Compact Models for Simulation of On-Chip ESD Protection Networks

Elyse Rosenbaum, *Fellow, IEEE*, Shudong Huang, Matthew Drallmeier, and Yujie Zhou

(Invited Paper)

***Abstract*— Technology scaling and increased data rates make it near impossible to achieve historic levels of electrostatic discharge (ESD) robustness. This heightens the need for pre-Si verification that a design's ESD level is above a critical value, below which the yield loss and the number of field returns are expected to be high. Transient simulation plays a role in ESD design verification and requires the availability of accurate compact models of the various semiconductor devices, which lie along the discharge path. The compact models included in a foundry process design kit (PDK) are not accurate at ESD current levels. This article describes compact models that have been developed in the ESD device research community. It reviews the measurements used to characterize ESD protection devices and acquire data for model parameter extraction. It is concluded that obtaining accurate measurement data is challenging and this impedes the widescale adoption of ESD compact models.**

***Index Terms*— Charged device model (CDM), compact models, electrostatic discharge (ESD) protection, silicon-controlled rectifier (SCR).**

I. Introduction

ON-CHIP electrostatic discharge (ESD) protection is a necessity for high-yield CMOS integrated circuits. Over the past 30 years, various persons have advocated for on-chip networks to be designed with the aid of circuit simulation [1], [2], [3], [4], [5], [6], [7], and they developed compact models for that purpose, e.g., [8], [9], [10], [11], [12], [13], [14], [15], [16]. The resources devoted to those modeling efforts pale in comparison to those expended in the development of foundry-supported models, such as BSIM [17] and VBIC [18], and the associated parameter extraction procedures. Due to the lack of a large-scale coordinated effort, ESD compact models are not available to many circuit designers, and most chips are designed without performing transient simulation of the on-chip protection circuits. In such cases, the on-chip protection is designed around prior art, foundry-provided data, and conservative assumptions. A programmable electrical rules checking (PERC) tool may be used to verify that there is a low impedance path for shunting ESD current between any pair of pins [19]. A conservative, simulation-free design approach might produce an ESD-robust design, especially in older process nodes, but it will not produce the minimum area ESD protection circuit; finding the minimum area solution requires simulation-based design optimization [20].

The most advanced wireline transceivers in development operate at data rates in the range of 112–224 Gb/s, corresponding to Nyquist frequencies in the range of 28–56 GHz [21], [22], [23], [24]. The total capacitance at the IO pin, contributed by the bondpad, ESD protection devices, and input or output transistors, is invariably too large to meet return loss specifications, necessitating the use of bandwidth extension circuits, such as T-coils, and placing tight limits on the allowable ESD capacitance. I_{T2}/C—thermal breakdown current per unit capacitance—is a figure of merit for ESD protection devices and its value decreases with technology scaling. Thus, a tight capacitance limit corresponds to a reduced ESD protection level in the advanced nodes, where the >100-Gb/s serial IOs are designed. This is not problematic if the ESD target level can be reduced [25], e.g., due to improved factory ESD control, but if the target level is reduced too aggressively, it will result in reduced yield and increased field returns.

Today, one can make a compelling argument that ESD transient simulation is needed to maintain acceptably high ESD protection levels without unduly compromising the IO performance. Accurate transient simulation enables a designer to determine how much of the ESD-induced pin voltage is applied to each transistor in the IO circuit and what fraction of the injected current flows through each device—not just the protection devices, but also the active circuit devices. If simulation reveals that the IO protection circuit will not meet the ESD targets when its devices are sized sufficiently small to satisfy the signal integrity specifications, the designer may use additional simulations to investigate and optimize design modifications. The transient simulation of IO protection circuits requires compact models; those models are the focus of this article.

The next section of this article reviews the measurement techniques used to characterize on-chip ESD protection devices. It describes the limitations of the techniques and how those compromise the thorough validation of ESD compact

Manuscript received 16 August 2023; accepted 24 August 2023. Date of publication 23 October 2023; date of current version 2 January 2024. This work was supported by the Semiconductor Research Corporation (SRC) Task through the Texas Analog Center of Excellence (TxACE) at The University of Texas at Dallas under Grant 2810.066. The review of this article was arranged by Editor H. Agarwal. *(Corresponding author: Elyse Rosenbaum.)*

The authors are with the Department of Electrical and Computer Engineering, University of Illinois Urbana-Champaign, Urbana, IL 61801 USA (e-mail: elyse@illinois.edu; shudong3@illinois.edu; mjd17@illinois.edu; zhou117@illinois.edu).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TED.2023.3320093.

Digital Object Identifier 10.1109/TED.2023.3320093

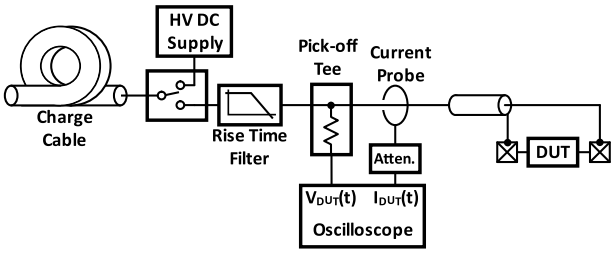

Fig. 1. TLP tester. All the transmission lines have a characteristic impedance of 50 Ω. The duration of the generated pulse is proportional to twice the length of the charge cable. The pulse transit time from the probes to the DUT and back is much less than the pulse duration. Commercial testers can deliver as much as 40 A to the DUT.

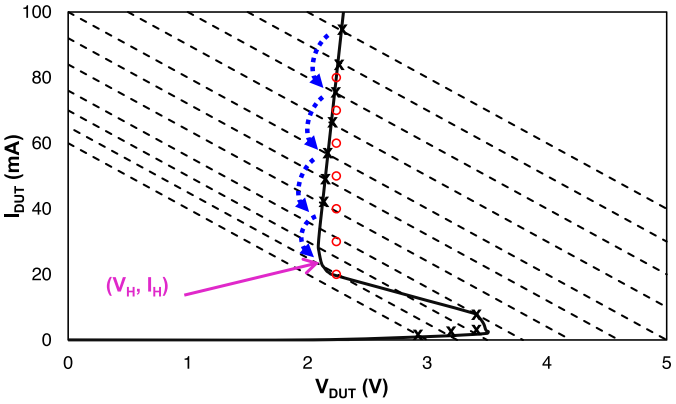

Fig. 2. Representative *I*–*V* characteristic of a device that undergoes snapback; the *I*–*V* curve resembles the letter S. A TLP tester samples (*V*, *I*) points marked by x, which lie at the intersection of its load line (dashed black lines) and the DUT's (quasi-) static *I*–*V*. The (*V*, *I*) points in red exist if the current density is nonuniform across the width of the DUT. The blue arrows illustrate the trajectory during a slow falling edge.

models, which in turn impedes the wide-scale adoption of those models. Section III presents examples of three ESD compact models developed by our research group. The models of the diode and silicon-controlled rectifier (SCR) are developed for dedicated protection devices. In contrast, the model presented for the MOS transistor may be applied to devices that are part of the active circuit. In the concluding section of this article, we briefly consider the roles that IC package modeling and full-chip ESD design checking play in ESD design verification.

II. ESD Device Characterization

Ordinarily, compact model parameters are derived from easy-to-perform dc current and C–V measurements. Ensuring that the model is accurate at gigahertz frequencies may require the C–V measurement data to be augmented by S-parameter measurement data, but in any case, the device dynamic response is measured by applying small signals that cause it to behave linearly. The instruments used for parameter extraction are widely available and the measurements are highly repeatable.

Extracting the parameters of an ESD device model is a far more challenging task. The I–V characteristic of an ESD protection device cannot be obtained from dc measurements nor can a C–V measurement be performed at a suitable dc bias point, because ESD protection devices are operated at higher current levels, i.e., larger power densities, than can be sustained under dc conditions. Protection devices are operated safely at very high power densities due to the short duration of an ESD event, which for component-level ESD ranges from about 1 ns to about 100 ns. The thermal failure current is a decreasing function of the stress duration until thermal steady-state is reached at about 100 μs [26], [27]. A further challenge is posed by the inability of the usual dc and ac (small signal) measurements to capture important dynamic effects, specifically, forward recovery and reverse recovery.

A. Transmission Line Pulse Testing

The I–V characteristic of an ESD protection device is measured using a transmission line pulse (TLP) tester [28], illustrated in Fig. 1. A TLP system produces single-shot current pulses of high amplitude, up to tens of amperes. A typical pulsewidth is 100 ns and rise time is 10 ns.

In principle, the parameters of a static model (low-frequency model) can be extracted from the pulse I–V data. However, it may not be possible to validate the accuracy of the device I–V model at all current levels, because many ESD protection devices have an S-shaped I–V characteristic, in which the applied voltage does not uniquely determine the static current. A representative S-shaped I–V curve is shown in Fig. 2. The TLP tester only samples (V, I) points that intersect with its 50-Ω load line and, if the load line intersects the I–V curve at multiple points, only one point is sampled, as indicated by the "x" on the plot in Fig. 2. Notably, the I–V curve of Fig. 2 is not sampled near its holding point, (V_H, I_H). The extent of the unsampled region depends on the ratio of the tester's 50-Ω output resistance to the ON-resistance (R_{on}) of the device under test (DUT); a very large ratio allows most of the I–V curve to be sampled. Poor resolution near the holding point is often encountered when characterizing MOS transistors, which have high R_{on} relative to a comparably sized SCR.

The holding point can be sampled by special instruments, e.g., a wafer-level human body model (HBM) tester [29]. Those testers produce pulses with very slow falling edge transients; during the pulse falling edge, the DUT moves along the quasi-static (QS) trajectory indicated by the blue arrows in Fig. 2. The holding point is read from an I–V curve that was constructed from a single transient measurement, using pairs of $(V(t), I(t))$. Note that a conventional TLP system is used only to obtain a quasi-static (pulse) I–V; despite its nanosecond-scale rise and fall times, it is not used to characterize the dynamic response of the DUT. This is a consequence of the pulse's nonzero transit time between the voltage and current probes and the DUT; the probes capture the superposition of the time-offset incident and reflected pulses. The time offset corrupts the accuracy of the measured voltage and current unless those quantities are constant, or they change very slowly.

A device with an S-shaped I–V characteristic can have a stable operating state in which only a portion of the

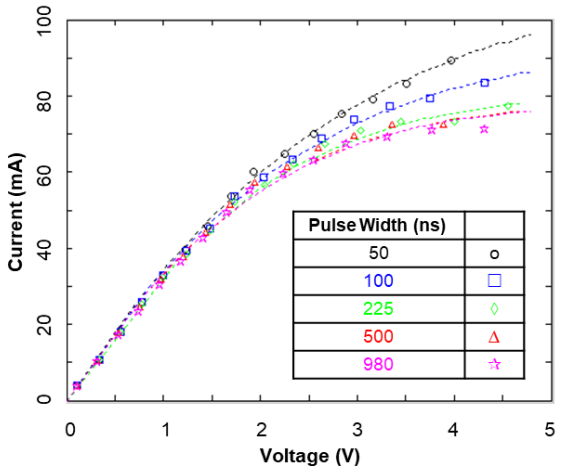

Fig. 3. Pulse I–V of a nonsilicided N^+ diffusion resistor [35]. The markers are measurement data; the dashed lines are simulation results. The compact model includes self-heating (1) and velocity saturation.

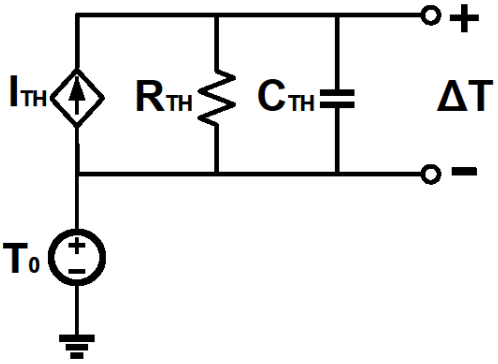

Fig. 4. Equivalent circuit used to solve for the temperature rise above the ambient temperature, ΔT. For a resistive element R with current I, the quantity I_{TH} is set equal to $V_R I$, where V_R is the voltage drop across the resistor.

device conducts the ESD current [30]. Nonuniform current conduction manifests itself in two ways. The red markers in Fig. 2 correspond to a case in which the current flows through only a fraction of the device width. When the current is increased, a larger fraction of the device width turns on, yielding a differential resistance of zero. Compact models describe terminal currents, not the current distribution inside the device; as a result, even though the zero-resistance branch can be easily measured with TLP [31], [32], compact models do not replicate it. The other manifestation of nonuniform conduction arises in devices with a multifinger layout, when some but not all the fingers turn on. The nonuniform triggering of the device fingers can be captured in simulation if the unequal p-well (or n-well) resistances of the individual fingers are modeled, as demonstrated in [33].

