

Non-uniform Bipolar Conduction in Single Finger NMOS Transistors and Implications for Deep Submicron ESD Design

Kwang-Hoon Oh, Charvaka Duvvury*, Craig Salling*, Kaustav Banerjee and Robert W. Dutton
 Center for Integrated Systems, Stanford University, CA 94305
 Phone: 650-723-9484; Fax: 650-725-7731; E-mail: okhoon@gloworm.stanford.edu
 *Silicon Technology Development, Texas Instruments, Dallas, TX 75265

ABSTRACT

This paper presents a detailed study of the non-uniform bipolar conduction phenomenon in single finger NMOS transistors and analyses its implications for deep submicron ESD design. It is shown that the uniformity of lateral bipolar triggering is severely degraded with device width (W) in advanced technologies with silicided diffusions and low resistance substrates, and that this effect can only be improved up to a maximum W by increasing the substrate bias. Additionally, the concept of *intrinsic second breakdown strength* is introduced, which is substrate bias independent and represents the maximum achievable breakdown current for a given technology.

INTRODUCTION

The non-uniform triggering in silicided, single finger NMOS transistors under ESD conditions was first reported by Scott et al. in 1986 [1], and subsequently the phenomenon of non-uniform triggering in multi-finger NMOS transistors and the implications for design of ESD protection was discussed by Polgreen et al. in 1992 [2]. However, in recent years the dependence of the ESD failure threshold current (I_{t2}), which is widely used for ESD-resistant strength monitoring, as a function of the single finger device width (W), has become quite an important criterion for optimum design of ESD protection circuits. In this regard, there has not been any significant work reporting investigations of this effect in detail nor any study reporting the impact for the advanced submicron technologies with low resistance substrates.

In this work extensive ESD characterization on deep submicron single finger NMOS transistors has been performed using I_{t2} measurements and emission microscopy (EMMI) analysis [3, 4] for both silicided and non-silicided cases. The correlation between I_{t2} measurements and EMMI analysis reveals the occurrence of a rather severe non-uniform lateral current conduction in single finger structures, especially for low resistance substrates in a silicided process. As shown in Fig.1, single finger devices show a significant width dependence of I_{t2} as measured in mA/ μm , and the effect is strongly dependent on the process. In this study, EMMI analysis is used to observe the bipolar (*npn*) conduction as a function of device width. Also, the impact of substrate bias is investigated as a means to reduce this effect.

EXPERIMENTS

A. I_{t2} Measurements

3.3 V NMOS transistors (for ESD protection) with 70 Å gate oxide thickness were used in this work. The devices were fabricated using a state-of-the-art 0.13 μm CMOS technology. These test devices were processed with two different CoSi₂ technologies, and also a non-silicided technology. Additionally, two types of epi-layers were used as a base (i.e. the substrate) material. For silicided_A and silicided_B processes, there is a difference in the effective substrate

resistance R_{sub} , due to a difference in P-well and substrate doping for the two processes. The measured substrate resistance for silicided_A and silicided_B processes is 6300 $\Omega\text{-}\mu\text{m}$ and 4800 $\Omega\text{-}\mu\text{m}$, respectively [5].

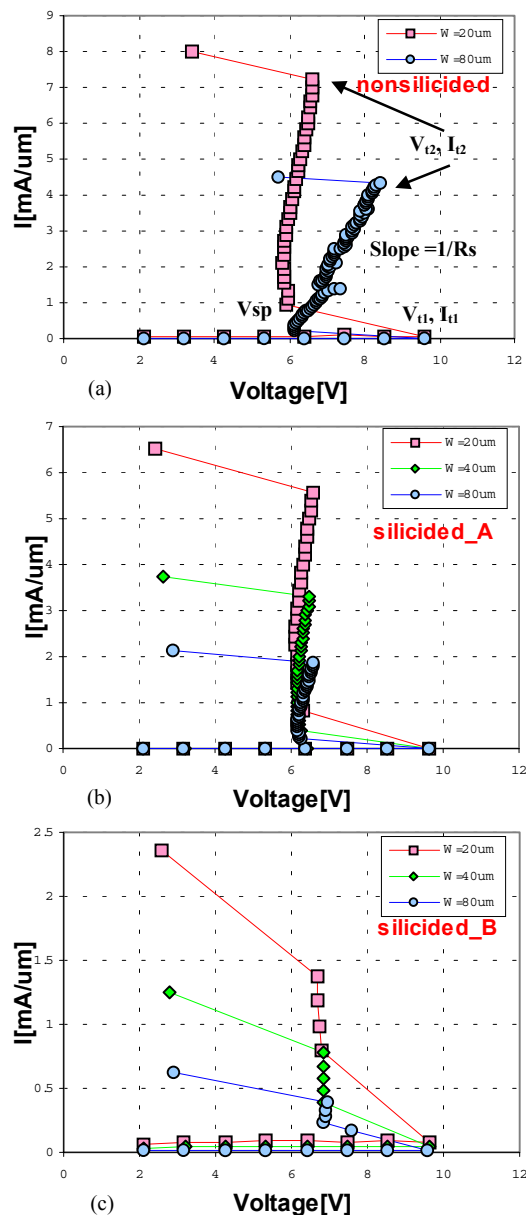


Figure 1. High pulsed current I-V curves with different finger widths for each technology where $L_{poly} = 0.5\mu\text{m}$. The current per unit finger width at second breakdown (I_{t2}) strongly depends on the finger width (W), which illustrates non-uniform bipolar current flow for (a) non-silicided (b) silicided_A and (c) silicided_B technologies.

Using an automated transmission line pulsing (TLP) system, I_{t2} measurements were performed with a pulse width of 210 ns for various test structures and sample I-V curves for a silicided and a non-silicided device are shown in Fig.1. During TLP measurements, the pulsed current is injected into the drain of the gate-grounded NMOS transistors with increased pulse magnitude at each step. This increase in the pulsed current at the drain terminal eventually induces impact ionization at the drain and substrate junction where the high electric field is elevated sufficiently enough to begin avalanche multiplication, resulting in the generation of electron-hole pairs. While the generated electrons are swept across the drain towards the drain contact, the generated holes drift towards the substrate contact giving rise to the substrate current, which also contributes as the base current of the parasitic *npn* transistors. The effective voltage rise in the substrate lowers the barrier of the emitter and base (source-substrate) junction and finally only a small section of the emitter and base junction turns on at the effective forward bias of ~ 0.7 V.

Once the parasitic lateral BJT is triggered at the current I_{t1} and the associated voltage V_{t1} , the bipolar device snaps back and the bias point moves to the minimum bias voltage V_{sp} showing a negative resistance due to the additional available carriers for avalanche multiplication. The operating mechanism in this condition is similar to that of the *npn* bipolar transistor in a self-biasing mode. However, beyond V_{sp} further increases in carrier density lowers the substrate resistance and the current gain β so that a higher substrate current is needed to maintain the on-state of an *npn* transistor. Therefore, a positive resistance region in the I-V curve appears again at high current level. In this region voltage starts to increase linearly with the current and the slope of the line also represents the inverse of the series resistance R_s of the parasitic BJT. Finally, the increased current is locally confined to a very small volume of the drain-substrate junction, dissipating large amount of power, and hence the peak temperature can exceed the critical temperature or the thermal limit. This results in the irreversible thermal damage to the device (at I_{t2} and V_{t2}) giving rise to large junction leakage current.

