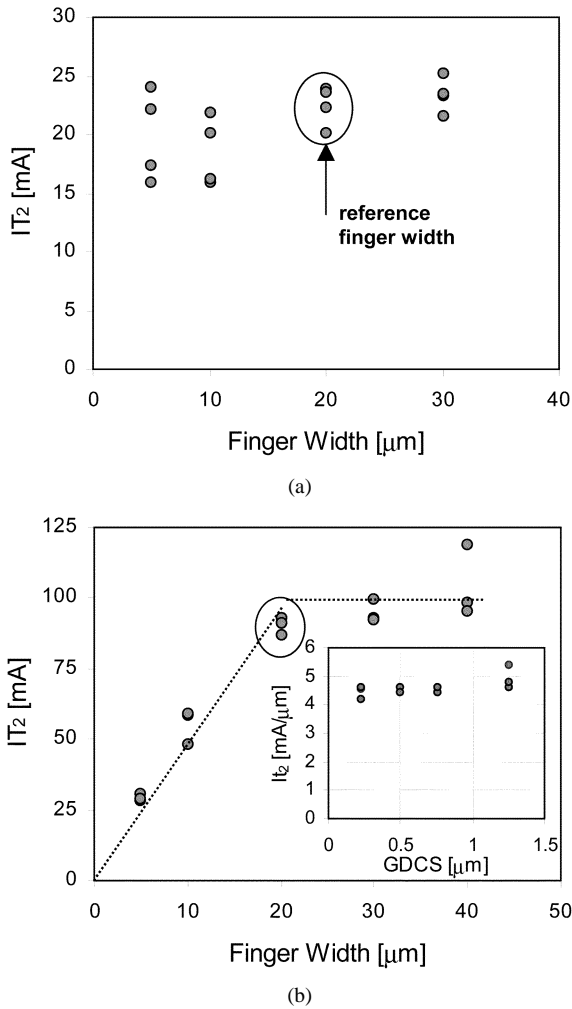


Fig. 6. Total failure threshold current (IT_2) with the finger width for (a) 1.5-V NMOS transistors and (b) 3.3-V NMOS transistors where GDCS/GSCS = $0.1 \mu\text{m}$. The total failure current can be scaled with finger width within only a limited finger width ($W \leq 20 \mu\text{m}$) and the inset shows that the 3.3-V device ($W = 20 \mu\text{m}$ and GSCS = $0.225 \mu\text{m}$) with nearly uniform current conduction shows negligible dependence of It_2 on GDCS.

drain side and source side contact spacing effects respectively, the following effects were investigated: 1) the influence of the ballasting ESD current distribution; 2) the improvement of the current driving capability of the lateral n-p-n transistor; and 3) the increase in thermal capacity due to the enlargement of power dissipating volume along with the increase in GSCS/GDCS. Each of these effects has been investigated in detail in the next section.

III. ANALYSIS AND DISCUSSION

A. Ballasting Current Distribution

As shown in Fig. 2, from the dc and TLP measurements for 1.5-V NMOS transistors, the drive current and the on-resistance show no differences between the structures with the minimum gate-to-contact spacing and the ones with increased gate-to-contact spacing as expected. First, to analyze the impact of the GDCS, for a fixed GSCS of $0.1 \mu\text{m}$, It_2 was measured for the structures with GDCS of 0.1 and $1 \mu\text{m}$, respectively. As can be observed from Fig. 7, consistent with Fig. 2 (inset), even with

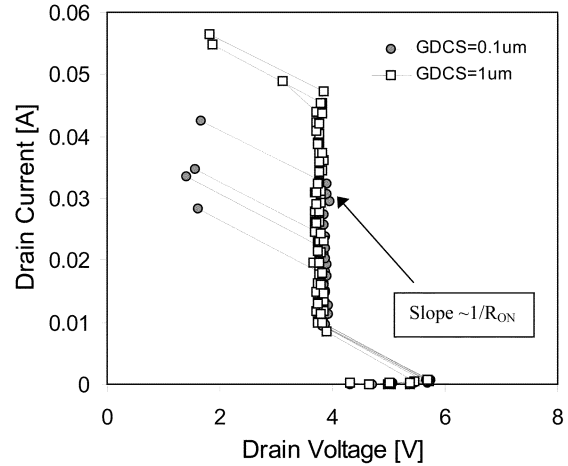


Fig. 7. High current I - V curves for the 1.5-V NMOS transistor with two different GDCS of $0.1 \mu\text{m}$ and $1 \mu\text{m}$ where GSCS is $0.1 \mu\text{m}$. The slope of high-current regions is almost identical, but It_2 is nearly doubled with the increased GDCS.

further increase in GDCS, the on-resistance remains unaffected. However, It_2 is nearly doubled with GDCS = $1 \mu\text{m}$.

It is instructive to note that, contrary to the experimental results presented in this work, it has been reported in the past that the on-resistance is changed depending on the silicide thickness and junction depth [4]. Since it is well known that the silicide contact resistance is a strong function of interfacial doping concentration (Silicide/Si), the silicide contact resistance is influenced by the variation of the silicide thickness [7]. Moreover, the junction depth can also affect the drain diffusion resistance. Therefore, it is likely that the impact of these factors was observed as slight changes in the on-resistance in [4], even with silicided structures.

