A Physics-based Compact Model for SCR Devices Used in ESD Protection Circuits

Robert Mertens and Elyse Rosenbaum
Department of Electrical and Computer Engineering
University of Illinois at Urbana-Champaign
1308 W. Main St., Urbana, IL 61801 USA

Achieving CDM-ESD reliability is increasingly difficult due to technology scaling, larger packages, and more complex mixed-signal designs [1]. Protection circuit design verification and optimization via circuit simulation becomes necessary. Recent works have presented layout-scalable, CDM-relevant compact models for SCR-based CMOS protection devices [2][3]. This paper presents a physics-based model that addresses shortcomings in the previous works; [2] employed a discontinuous model, while [3,4] incorrectly describes the resistive voltage drop across the N- and P-well as a reverse bias on the N-well/P-well junction and has a large parameter extraction burden (49 parameters, compared to 31 in this work). This work also describes a previously unnoted effect that influences the SCR I-V.

A schematic representation of the model is shown in Fig. 1. It includes cross-coupled NPN and PNP transistors. Each transistor uses a modified Ebers-Moll model and includes non-linear junction and diffusion capacitances, linear emitter and base resistances, non-linear collector resistance, and avalanche current across the N-well/P-well junction. The model equations are summarized in Table 1. This model has been implemented in Verilog-A and simulated using Spectre. The model in [3,4] also represented the SCR as two cross-coupled BJTs; however, its collector resistors are oriented such that they cannot contribute to the voltage drop between the anode and cathode. Thus in order to reproduce the on-branch of a measured I-V, the model must erroneously reverse bias the N-well/P-well junction. In [3,4], the BJTs are modeled using the diffusion based Gummel-Poon model. Since the SCR's on-state current is dominated by drift, the model requires highly empirical extensions. In contrast, the model in [2] represents the on-state SCR as a PIN diode; this provides a valid representation of the drift dominated current conduction in an on-state SCR. However, the PIN diode representation does not capture the device behavior in the off-state, requiring the use of a two-part model with its associated discontinuities. The model presented here is continuous while providing a physically accurate description of voltage drops between the anode and cathode.

In this work, the equations used to represent each transistor differ from the Ebers-Moll model by allowing the link current of each transistor to influence the current gain of the other. A majority carrier current flowing through the collector of one transistor will induce an electric field in the base of the other transistor, as illustrated in Fig. 2 for a diode-triggered SCR (DTSCR). The electric field induced in the base affects the minority carrier transport through this region and the current gain. In Fig. 3, the pulsed I-V characteristic of a 65nm DTSCR is compared to that from simulations in which the current gain parameter $ß_{FN}$ was either fixed at its measured value $ß_{N0}$ or allowed to vary using the model in Table 1; $ß_{N0}$ is obtained by characterizing the 3-terminal NPN. Simulation results match the measurement results only when the model of Table 1 is used, in which $ß_{FN}$ is a function of the PNP link current. Alternatively, correct simulation of the DTSCR I-V characteristic may be achieved using a constant, empirical value of $ß_{FN}=ß_{N0}^{(tuned)}$. However, this will provide incorrect simulation results if the core SCR is combined with a trigger circuit that injects current into the P-well, e.g., the GGSCR [5]; specifically, the value of I_{t1} will be incorrect.

The new model captures important transient effects in SCR-based protection circuits. Fig. 4(a) presents the measured and simulated transient responses of a 65nm DTSCR to pulses from a VFTLP tester; in Fig. 4(b), the peak voltage during device turn-on is plotted as a function of the steady-state current. The peak voltage is predicted correctly by simulation only when avalanche breakdown of the N-well/P-well junction is included in the model. For completeness, the I-V curves are shown in Fig. 4(c).

A single set of parameters may be used to represent devices with different layout spacings, as most parameters scale linearly with device dimensions (including τ_R, which represents the drift transit time at high currents). Fig. 5 shows the measured, pulsed I-V characteristics of 130nm DTSCRs with a variety of layout spacings and the corresponding simulation results; here, the model of Table 1 was enhanced to include self-heating in the emitter resistances [2]. The model well represents the dependence of I_{t1} and V_{Hold} on anode-to-cathode spacing (Fig. 5(a)), as the collector resistances, current gains, and charge storage parameters (τ_R and Q_{C0}) vary with this dimension. Changing the well tap spacings changes the base resistances; this will change the balance of current between the NW/PW junction and one or the other base resistance, depending on the trigger circuit implementation, thereby modulating the holding voltage as shown in Fig. 5(b). In this work, the holding voltage dependence on well tap spacing is modeled from first principles; prior models describe this behavior using purely empirical relationships [2-4]. The layout-scalable version of the model takes into account the effect of the vertical PNP formed by the anode, N-well, and P-substrate. As shown in Fig. 3, the current from the vertical PNP does not flow in proximity to the cathode diffusion, so it cannot effectively forward bias the P-well/cathode junction. An effective resistance is used for $R_{B,N}$, which changes both with P-well tap spacing and anode-to-cathode spacing, thereby affecting the ratio of lateral PNP current to vertical PNP current [2]. The full-length paper will present geometric equations scaling in detail.