It is customary to measure the quasi-static I–V characteristic of an ESD protection device using a variety of pulse widths. The I–V measurements are taken up to the failure point, to quantify the failure current for different duration ESD events. The device fails due to current filamentation and thermal runaway [34], effects not readily captured in a compact model; however, the usual purpose of ESD circuit simulation is to check whether all the devices remain biased within safe limits rather than to simulate the dynamics of failure. Pulse I–V data show that the ON-resistance of a protection device increases with pulsewidth and current density; this is a consequence of Joule heating in the device series resistance. An ESD protection circuit should be designed such that its devices are operated well below their failure currents, to avoid the increased ON-resistance and degraded voltage clamping that occur at the highest current levels, and an ESD compact model should include self-heating. That effect can be significant, as evident in the pulse I–V characteristic of an N^+-diffusion resistor shown in Fig. 3.

Self-heating is not the only effect that may degrade a device's ON-resistance (R_{on}) at high current densities; velocity saturation or space charge limited conduction may also play a role. However, self-heating is the sole cause of the pulsewidth dependency. For ohmic conduction, a temperature-dependent series resistance $R(T)$ may be modeled as shown in the following:

$$R(T) = R_0\left(1 + \frac{T - T_0}{T_0}\right)^{\chi}, \tag{1}$$

where R_0 and χ are the fitting parameters, and T_0 is the ambient temperature.

In (1), the quantity $T - T_0$ is the temperature rise, denoted as ΔT. The time-dependent ΔT is solved for by the simulator using a thermal equivalent circuit. A one-pole thermal circuit is depicted in Fig. 4. On the time scales of interest, analytic and experimental works suggest that initially $\Delta T \propto \sqrt{t}$, where t is time, later transitioning to $\Delta T \propto \ln(t)$ [26], [27]. Clearly, the functional form of the simulated temperature rise, $\Delta T \propto 1 - e^{-t/R_{\rm th}C_{\rm th}}$, cannot fit the true function over a very wide range of times. Therefore, it may be advisable to adopt an RC model with more than one time constant (i.e., pole) if the model will be used to simulate ESD events whose durations differ by more than one or two orders of magnitude [36].

It was noted earlier that devices with an S-shaped I–V characteristic may operate in a stable state of nonuniform conduction. However, it is also possible that the current gradually spreads out across the entire device width, presumably due to thermal gradients [32]. An example of that time evolution is seen in [37, Fig. 6], where the SCR ON-resistance is observed to decrease during the first 20 ns of each current pulse. That finding is consistent with the current becoming distributed across a larger fraction of the device width. Dynamic changes in current density are not captured by today's compact models.

B. Dynamic Effects

Protection devices with an S-shaped I–V characteristic are referred to as snapback devices, because the voltage across the device decreases when the device is triggered from its high-impedance state to its low-impedance state. The trigger point is denoted as (V_{t1}, I_{t1}). Displacement current may help to switch a snapback device from its OFF-state into the ON-state, in which case V_{t1} is a function of the pulse rise time, a phenomenon referred to as dV/dt triggering [9], [38]. TLP I–V measurements usually are performed with a pulse

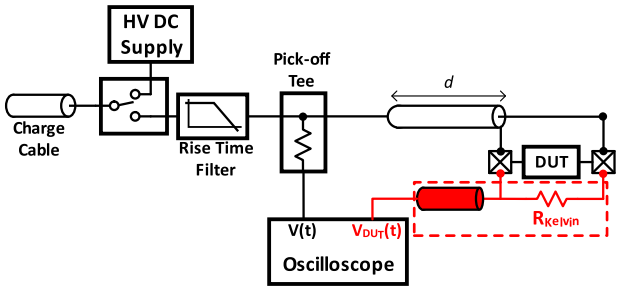

Fig. 5. VFTLP tester. The reflected pulse arrives at the voltage probe 2τ seconds after the incident pulse ($\tau \propto d$). VFTLP was developed as a time-domain reflectometry system, in which the DUT voltage and current are calculated from the measured pulses using $V_{\mathrm{DUT}}(t) = V_{\mathrm{inc}}(t-\tau) + V_{\mathrm{ref}}(t+\tau)$ and $I_{\mathrm{DUT}}(t) = (1/50)[V_{\mathrm{inc}}(t-\tau) - V_{\mathrm{ref}}(t+\tau)]$. Optionally, a Kelvin probe (shown in red) may be placed at the DUT for direct measurement of $V_{\mathrm{DUT}}(t)$.

rise time of 10 ns; however, a variable rise time may be used to investigate *dV/dt* triggering. *dV/dt* triggering is initiated with the device in the OFF-state and the relevant capacitances, primarily junction capacitances, may be extracted using C–V measurements.

The ON-resistance of protection devices that operate with ambipolar conduction, e.g., diodes and SCRs, decreases on a time scale measured in hundreds of picoseconds; this is referred to as forward recovery and examples of the recovery transients will be shown in Sections III-A and III-B. Forward recovery results from conductivity modulation, which occurs if the density of majority carriers in a quasi-neutral region significantly exceeds the doping concentration; the amount of charge stored in a device cannot change instantly, and thus there is a delay before the conductivity reaches its final value. The device ON-resistance transitions to a lower value in the same time interval that the charging current flows, i.e., when the device is not operating statically. Thus, forward recovery cannot be extracted from a quasi-static TLP I–V curve. However, it is critically important that the phenomenon be modeled, because it compromises the voltage clamping ability of the protection device; a large transient voltage overshoot is associated with gate oxide failure in input transistors [39], [40], [41].

Reverse recovery is a large-signal non-quasi-static (NQS) effect, which cannot be captured by dc or ac measurements. The reverse recovery current flows until the density of excess carriers stored in the device is reduced to its static value. In the context of ESD simulation, the importance of reverse recovery is less certain than that of forward recovery. However, reverse recovery must be included in a completely general model; a general model can be used to represent the device response to bipolar as well as unipolar ESD current injections.

To characterize forward and reverse recovery in ESD protection devices requires instrumentation beyond a conventional TLP tester.

C. Very-Fast TLP Tester

A modified version of the TLP tester, named very-fast TLP (VFTLP), is shown in Fig. 5. VFTLP was developed to measure the time-varying $V_{\mathrm{DUT}}(t)$ and $I_{\mathrm{DUT}}(t)$, rather than just the quasi-static I–V [42]. The distance d from the probes to the DUT is made intentionally long, so that the incident and reflected pulses do not overlap; VFTLP pulses are usually just a few nanoseconds in duration. In essence, VFTLP is a time-domain reflectometry measurement system. The resultant transient waveforms are not without errors; the accuracy is limited by the sampling rate of the oscilloscope and the precision of the cable delay measurement. More recent works, e.g., [43], established that the VFTLP voltage transient can be measured with higher precision by placing a high-impedance probe directly in parallel with the DUT, i.e., the Kelvin probe inside the dashed box of Fig. 5. The Kelvin probe also mitigates measurement error due to contact resistance. A VFTLP tester is constructed with high-bandwidth components, enabling it to deliver pulses with rise time as short as 100 ps.

If a high resolution measurement of the current transient is desired, e.g., to measure reverse recovery, the oscilloscope can be placed in series with the DUT, such that $I_{\mathrm{DUT}}(t) = (V_{\mathrm{meas}}(t)/Z_o)$, where it is assumed that the input impedance of the oscilloscope is matched to the characteristic impedance of the cable, Z_o, which is 50 Ω. Simburger et al. [43] provide a thorough description of the various TLP configurations.

VFTLP-based reverse recovery measurement is imperfect for compact model parameter extraction and model validation. Ideally, a bias-tee would be used to apply a dc reverse bias to the DUT along with a VFTLP pulse. The pulse would drive the DUT into the ON-state and then back into the OFF-state, at which time reverse recovery would be observed. Unfortunately, TLP and VFTLP systems do not produce a well-controlled and smooth falling edge transient. To quickly drive the device from forward to reverse bias with a well-defined voltage transition time requires one to use the first edge of the VFTLP pulse. Therefore, as shown in [43] and [44], a dc forward bias is applied to the DUT, and then the device is driven into reverse bias by a reverse polarity VFTLP pulse applied through a bias-tee. To avoid thermal failure, the magnitude of the dc forward bias is limited to values that are smaller than the typical forward bias during ESD, which impedes the ESD model verification.

D. Parameter Extraction Using Dynamic Waveforms

In most ESD research laboratories, TLP and VFTLP testers are the only instruments that can provide the short-duration high-current injections needed to characterize ESD devices. The transient voltage response of the device to a VFTLP current pulse is used to extract model parameters related to forward recovery, and the transient current response to an injected voltage pulse is used to extract model parameters related to reverse recovery [44], [45]. The model parameters are extracted by fitting the simulated waveform to the measured one. This procedure is sound if and only if the incident pulse is represented correctly in simulation. First, the simulation model of the VFTLP system must correctly represent the frequency responses of the voltage probe, current probe, and oscilloscope. In particular, the Kelvin probe of Fig. 5 has a high-pass characteristic that can amplify a forward recovery transient.

Second, the VFTLP tester must be modeled accurately. The pulses produced by VFTLP testers are not perfect square (or trapezoidal) pulses. The rising edge does not have a constant slew rate, and there is some ringing after the rising edge transient. If the VFTLP tester is represented as an ideal square or trapezoidal pulse source in simulation and the model parameters are optimized such that the simulated results fit the measured ones, the corresponding parameters are inaccurate.

A VFTLP tester's nonidealities can be captured in simulation by using a numeric model of the tester [46]. To obtain that model, the 50-Ω tester is discharged into a 50-Ω load and $V_{\text{pulse}}(t)$ is recorded. In simulation, the tester is represented by a Thevenin equivalent circuit consisting of a voltage source, $2 \cdot V_{\text{pulse}}(t)$, in series with a 50-Ω resistor.

TLP and VFTLP measurement data are used for both parameter extraction and model validation. This means that the model only is proven to replicate the device's response to a square pulse; the model has not been proven to generalize to other stimuli. Earlier, the availability of wafer-level HBM testers was remarked upon and those might seem to provide an opportunity for improved model validation. However, the HBM transient is so slow that the device behaves quasi-statically.

A charged device model (CDM) ESD transient has a very high slew rate, which can exceed 10^9 A/s, suggesting that a CDM impulse might be highly suitable for validating the compact model of an ESD protection device. However, the CDM test standard, ANSI/ESDA/JEDEC JS-002, was developed for field-induced CDM (FICDM) testers. FICDM is a one-pin test and it does not provide a means to perform a voltage measurement. Therefore, one cannot observe the transient response of an ESD device using a CDM tester.

Capacitively-coupled TLP [47] and wafer-level CDM [48] testers may be used to inject high-current nonsquare pulses into wafer-level test structures, but those instruments are not in wide commercial distribution and are found in just a few research laboratories. One can create decaying oscillatory pulses by inserting an *LC* tank at the output of a VFTLP tester; the resultant system is named oscillatory TLP (OTLP) [49]. OTLP has been used to observe charge storage effects in SCRs [50] and will be used in the next section to help validate a compact model of an ESD diode.

III. Compact Models

In this section, examples of physics-based ESD compact models of diodes, MOSFETs, and SCRs are presented. The models were implemented in Verilog-A and simulations were performed using commercial circuit simulators, e.g., Specter [51]. As detailed in Section II, measurement challenges limit one's ability to fully validate the transient model of an ESD device. Thus, one could make a reasonable case for using empirical models, curve fit to the available data. However, the physics-based modeling approach has provided excellent insight into the inner workings of ESD protection devices and enabled ESD device designers to improve performance and adapt to new technologies.

In an on-chip ESD protection network, diodes are used as rectifiers. The diode is intended to be reverse biased under normal operating conditions. The SCRs and MOSFETs inside a protection circuit are used as switches. Those devices may be switched from the OFF-state (blocking state) to the ON-state by a trigger circuit that has detected the ESD transient. The diodes and/or SCRs inside a protection network are dedicated protection devices. However, under normal operating conditions, those devices are not ideal open circuits, and the OFF-state impedance of the IO protection devices must be included in signal integrity simulations. Reasonably, one may optimize the compact model of a dedicated protection device, such that it accurately represents the OFF-state impedance, the static current in all bias ranges, and the dynamic response of the device to ESD-like transients.