According to the measured data shown in Fig. 1, the application of silicided technology shows no significant differences in V_{t1} since the *npn* transistor triggering voltage V_{t1} is dominantly determined by the avalanche multiplication across the drain-substrate junction. Meanwhile, the differences in R_{sub} for two silicided processes (A and B) could give a minor effect in V_{t1} . Nevertheless, overall results indicate that the doping profile and the associated junction characteristics are similar for the three different technologies. However the series resistance R_s for silicided and non-silicided devices is clearly showing the significant differences as expected. In addition, the holding voltage V_{sp} of both the silicided_A process and the non-silicided process is almost the same, which implies that the existence of silicide diffusion in devices has no observable impact on the snapback behavior of the parasitic BJT's. However, when the two silicided processes are compared, the device with low R_{sub} (silicided_B) shows higher V_{sp} due to the reduction in current gain β , as predicted in [6]. The high current I-V curves in Fig. 1 show consistency with the ESD performance of an advanced silicide process, but the results from the different finger width devices for the same process show a strong width dependency, especially in the high current regime. This clearly indicates that the current flows non-uniformly over the entire finger width (W) at a given bias condition. In order to examine the dependence of ESD strength on the design parameters such as L_{poly} and W for the three processes, TLP tests were performed for various structures.

As shown in Fig. 2, I_{t2} values are moderately influenced by the channel length L_{poly} and strongly influenced by the finger width W for all device groups. It was observed that for a given technology, the channel length dependence of I_{t2} is consistent with the current gain β and R_{sub} dependence of a lateral *npn* transistor depending on the process as described in [6, 7], but only a limited portion of the finger is

effective for ESD current conduction as apparent from the channel width dependence. As shown in Fig. 2 (a), the improvement of I_{t2} for silicided_A process is rather strong and it seems to be caused by the rapid change in β with L_{poly} associated with the lower effective doping concentration (high R_{sub}) in the P-well and substrate. The width dependence of I_{t2} is a clear evidence of the strong non-uniformity of bipolar conduction, even in single finger NMOS transistors. Although this was previously reported [1, 2], it has been usually overlooked in ESD protection designs. For all the cases, I_{t2} of non-silicided devices shows the highest value, but the difference in I_{t2} for each process seems to decrease as W and L_{poly} shrink.

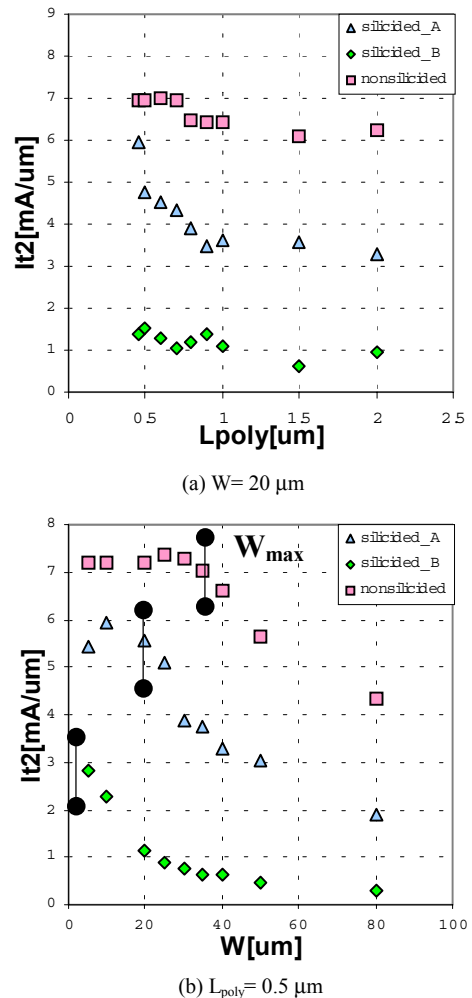


Figure 2. Variation of technology dependent I_{t2} with (a) the channel length and (b) the finger width for single finger NMOS transistors. The roll-off of I_{t2} with W indicates that the failure current essentially remains constant as W is further increased beyond this point, and this is called W_{max} . (indicated by a circled bar)

B. Experimental Setup and EMMI Analysis

In order to visualize the strong non-uniformity of lateral bipolar current conduction, the spatial distribution of ESD current was observed by EMMI for the silicided_A and non-silicided 20 μm and 80 μm wide single finger ESD NMOS transistors. The nominal gate drawn length was 0.5 μm for the device under test and this was the baseline device for characterizing ESD robustness of the 3.3 V NMOS transistors.

EMMI is a widely used technique for wafer level reliability and yield analysis for semiconductor devices. These analyses have been performed by collecting the visible and near infrared wavelength (390 ~ 1000 nm) photons emitted under device operation. General mechanisms of light emission in devices are electron-hole recombination, intraband transition, thermal radiation and tunneling currents in oxide. For NMOS transistors, the radiative intraband transitions are the predominant emission mechanism under high electric field and currents. Thereby the generated photons, transmitted through the overlying layers such as dielectrics and metal interconnections, are detected and this is referred as front-side light emission analysis. In this analysis, the FA-1000 model of EMMI (manufactured by Alpha Innotech Corporation) was utilized. Fig. 3 shows the schematic diagram for the experimental setup of EMMI in this work.

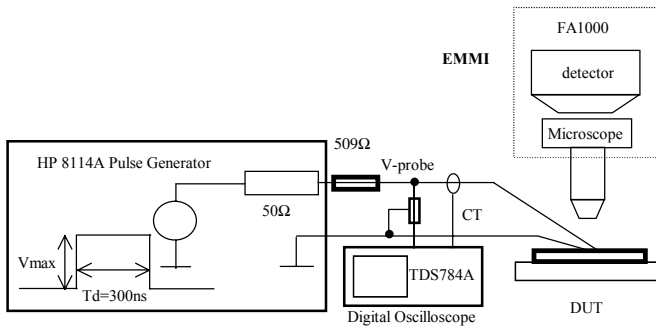
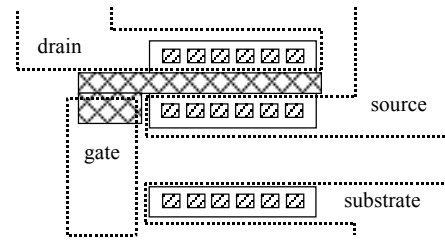


Figure 3. EMMI setup for observing the spatial distribution of ESD currents.