Nevertheless, based on the empirical results presented in this work, it can be conjectured that the increase of GDCS alleviates the current localization problem by further expanding the turned on portion of the finger width, which appears only for devices with strongly nonuniform current distributions under ESD stress. For devices with various finger widths, total failure threshold currents, IT_2 [mA], were measured for the low and high voltage transistors. As can be seen in Fig. 6, the total failure threshold current does not scale with the drawn finger width for the 1.5-V transistor. For the 3.3-V transistor, IT_2 scales only for a limited range of widths. This implies that the ESD current distribution for 1.5-V NMOS devices is highly nonuniform while it is nearly uniform for the narrow 3.3-V NMOS transistors ($W \leq 20 \mu\text{m}$). The It_2 data shown in Fig. 6(b) support the observation that only devices with uniform ESD current distribution do not show dependence on GDCS. Thus, it can be concluded that the increase in GDCS helps spread out the ESD current more uniformly along the finger width, which leads to improvement in the effective value of It_2 , though the increase in the ballasting resistance with the GDCS is hardly noticeable. This is apparent from Fig. 3 and from the total failure threshold current (IT_2) data in Fig. 6(a) for the 1.5-V NMOS devices where the ESD current distribution is strongly nonuniform. In addition, for low voltage transistors, with adequate external substrate bias, the dependence of It_2 on GDCS disappears as shown in Fig. 5. This result also supports the argu-

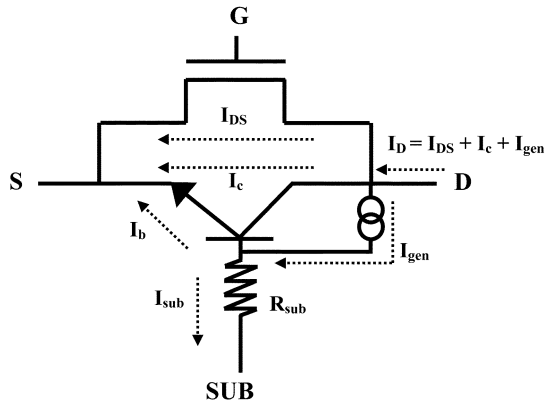


Fig. 8. Equivalent circuit of the NMOS including the parasitic lateral n-p-n bipolar transistor when the lateral n-p-n is on. I_{DS} is the channel current, I_{gen} is the avalanche-generated current, and I_{sub} is the substrate current.

ment that the increase in GDCS alleviates current localization for low voltage transistors. In other words, it seems that for the salicided devices, the shorter the gate-to-drain contact spacing is, the stronger the current localization. Therefore for the devices requiring higher ESD strength, minimum gate-to-contact spacing should be avoided unless substrate bias can be used in the protection circuit design [8].

B. Characteristics of the Lateral n-p-n

In general, the ESD hardness of NMOS devices can be described in terms of the primary device parameters of the parasitic lateral n-p-n transistors, such as the current gain (β), avalanche multiplication factor (M), and effective substrate resistance (R_{sub}) [9]–[11]. When the lateral n-p-n turns on (Fig. 8), the M , β , and R_{sub} are given by [10]

$$M = \frac{(\beta + 1)I_D}{\beta(I_D - I_{sub})} \quad (1)$$

$$\beta = \frac{1}{(M - 1) - M \left(\frac{I_{sub}}{I_D} \right)} \quad (2)$$

$$R_{sub} \approx \frac{0.8}{I_{sub}}. \quad (3)$$

Despite variation of the GSCS/GDCS, R_{sub} can be assumed to remain constant since the substrate doping concentration remains unchanged. In addition, the substrate contact is designed 20 μm away from the STI boundary at the source side and the maximum variation in the distance of substrate contact from the gate poly edge can be 0.65 μm due to increase in GSCS. This 0.65- μm variation in the distance of substrate contact will have a negligible impact on the effective substrate resistance value. Therefore, the current gain and avalanche multiplication are of primary interest for studying the impact of variations of gate-to-contact spacing on I_{t2} . The lateral n-p-n operation depends on a combination of M , β , and R_{sub} for a given power dissipation. For a given ESD current, less avalanche multiplication and higher current gain is preferable for higher ESD strength, since strong avalanche multiplication results from high fields which in turn, leads to locally higher temperatures. The variation of the avalanche multiplication for the test structures can be observed by employing two-dimensional (2-D) device (MEDICI) simulation. It should be noted that in the test structures used in this

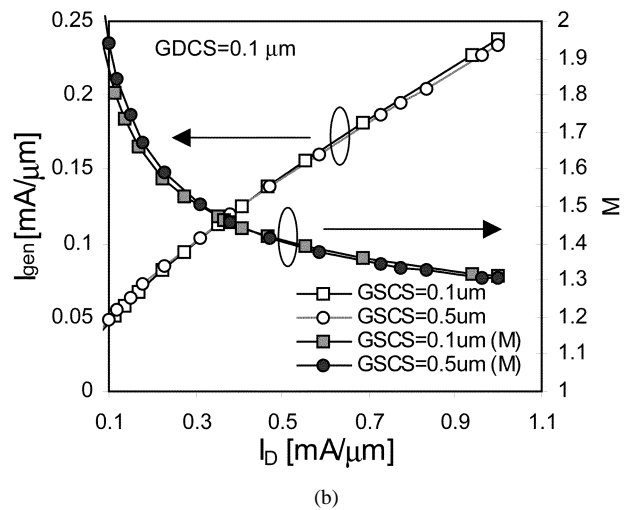
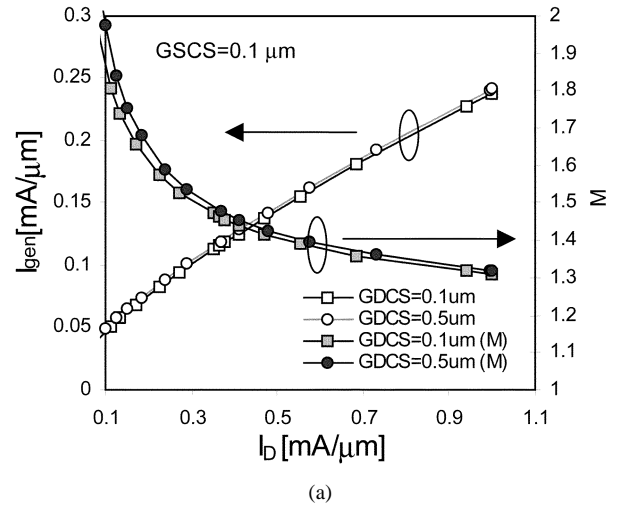