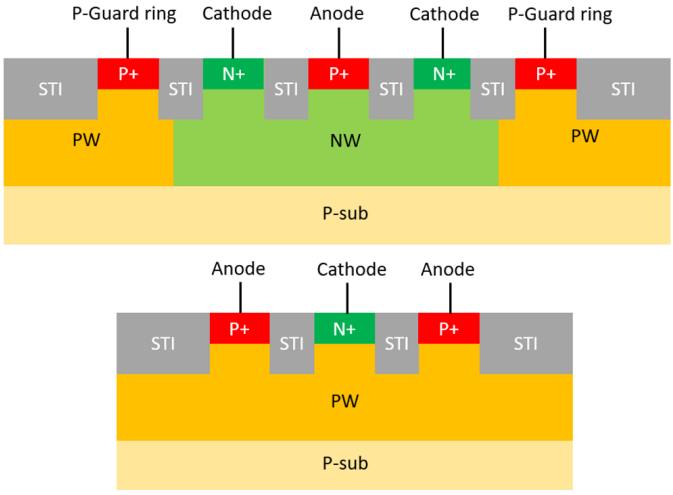

Fig. 6. Cross-sectional illustrations of n-well (top) and p-well (bottom) diodes.

In contrast, some of the MOSFETs along the discharge path may not be dedicated protection devices. Those devices may be part of the active circuit, and thus a different modeling approach is warranted.

In all cases, the model equations must be numerically well conditioned, e.g., lacking singularities, so that convergence problems do not arise during circuit simulation [16], [52].

A. ESD Diode in a Low-Voltage CMOS Technology

ESD protection diodes are p-n junction devices implemented as either a P^+-diffusion to n-well junction (n-well diode) or a p-well to N^+-diffusion junction (p-well diode), as illustrated in Fig. 6. Separate models are constructed for the two devices because the n-well diode includes a parasitic p-n-p transistor.

The SPICE diode model is not a suitable ESD model, even for the p-well diode, because it does not include forward recovery, which may impact the device's ability to provide protection. There are specialized diode models that include forward and/or reverse recovery, but those were developed for high-voltage p-i-n diodes [53], [54]. The reverse recovery transient of a p-i-n diode is very pronounced, and it needs to be captured in simulation because it reduces power converter efficiency [55]. The reverse recovery transient for a low-voltage CMOS diode is much briefer than for a p-i-n diode, and it can be observed only if the measurement setup is carefully designed to minimize stray inductance [44]. The effect of

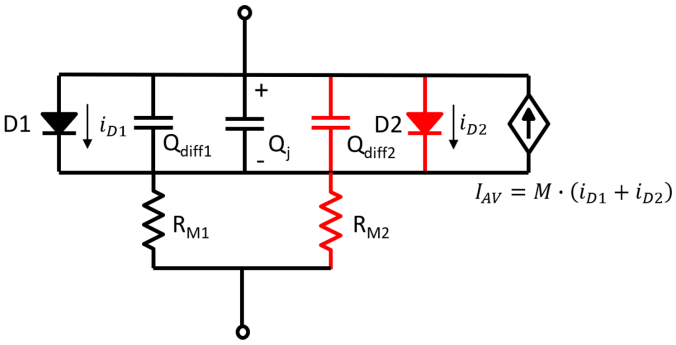

Fig. 7. Schematic representation of p-well diode model. The elements in red are not part of the model given by the equations of Table I.

reverse recovery on ESD robustness is uncertain, although it has been linked to ESD failures during machine model testing [56]. Bipolarity current may also be injected into an IO pin during IEC 61000-4-2 system-level ESD due to the impedance mismatch between the ESD gun and the equipment under test [46], suggesting that reverse recovery should be included in the compact model.

Forward recovery in low-voltage ESD diodes is modeled by including a conductivity-modulated resistor in series with the rectifying p-n junction [12], [57]. Reverse recovery is an NQS effect [54]. A physically rigorous modeling approach based on the finite difference solution of the current continuity equation is inconsistent with a compact modeling approach that prioritizes computational efficiency. Therefore, the above-cited p-i-n diode models use semi-empirical NQS models to capture reverse recovery.

Forward recovery can be simulated accurately using a quasi-static charge control model of the rectifying p-n junction [12] because the forward recovery transient is dominated by the response of the conductivity-modulated series resistance. The speed at which the resistance transitions to a lower value is determined by the rate at which the external circuit can provide charge to the diode. In contrast, the reverse recovery transient is determined solely by the diode's intrinsic dynamics (diffusion and recombination), requiring a more physical treatment of those dynamics than a quasi-static model can offer. In fact, if a quasi-static model is optimized to fit the forward recovery transient, it greatly overestimates the magnitude and duration of the reverse recovery transient [44]. Therefore, an NQS model of the ESD diode is adopted, inspired by those used for p-i-n diodes [53].

1) P-Well Diode Model: A schematic representation of the p-well diode model is provided in Fig. 7, and the model equations are summarized in Table I. For convenience, a simplified form of the expression for the static current (5) is presented in the table; it is valid when there is high-level injection (HLI) into the quasi-neutral "base" region of the diode, the usual case during ESD.

The minority carrier density at the junction side of the quasi-neutral region in the p-well, which is proportional to Q_F, is assumed to follow the instantaneous junction voltage, as reflected in (4) and (5). In contrast, it takes a nonzero amount of time for the total amount of charge stored in the quasi-neutral region, Q_D, to reach its static value, which is

TABLE I

SELECTED EQUATIONS FOR THE P-WELL DIODE MODEL. V_T DENOTES THE THERMAL VOLTAGE. EQUATION (8) [58] IS VALID ONLY FOR $V_j < 0$. FOR $V_j \geq 0$, M IS SET TO ZERO. EQUATION (10) IS VALID ONLY FOR $V_j < 0.5\ V_{bi}$. THE FUNCTION S IS GIVEN IN [46]. MODEL PARAMETERS ARE BOLDED

$0 = \frac{dQ_D}{dt} + \frac{Q_D}{\boldsymbol{\tau}} - i_D$	(2)
$i_D = \frac{Q_F - Q_D}{\boldsymbol{T_M}}$	(3)
$Q_F = (\tau + T_M) I_F$	(4)
$I_F = \boldsymbol{I_S}\left[\exp\left(\frac{V_j}{2V_T}\right) - 1\right]$	(5)
$V_j = V_D - i_D R_D$	(6)
$R_D = \boldsymbol{R_0} + \frac{\boldsymbol{R_{m0}}}{1 + \frac{R_{m0} Q_D}{2 T_M V_T}}$	(7)
$M = S \cdot \boldsymbol{\alpha_{cr}} \frac{-V_j}{1 - \boldsymbol{m}} \exp\left(-A_{AVL} \cdot (V_{bi} - V_j)\right)$	(8)
$A_{AVL} = \frac{\boldsymbol{A_{cr}}(1-m)}{C_{j0} V_{bi}^{\ m}}$	(9)
$C_j = \frac{\boldsymbol{C_{j0}}}{\left(1 - \frac{V_j}{\boldsymbol{V_{bi}}}\right)^m}$	(10)

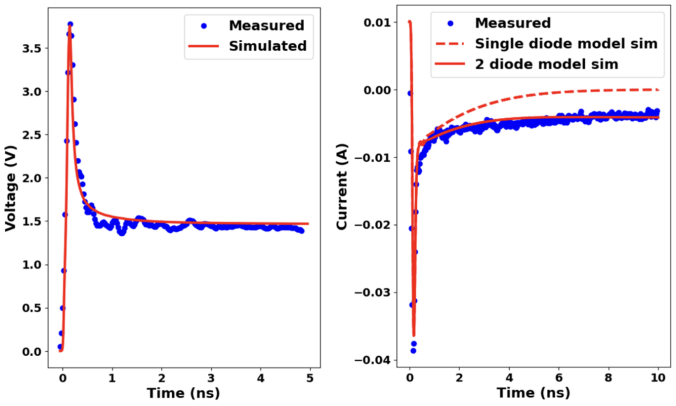

Fig. 8. 65-nm CMOS p-well diodes were characterized using VFTLP. (a) Forward recovery, 2-A injected current. (b) Reverse recovery. The same diode was simulated using the model of Table I (dashed line). The residual reverse current is simulated more accurately when the augmented model of [46] is used (solid line); the two models give identical results for the forward transient.

reflected in (2) and (3). Physical considerations suggest that the parameter T_M should be close in value to the transit time τ, and that expectation is borne out when the model parameters are optimized.

The diode series resistance R_D, (7), is the sum of a constant resistance R_0 and a conductivity-modulated resistance; the constant component models the contact resistance and the resistance of the heavily doped N^+ and P^+ diffusions. The conductivity-modulated component decreases as the charge Q_D builds up.

Fig. 8 shows the examples of forward and reverse recovery transients, both measured and simulated. Additional examples may be found in [46], and those demonstrate that the amplitude

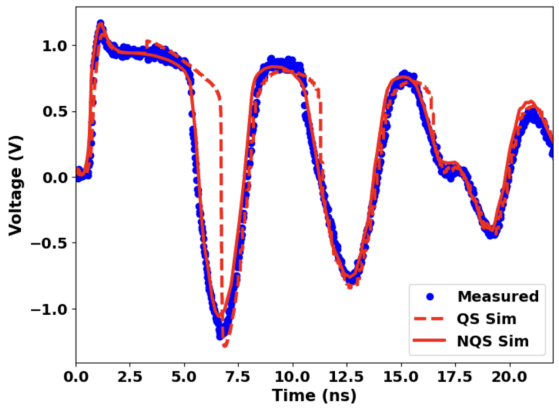

Fig. 9. Measured and simulated diode response to the OTLP waveform. The NQS model produces a more accurate result than the quasi-static (QS) model does.

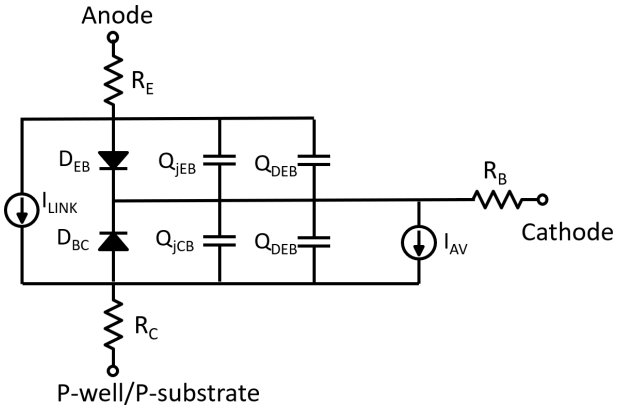

Fig. 10. Schematic view of the n-well diode model.

of the forward recovery transient is a decreasing function of the pulse rise time. In Fig. 8(b), the measured current remains nonzero significantly longer than in simulations that use the model of Table I. The residual current is small in magnitude and might reasonably be neglected. It is caused by charge that was stored deep in the p-well, away from the p-n junction [46]; current flows until all the excess carriers have been removed by diffusion and recombination. The residual current can be modeled by adding a second diode in parallel with the one described in Table I [46]. The primary diode carries most of the static current, and it therefore controls the forward recovery transient and initial reverse recovery impulse. The secondary diode has a longer base length and will store charge farther from the junction, leading to an extended reverse recovery. As shown in Fig. 8(b), the augmented model well represents the prolonged reverse recovery.

When a reverse bias is applied to the p-n junction, the current passing through it undergoes avalanche multiplication. The current-controlled current source I_{AV} allows the model to represent avalanche breakdown, which can occur during ESD. The reverse bias on the junction is large when the residual recovery current flows and avalanche multiplication of that current is discernible [46]. There is a delay of a few nanoseconds before the ESD diode's low-level reverse leakage current gets amplified; time-delayed avalanche multiplication was reported previously for high-voltage diodes [59]. The delay is a function of the seed current and the junction reverse bias. The function S, described in [46], models the delay and its inclusion improves the model fitting.

The ON-resistance extracted from the high-current portion of the diode pulse I–V curve is an increasing function of the pulsewidth [15]. This effect is included in the model by using (1) to make R_0 and R_{m0} functions of temperature. The separate effects of R_0 and R_{m0} cannot be determined unless one has available test structures with varying anode to cathode spacing, i.e., different R_{m0}. If suitable test structures are unavailable, the temperature dependence is assigned wholly to R_0.

OTLP was used to inject decaying sinusoidal pulses into a 65-nm p-well diode. The measurement results are well replicated in simulations that use the NQS model of Table I, as shown in Fig. 9. However, if the NQS model is replaced by a charge control quasi-static model, the simulated voltage waveform no longer matches the measurement results, even though the quasi-static model includes a conductivity-modulated series resistance. This indicates that an NQS compact model is more suitable for the simulation of bipolarity large-signal fast transients.