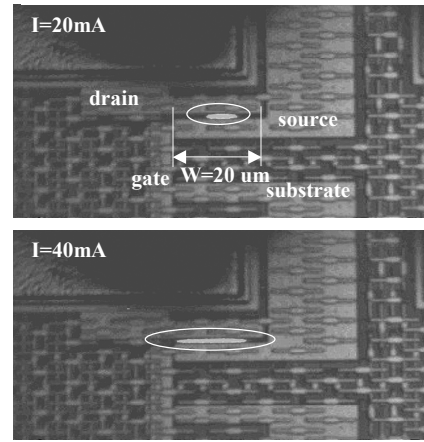
Using the HP8114A pulse generator with the series resistance of 509 Ω as a source of the pulsed current to the drain, bipolar conduction currents were monitored by a digital oscilloscope and the EMMI images were taken as well. Once the snapback occurs by the current stress beyond the threshold triggering current I_{t1} for NMOS transistors, the bipolar conduction currents could increase rapidly to the critical current value associated with a thermal limit with even small increment in the drain bias, which could cause thermal failure of the transistor under test. Thus in order to avoid any thermal damage by self-heating, which is normally seen as a bright spot in EMMI, a short duration pulsed bias with low frequency is required. In this experiment, the pulsed bias with duration of 300 ns was applied at a frequency of 400 Hz. The emitted photons during each current pulse were integrated over the exposure time of 6 min.

In Fig. 4, the observed transistor turned-on widths are shown at current levels of 20 mA and 40 mA for the 20 μm wide transistors with silicided_A and non-silicided process. For devices with low R_{sub} (silicided_B), no optimum condition has been found for taking images with the reasonable value of exposure time and current stress. Once the snapback happened for the silicided_B devices, subsequently it failed causing a resistive current path, which seemed to be attributed to the abnormal sensitivity to the drain current, which can be inferred from nearly zero series resistance values in Fig. 1 (c). However the spatial distribution of bipolar conduction current for silicided_B process can be expected to be similar to that of silicided_A and non-silicided devices based on the correlation of EMMI images and I_{t2} measurements for silicided_A and non-silicided devices. As shown in Fig. 4, initially only a small section of the finger is turned on (in the area indicated by an oval) for both devices. The turned-on location is observed to be random through repetitive tests, which is believed to result from the inhomogeneities in processing and device behavior. The turned-on width expands with increase in the bias current and it is

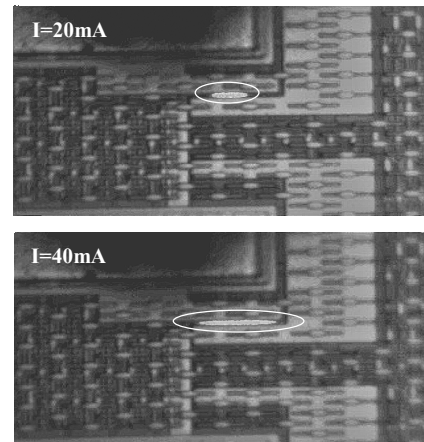
observed that most of the finger width is turned on at the current of 40 mA. This result shows strong correlation with the I_{t2} measurements shown in Fig. 2 where the failure currents ($I_{t2} \cdot W$) are linearly scaled up to the finger width of 20 μm for both processes.



(a) test structure schematic



(b) silicided_A device

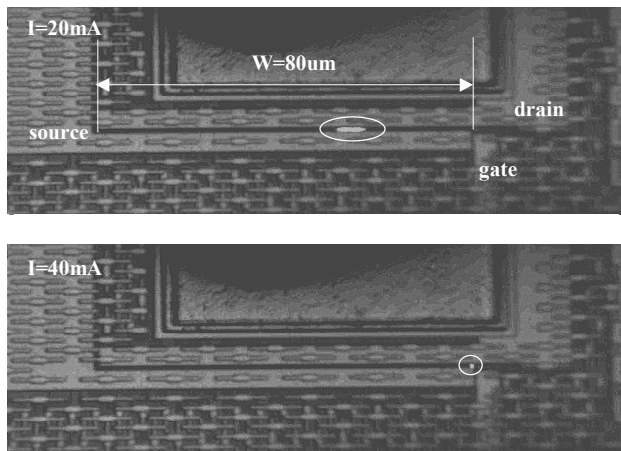


(c) non-silicided device

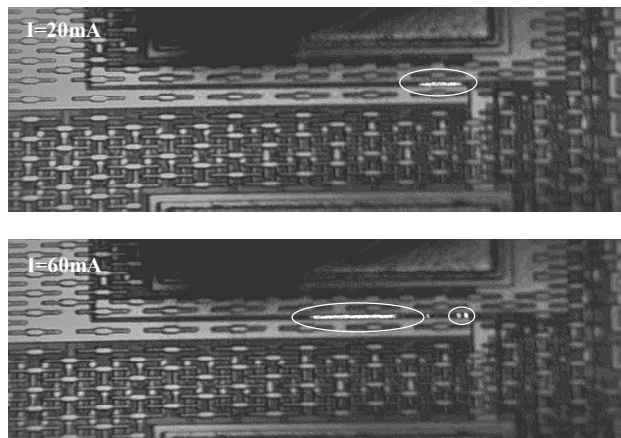
Figure 4. Emission microscopy images that show spatial extent of lateral current conduction at different current levels for $W/L=20/0.5$ μm (a) test structure schematic (b) silicided_A and (c) non-silicided devices. (Pulse width= 300 ns, $f=400$ Hz, exposure time= 6 min)

EMMI images for the 80 μm wide device are shown in Fig. 5. The initial turned-on location is randomly placed as observed in the 20 μm wide device. However, both silicided_A and non-silicided devices failed (shown by circled areas) or were partially damaged (at circled area on the right corner of the finger), before the bipolar conduction width extended to the whole 80 μm finger width, at 40 mA and

60 mA, respectively. For 80 μm wide devices, full triggering of the *npn* transistor has not been observed throughout repetitive tests for the silicided_A and non-silicided processes. This is related to observations from I_{t2} measurements in Fig. 2 and the severe degradation of I_{t2} with increase in single finger width is in line with the current locking phenomenon as seen in EMMI images for both processes. Hence the maximum turn-on width W_{max} of the *npn* transistor under ESD stress is equal to the roll-off point in Fig. 2 (b), the estimated value for silicided_A and non-silicided devices are about 20 μm and 40 μm , respectively. This clearly illustrates the I_{t2} correlation between EMMI and TLP data. In practice, the designed finger width above W_{max} causes no improvement in I_{t2} . Thus the obvious way of improving ESD performance is to enlarge the turned-on width of single finger transistors. Accordingly, in the following section we investigate the substrate bias effect on non-uniform bipolar conduction as a possible solution for current locking phenomenon in single finger NMOS transistors.