Fig. 9. Avalanche-generation current (I_{gen}) and multiplication factor (M) for the variation of (a) the GDCS and (b) GSCS.

study, as the contact spacing is increased, the source or drain area did not remain constant since the distance from the edge of the gate poly to the STI boundary on the source and drain side also simultaneously increased (Fig. 1). In addition, in the simulations, the silicide layers are treated as virtual electrodes and the effective substrate resistance of 5 $\text{k}\Omega/\mu\text{m}$ is attached to the bottom substrate contact. For the thermal boundary conditions, a thermal electrode is defined at the bottom of the substrate and the temperature of this thermal electrode is assumed to be the same as the ambient temperature. It is important to note that we assume the simple thermal boundary conditions for the simulations since the results of the thermal simulations are intended to provide only a relative comparison between different structures for a given thermal boundary condition. Furthermore, from the thermal point of view, the size of the simulation structure should be fixed, otherwise the thermal boundary conditions change. Using the dc current sweep simulation mode, high-current characteristics were reproduced for the structures with minimum gate-to-contact spacings (GSCS/GDCS = 0.1 μm) and also for longer GDCS (=0.5 μm) or GSCS (=0.5 μm). As shown in Fig. 9, the avalanche-generated current (I_{gen}) and multiplication factor (M) are compared for the three cases. As the drain

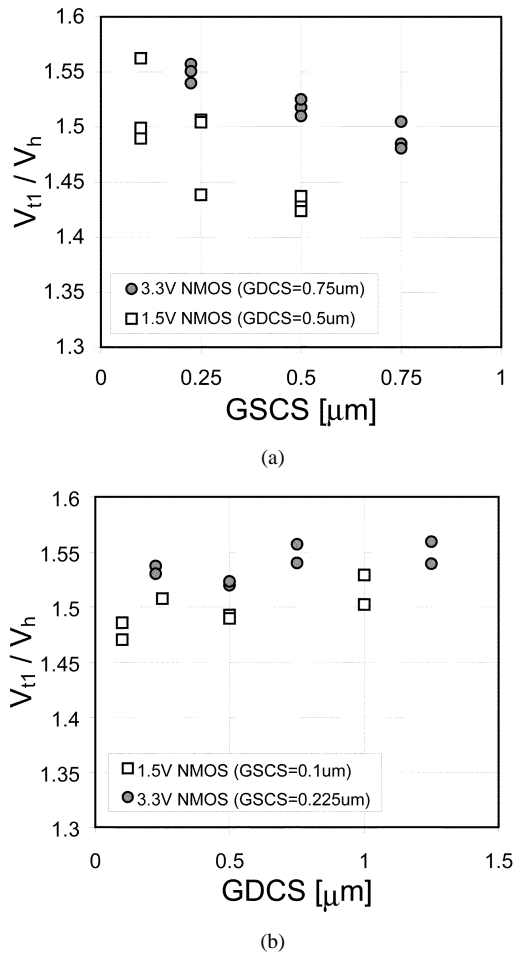


Fig. 10. To track the current gain of a lateral n-p-n transistor, the triggering voltage (V_{t1}) and holding voltage (V_h) are measured for the 1.5-V and 3.3-V NMOS transistors. The data clearly show that the current gain is reduced by (a) increase in GSCS while it (b) remains nearly constant with GDCCS.

current increases the generation current increases while the multiplication factor decreases rapidly, regardless of the gate-to-contact spacing. Based on the simulation results, variations of the gate-to-contact spacing appear to have no impact on the avalanche process.

The increase in the current gain of the lateral n-p-n transistor can result in an improvement of It_2 by conducting more current for a given ESD stress. In order to track the current gain of a lateral n-p-n transistor with the variation of GSCS and GDCCS, the ratio of the triggering voltage (V_{t1}) and holding voltage (V_h) is monitored for various test structures as shown in Fig. 10, since the current gain of a lateral n-p-n transistor (for a gate grounded NMOS) in a self-biasing mode is proportional to the V_{t1}/V_h ratio [12]

$$(1 + \beta)^{1/n} = \frac{\beta V_{CBO}}{\beta V_{CEO}} \approx \frac{V_{t1}}{V_h} \quad (4)$$

where BV_{CBO} is the common-base breakdown voltage (with emitter open-circuited), BV_{CEO} is the breakdown voltage for the common-emitter configuration (with base open-circuited), and n is a constant. For a constant GDCCS of both the low- and high-voltage devices, the ratio of V_{t1} and V_h decreases with increase in GSCS. However, compared with the V_{t1}/V_h values