[1] JEDEC, *Recommended ESD-CDM Target Levels JEP157*, Oct. 2009.
[2] J. Di Sarro et al., IEEE T-ED, pp. 3275–3286, Dec. 2010.
[3] A. Romanescu et al., EOS/ESD Symp., pp. 179-186, 2011.
[4] A. Romanescu et al., EOS/ESD Symp., pp. 21-30, 2010.
[5] C.C. Russ et al., EOS/ESD Symp., pp. 22- 31, 2001.
[6] D. Pogany et al., IEDM, pp. 16.6.1-16.6.4, 2009

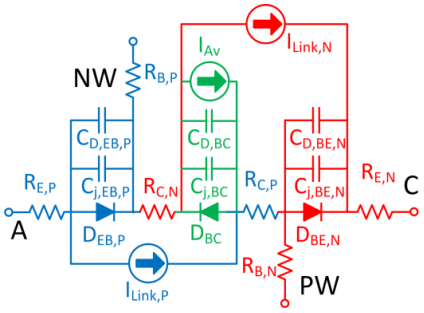

Fig. 1. Circuit representation of model. Components associated with the NPN and PNP are labeled with 'N' and 'P,' respectively. Components associated with the shared bases/collectors created by the NW/PW junction are labeled with 'BC.' The potential drop between the A and C terminals has three components: diodes drops, resistive drops across the conductivity modulated collector resistances, and ohmic drops across the emitter resistances.

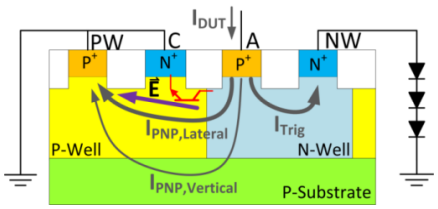

Fig. 2. The lateral portion of the PNP's collector current induces an electric field in the NPN transistor's base which aids electron transport across this region, thereby increasing the current gain of the NPN transistor.

Table 1. Selected equations for SCR model. Ebers-Moll equations are used for the NPN and PNP transistors. Each junction capacitor is modeled using the expression in the first column of row five. Avalanche current is modeled using the Miller multiplication expression. At voltages near the breakdown voltage, this expression is replaced with linear function to avoid convergence issues associated with the singularity. The link current through one transistor affects the current gain of the other, as explained in Fig. 2.

$I_{Link,N} = \boldsymbol{I_{SN}}\left[exp\left(\frac{V_{BE,N}}{V_T}\right) - exp\left(\frac{V_{BC,N}}{V_T}\right)\right]$	$I_{Link,P} = \boldsymbol{I_{SP}}\left[exp\left(\frac{V_{EB,P}}{V_T}\right) - exp\left(\frac{V_{CB,P}}{V_T}\right)\right]$
$I_{D,BE,N} = \frac{\boldsymbol{I_{SN}}}{\boldsymbol{\beta_{FN}}}\left[exp\left(\frac{V_{BE,N}}{V_T}\right) - 1\right]$	$I_{D,EB,P} = \frac{\boldsymbol{I_{SP}}}{\boldsymbol{\beta_{FP}}}\left[exp\left(\frac{V_{EB,P}}{V_T}\right) - 1\right]$
$I_{D,BC} = \boldsymbol{I_{SR}}\left[exp\left(\frac{V_{BC}}{V_T}\right) - 1\right]$	$Q_{D,EB,P} = \boldsymbol{I_{SP}\tau_{FP}}\left[exp\left(\frac{V_{EB,P}}{V_T}\right) - 1\right]$
$Q_{D,BE,N} = \boldsymbol{I_{SN}\tau_{FN}}\left[exp\left(\frac{V_{BE,N}}{V_T}\right) - 1\right]$	$Q_{D,BC} = \boldsymbol{I_{SR}\tau_R}\left[exp\left(\frac{V_{BC}}{V_T}\right) - 1\right]$
$C_j = \frac{\boldsymbol{C_{J0}}}{\left(1 - \frac{V_j}{\boldsymbol{\varphi}}\right)^{\boldsymbol{m_j}}}$	$I_{Av} = \boldsymbol{I_{SR}}\left(\frac{1}{1 - \left(\frac{V_{Av}}{\boldsymbol{BV_{NW-PW}}}\right)^{\boldsymbol{m_{Av}}}} - 1\right)$
$R_{C,N} = \frac{\boldsymbol{R_{C,N0}}}{1 + \frac{Q_{D,BC}}{\boldsymbol{Q_{C,N0}}}}$	$R_{C,P} = \frac{\boldsymbol{R_{C,P0}}}{1 + \frac{Q_{D,BC}}{\boldsymbol{Q_{C,P0}}}}$
$\beta_{FN} = \boldsymbol{\beta_{N0}} + \frac{I_{Link,P}}{\boldsymbol{I_{\beta P}}}\left(\frac{1}{1 + \frac{I_{Link,P}}{\boldsymbol{I_{\beta PSat}}}}\right)$	$\beta_{FP} = \boldsymbol{\beta_{P0}} + \frac{I_{Link,N}}{\boldsymbol{I_{\beta N}}}\left(\frac{1}{1 + \frac{I_{Link,N}}{\boldsymbol{I_{\beta NSat}}}}\right)$