(a) silicided_A device

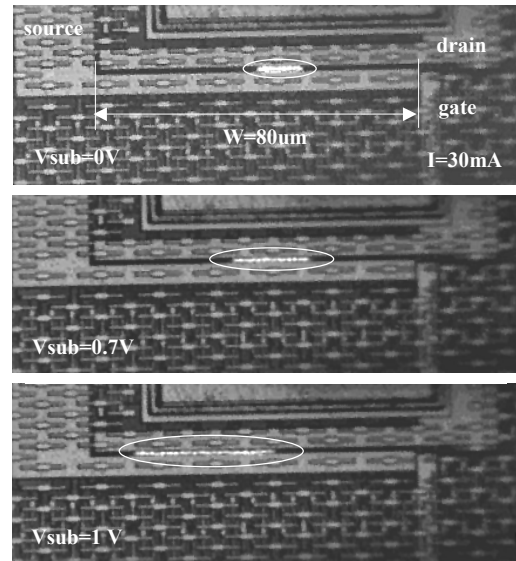


(b) non-silicided device

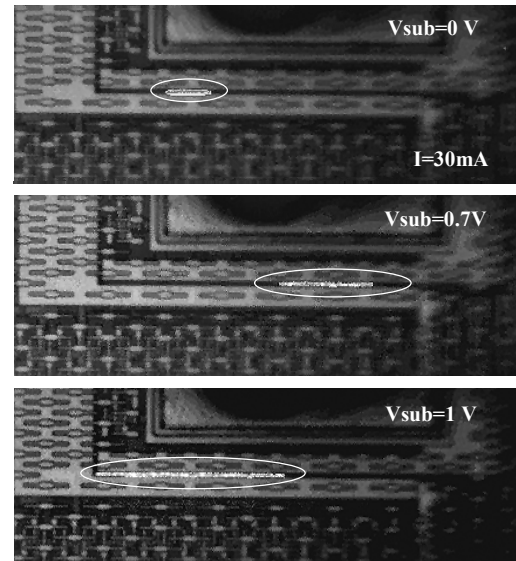
Figure 5. Emission microscopy images that show spatial extent of lateral current conduction at different current levels for $W/L= 80/0.5 \mu\text{m}$ (a) silicided_A (b) non-silicided devices. (Pulse width= 300 ns, $f= 400 \text{ Hz}$, exposure time= 6 min)

C. Substrate Bias Effect on Uniform *npn* Triggering

In order to investigate the impact of substrate bias on the uniformity of bipolar conduction, we applied the forward substrate bias to the emitter and base (source and substrate) junction of the lateral *npn* transistor. By applying an external substrate bias, the local potential of the substrate could be raised and consequently the bipolar transistor could be triggered by a small ohmic drop in the substrate rather than the value needed ($\sim 0.7 \text{ V}$) in a self-biasing mode without an external substrate bias. The influence on the spatial distribution of bipolar conduction by a substrate bias was also visualized through EMMI analysis as shown in Fig. 6.



(a) silicided_A device



(b) non-silicided device

Figure 6. Emission microscopy images that show the spatial extent of lateral current conduction with different substrate bias for $W/L= 80/0.5 \mu\text{m}$ (a) silicided_A and (b) non-silicided devices. (Pulse width= 300 ns, $f= 400 \text{ Hz}$, exposure time= 6 min)

Injecting a constant pulsed current of 30 mA, an external substrate bias was applied as $V_{sub}=0$ V, 0.7 V and 1 V. The transistor turned-on width is seen to spread out with increase in substrate bias for both silicided_A and non-silicided devices. Total turned-on width is enlarged by 3 to 4 times for both devices with $V_{sub} = 1$ V. On the contrary, the intensity of bipolar conduction diminishes as the turned-on width increases since the total injected current remains constant. Accordingly, this suggests that the decreasing local current density for a given injection current could lower the temperature rise caused by the Joule heating in NMOS transistors and as a result temperature distribution during bipolar conduction would become smoother. However, even with the substrate bias, full *npn* transistor triggering was not observed for the 80 μm wide transistors. Nevertheless, EMMI images illustrate that the substrate bias can increase the effective finger width under ESD stress by enlarging the lateral bipolar conduction width, which can result in the improvement of I_{t2} as demonstrated in Fig. 7. The positive impact of substrate bias on I_{t2} was reported before for NMOS transistors of a 0.35 μm technology [8]. In Fig. 7, we show this behavior for two different silicided processes, which also confirm the EMMI results shown in Fig. 6. We obtain new insight that second breakdown triggering current I_{t2} improves dramatically with the substrate bias by widening the turned-on width of the lateral *npn* transistor. This means that the possible origin of non-uniform bipolar conduction by the inhomogeneous nature of the local substrate potential, which results from the variation in current gain β , substrate resistance R_{sub} and substrate current by impact ionization, can be screened by an external substrate bias. Furthermore, based on I_{t2} values for the two silicided devices, 5.5 to 6 mA/ μm with $V_{sub} > 1$ V in Fig. 7, ESD performance of device can become rather insensitive to the statistical variations that may arise from process conditions and base material properties.

It should be noted that I_{t2} for the silicided_B process is substantially improved as V_{sub} changes from 0.85 V to 1 V. This seems to be the consequence of the fact that the substrate bias dependence of I_{t2} (or W_{max}) is a strong function of the single finger width W and this becomes stronger in the devices with lower R_{sub} (data shown in Fig. 11, later in the paper). Since the substrate bias is forward biased to the emitter and base junction, the *npn* bipolar triggering voltage (V_{ti}) reduces with increase in the substrate bias and eventually bipolar turns on without snapback when the effective emitter base junction bias is about 0.7 V. This corresponds to 0.85 V for an external substrate bias and any small base current can trigger the *npn* transistor beyond this value.

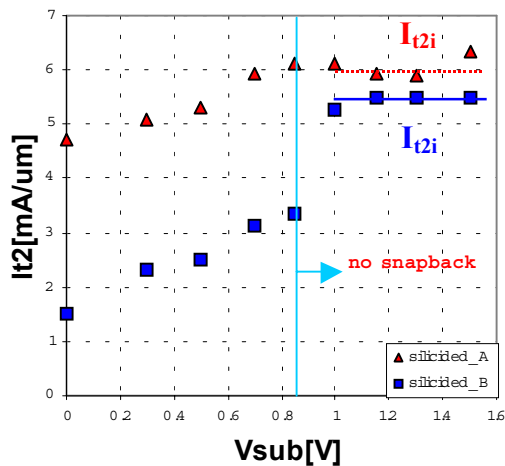


Figure 7. Second breakdown triggering current I_{t2} with substrate bias for silicided devices (A and B). I_{t2} is approaching its intrinsic value I_{t2i} as substrate bias is increased. $W/L=20/0.5$ μm

For both silicided devices, I_{t2} saturates right above the substrate bias of 0.85 V. This implies that the bipolar conduction width is locked and the local substrate potential near emitter and base junction is not altered with higher substrate bias. The associated I_{t2} value at this substrate bias will be the maximum achievable second breakdown triggering current for both technologies. This current is substrate bias independent as seen in Fig. 7 and can be named the *intrinsic second breakdown triggering current* I_{t2i} .

PHYSICAL MODELING OF NON-UNIFORM CONDUCTION

Based on the measurements and device simulation results, a simple physical model for non-uniform bipolar conduction is proposed.

A. Current Localization

As the experimental data illustrate, uniformity of bipolar conduction is determined by the triggered portion over the entire finger width. The single finger transistor can be considered as a parallel-connected network of narrow (segmented) *npn* transistors as shown in Fig. 8. Assuming each narrow *npn* transistor has slightly different intrinsic characteristics, which comes from the inherent statistical variations, the location of the triggered *npn* transistor is expected to be uncertain.