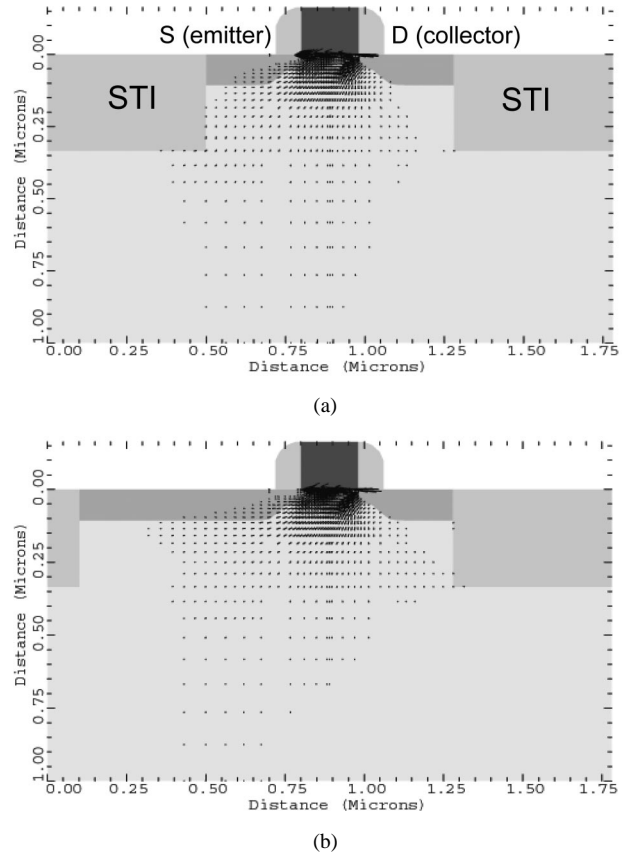


Fig. 11. Base current (hole current) density vector at $I_D = 1 \text{ mA}/\mu\text{m}$ for the device with (a) minimum gate to contact spacing, $GSCS/GDCCS = 0.1 \mu\text{m}$ and (b) increased gate to source contact spacing, $GSCS = 0.5 \mu\text{m}$. As GSCS increases, wider emitter and base junction is utilized for the current conduction and this results in a drop in the current gain due to increase in base current for a given collector current.

for GSCS variations, the V_{t1}/V_h values are nearly independent of the GDCCS variations. As described earlier, for the test structures, size of the source/drain is increased as the GSCS/GDCCS increases. Therefore, the effective area of the emitter (source) of the lateral n-p-n is enlarged with increase in the GSCS, and in turn, the effective current path is also increased. As shown in Figs. 11 and 12, for a given collector current, with the increased effective size of the emitter, more base current (hole current component) flows into the emitter for a given generation current (Fig. 11), which results in a slight decrease in the current gain (Fig. 12).

However, despite the decreased current gain with GSCS, improvement in It_2 is observed with GSCS. Consequently, the increase of It_2 with the GSCS is not attributed to the lateral n-p-n current gain. These results suggest that the mechanism for improvement of It_2 with the gate-to-source contact spacing variations is not dependent on the efficiency of lateral parasitic n-p-n transistor. In fact, the efficiency of the lateral n-p-n for conducting ESD current seems to be degraded or unchanged with increases in gate-to-source contact spacing. Therefore the main cause of the improvement in It_2 due to increase in GSCS is still unclear. Hence, to gain better insight into the physical behavior, we next explored possible thermal effects involved in the gate-to-source contact spacing by performing electrothermal simulations.

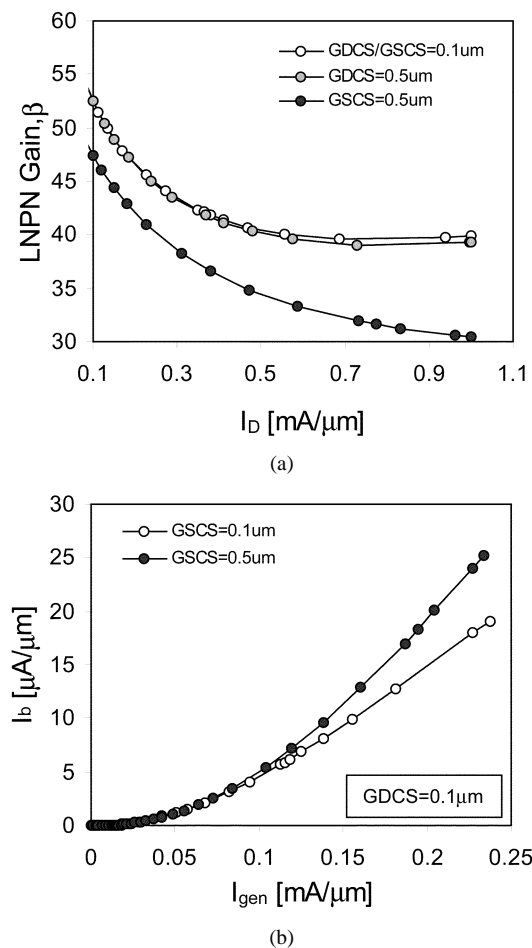


Fig. 12. (a) Current gain (β) versus the drain current (I_D) and (b) the base current (I_b) versus the generation current (I_{gen}) for the device with different gate-to-contact spacings.