The model equations indicate that the most effective way to limit transient voltage overshoot during a fast rise time ESD event is to minimize the unmodulated value of the diode resistance, e.g., by minimizing the spacing between the N^+ and P^+ diffusions.

2) N-Well Diode Model: Arguably, the n-well diode must be modeled as a three-terminal p-n-p even though, at ESD current levels, the common-emitter current gain is much less than 1, and the current flows primarily from the P^+ emitter to the N^+ base, as if the device were a two-terminal diode. The p-n-p representation is needed for proper modeling of the leakage current under normal operating conditions. Furthermore, if the on-chip protection includes a string of n-well diodes, the turn-on voltage of that string is reduced due to the transistor action [60].

Several prior works use the SPICE Gummel–Poon (SGP) model to represent the n-well ESD diode [3], [15], [61]. However, that model was developed for carefully designed short-base n-p-n transistors rather than a parasitic p-n-p. It has been shown that the SGP model cannot be well fit to the parasitic p-n-p I–V over a wide range of current levels [62]. An alternative model is presented here; its schematic representation, Fig. 10, is identical to that of the SGP model, but the model equations, Table II, are different. One significant difference is that an NQS formulation is used to represent the instantaneous value of the charge stored in the quasi-neutral base, Q_{DEB}.

Inspection of Table II reveals that the ideality factor of the junction current smoothly changes from 1 to 2 at high current levels, governed by (11), (17), and (18). According to the model, the n-well diode's anode and cathode (i.e., p-n-p emitter and base) currents have their ideality factors increased due to HLI, consistent with measurement data. The SGP model, instead, predicts that only the anode current is affected by HLI. The SGP model assumes that the base current supports minority carrier injection into the heavily

TABLE II
N-WELL DIODE MODEL. THE EQUATIONS FOR THE EMITTER-BASE JUNCTION ARE LISTED. THE MODEL INCLUDES ANALOGOUS EQUATIONS FOR THE COLLECTOR–BASE JUNCTION

$I_F = \theta_1 \boldsymbol{I_{sh}}\left[exp\left(\frac{V_{eb}}{2V_T}\right)-1\right] + \theta_2 \boldsymbol{I_{sl}}\left[exp\left(\frac{V_{eb}}{V_T}\right)-1\right]$	(11)
$Q_F = I_F(\tau_F + T_{MF})$	(12)
$0 = \frac{dQ_{DEB}}{dt} + \frac{Q_{DEB}}{\boldsymbol{\tau_F}} - i_{DF}$	(13)
$i_{DF} = \frac{Q_F - Q_{DEB}}{\boldsymbol{T_{MF}}}$	(14)
$i_{Link} = i_{DF} - i_{DR}$	(15)
$i_{DEB} = \frac{i_{DF}}{\boldsymbol{\beta_F}}$	(16)
$\theta_1 = \frac{1}{2}\left[1 + \tanh\left(\boldsymbol{K_1} \cdot (V_{eb} - \boldsymbol{V_1})\right)\right]$	(17)
$\theta_2 = \frac{1}{2}\left[1 + \tanh\left(\boldsymbol{K_2} \cdot (V_{eb} - \boldsymbol{V_2})\right)\right]$	(18)
$R_B = \boldsymbol{R_{B1}} + \frac{\boldsymbol{R_{B0}}}{1 + \frac{R_{B0}Q_{DEB}}{2T_{MF}V_T}}$	(19)

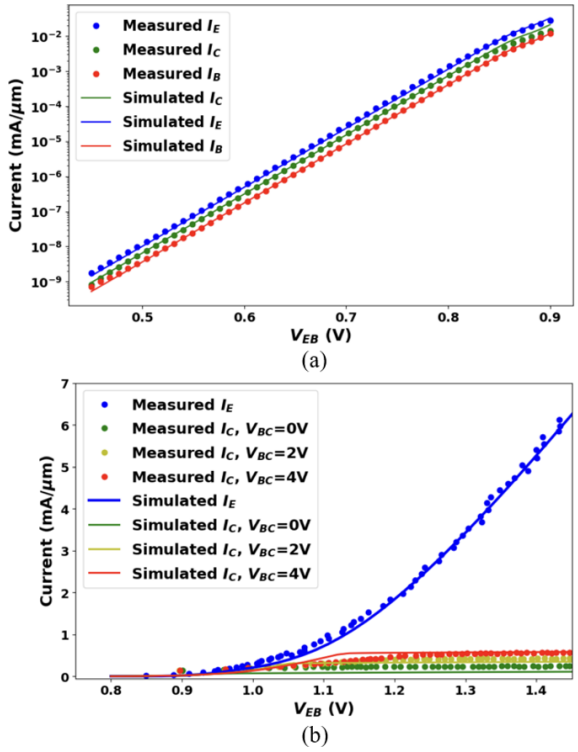

Fig. 11. Measured and simulated I–V of a 65-nm n-well diode (p-n-p). (a) DC I–V for $V_{BC} = 0$. (b) TLP I–V. The indicated values of V_{EB} and V_{BC} are the externally applied biases.

doped emitter. However, for the parasitic p-n-p of the n-well diode, the base current primarily supports recombination in the quasi-neutral base.

DC and pulse I–V characteristics of an n-well diode are shown in Fig. 11(a) and (b), respectively. Overall, the simulations and measurement data are in good agreement. However, in Fig. 11(b), there is a small range of current levels in which the simulated collector current is noticeably inaccurate.

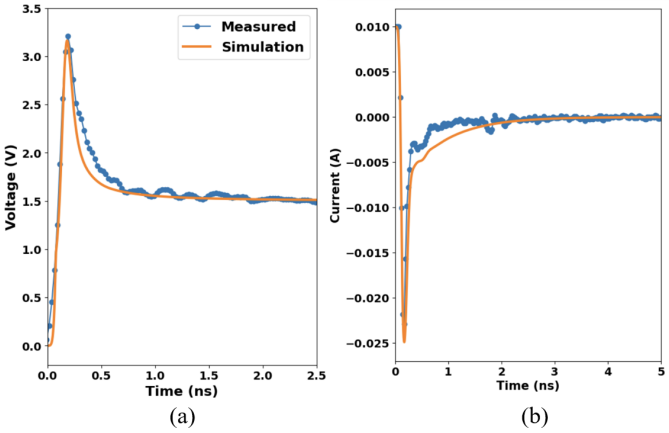

Fig. 12. Measured and simulated (a) forward recovery for a 1.25-A current injection and (b) reverse recovery of the n-well diode.

This results from having modeled only one component of the base current. The error occurs at current levels so high that the collector current has become small due to gain roll-off, and thus the error does not have a significant effect on an ESD simulation. Furthermore, the model accuracy improves as the bias is increased further, due to the onset of quasi-saturation. Unlike the base resistance, the collector resistance does not undergo conductivity modulation and the $I_C R_C$ voltage drop can raise the collector voltage above the base voltage, thereby forward biasing the collector–base junction. Once the collector–base junction becomes forward-biased, any increase in i_{DF} produces a proportionate increase in i_{DR}, causing the net collector current (i_{Link}) to become near constant; this is referred to as quasi-saturation.

Examples of an n-well diode's forward and reverse recovery transients are provided in Fig. 12(a) and (b); both measurement and simulation results are shown. Forward recovery and reverse recovery are simulated accurately due to the inclusion of a conductivity-modulated base resistance and the NQS model formulation, respectively. In contrast to the p-well diode, a residual reverse current does not flow through the n-well diode following the initial reverse recovery impulse, likely due to the n-well diode having a smaller volume extrinsic base.

During an ac simulation, the large-signal Verilog-A model is linearized by the simulator. In [62], the n-well diode model is used to simulate S11 (return loss) of a 65-nm n-well diode, and it is shown that the simulations match measurement results over the range 1–30 GHz for a variety of reverse biases. This demonstrates that the OFF-state loading provided by the ESD protection device is simulated accurately.

B. Silicon-Controlled Rectifier (SCR)

SCRs are mostly used as IO protection devices. The implementation of an SCR in a low-voltage CMOS technology is shown in Fig. 13. Due to its p-n-p-n structure, an SCR often is represented by cross-coupled p-n-p and n-p-n transistors. The dc impedance of an SCR is normally high; however, if the anode-to-cathode voltage is raised above the trigger voltage V_{t1}, the device transitions to a low-impedance ON-state.

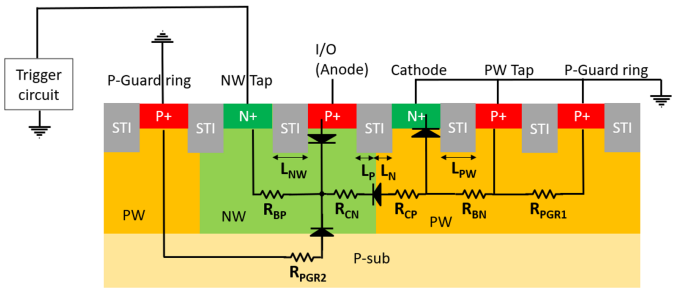

Fig. 13. Cross section of a CMOS SCR, configured as an n-well triggered device and placed between an IO pad and ground. Key device spacings are labeled. The anode to cathode spacing S_{AC} is given by the sum $L_N + L_P$.

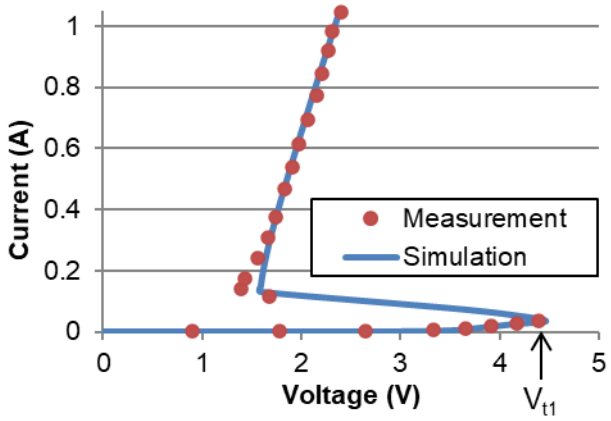

Fig. 14. 10-ns pulse I–V (t_{rise} = 0.3 ns) of a 65-nm DTSCR. The cathode, p-well tap, and P guard ring are grounded. The trigger circuit is a string of three n-well diodes connected from the n-well tap to ground. The trigger voltage V_{t1} is marked.

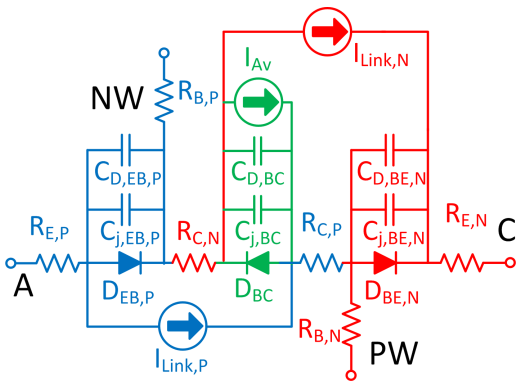

Fig. 15. Schematic view of SCR compact model. Components in red are associated with the n-p-n, those in blue with the p-n-p. Green is used for components that are common to (shared by) both transistors.

In low-voltage CMOS circuits, a trigger circuit is connected to either the SCR's n-well tap or its p-well tap to reduce its trigger voltage to a suitably low value. The pulse I–V of a diode-triggered SCR (DTSCR) is shown in Fig. 14.

The schematic representation of an SCR compact model [52] is provided in Fig. 15. It is noted that the collector resistance of the n-p-n sits between the p-n-p's base–emitter and base–collector junctions, and the p-n-p collector resistance sits between the junctions of the n-p-n. The placement of the base and collector resistors inside the schematic is very important. The correct placement of those resistors allows the model to be scalable, i.e., the model parameters are physically valid functions of the device layout spacings.