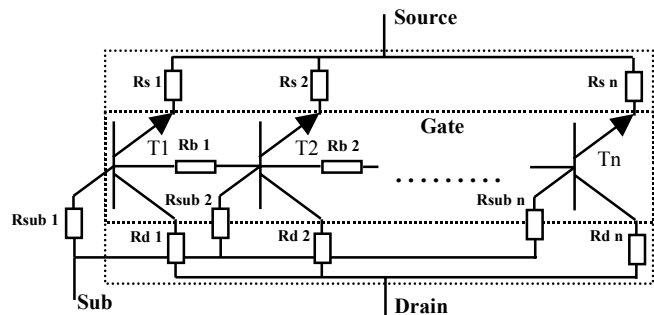


Figure 8. Schematic of the segmented *npn* transistors for a single finger NMOS transistor. Each *npn* transistor has different intrinsic characteristics due to the statistical variations. R_{sub} , R_s , R_d and R_b denote the parasitic resistance in the substrate, the source, the drain and the base, respectively.

According to the study by Russ et al. [4], avalanche multiplication starts at the corner of the drain structure where the field is highest due to the spherical junction curvature and avalanche current is uniformly distributed along the channel width before the snapback of an *npn* transistor occurs. However, in this study using shallow trench isolated devices, we haven't observed any location preference of avalanche multiplication, but rather observed randomness in the turned-on location along the finger width.

The parasitic bipolar triggering mechanism can be well described [7] in terms of three main device parameters such as the current gain β , the substrate resistance R_{sub} , and the multiplication factor M . The substrate current (hole current) provided by impact ionization is strongly dependent on the doping profile and field distribution of the drain structure so the effective forward bias due to the local substrate potential for each segmented *npn* transistor cannot be the same, even if R_{sub} is assumed to be constant. Assuming that a small portion of the finger width (source-substrate junction) is sufficiently (> 0.7 V) forward biased initially, the segmented transistors within this portion can be turned on and this assumption is supported by the EMMI observa-

tion that the location of initially turned-on segment is arbitrary over the entire finger width. Once the *npn* transistors turn on, to maintain the on-state of the transistors, the snapback condition of $\beta \cdot (M-1) > 1$ [9] should be satisfied. Both β and M are functions of the injected drain current [7] so the location of bipolar conduction should be dependent on the current level as seen by EMMI analysis as long as the dependence of β and M on the drain current varies for all the segmented *npn* transistors. As the injected drain current increases, a large number of segmented transistors should turn on while satisfying the snapback condition because the maximum current capability for each segmented transistor is limited. This is supported by the observation that the turned-on width in EMMI images spreads out with increase in the injection current.

To represent 3D localized non-uniform bipolar conduction behavior, an NMOS transistor structure with a common drain was devised including extreme asymmetries and simulated using 2D device simulator *MEDICI*. In Fig. 9, each of the half transistors (TR1 or TR2) represents a bisection of a single finger transistor and has different channel length ($L_{poly} = 0.5 \mu\text{m}$ for TR1 and $L_{poly} = 0.4 \mu\text{m}$ for TR2) and source series resistance (none for TR1 and 5Ω to the source of TR2) to reflect the effect of statistical variations in β and the local substrate potential ($\sim I_{sub} \cdot R_{sub}$), respectively. Adjusting the device parameters used in simulation, current competition between the two transistors is illustrated as shown in Fig. 9.

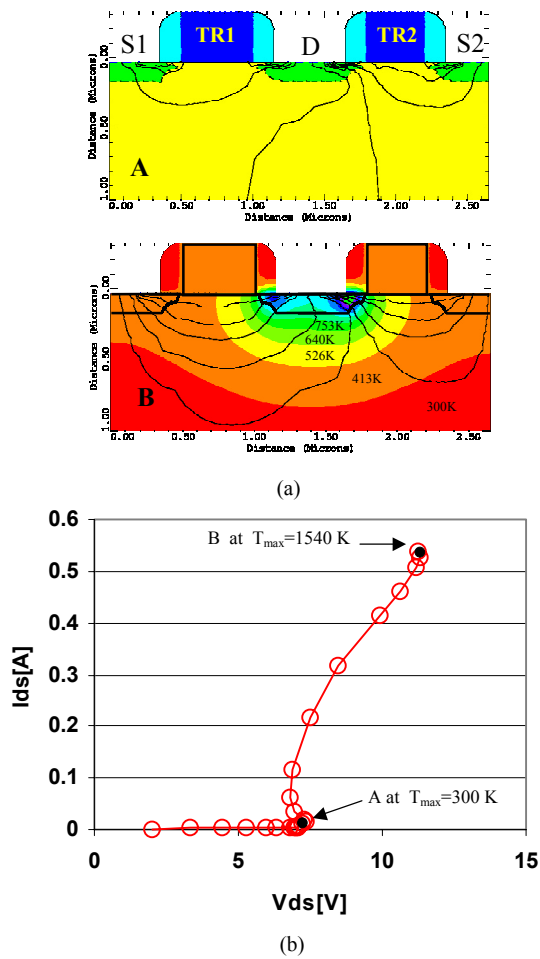


Figure 9. Current localization for a 3D like NMOS transistor [10] (a) current flowlines and temperature distribution at bias conditions A and B. Each step of temperature contour is 113 K (b) I-V curve by a transient simulation. Initially TR2 turns on first but TR1 conducts more bipolar current at high current level. Each transistor has $W = 40 \mu\text{m}$.

Initially the transistor TR2 triggers (at A in I-V curve of (b)) and snaps back first since it has higher current gain (due to shorter channel length). However as the drain current increases, the two transistors TR1 and TR2 compete for current conduction since the effective forward bias for the corresponding *npn* transistors changes depending on the injected current level due to differences in the source series resistance and the extent of current gain degradation due to increase in the drain current. At high current levels (at B in I-V curve of (b), near thermal limit), more current flows through TR1, which also suggests that the location of the turned-on width in the NMOS transistor moves along the finger width (as shown in Fig. 5 (b)); similar behavior has been previously reported in conjunction with thermal instability [4]. The simulation results confirm that the inequality of intrinsic characteristics of each segmented transistor causes asymmetry in current conduction, and subsequently results in current localization to become dependent on the injected drain current level. In addition, as discussed in the next section, it provides an understanding of how the external substrate bias influences the spatial distribution of bipolar conduction as observed by EMMI analysis.