C. Thermal Effects

Before investigating any thermal effects involved in the mechanism of It_2 improvement with gate-to-source contact spacings, it is important to discern any possible role of the uniformity of current distribution (arising solely due to increase in GSCS or the effective emitter area of the lateral n-p-n transistor) as the primary physical mechanism responsible for the improvement. As shown in Fig. 6(b), the current distribution is nearly uniform for the 20 μm wide high-voltage transistors and increase in GDCS has no impact on It_2 since the current is already uniformly distributed. Using the same argument, increase in GSCS should also have little impact on the uniformity and on It_2 . Nevertheless, the It_2 data for high voltage devices in Fig. 3 clearly show the impact of GSCS even though the current is already uniform (since $W = 20 \mu\text{m}$). Hence, it can be concluded that the observed improvement of It_2 with GSCS is not primarily due to any improved uniformity of current distribution. The above arguments are particularly valid for the high-voltage (3.3 V) transistors.

In order to investigate any increase in thermal capacity due to enlargement of power dissipating volume along with the increase in GSCS/GDCS, the temperature distributions in the device have been compared using electrothermal simulations. Current flowlines and temperature distribution contours for

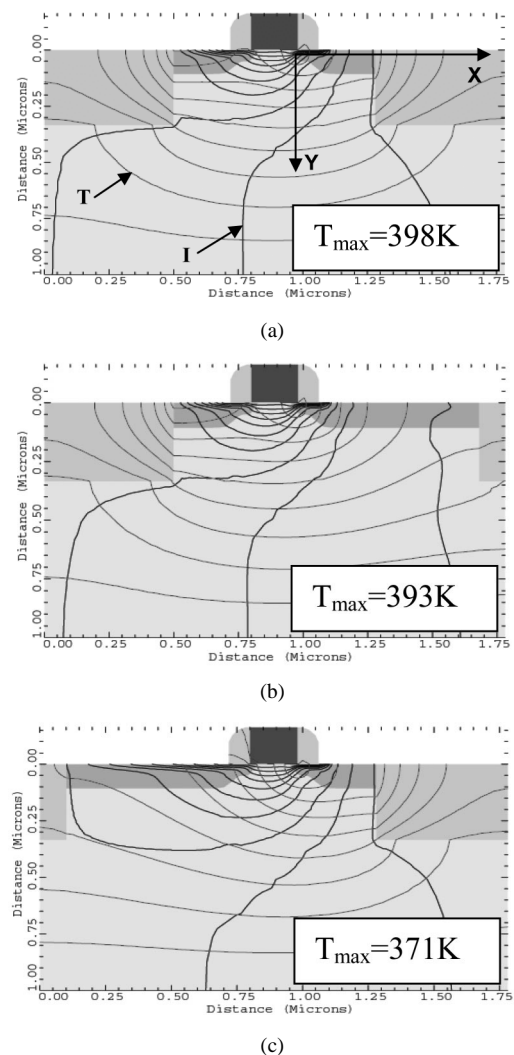


Fig. 13. Current flowlines and temperature distribution contours at the drain current of 1 $\text{mA}/\mu\text{m}$ for the three different structures are shown. (a) GSCS = 0.1 μm and GDCS = 0.1 μm , (b) GSCS = 0.1 μm and GDCS = 0.5 μm , and (c) GSCS = 0.5 μm and GDCS = 0.1 μm .

a drain current (I_D) of 1 $\text{mA}/\mu\text{m}$ in the three different structures are shown in Fig. 13. According to the simulation results, the location of the maximum temperature in the device remains the same, despite differences in the maximum temperature value itself. Due to the higher thermal resistance of STI structures and reduced thermal conductivity of upper passivation layers, the heat generated in the device under the ESD stress is confined and mostly dissipated through the substrate. Therefore, changes in the STI boundary associated with variations of GSCS/GDCS influence the overall temperature distribution and the peak temperature in the device as well. In Fig. 14, we show the effect of the gate-to-contact spacings on the temperature distribution. The temperature for the device with longer GSCS ($=0.5 \mu\text{m}$) is significantly lower than that of the two other structures with GDCS of 0.1 and 0.5 μm . The augmented power dissipating volume for the larger GSCS results in a lower peak temperature for a given drain current. Thus for the device with larger GSCS, a higher ESD failure threshold can be obtained due to a reduction in the peak temperature. Therefore, under ESD conditions, the maximum temperature of the device for a given

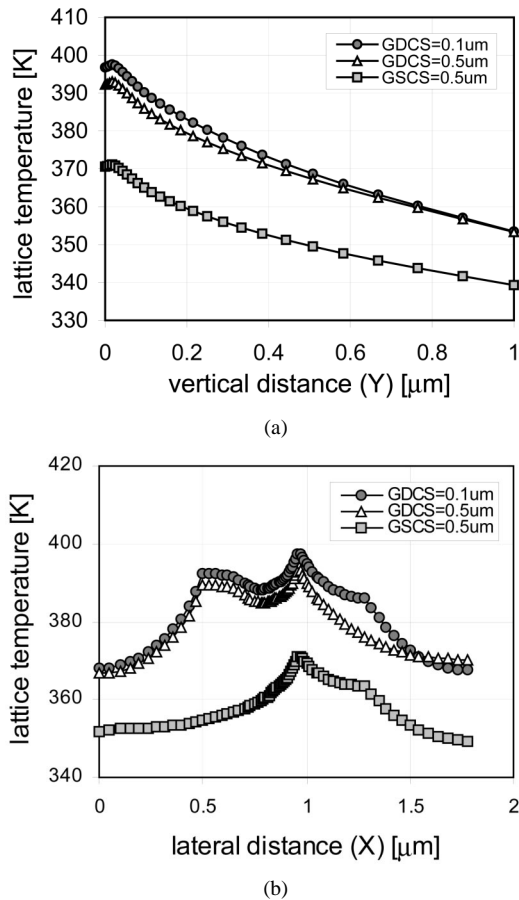


Fig. 14. Temperature distribution along the x and y direction [as indicated in Fig. 13(a)] at the drain current of $1 \text{ mA}/\mu\text{m}$ for the three different structures: (a) vertical temperature distribution and (b) lateral temperature distribution.