TABLE III
SELECTED MODEL EQUATIONS FOR AN N-WELL TRIGGERED SCR [52]. PARAMETERS ARE BOLDED. THE P-N-P IS MODELED BY ANALOGS OF (20)–(25)

$I_{Link,N} = \boldsymbol{I_{SN}}\left[exp\left(\frac{V_{be,N}}{V_T}\right) - exp\left(\frac{V_{bc,N}}{V_T}\right)\right]$	(20)
$I_{D,BE,N} = \frac{I_{SN}}{\beta_{FN}}\left[exp\left(\frac{V_{be,N}}{V_T}\right) - 1\right]$	(21)
$\beta_{FN} = \boldsymbol{\beta_{N0}} + \frac{I_{Link,P}}{\boldsymbol{I_{\beta P}}}\left(\frac{1}{1 + \frac{I_{Link,P}}{\boldsymbol{I_{\beta PSat}}}}\right)$	(22)
$Q_{D,BE,N} = I_{SN}\boldsymbol{\tau_{FN}}\left[exp\left(\frac{V_{be,N}}{V_T}\right) - 1\right]$	(23)
$R_{C,N} = \frac{\boldsymbol{R_{C,N0}}}{1 + \frac{Q_{D,BC}}{\boldsymbol{Q_{C,N0}}}}$	(24)
$R_{E,N} = \boldsymbol{R_{E,N0}}\left(1 + \frac{\Delta T}{T_0}\right)^{\chi}$	(25)
$I_{D,BC} = \boldsymbol{I_{SR}}\left[exp\left(\frac{V_{bc}}{V_T}\right) - 1\right]$; $Q_{D,BC} = \boldsymbol{\tau_R} \cdot I_{D,BC}$	(26)
$I_{Av} = \left(I_{SR} + I_{Link,N} + I_{Link,P}\right)\left(\frac{1}{1 - \left(\frac{V_{bc}}{\boldsymbol{BV}}\right)^{\boldsymbol{m_R}}} - 1\right)$	(27)
$R_{B,P} = \boldsymbol{R_{B,Pmin}} + \frac{\boldsymbol{R_{B,P0}} - R_{B,Pmin}}{1 + \frac{Q_{D,BE,P}}{\boldsymbol{Q_{B,P0}}}}$	(28)

The equations for the SCR compact model are listed in Table III. The individual bipolar transistors are represented by simple charge control models, such as would be derived from the Ebers–Moll model. Using a more advanced bipolar transistor model that includes gain roll-off would compromise the accuracy of the SCR model [63]. When the SCR is in its ON-state, the shared collector–base junction of the two transistors is forward biased and high current transistor effects, such as base pushout, do not occur.

When the SCR is latched into its ON-state, its three p-n junctions are forward-biased and the n-well and p-well will undergo conductivity modulation. The equations of Table III are specific to the case of an n-well triggered SCR, and Fig. 13 is used to ascertain which of its resistances undergo conductivity modulation. $R_{C,N}$ and $R_{C,P}$ lie along the main current path, where there are excess minority carriers; both of those resistances are conductivity modulated, as indicated in (24). $R_{B,N}$ is in parallel with the base–emitter junction of the n-p-n, whereas $R_{B,P}$ is in series with the base–emitter junction of the p-n-p. As a result, in [63], it was argued that only $R_{B,P}$, (28), undergoes conductivity modulation; forward recovery in $R_{B,P}$ is evident in an n-well triggered SCR's turn-on transient. $R_{B,N}$ cannot experience conductivity modulation until after the cathode/p-well junction becomes forward biased, at which point its effect on the device I–V would be small. Thus, a consideration of conductivity modulation in $R_{B,N}$ is mostly academic.

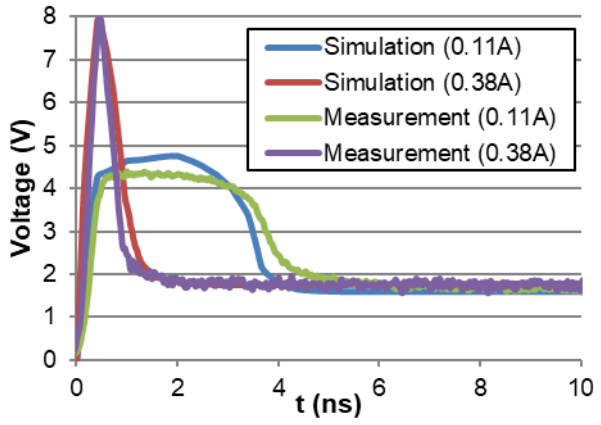

Fig. 16. Voltage response of a 65-nm DTSCR to injected current pulses (t_{rise} = 300 ps and t_{width} = 10 ns). The current levels of the two pulses are moderately low, 1.5 and 5 mA/μm, yet the overshoot is large, which is typical of SCR-based protection. The imperfect simulation of the smaller amplitude transient is attributed to the imprecise representation of the injected pulse, i.e., the tester model. Data from [52].

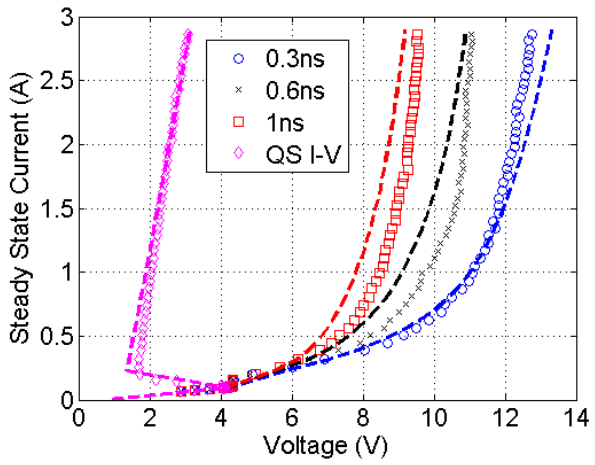

Fig. 17. 130-nm SCR; a resistor from the n-well tap to ground comprises the trigger circuit (RTSCR). Markers show VFTLP measurement data for risetimes ranging from 300 ps to 1 ns. Simulation results are plotted with dashed lines. The leftmost curve is the quasi-static pulse I–V. The other curves show the peak voltage. Data from [61].

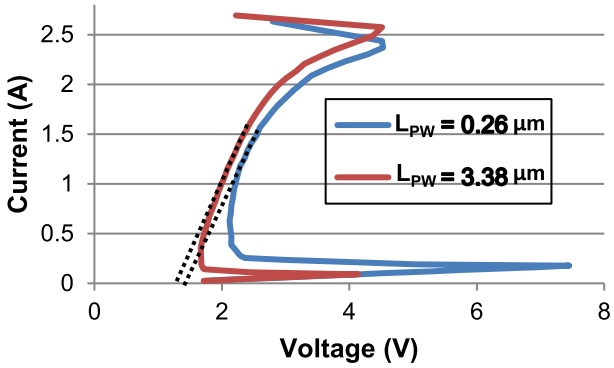

Fig. 18. Measured, 100-ns pulse I–V of 130-nm RTSCR devices; t_{rise} = 10 ns. L_N, L_P, and L_{NW} are fixed. L_{PW} has a large effect on V_{t1} and a small effect on V_H; the extrapolated V_H is at the x-intercepts of the dashed lines.

The values of β_N and β_P affect the SCR I–V only before it reaches the latched-on state. In (22), β_N is an increasing function of the p-n-p link current and that formulation has been shown to improve the agreement between measurement and simulation, relative to a model with constant β_N [52]. The gain of one bipolar transistor is affected by the current in the other because the base of one is coincident with the collector of the other. The majority carrier drift current in $R_{C,P}$ is supported by an electric field that aids minority carrier transport in the base of the n-p-n, thereby increasing β_N. The data plot in Fig. 14 confirms that the equations of Table III accurately model an SCRs pulsed I–V.

Fig. 16 demonstrates that the model of Table III well represents the transient response of an SCR to an injected current pulse. For an n-well triggered SCR, the good fit is due to having implemented $R_{B,P}$ as a conductivity-modulated resistor and having modeled the avalanche current using (27). When a current pulse is injected into the anode of an n-well triggered SCR, the voltage at the anode rises until the trigger circuit turns on and provides base current for the p-n-p. The voltage across the device then decays toward its static value over an interval termed the turn-on time. The turn-on time is determined by the intrinsic SCR and is an increasing function of anode to cathode spacing, S_{AC} [64]. However, the peak voltage is determined by devices along the current path through the trigger circuit [65]. The trigger circuit is connected in series with the anode to n-well junction diode, which undergoes forward recovery. In the case of a DTSCR, the diodes in the trigger circuit also experience forward recovery. As a result, the amplitude of the voltage overshoot increases dramatically with decreasing ESD pulse rise time. However, the overshoot is not unbounded; ultimately, it is limited by avalanche breakdown of the SCR's n-well to p-well junction.

The avalanche breakdown voltage depends on the magnitude of the "seed current" that undergoes impact ionization. In an n-well triggered SCR that has not yet entered the latched-on state, the current passing through the n-well/p-well junction consists of not just the junction leakage current but also the p-n-p collector current [61]. The p-n-p collector current takes time to build up, and thus the seed current available to initiate avalanche breakdown is an increasing function of the pulse rise time. The model predicts that the avalanche breakdown voltage of the n-well/p-well junction is a decreasing function of the injected pulse rise time, which is confirmed by the measurement data shown in Fig. 17.

The ON-state resistance of the SCR increases with pulsewidth; the effect is most noticeable at high current levels. The current density is highest in the P^+ anode and the N^+ cathode, and, accordingly, the expression for $R_{E,N}$, (25), includes self-heating, likewise the expression for $R_{E,P}$.

The SCR model of Table III is made scalable by expressing the model parameters (bolded) as functions of the device layout spacings. In [52], it is demonstrated that the parameters have the expected layout dependencies.

The measurements in [63]—an example is in Fig. 18—show that 1) the holding voltage (V_H) is a function of the anode to cathode spacing ($S_{AC} = L_N + L_P$), 2) the SCR ON-resistance is independent of S_{AC}, 3) the trigger voltage (V_{t1}) of the n-well triggered SCR is a decreasing function of p-well tap spacing

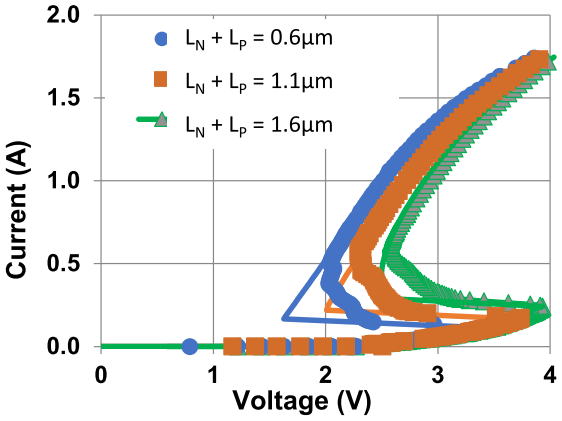

Fig. 19. 100-ns pulse I–V of three 130-nm DTSCRs with variable S_{AC} [50], [52]. Markers are measurement data; lines are the model. The scalable model fits the measured data except near the holding point, where the current is distributed nonuniformly across the SCR width.

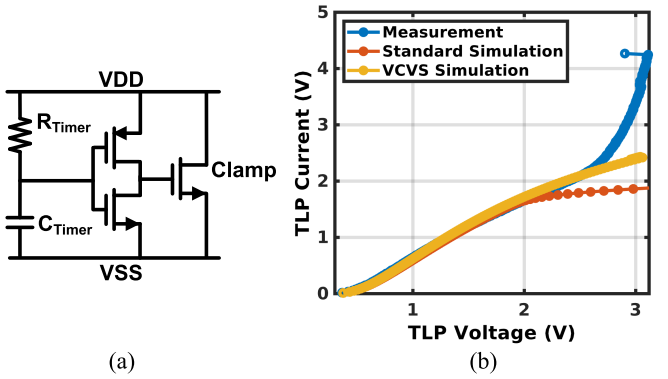

Fig. 20. (a) Schematic of an active rail clamp [66] in 65-nm CMOS. (b) 100-ns TLP I–V characterization is performed by grounding VSS and injecting pulses to VDD; the figure provides a comparison of measurement and simulations. In one simulation, a VCVS was inserted to mitigate the clamp's artificially high I_G. Alternatively, if the PDK model is BSIM4v5, the simulation can be improved by changing the option for IgbMod.

L_{PW}, and 4) the holding voltage is a weakly decreasing function of L_{PW}. Those observations are consistent with the model and are explained in order as follows. 1) In the scalable model, the anode to cathode spacing determines the zero-current collector resistances, $R_{C,N0}$ and $R_{C,P0}$. The resistors $R_{C,N}$ and $R_{C,P}$ are linearly proportional to their zero-bias values but inversely proportional to the collector current, and thus, the $I_C R_C$ voltage drops over the middle portion of the ON-state SCR are invariant with the current level. That constant voltage drop is a linear function of S_{AC} and constitutes most of the holding voltage. 2) The device ON-resistance comes from the terms $R_{E,N}$ and $R_{E,P}$, which are independent of S_{AC}. 3) A larger L_{PW} increases $R_{B,N}$, allowing the n-p-n base–emitter junction to be forward-biased at a lower current level, thereby reducing V_{t1}. 4) A larger L_{PW} steers a larger fraction of the ON-state current into the base–emitter junction of the n-p-n, thereby increasing conductivity modulation in the wells and slightly lowering the voltage drop across the device, which is V_H. The data plot in Fig. 19 confirms that the scalable model accurately predicts the measurement results, other than near the holding point. Near the holding point, only a portion of the device width is conducting, and that effect is not captured in a compact model, per the discussion in Section II.