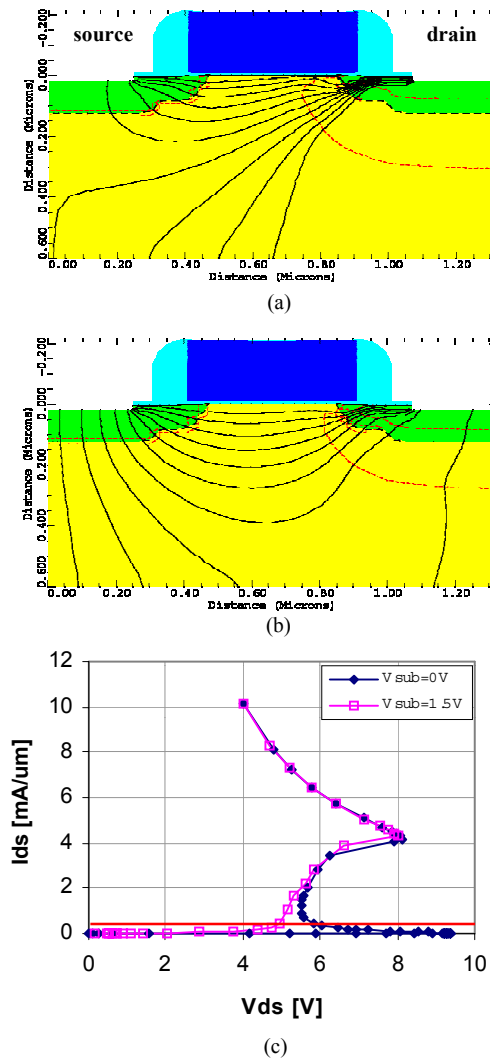


Figure 10. Current flowlines for the gate-grounded NMOS with $L_{poly} = 0.5 \mu\text{m}$ at the drain current of $0.5 \text{ mA}/\mu\text{m}$ (a) without substrate bias (b) with substrate bias of 1.5 V and (c) DC high current I-V curve (the horizontal line indicates $0.5 \text{ mA}/\mu\text{m}$).

B. Substrate Bias Effect

When a lateral *npn* transistor turns on in a self-biasing mode, only part of the emitter base junction is forward biased as shown in Fig. 10 (a). It should be noted that the local substrate potential distribution of the emitter base junction is strongly influenced by an external substrate bias. However, understanding the effect of external substrate bias requires 3D device simulation to account for non-isotropic device structures and properties along the width of the NMOS transistor, which is nontrivial. Hence 2D device simulation is employed to understand 3D behavior. In Fig.10, DC high current characteristics of a gate-grounded NMOS transistor with $L/W=0.5/1\ \mu\text{m}$ is simulated for two different substrate bias conditions, $V_{sub}=0\ \text{V}$ and $V_{sub}=1.5\ \text{V}$, and with $R_{sub}=5000\ \Omega\text{-}\mu\text{m}$ in each case. The current flowlines can be seen in Fig. 10 (a) and (b) at the same current level of $0.5\ \text{mA}/\mu\text{m}$ where the *npn* transistor turns on. As clearly shown in Fig. 10 (b), the current flows more deeply into the substrate with a substrate bias. The altered local substrate potential can change the snapback triggering point and eventually turn on the *npn* transistor without snapback as shown in Fig. 10 (c). However, I-V curves at high current levels show no significant differences between the two substrate biasing conditions. The second snapback at the drain current of $\sim 4\ \text{mA}/\mu\text{m}$ in the I-V curves is attributed to a rapid increase in the base current resulting from increase in the current component associated with thermally generated carriers while the carriers generated by impact ionization decrease with temperature rise.

From the results presented above, it can be deduced that there are significant differences between the operating modes of the parasitic lateral *npn* transistor with and without the substrate bias. In the case of no substrate bias, the lateral *npn* transistor operates solely in a self-biasing mode, which requires sufficient substrate current by impact ionization to maintain the forward bias for the emitter and base

junction (source and substrate) and the base current is also supplied by impact ionization (Fig. 10 (a)). Hence sufficient avalanche multiplication is necessary for lateral *npn* conduction.

On the other hand, under adequate external substrate bias ($V_{sub} > \sim 0.85\ \text{V}$ in the measurement) the lateral *npn* transistor operates in a normal biasing mode (common emitter). Even in the absence of the drain injection current, the source-substrate and the drain-substrate junctions are fully turned on. However, both parasitic diodes in an NMOS transistor have a relatively long base so most of the injected carriers from the source and drain are recombined resulting in small diffusion currents. As the drain current and the associated drain bias increase, the drain-substrate junction eventually turns off and thus the lateral *npn* transistor operates under normal bias conditions. However, the bipolar conduction current increases very slowly with the increase in the drain current (the associated drain bias) until impact ionization initiates. Before the avalanche multiplication occurs, most of the drain current needs to be supported by the electrons injected from the source (emitter), but the injected carriers contributes to the source-substrate diode current rather than to the drain (collector) current due to the low current gain β of the lateral *npn* transistor. Further increase in the drain current (I_D) induces impact ionization and eventually the avalanche multiplication process becomes regenerative as in the self-biasing mode.

Since the emitter base junction is already fully turned on (shown in Fig. 10 (b)), small value of generation current ($I_{gen} \sim (I/M)I_D$) is sufficient to initiate the regenerative process and accordingly the threshold value of M should be lower than the value needed in a self-biasing mode. As a result, the drain current flows through the low field area in the drain-substrate junction. Fig. 10 (b) clearly shows that relatively wide area of the drain junction is utilized for the same current conduction, compared to Fig. 10 (a) due to the fact that less impact ionization current is required with $V_{sub}=1.5\ \text{V}$.