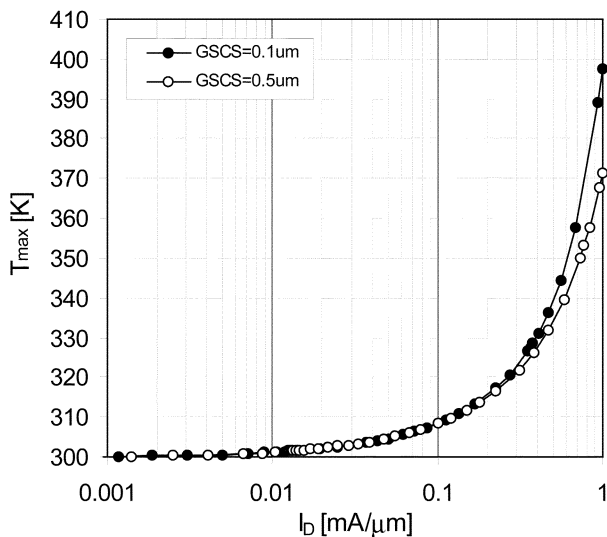


Fig. 15. Simulated maximum temperature for the two different test structures with injected drain current. The maximum temperature increases more rapidly as the power dissipating volume decreases for the shorter gate-to-source contact spacing. $\text{GDCS} = 0.1 \mu\text{m}$.

drain current is higher due to reduction in the effective source size resulting from smaller GSCS. The simulated maximum temperature with the drain current is also plotted in Fig. 15.

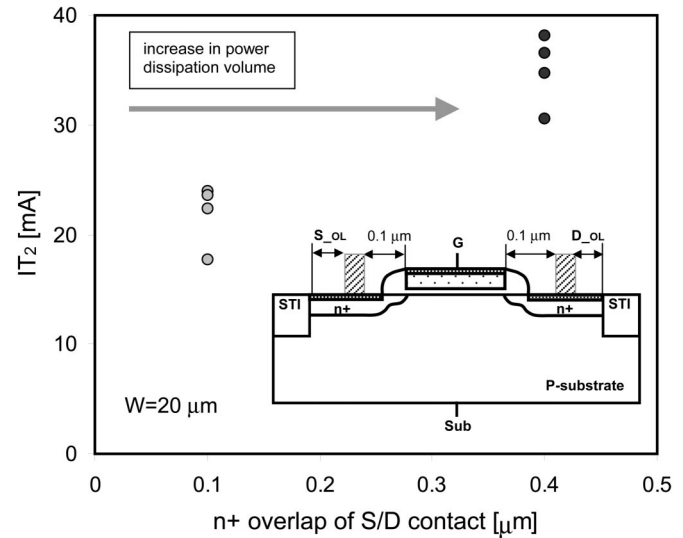


Fig. 16. Total failure threshold current (IT_2) for the two different test structures having different n^+ overlap lengths of S/D contact of $0.1 \mu\text{m}$ and $0.4 \mu\text{m}$ (D_{OL} and S_{OL}). The inset shows the schematic of the test structure. $\text{GDCS}/\text{GSCS} = 0.1 \mu\text{m}$.

The maximum temperature increases more rapidly with the drain current as the power dissipating volume decreases for the shorter gate-to-source contact spacing. Note that a significant difference in the maximum temperature can be expected at higher drain currents. To provide further support for the thermal effect involved in the gate-to-contact spacing, the total failure threshold currents (IT_2) of the 1.5-V test devices having two different power dissipating volumes (arising due to increased n^+ overlap length of S/D contacts) are shown in Fig. 16. It clearly indicates the dependence of IT_2 on the size of the power dissipating volume. For the device with S_{OL}/D_{OL} of $0.1 \mu\text{m}$, the total failure current IT_2 is less than 25 mA. However, with increase in the power dissipating volume due to the extension of S_{OL}/D_{OL} , significant improvement in the failure current can be obtained. It should be noted that this improvement of IT_2 for low-voltage transistors is not attributed to any improved uniformity of current distribution since the gate-to-source/drain contact spacing is unchanged for both the test structures in Fig. 16. The experimental result in Fig. 16 agrees with the predictions based on electrothermal simulations (Figs. 13–15). Both sets of data suggest that thermal effect is the root cause of IT_2 improvement with increase in GSCS. Therefore, it can be concluded that the observed improvement of IT_2 with GSCS for the low voltage transistors is also primarily due to thermal effects.

Also, it should be noted from Fig. 5(b) and (c) that both GSCS and GDCS have little impact on IT_2 under substrate bias condition, which makes the current distribution more uniform for the low voltage transistors. As shown in [2], the current flowlines, with substrate bias, spread out more uniformly and deeper into the substrate compared with the flowlines without substrate bias. Hence, Fig. 5 implies that the volume associated with current (or temperature) distribution is significantly increased by substrate bias and that this volume is substantially larger than the one arising due to increase in GSCS. This is reflected in the

dramatic improvement of It_2 in Fig. 5(b) and (c), while showing almost no sensitivity to GSCS and GDCS.