From the model, it is concluded that to minimize the transient voltage overshoot for an n-well triggered SCR, one should minimize the spacing from the anode to the n-well tap (L_{NW}). To prevent V_{t1} from being significantly larger than the turn-on voltage of the trigger circuit, I_{t1} should be made small. That is achieved by using a nonminimum spacing from the cathode to the p-well tap (L_{PW}). A small anode to cathode spacing minimizes the turn-on time, further reducing the stress applied to the circuit being protected.

The transient model presented in Table III is quasi-static. OTLP measurements of SCR devices, found in [49], suggest that an NQS model may be needed for accurate simulation of large-amplitude bipolarity ESD pulses. The NQS model formulation used for ESD diodes may be a starting point for the development of an NQS SCR model.

C. N-Channel MOSFET

It is tempting to use the PDK MOSFET model to simulate any MOS transistor that is not a dedicated protection device. Sometimes, the PDK transistor model is even used to represent transistors in the latter category, e.g., the MOSFETs comprising an active rail clamp circuit, shown in Fig. 20(a). However, the PDK MOSFET model underpredicts the rail clamp's conductivity when the induced V_{DD} exceeds the transistor's intended operating voltage by more than a few 100 mV, as shown in Fig. 20(b). That occurs because the PDK model greatly overestimates the gate current of the clamp device, which, in turn, increases the V_{SD} of the pMOS and reduces the overdrive voltage applied to the clamp. That error can be mitigated by inserting a voltage-controlled voltage source (VCVS) to decouple the pMOS V_{SD} from the gate current of the clamp device; as shown in Fig. 20(b), the resultant simulations are accurate to about 2.5 V. Ultimately, the measured I–V curve diverges from the simulated one because the parasitic n-p-n in parallel with the clamp turns on and the device breaks down, an effect that is not captured in the PDK MOSFET model. It may be concluded that when simulating ESD, one must use a MOSFET model that correctly represents drain breakdown.

Efforts to analyze and model MOSFET drain breakdown and the consequent snapback date back at least 40 years (e.g., [67]). Those efforts facilitated the development of models that provide accurate results when an ESD event is simulated; examples include [2], [8], [68], and [69]. The essential elements of an ESD MOSFET model are the channel current, a lateral n-p-n, the series resistance of the drain and source regions, a current-controlled current source to model avalanche multiplication, and a resistor to model the p-well resistance between the base of the n-p-n and the body contact. A representative implementation is shown in Fig. 21 [8], [69], [70]. In the figure, the MOSFET is represented by a PDK model; however, it is not strictly necessary to make use of the PDK MOSFET model. A simple model of the MOSFET channel current is adequate for ESD simulations [69].

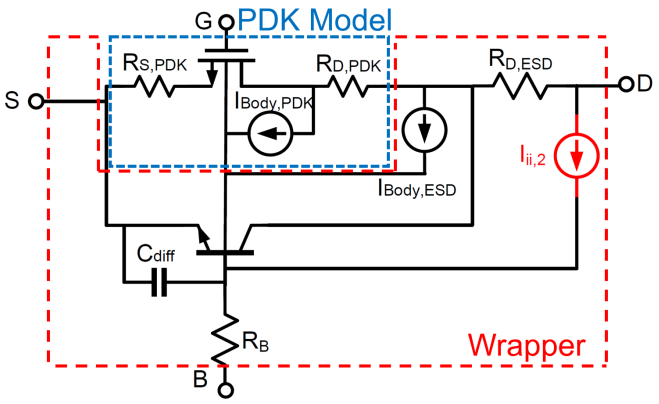

Fig. 21. Schematic representation of an ESD MOSFET model. In this implementation, the PDK model is used to represent the channel current, while the "wrapper" models the ESD response. $R_{D,ESD}$ will be sizable if the transistor includes a drain ballasting resistance. $I_{ii,2}$ is drawn in red to emphasize that it is not included in prior works but its inclusion will be advocated for herein.

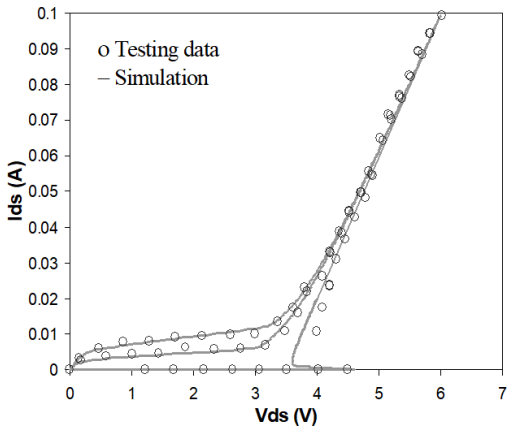

Fig. 22. 100-ns pulse I–V of a 130-nm NMOSFET for V_{GS} of 0, 0.8, and 1 V [69].

Many ESD MOSFET models, e.g., [8], [68], [69], employ a simple model of the parasitic n-p-n, which does not include HLI effects, yet those models well match the measured I–V, as demonstrated in Fig. 22. That outcome can be understood by performing circuit analysis of the schematic in Fig. 21, setting all the source and drain resistances to zero for convenience. Following the approach in [71], an Ebers–Moll model of the n-p-n is adopted, Kirchhoff's current law is applied at each node in the circuit, the Miller model is used to model avalanche multiplication, and the trigger voltage is obtained by setting $(\partial I_D/\partial V_{DS})|_{V_{t1}} \to \infty$. The analysis yields

$$V_{t1} = \left[1 - \frac{\alpha_F}{1 + \frac{V_T}{R_B I_e}}\right]^{\frac{1}{n}} \cdot \mathrm{BV_{CBO}}, \tag{29}$$

where $\mathrm{BV_{CBO}}$ is the drain junction breakdown voltage measured with the source terminal open-circuited, I_e is the n-p-n's emitter current, α_F is the common-base current gain of the n-p-n, and n is a parameter of the Miller model, typically about 4. Measurement data, e.g., Fig. 22, show that the breakdown voltage of the n-p-n is modulated by V_{GS}, and the following analysis of (29) confirms that the V_{GS} dependency is captured by the model of Fig. 21. Furthermore, (29) predicts that V_{t1} is a decreasing function of R_B, consistent with measurement results [38].

In (29), the V_{GS} dependency resides in the emitter current, which is given by

$$I_e = M \cdot I_c + (M - 1) \cdot I_{DS}, \tag{30}$$

where M denotes the multiplication factor. In (30), I_{DS}—the MOS channel current—is a strongly increasing function of V_{GS} when the MOSFET is biased in subthreshold, and thus (29) correctly predicts that V_{t1} is lowered when V_{GS} is raised above zero. In Fig. 22, snapback is observed only for $V_{GS} < V_{th}$, where V_{th} is the MOSFET threshold voltage, and this result is expected given (29) and (30). The n-p-n is in cutoff at bias points below V_{t1}; the n-p-n turns on and I_c becomes sizable at V_{t1}. The sudden increase in I_c and thus I_e reduces the right-hand side of (29), causing V_{DS} to snapback from the value V_{t1} to a lower V_H. However, snapback does not occur if I_{DS} is sufficiently large to cause the right-hand side of (29) to be near its minimum value even when I_c is small; this occurs when the MOSFET is biased above threshold.

The analysis that produced (29) suggests that above the trigger point, the current in the device is limited primarily by the series resistance, such that $I_D = (V_{DS} - V_H)/(R_S + R_D)$, and V_{be} will hardly change. Thus, the n-p-n model needs to be valid up to the point at which the device breaks down, and a simple model appears to suffice.

As indicated in Fig. 21, the outer portion of an ESD MOSFET model built around a PDK model is called the "wrapper" (or "shell" [70]). The advanced MOSFET model in the PDK has many appealing features: it is scalable, it represents the many parasitic capacitances of a modern device, and it is sufficiently accurate for analog circuit design. Series resistances are included in the PDK model, but some authors choose to augment or replace those with resistors in the wrapper, e.g., $R_{D,ESD}$ [72]. Although the PDK model includes the body current, it is customary for the ESD MOSFET model to override it with a model for avalanche multiplication that has been optimized for ESD conditions [8], [68], [69], [72], shown as $I_{BODY,ESD}$ in Fig. 21. Prior works did not investigate whether the inclusion of the ESD wrapper compromises the PDK model accuracy at normal bias conditions; in fact, Li et al. [70] proposed to activate the wrapper only for ESD simulations.

Meng and Rosenbaum [73] demonstrated that a MOSFET model composed of the PDK model and an ESD wrapper can perform very well in diverse ESD simulations. The model replicates the "snap-up" that occurs if a significant amount of charge remains on the capacitances inside an ESD tester when the protection device turns off; this is shown in Fig. 23 [73]. The model was used to study a related phenomenon—the retriggering and oscillations that may occur during the tail end of an HBM test and that can damage the DUT. Circuit simulations revealed that oscillation is most likely to occur if the protection device has a high holding current [73].

We find that constructing a wrapper model according to the procedures outlined in [72] can yield a good representation of the measured TLP and VFTLP I–V curves. However, when the device model is used to simulate the behavior of the

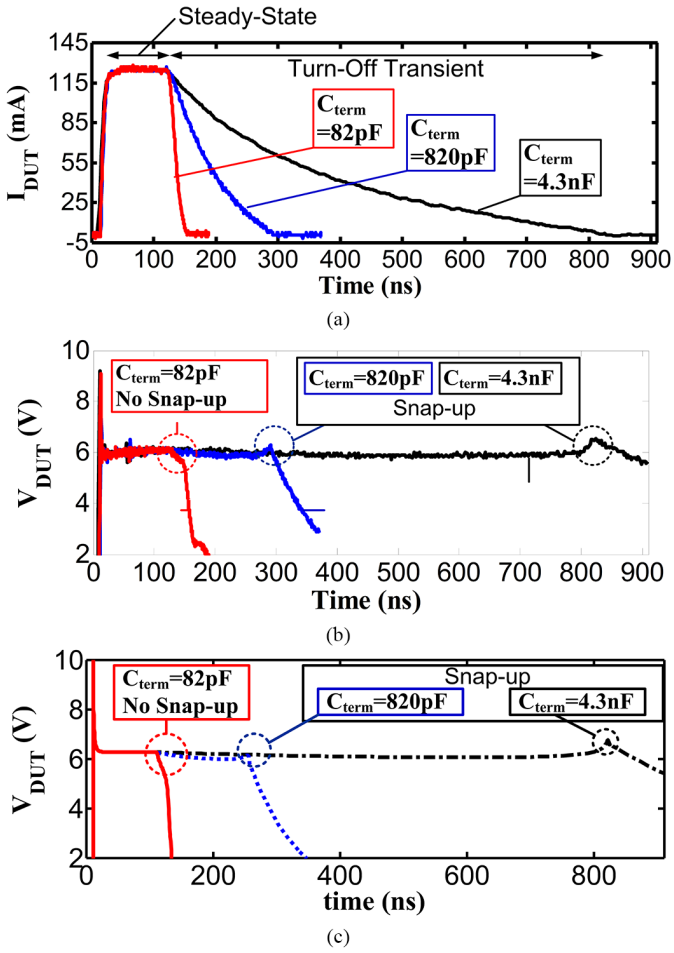

Fig. 23. Charge cable in a TLP tester may be terminated by an *RC* load to obtain a slow falling edge transient [74]; here, the tester was configured with three different capacitive terminations. The DUT is a grounded-gate 5-V nMOS (GGNMOS) in a 90-nm process. (a) Measured current. (b) Measured voltage (nMOS V_{DS}). (c) Simulated voltage. The simulated trigger voltage is about 10 V, which is slightly higher than in the transient measurement but consistent with the V_{t1} of the pulse *I*–*V*. The oscilloscope sampling rate is not high enough to capture the maximum V_{DUT} during the transient measurement. Overall, the measurement and simulation results are very consistent.

nMOS under normal operating conditions, the accuracy of the simulation is compromised. The error is largest when the transistor is biased in the triode region. This is unsurprising given that the series resistances have their largest effect in the triode region, but those resistor values were optimized for high-current ESD conditions.