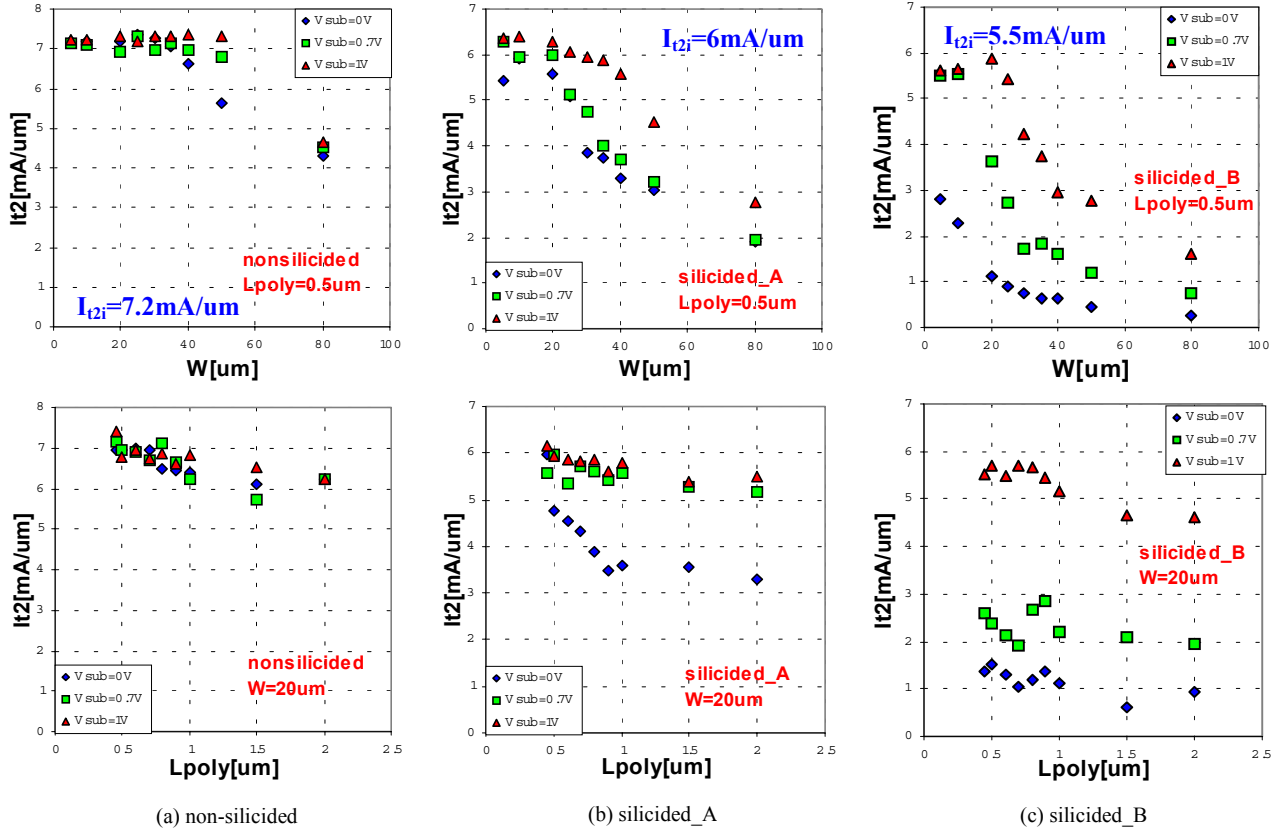


Figure 11. Second breakdown triggering current, I_2 for various technologies and test structures with different substrate bias.

Assuming the multiplication factor M for each segmented npn transistor at the onset of triggering follows a normal distribution, more number of the segmented npn transistors can be turned on under substrate bias due to lower value of M . Therefore 2D simulation results qualitatively suggest that the lateral bipolar conduction could take place over a wider area of the base-collector junction along the finger width under an adequate substrate bias. However, to understand the detailed underlying mechanism of non-uniform bipolar conduction in the entire range of I-V characteristics, a full 3D model that includes thermal effects is required.

RESULTS AND DISCUSSION

As is well known, ESD strength of silicided technology is lower than that of non-silicided technology due to either the reduction in emitter efficiency [8] or early current localization associated with the reduced series resistance [11], which has also been observed by EMMI analysis and shows the different bipolar turned-on widths at the same current level. Therefore, these facts indicate that wider current conduction is associated with a higher ESD failure threshold.

According to Fig. 2 (b), the maximum transistor turned-on width, W_{max} decreases as ESD failure threshold of technology decreases. For instance, the maximum current conduction width at ESD failure threshold is about $35\ \mu\text{m}$ for non-silicided devices while it is about $20\ \mu\text{m}$ for a silicided process (A) and less than $5\ \mu\text{m}$ for a silicided process with low substrate resistance (B). This means the ESD breakdown strength can be linearly scaled only up to W_{max} and that I_{t2} drops off rapidly beyond W_{max} . It further implies that the uniformity in bipolar conduction is the main physical mechanism for high ESD robustness. This obviously places a severe restriction in determining the useful width of a single finger in a multi-finger structure for ESD protections. From a practical design point of view, the minimum value for W_{max} should be at least $30\ \mu\text{m}$ with a minimum I_{t2} (@210 ns) of $4\ \text{mA}/\mu\text{m}$. This will ensure that $> 8\ \text{V}/\mu\text{m}$ for HBM is available for the design of multi-finger protection devices.

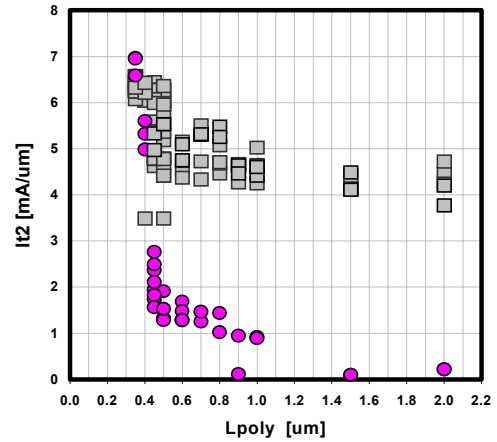


Figure 12. I_{t2} vs. L_{poly} for silicided_A (square: high substrate resistance) and silicided_B (circle: low substrate resistance) devices where $W=20\ \mu\text{m}$.

Fig. 11 illustrates the dependence of I_{t2} values on W and L_{poly} for various substrate biases in the three technologies. It can be observed that W_{max} increases with substrate bias and that the I_{t2} values for silicided devices approach that of non-silicided devices with substrate bias. However, it can also be observed that the values of I_{t2} for $W < W_{max}$ remain almost independent of the substrate bias within the scatter of data. Hence, this value of I_{t2} for finger widths less than W_{max} can be thought of as the maximum obtainable I_{t2} (this was defined as I_{t2i} earlier) under uniform bipolar conduction for a given process technology and ranges from 5.5 to $7.2\ \text{mA}/\mu\text{m}$ for the three processes, which is determined by process effects such as silicide/non-silicide process, thickness of silicide, source/drain engineering and substrate resistance.

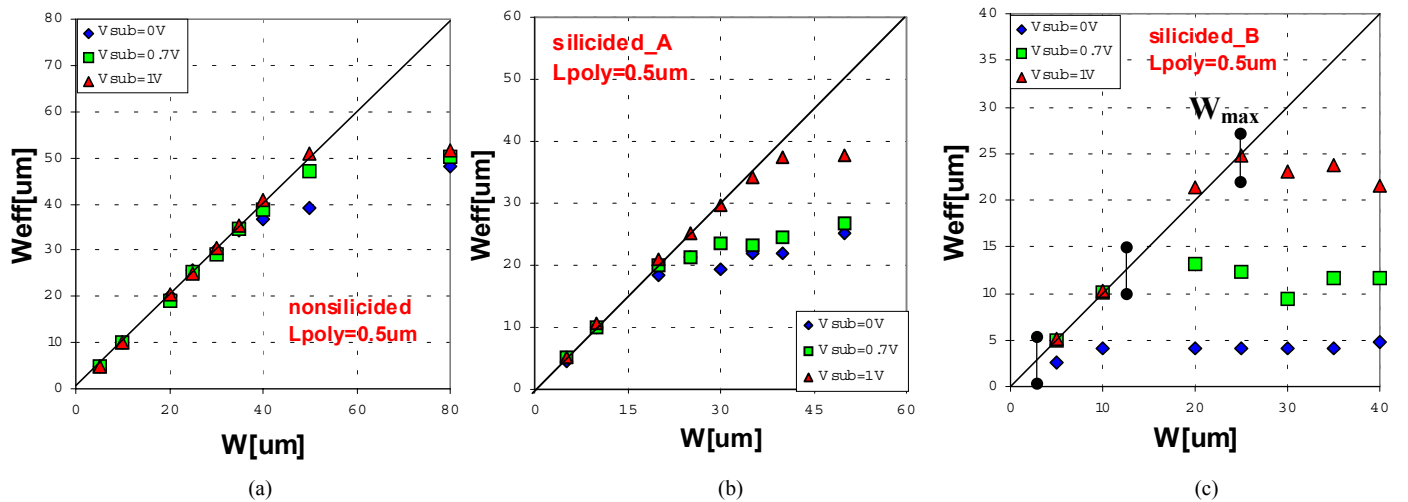


Figure 13. Effective finger widths vs. designed finger widths for different substrate bias (a) non-silicided device (b) silicided_A device and (c) silicided_B device.