Finally, the overall physical mechanisms involved in the gate-to-contact spacing can be summarized. Although the changes in ballast resistance cannot be observed directly, increases in the gate-to-drain contact spacing are effective in mitigating severe nonuniform ESD current conduction. The increase in the GDCS improves It_2 for devices with nonuniform ESD current distributions, primarily the low-voltage (1.5 V) transistors used in this work. Despite reduction in current gain of the lateral n-p-n transistor, increases in GSCS lead to higher It_2 primarily due to thermal effects arising from increase in the power dissipation volume. This implies that It_2 for salicided deep submicron devices with STI is sensitive to the thermal capacity of the structures; and that the lateral n-p-n model is not sufficient for describing the device behavior of these devices under ESD conditions. Therefore, analysis of ESD behavior of advanced devices should consider both thermal effects and the nonuniform bipolar conduction. Based on this study, it is recommended that the minimum gate-to-contact spacing should be avoided for the design of protection devices. However, the minimum contact spacing can be used if sufficient substrate bias can be supplied to the NMOS device under ESD conditions, because the substrate-triggered lateral n-p-n transistor is independent of the gate-to-contact spacings as confirmed experimentally in this work. Furthermore, on-going experimental work also show that the results obtained from the single finger structure correlate very well with the results from multifinger structures provided the multifingers are uniformly triggered.

IV. CONCLUSION

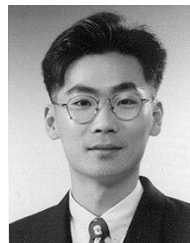
Improvement of ESD failure threshold with the gate-to-contact spacing for fully silicided NMOS transistors have been investigated. The results provide new insight into ESD design rules for deep submicron technology based on detailed experimental and simulation results. It has been shown that the reduction in current localization and increase in the power dissipation volume with increases in the gate-to-contact spacing are the primary factors influencing improvement of ESD performance. It has also been established that substrate biasing can help eliminate the impact of the gate-to-contact spacing on the ESD robustness. Results from this work suggest that even for silicided processes, the gate-to-contact spacing should be carefully engineered to achieve efficient and robust ESD protection designs.

ACKNOWLEDGMENT

The authors wish to thank V. Gupta and Dr. A. Amerasekera of Texas Instruments, Inc., Dallas, TX, for their insight in the design of test structures.

REFERENCES

- [1] A. Amerasekera and C. Duvvury, *ESD in Silicon Integrated Circuits*. New York: Wiley, 1995.
- [2] K.-H. Oh, C. Duvvury, C. Salling, K. Banerjee, and R. W. Dutton, "Non-uniform bipolar conduction in single finger NMOS transistors and implications for deep submicron ESD design," in *Proc. IEEE Int. Reliability Physics Symp.*, 2001, pp. 226–234.
- [3] A. Amerasekera, W. Abeelen, L. Roozendaal, M. Hannemann, and P. Schofield, "ESD failure modes: Characteristics, mechanisms, and process influences," *IEEE Trans. Electron Devices*, vol. 39, pp. 430–436, Mar. 1992.
- [4] G. Notermans, A. Heringa, M. Dort, S. Jansen, and F. Kuper, "The effect of silicide on ESD performance," in *Proc. IEEE Int. Reliability Physics Symp.*, 1999, pp. 154–158.
- [5] K. Verhaege and C. Russ, "Wafer cost reduction through design of high performance fully silicided ESD devices," in *Proc. EOS/ESD Symp.*, 2000, pp. 18–28.
- [6] K.-H. Oh, C. Duvvury, K. Banerjee, and R. W. Dutton, "Investigation of gate to contact spacing effect on ESD robustness of salicided deep submicron single finger NMOS transistors," in *Proc. IEEE Int. Reliability Physics Symp.*, 2002, pp. 148–155.
- [7] K. Banerjee, A. Amerasekera, G. Dixit, and C. Hu, "Temperature and current effects on small-geometry-contact resistance," in *IEDM Tech. Dig.*, 1997, pp. 115–118.
- [8] C. Duvvury, S. Ramaswamy, A. Amerasekera, R. Cline, B. Anderesen, and V. Gupta, "Substrate pump NMOS for ESD protection application," in *Proc. EOS/ESD Symp.*, 2000, pp. 7–17.
- [9] A. Amerasekera, S. Ramaswamy, M.-C. Chang, and C. Duvvury, "Modeling MOS snapback and parasitic bipolar action for circuit-level ESD and high current simulations," in *Proc. IEEE Int. Reliability Physics Symp.*, 1996, pp. 318–326.
- [10] A. Amerasekera, V. Gupta, K. Vasanth, and S. Ramaswamy, "Analysis of snapback behavior on the ESD capability of sub-0.20 μm NMOS," in *Proc. IEEE Int. Reliability Physics Symp.*, 1999, pp. 159–166.
- [11] A. Amerasekera, V. McNeil, and M. Rodder, "Correlating drain junction scaling, silicide thickness, and lateral NPN behavior, with the ESD/EOS performance of a 0.25 μm CMOS process," in *IEDM Tech. Dig.*, 1996, pp. 893–896.
- [12] S. M. Sze, *Physics of Semiconductor Devices*. New York: Wiley, 1981.