Simulation-based co-design of ESD protection and IO circuits requires a transistor model that is accurate for both normal and ESD-like bias conditions. Therefore, the specific wrapper modeling approach illustrated in Fig. 21 is proposed. The PDK model is kept intact. As indicated in the figure, the PDK model will represent the channel current, the source and drain series resistances, and the body current (i.e., impact ionization current and gate-induced drain leakage current) under normal operating conditions. The $I_{Body,ESD}$ current source is used to compensate for any inaccuracy of the PDK body current model, $I_{Body,PDK}$, that arises when a higher-than-normal drain voltage is applied to the device. If the transistor has been made ESD-robust by the inclusion of a silicide-blocked drain resistor, a significantly larger reverse bias may be applied to the drain junction near the contact than near the gate edge, and $I_{ii,2}$ can be included to model the avalanche multiplication of the junction leakage. The devices inside the wrapper should be represented by functions that have very small values under normal operating conditions and that become increasingly significant as V_{DS} goes above the nominal supply voltage.

D. ESD Failure Prediction

Compact models, such as those described in Sections III-A to III-C, allow one to simulate the node voltages and branch currents that arise when an ESD stimulus is applied to one of the circuit ports. A prediction as to whether the chip is ESD robust requires one to determine whether any of those currents or voltages are large enough to damage the circuit.

Most ESD failures are attributed to thermal failure or dielectric breakdown in a semiconductor device. Thermal failure is associated with second breakdown, which occurs when a critical temperature, T_{crit}, is reached inside the device [75].

One approach to failure detection is to have the simulator activate a flag if the voltage across a MOS transistor gate oxide exceeds the specified breakdown voltage or if the current in any device exceeds its specified failure current. Breakdown voltage and failure current values are obtained from pulse measurements of suitable duration, e.g., 1 ns for CDM simulations and 100 ns for HBM.

Alternatively, one may implement a time-dependent model for each breakdown mechanism. Models of time-dependent dielectric breakdown allow a failure prediction to be made for arbitrary pulse durations [76], [77] and may even be formulated to comprehend the time-varying nature of the ESD-induced stress [78], [79]. A dynamic model of self-heating may be used to check whether the simulated temperature exceeds T_{crit} [36]. Alternatively, the device model may be constructed, such that the thermal runaway process will be captured in circuit simulation, e.g., by including the temperature dependencies of the carrier concentrations, impact ionization rate, and so on, in the model equations [80]; the self-heating model updates the device temperature at each time step in the simulation.

IV. Discussion and Conclusion

In Section III, it was established that the ESD device research community has cultivated a sufficient understanding of device physics to develop compact models of ESD protection devices. Measurement and parameter extraction challenges, described in Section II, remain, but those likely can be overcome by allocating more resources to the development and modeling of testers.

The CDM test standard has the most real-world relevance of all the component-level ESD test standards [25]. Compact models suitable for transient simulation are necessary but not sufficient for successful CDM ESD simulations. Forward recovery transients compromise the ability of a protection device to protect an active circuit. The amplitude of the

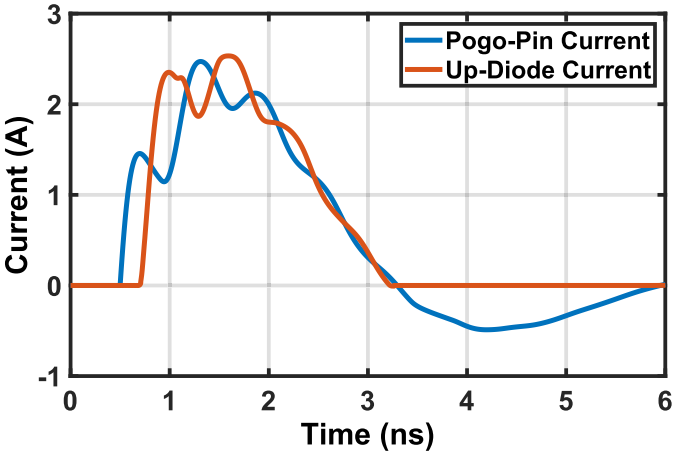

Fig. 24. Simulated current at the pogo pin of a CDM tester and at an on-chip receiver; for the latter case, the current in the ESD diode from the input pad to VDD is plotted. The chip is in a BGA package; the package model was extracted from electromagnetic simulation. The slew rate of the current injected into the protection devices is extremely high.

transient is a function of the injected pulse's di/dt. The characteristics of the injected CDM pulse are determined not only by the tester and the on-chip protection but also the IC package [81]. The pulse delivered from an FICDM tester to an IC component is a distorted version of the pulse measured using a calibration coin; furthermore, analysis indicates that the pulse that appears on-die, which cannot be measured directly, is different from the one measured at the tester [82], [83]. The pulse shape is affected by the delay along the package signal trace. An example simulation is shown in Fig. 24 and high slew rate transients are evident. It is concluded that the circuit model used for CDM simulation must contain extracted RF models of the package and the on-chip power/ground buses, accurate compact models of the protection devices, and a model of the tester.

Finally, it must be mentioned that transient simulation using physics-based nonlinear compact models is appropriate for IO circuit optimization, but a different or hybrid approach is required to validate the design of a full-chip protection network; see, for example, [5], [7], [84], [85]. Transient simulation with nonlinear models is too computationally expensive for full-chip verification because the netlist is very large [86]. The chip model must include all the pad cells and power buses; furthermore, the simulator must check the induced voltages at power domain crossing circuits, as those are known to be ESD failure sites [87].

Acknowledgment

Dr. Robert Mertens, Dr. Junjun Li, Dr. Kuo-Hsuan Meng, and Dr. Sopan Joshi are thanked for their contributions to this body of work.

References

[1] C. H. Diaz, S.-M. Kang, and C. Duvvury, "Circuit-level electrothermal simulation of electrical overstress failures in advanced MOS I/O protection devices," *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst.*, vol. 13, no. 4, pp. 482–493, Apr. 1994, doi: 10.1109/43.275358.

[2] S. G. Beebe, "Simulation of complete CMOS I/O circuit response to CDM stress," in *Proc. Electr. Overstress/Electrostatic Discharge Symp.*, Reno, NV, USA, 1998, pp. 259–270, doi: 10.1109/EOSESD.1998.737046.

[3] C. A. Torres, J. W. Miller, M. Stockinger, M. D. Akers, M. G. Khazhinsky, and J. C. Weldon, "Modular, portable, and easily simulated ESD protection networks for advanced CMOS technologies," in *Proc. Elect. Overstress/Electrostatic Discharge Symp.*, Portland, OR, USA, 2001, pp. 81–94.

[4] M. Etherton, J. Willemen, W. Wilkening, N. Qu, S. Mettler, and W. Fichtner, "Verification of CDM circuit simulation using an ESD evaluation circuit," in *Proc. Elect. Overstress/Electrostatic Discharge Symp.*, Anaheim, CA, USA, 2005, pp. 1–10.

[5] J.-J. Hajjar, T. Weyl, Y. Zhou, and S. Parthasarathy, "CDM event simulation in SPICE: A holistic approach," in *Proc. EOS/ESD Symp.*, Anaheim, CA, USA, 2011, pp. 1–8.

[6] D. Abessolo-Bidzo, T. Smedes, and P. C. de Jong, "A study of the effect of remote CDM clamps in integrated circuits," in *Proc. 37th Electr. Overstress/Electrostatic Discharge Symp. (EOS/ESD)*, Reno, NV, USA, Sep. 2015, pp. 1–8, doi: 10.1109/EOSESD.2015.7314787.

[7] K.-H. Meng, V. Shukla, and E. Rosenbaum, "Full-component modeling and simulation of charged device model ESD," *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst.*, vol. 35, no. 7, pp. 1105–1113, Jul. 2016, doi: 10.1109/TCAD.2015.2495196.

[8] 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. Int. Rel. Phys. Symp. (RELPHY)*, Dallas, TX, USA, 1996, pp. 318–326, doi: 10.1109/RELPHY.1996.492137.

[9] C. Russ, K. Verhaege, K. Bock, P. J. Roussel, G. Groeseneken, and H. E. Maes, "A compact model for the grounded-gate nMOS behaviour under CDM ESD stress," in *Proc. Electr. Overstress/Electrostatic Discharge Symp.*, Orlando, FL, USA, 1996, pp. 302–315, doi: 10.1109/EOSESD.1996.865157.

[10] V. Vassilev et al., "Advanced modelling and parameter extraction of the MOSFET ESD breakdown triggering in the 90 nm CMOS node technologies," in *Proc. Electr. Overstress/Electrostatic Discharge Symp.*, Grapevine, TX, USA, Sep. 2004, pp. 1–9, doi: 10.1109/EOSESD.2004.5272628.

[11] S. Joshi and E. Rosenbaum, "Compact modeling of vertical ESD protection NPN transistors for RF circuits," in *Proc. Elect. Overstress/Electrostatic Discharge Symp.*, Charlotte, NC, USA, 2002, pp. 292–298.

[12] J.-R. Manouvrier et al., "A physics-based compact model for ESD protection diodes under very fast transients," in *Proc. 30th Elect. Overstress/Electrostatic Discharge Symp.*, Tucson, AZ, USA, 2008, pp. 67–75.

[13] A. Romanescu et al., "A novel physical model for the SCR ESD protection device," in *Proc. Elect. Overstress/Electrostatic Discharge Symp.*, Reno, NV, USA, 2010, pp. 1–10.

[14] H. Li, M. Miao, Y. Zhou, J. A. Salcedo, J.-J. Hajjar, and K. B. Sundaram, "Modeling and simulation of comprehensive diode behavior under electrostatic discharge stresses," *IEEE Trans. Device Mater. Rel.*, vol. 19, no. 1, pp. 90–96, Mar. 2019, doi: 10.1109/TDMR.2018.2882454.

[15] S. Khandelwal and D. Bavi, "ASM-ESD—A comprehensive physics-based compact model for ESD diodes," in *Proc. IEEE Int. Rel. Phys. Symp. (IRPS)*, Dallas, TX, USA, Mar. 2022, pp. 5C.1-1–5C.1-6, doi: 10.1109/IRPS48227.2022.9764453.

[16] Y. Zhou, J. A. Salcedo, and J.-J. Hajjar, "Modeling of high voltage devices for ESD event simulation in SPICE," *Microelectron. J.*, vol. 43, no. 5, pp. 305–311, May 2012, doi: 10.1016/j.mejo.2011.12.010.

[17] Y. S. Chauhan et al., "BSIM—Industry standard compact MOSFET models," in *Proc. ESSCIRC*, Bordeaux, France, 2012, pp. 30–33, doi: 10.1109/ESSCIRC.2012.6341249.

[18] C. C. McAndrew et al., "VBIC95, the vertical bipolar inter-company model," *IEEE J. Solid-State Circuits*, vol. 31, no. 10, pp. 1476–1483, 1996, doi: 10.1109/4.540058.

[19] J. Karp, M. J. Hart, M. Fakhruddin, V. Kireev, L. Horwitz, and M. Hogan, "FinFET MPSoC 32 Gb/s transceivers: Custom ESD protection and verification," in *Proc. IEEE 59th Int. Midwest Symp. Circuits Syst. (MWSCAS)*, Abu Dhabi, UAE, Oct. 2016, pp. 1–4, doi: 10.1109/MWSCAS.2016.7869952.

[20] J. W. Miller, M. Etherton, M. G. Khazhinsky, M. Stockinger, and J. C. Weldon, "Comprehensive ESD protection for flip-chip products in a dual gate oxide 65 nm CMOS technology," in *Proc. Elect. Overstress/Electrostatic Discharge Symp.*, Tucson, AZ, USA, 2006, pp. 186–195.

[21] J. Kim et al., "A 112 Gb/s PAM-4 56 Gb/s NRZ reconfigurable transmitter with three-tap FFE in 10-nm FinFET," *IEEE J. Solid-State Circuits*, vol. 54, no. 1, pp. 29–42, Jan. 2019, doi: 10.1109/JSSC.2018.2874040.

[22] E. Groen et al., "A 10-to-112 Gb/s DSP-DAC-based transmitter with 1.2 Vppd output swing in 7nm FinFET," in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, San Francisco, CA, USA, Feb. 2020, pp. 120–122, doi: 10.1109/ISSCC19947.2020.9063130.

[23] J. Im et al., "A 112-Gb/s PAM-4 long-reach wireline transceiver using a 36-way time-interleaved SAR ADC and inverter-based RX analog front-end in 7-nm FinFET," *IEEE J. Solid-State Circuits*, vol. 56, no. 1, pp. 7–18, Jan. 2021, doi: 10.1109/JSSC.2020.3024261.