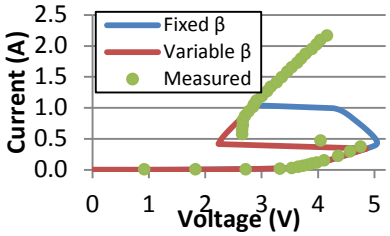

Fig. 3. Measured and simulated I-V curves for a DTSCR in 65nm CMOS. Using the measured β_{N0} of the 3-terminal NPN yields a poor fit where both transistors are active, i.e. the NDR region. The measurement data suggest $R_{on}\approx 0$ near the holding point; this is attributed to non-uniform conduction across the device width [6], which existing models—[2-4] and this work—don't replicate.

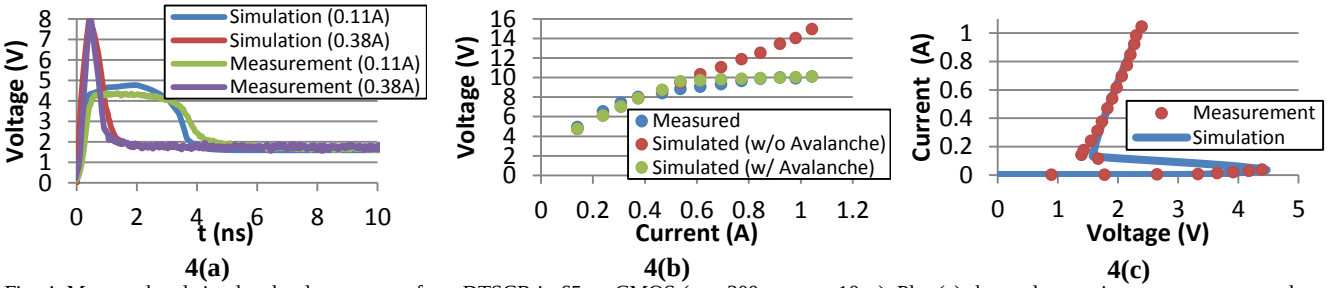

Fig. 4. Measured and simulated pulse response for a DTSCR in 65nm CMOS (t_{rise}=300ps, t_{width}=10ns). Plot (a) shows the transient responses at steady-state currents of 0.11A (1.5mA/μm) and 0.38A (5mA/μm) Plot (b) shows the peak voltage observed as a function of the steady-state current; simulation is performed with and without NW/PW avalanching modeled. V_{peak} > dc-V_{t1} is observed due to the pulse's short risetime. Plot (c) shows the pulsed I-V characteristic for the device.

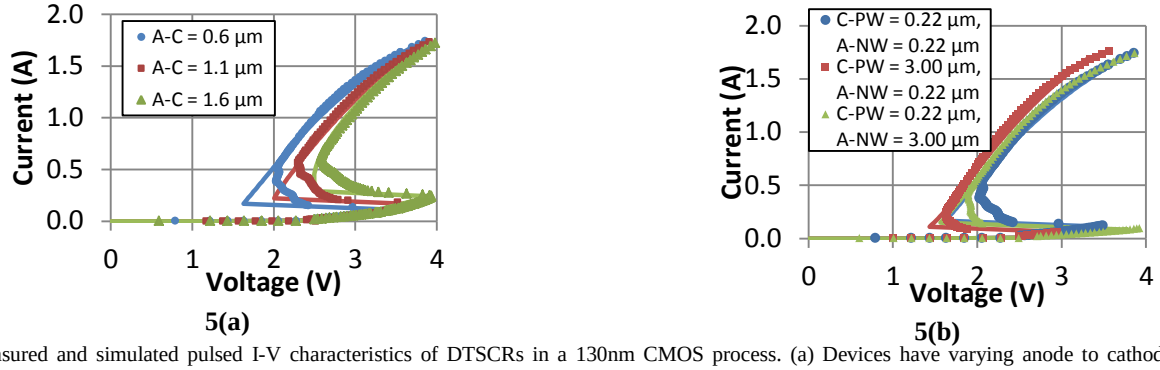

Fig. 5. Measured and simulated pulsed I-V characteristics of DTSCRs in a 130nm CMOS process. (a) Devices have varying anode to cathode spacing; geometric scalability is achieved by using measured β values and linearly extrapolated R_{C0}, Q_{C0}, and τ_R values. (b) Devices with varying well tap spacings. R_B is a non-linear function of well tap spacing.