In the mean time it should be noted that for the silicided_B devices the I_{I_2} values with $V_{sub} = 0$ V are expected to converge with $I_{I_{2i}}$ for $W < 5$ μm . In addition, I_{I_2} vs. L_{poly} data also suggest that I_{I_2} is approaching $I_{I_{2i}}$ as L_{poly} decreases. This behavior is observed clearly for non-silicided and silicided_A process technologies. However this is not obvious for silicided_B process shown in Fig. 11. To confirm the argument that the maximum I_{I_2} can be obtained through the uniform bipolar conduction as either the finger width (W) or the channel length (L_{poly}) decreases, I_{I_2} was also measured for devices with $L_{poly} < 0.5$ μm . As shown in Fig. 12, the distribution of I_{I_2} for silicided_B devices demonstrates great improvement in I_{I_2} for $L_{poly} < 0.5$ μm even without a substrate bias. Overall the data suggest that bipolar conduction can be uniform within the finger dimension $L_{poly} < 0.3$ μm and $W < 5$ μm , though this dimension is not practically useful.

Among device parameters for ESD characterization such as the current gain β , the substrate resistance R_{sub} and the multiplication factor M , the β and R_{sub} are functions of L_{poly} . There is less chance for R_{sub} to change so abruptly around $L_{poly} \sim 0.45$ μm since in general the resistance has a linear dependency on the length. Hence the main cause in the dramatic improvement of I_{I_2} seems to be the change of β , which must be involved with 3D behavior. Amerasekera et al. [7] showed that the current gain dependence on L_{poly} is strongly process dependent where $1/L_{poly}$ dominates its dependence.

In this regard, for silicide_B process, L_{poly} dependence of β seems sharper than that of silicided_A process and thus in the range of $L_{poly} < \sim 0.45$ μm , increasing number of segmented transistors are activated to conduct more current by means of satisfying the snapback condition discussed earlier. Based on the value of $I_{I_{2i}}$ for each technology, the effective finger width W_{eff} , for bipolar conduction can be determined from the relation, $W_{eff} = [(I_{I_2}/I_{I_{2i}}) W]$. Fig. 13 shows plot of W_{eff} vs. W for various substrate bias and technologies at a given value of L_{poly} . At 1 V of substrate bias, W_{max} is 50 μm , 35 μm , and 25 μm for the non-silicided, silicided_A, and silicided_B technologies respectively, showing 2 to 5 times improvement over the W_{max} values for the zero substrate bias case.

These experimental results can be explained by the proposed physical model consisting of a discrete set of lateral bipolar *npn* transistors within a single finger transistor. The actual number of these *npn* devices that simultaneously turn on seems to be reduced with lower substrate resistance R_{sub} and lower current gain β because of the reduced probability of meeting the snapback condition, but can be improved with applied substrate bias by enlarging the conduction width resulting from the local substrate potential increment. In addition, the substrate bias scheme can extend the ESD design capability beyond its design and technology limit and one of its applications has been recently implemented [12]. More recently, the uniform substrate bias trigger effect that was reported here for single finger transistors was effectively demonstrated through evaluation of a multi-finger test circuit [5].

CONCLUSIONS

Non-uniform bipolar conduction for single finger NMOS transistors results in severe reduction in ESD protection strength. The effect of substrate bias to overcome the decrease in I_{I_2} and achieve practical design has been demonstrated. Finally, insight regarding *intrinsic second breakdown triggering current* $I_{I_{2i}}$ has been introduced, which can be used to generate efficient ESD design guidelines for deep submicron processes.

ACKNOWLEDGEMENTS

This research was supported through a customized SRC project funded by Texas Instruments Inc. The authors would also like to acknowledge Ajith Amerasekera and David Scott for their valuable comments, Vikas Gupta for his early contributions to the I_{I_2} behavior in this technology node, and Jon Brodsky for his help with the EMMI analysis.

REFERENCES

1. D. Scott, J. Hall and G. Giles, "A lumped element model for simulation of ESD failures in silicided devices," *EOS/ESD Symp. Proc.* pp. 41 – 47, 1986.
2. T. Polgreen and A. Chatterjee, "Improving the ESD failure threshold of silicided n-MOS output transistors by ensuring uniform current flow," *IEEE TED.* Vol. 39, no. 2, pp. 379 – 388, 1992.
3. P. Salome, C. Leroux, J. P. Chante, P. Crevel and G. Reimbold, "Study of a 3D phenomenon during ESD stresses in deep sub-micron CMOS technologies using photon emission tool," *IRPS. Proc.* pp. 325 – 332, 1997.
4. C. Russ, K. Bock, M. Rasras and I. D. Wolf, "Non-uniform triggering of gg-nMOST investigated by combined emission microscopy and transmission line pulsing," *EOS/ESD Symp. Proc.* pp. 177 – 186, 1998.
5. C. Salling, J. Hu, J. Wu, C. Duvvury, R. Cline, and R. Pok, "Development of substrate-pumped nMOS protection for a 0.13 μm technology," (submitted to the 2001 *EOS/ESD Symposium*).
6. V. Gupta, A. Amerasekera, S. Ramaswamy, and T. Taso, "ESD-related process effects in mixed-voltage sub-0.5 μm technologies," *EOS/ESD Symp. Proc.* pp. 161 – 169, 1998.
7. A. Amerasekera, V. Gupta, K. Vasanth and S. Ramaswamy, "Analysis of snapback behavior on the ESD capability of sub-0.20 μm NMOS," *IRPS. Proc.* pp. 159 – 166, 1999.
8. A. Amerasekera, C. Duvvury, V. Reddy and M. Rodder, "Substrate triggering and silicide effects on ESD performance and protection circuit design in deep submicron CMOS processes," *IEDM. Tech. Dig.* pp. 547 – 550, 1995.
9. R. W. Dutton, "Bipolar transistor modeling of avalanche generation for computer circuit simulation," *IEEE TED.* ED-22, pp. 334 – 338, 1975.
10. C-H. Choi, Z. Yu and R. W. Dutton, "Elimination of non-simultaneous triggering effects in finger-type ESD protection transistor using heterojunction buried layer," *SISPAD. Proc.* pp. 304 – 307, 1998.
11. G. Notermans, A. Heringa, M. V. Dort, S. Jansen and F. Kuper, "The effect of silicide on ESD performance," *IRPS. Proc.* pp. 154 – 158, 1999.
12. C. Duvvury, S. Ramaswamy, A. Amerasekera and R. A. Cline, "Substrate pump NMOS for ESD protection application," *EOS/ESD Symp. Proc.* pp. 1A.2.1-1A.2.11, 2000.