Kwang-Hoon Oh (S'92) received the B.S. and M.S. degrees in electrical engineering from Seoul National University, Seoul, Korea, in 1990 and 1992, respectively. He is currently pursuing the Ph.D. degree at Stanford University, Stanford, CA.

From 1992 to 1997, he was with Samsung Electronics, Puchon, Korea, where he was engaged in the design and development of power MOSFETs and IGBTs. During 2000–2001, he held summer research positions at Texas Instruments, Inc., Dallas, TX, where he focused on the modeling of ESD reliability for advanced CMOS technologies. His research interests are in the area of device simulation, characterization, and electrothermal and reliability modeling for advanced deep submicron CMOS technologies with applications to IC circuits.

Mr. Oh is a member of the IEEE Electron Devices Society.



Charvaka Duvvury (SM'01) received the Ph.D. degree in engineering science from the University of Toledo, Toledo, OH.

After working as a Postdoctoral Fellow in physics at the University of Alberta, Alberta, ON, Canada, he joined Texas Instruments, Dallas, TX, in 1977. He initially worked in the Houston DRAM Group as a Design/Product Engineer for 4K/16K DRAMS. He then was part of the first 256K CMOS DRAM design and the Advanced Development Group that worked on the 1 Meg DRAM with specific contributions in

DRAM circuit design, transistor modeling, and reliability studies. He joined the Semiconductor Process and Device Center, Dallas, in 1988, where his work was on the transistor modeling of CMOS/BiCMOS technologies and development of ESD protection for high voltage designs and submicron CMOS technologies. He was elected Senior Member of Technical Staff in 1990, Distinguished Member of Technical Staff in 1997, and Texas Instruments Fellow, also in 1997. His current work is on ESD development for deep submicron CMOS technologies. He has published over 65 papers in technical journals and conferences and holds 25 patents with several pending. He has coauthored books on hot carriers (New York: Van Nostrand Reinhold, 1992), modeling of electrical overstress (Boston, MA: Kluwer, 1994), and ESD reliability phenomena and protection design (New York: Wiley, 1995).

Dr. Duvvury is a recipient of the Outstanding Contributions Award from the EOS/ESD Symposium (1990), Outstanding Mentor Award from the SRC (1994), several Best Paper Awards from the EOS/ESD Symposium, and Outstanding Paper Award from the International Reliability Physics Symposium. He has been very active in the EOS/ESD Symposium, where he was the Technical Program Chairman of the 1992 Symposium and was the General Chairman of the 1994 ESD Symposium. He is currently a member of the ESD Association Board of Directors, promoting university education and research in ESD. He is also a member of Eta Kappa Nu and Sigma Xi.



Kaustav Banerjee (M'99) received the Ph.D. degree in electrical engineering and computer sciences from the University of California, Berkeley, in 1999.

He was with Stanford University, Stanford, CA, from 1999 to 2002 as a Research Associate at the Center for Integrated Systems. In July 2002, he joined the Faculty of the Department of Electrical and Computer Engineering, University of California, Santa Barbara, as an Assistant Professor. His research interests include nanometer scale circuit effects and their implications for high-performance/low-power VLSI and mixed-signal designs and their design automation methods. He is also interested in some exploratory interconnect and circuit architectures such as 3-D ICs, integrated optoelectronics, and nanotechnologies such as single electron transistors. He co-advises several doctoral students at Stanford University, University of Southern California, Los Angeles, and the Swiss Federal Institute of Technology (EPFL), Lausanne, Switzerland. From February 2002 to August 2002, he was a Visiting Professor at the Circuit Research Labs of Intel, Hillsboro, OR. In the past, he has also held summer/visiting positions at Texas Instruments, Inc., Dallas, TX, and EPFL-Switzerland, and has consulted for several EDA companies in the Silicon Valley. He has authored or coauthored over 70 technical papers in archival journals and refereed international conferences and has presented numerous invited talks and tutorials.

Dr. Banerjee served as Technical Program Chair of the 2002 IEEE International Symposium on Quality Electronic Design (ISQED '02), and is the Conference Vice-Chair of ISQED '03. He has also served on the technical program committees of the ACM International Symposium on Physical Design, the EOS/ESD Symposium, and the IEEE International Reliability Physics Symposium. He is the recipient of a Best Paper Award at the 2001 Design Automation Conference.



Robert W. Dutton (F'84) received the B.S., M.S., and Ph.D. degrees from the University of California, Berkeley, in 1966, 1967, and 1970, respectively.

He is currently Professor of electrical engineering, Stanford University, Stanford, CA, and Director of Research at the Center for Integrated Systems. He has held summer staff positions at Fairchild, Bell Telephone Laboratories, Hewlett-Packard, IBM Research, and Matsushita during 1967, 1973, 1975, 1977, and 1988, respectively. His research interests focus on integrated circuit processes, device and circuit technologies (especially the use of computer-aided design (CAD) in device scaling and for RF applications). He has published more than 200 journal articles and graduated more than four dozen doctoral students.

Dr. Dutton was Editor of IEEE TRANSACTIONS ON COMPUTER-AIDED DESIGN from 1984 to 1986, the winner of the 1987 IEEE J. J. Ebers and 1996 Jack Morton Awards, the 1988 Guggenheim Fellowship to study in Japan, was elected to the National Academy of Engineering in 1991, and was also been honored with the Jack A. Morton Award in 1996 and the C&C Prize (Japan) in 2000.