[24] A. Khairi et al., "A 1.41-pJ/b 224-Gb/s PAM4 6-bit ADC-based SerDes receiver with hybrid AFE capable of supporting long reach channels," *IEEE J. Solid-State Circuits*, vol. 58, no. 1, pp. 8–18, Jan. 2023, doi: 10.1109/JSSC.2022.3211475.

[25] *Recommended ESD-CDM Target Levels*, JEDEC Publication 157A, Joint Electron Device Eng. Council, Arlington, VA, USA, 2022.

[26] V. M. Dwyer, A. J. Franklin, and D. S. Campbell, "Thermal failure in semiconductor devices," *Solid-State Electron.*, vol. 33, no. 5, pp. 553–560, 1990, doi: 10.1016/0038-1101(90)90239-B.

[27] G. Boselli, A. Salman, J. Brodsky, and H. Kunz, "The relevance of long-duration TLP stress on system level ESD design," in *Proc. Elect. Overstress/Electrostatic Discharge Symp.*, Reno, NV, USA, 2010, pp. 1–10.

[28] T. Maloney and N. Khurana, "Transmission line pulse technique for circuit modeling and ESD phenomena," in *Proc. EOS/ESD Symp.*, 1985, pp. 49–54.

[29] M. Scholz et al., "Calibrated wafer-level HBM measurements for quasi-static and transient device analysis," in *Proc. 29th Electr. Overstress/Electrostatic Discharge Symp. (EOS/ESD)*, Anaheim, CA, USA, Sep. 2007, pp. 2A.2-1–2A.2-6, doi: 10.1109/EOSESD.2007.4401736.

[30] M. Meixner, P. Rodin, E. Schöll, and A. Wacker, "Lateral current density fronts in globally coupled bistable semiconductors with S- or Z-shaped current voltage characteristics," *Eur. Phys. J. B*, vol. 13, no. 1, pp. 157–168, Jan. 2000, doi: 10.1007/s100510050019.

[31] M. Litzenberger, R. Pichler, D. Pogany, E. Gornik, K. Esmark, and H. Gossner, "Influence of layout parameters on triggering behavior in 0.35 μm and 0.18 μm process gg-nMOS ESD protection devices," in *Proc. 31st Eur. Solid-State Device Res. Conf.*, Nuremberg, Germany, 2001, pp. 335–338, doi: 10.1109/ESSDERC.2001.195269.

[32] D. Pogany et al., "Nonlinear dynamics approach in modeling of the on-state-spreading–related voltage and current transients in 90 nm CMOS silicon controlled rectifiers," in *IEDM Tech. Dig.*, Baltimore, MD, USA, Dec. 2009, pp. 1–4, doi: 10.1109/IEDM.2009.5424308.

[33] K.-H. Meng and E. Rosenbaum, "Layout-aware, distributed, compact model for multi-finger MOSFETs operating under ESD conditions," in *Proc. 35th Elect. Overstress/Electrostatic Discharge Sympo*, Las Vegas, NV, USA, 2013, pp. 1–8.

[34] M. Shrivastava, H. Gossner, M. S. Baghini, and V. R. Rao, "Part II: On the three-dimensional filamentation and failure modeling of STI type DeNMOS device under various ESD conditions," *IEEE Trans. Electron Devices*, vol. 57, no. 9, pp. 2243–2250, Sep. 2010, doi: 10.1109/TED.2010.2055278.

[35] J. Li, "Characterization, design and modeling of on-chip electrostatic discharge protection devices," Ph.D. dissertation, Dept. Elect. Comput. Eng., Univ. Illinois Urbana-Champaign, Urbana, IL, USA, 2004.

[36] Y. Zhou, M. Miao, J. A. Salcedo, J.-J. Hajjar, and J. J. Liou, "Compact thermal failure model for devices subject to electrostatic discharge stresses," *IEEE Trans. Electron Devices*, vol. 62, no. 12, pp. 4128–4134, Dec. 2015, doi: 10.1109/TED.2015.2491223.

[37] Z. Chen, R. Mertens, C. Reiman, and E. Rosenbaum, "Improved GGSCR layout for overshoot reduction," in *Proc. IEEE Int. Rel. Phys. Symp.*, Monterey, CA, USA, Apr. 2015, pp. 3F.2.1–3F.2.8, doi: 10.1109/IRPS.2015.7112720.

[38] J. Li, H. Li, R. Barnes, and E. Rosenbaum, "Comprehensive study of drain breakdown in MOSFETs," *IEEE Trans. Electron Devices*, vol. 52, no. 6, pp. 1180–1186, Jun. 2005, doi: 10.1109/TED.2005.848858.

[39] J. Wu, P. Juliano, and E. Rosenbaum, "Breakdown and latent damage of ultra-thin gate oxides under ESD stress conditions," in *Proc. Electr. Overstress/Electrostatic Discharge Symp.*, Anaheim, CA, USA, 2000, pp. 287–295, doi: 10.1109/EOSESD.2000.890088.

[40] M. P. J. Mergens et al., "Speed optimized diode-triggered SCR (DTSCR) for RF ESD protection of ultra-sensitive IC nodes in advanced technologies," *IEEE Trans. Device Mater. Rel.*, vol. 5, no. 3, pp. 532–542, Sep. 2005, doi: 10.1109/TDMR.2005.853510.

[41] S. Thijs, A. Griffoni, D. Linten, S.-H. Chen, T. Hoffmann, and G. Groeseneken, "On gated diodes for ESD protection in bulk FinFET CMOS technology," in *Proc. EOS/ESD Symp.*, Anaheim, CA, USA, 2011, pp. 1–8.

[42] H. Gieser and M. Haunschild, "Very fast transmission line pulsing of integrated structures and the charged device model," *IEEE Trans. Compon., Packag., Manuf. Technol., C*, vol. 21, no. 4, pp. 278–285, Oct. 1998, doi: 10.1109/3476.739177.

[43] W. Simburger, D. Johnsson, and M. Stecher, "High current TLP characterisation: An effective tool for the development of semiconductor devices and ESD protection solutions," in *Proc. ARMMS RF Microw. Soc. Conf.* Wyboston, U.K.: Wyboston Lakes, 2013. [Online]. Available: https://www.armms.org/media/uploads/03_armms_nov12_wsimbuerger.pdf

[44] A. Ayling, S. Huang, and E. Rosenbaum, "Sub-nanosecond reverse recovery measurement for ESD devices," in *Proc. IEEE Int. Rel. Phys. Symp. (IRPS)*, Dallas, TX, USA, Apr. 2020, pp. 1–8, doi: 10.1109/IRPS45951.2020.9129596.

[45] J.-R. Manouvrier, P. Fonteneau, C.-A. Legrand, P. Nouet, and F. Azais, "Characterization of the transient behavior of gated/STI diodes and their associated BJT in the CDM time domain," in *Proc. 29th Electr. Overstress/Electrostatic Discharge Symp. (EOS/ESD)*, Anaheim, CA, USA, Sep. 2007, pp. 3A.2-1–3A.2-10, doi: 10.1109/EOSESD.2007.4401748.

[46] S. Huang and E. Rosenbaum, "Compact model of ESD diode suitable for subnanosecond switching transients," in *Proc. IEEE Int. Rel. Phys. Symp. (IRPS)*, Monterey, CA, USA, Mar. 2021, pp. 1–7, doi: 10.1109/IRPS46558.2021.9405149.

[47] H. Wolf, H. Gieser, W. Stadler, and W. Wilkening, "Capacitively coupled transmission line pulsing CC-TLP—A traceable and reproducible stress method in the CDM-domain," in *Proc. Elect. Overstress/Electrostatic Discharge Symp.*, Las Vegas, NV, USA, 2003, pp. 1–8.

[48] B. C. Chou, T. J. Maloney, and T. W. Chen, "Wafer-level charged device model testing," in *Proc. 30th Elect. Overstress/Electrostatic Discharge Symp.*, Tucson, AZ, USA, 2008, pp. 115–124.

[49] J. P. Di Sarro and E. Rosenbaum, "Oscillatory transmission line pulsing for characterization of device transient response," *IEEE Electron Device Lett.*, vol. 30, no. 2, pp. 168–170, Feb. 2009, doi: 10.1109/LED.2008.2009361.

[50] J. P. Di Sarro and E. Rosenbaum, "A scalable SCR compact model for ESD circuit simulation," *IEEE Trans. Electron Devices*, vol. 57, no. 12, pp. 3275–3286, Dec. 2010, doi: 10.1109/TED.2010.2081674.

[51] *Cadence*. Accessed: Sep. 9, 2023. [Online]. Available: https://www.cadence.com/en_US/home/tools/custom-ic-analog-RF-design/circuit-simulation.html

[52] R. Mertens and E. Rosenbaum, "A physics-based compact model for SCR devices used in ESD protection circuits," in *Proc. IEEE Int. Rel. Phys. Symp. (IRPS)*, Monterey, CA, USA, Apr. 2013, pp. 2B.2.1–2B.2.7, doi: 10.1109/IRPS.2013.6531947.

[53] C. L. Ma and P. O. Lauritzen, "A simple power diode model with forward and reverse recovery," *IEEE Trans. Power Electron.*, vol. 8, no. 4, pp. 342–346, Oct. 1993, doi: 10.1109/63.261002.

[54] K. J. Tseng and S. Pan, "Modified charge-control equation for simulation of diode reverse recovery," *Electron. Lett.*, vol. 32, no. 4, pp. 404–405, 1996, doi: 10.1049/el:19960210.

[55] M. M. Jovanovic, "A technique for reducing rectifier reverse-recovery-related losses in high-power boost converters," *IEEE Trans. Power Electron.*, vol. 13, no. 5, pp. 932–941, Sep. 1998, doi: 10.1109/63.712314.

[56] J. Whitfield et al., "ESD MM failures resulting from transient reverse currents," in *Proc. IEEE Int. Rel. Phys. Symp.*, San Jose, CA, USA, Mar. 2006, pp. 136–139, doi: 10.1109/RELPHY.2006.251205.

[57] J. Willemen et al., "Characterization and modeling of transient device behavior under CDM ESD stress," in *Proc. Elect. Overstress/Electrostatic Discharge Symp.*, Las Vegas, NV, USA, 2003, pp. 1–10.

[58] W. J. Kloosterman and H. C. De Graaff, "Avalanche multiplication in a compact bipolar transistor model for circuit simulation," *IEEE Trans. Electron Devices*, vol. 36, no. 7, pp. 1376–1380, Jul. 1989, doi: 10.1109/16.30944.

[59] D. Johnsson, D. Pogany, J. Willemen, E. Gornik, and M. Stecher, "Avalanche breakdown delay in ESD protection diodes," *IEEE Trans. Electron Devices*, vol. 57, no. 10, pp. 2470–2476, Oct. 2010, doi: 10.1109/TED.2010.2058790.

[60] S. Dabral, R. Aslen, and T. Maloney, "Designing on-chip power supply coupling diodes for ESD protection and noise immunity," *J. Electrostatics*, vol. 33, no. 3, pp. 357–370, 1994, doi: 10.1016/0304-3886(94)90039-6.

[61] R. Mertens and E. Rosenbaum, "Separating SCR and trigger circuit related overshoot in SCR-based ESD protection circuits," in *Proc. 35th Elect. Overstress/Electrostatic Discharge Symp.*, Las Vegas, NV, USA, 2013, pp. 1–8.

[62] S. Huang and E. Rosenbaum, "Physics-based compact model of N-well ESD diodes," presented at the Elect. Overstress/Electrostatic Discharge Symp., Riverside, CA, USA, Oct. 4, 2023.

[63] R. Mertens and E. Rosenbaum, "Physical basis for CMOS SCR compact models," *IEEE Trans. Electron Devices*, vol. 63, no. 1, pp. 296–302, Jan. 2016, doi: 10.1109/TED.2015.2502951.

[64] J. Di Sarro, K. Chatty, R. Gauthier, and E. Rosenbaum, "Study of design factors affecting turn-on time of silicon controlled rectifiers (SCRS) in 90 and 65 nm bulk CMOS technologies," in *Proc. IEEE Int. Rel. Phys. Symp.*, San Jose, CA, USA, Mar. 2006, pp. 163–168, doi: 10.1109/RELPHY.2006.251210.

[65] W.-Y. Chen, E. Rosenbaum, and M.-D. Ker, "Diode-triggered silicon-controlled rectifier with reduced voltage overshoot for CDM ESD protection," *IEEE Trans. Device Mater. Rel.*, vol. 12, no. 1, pp. 10–14, Mar. 2012, doi: 10.1109/TDMR.2011.2171487.

[66] R. Merrill and E. Issaq, "ESD design methodology," in *Proc. Elect. Overstress/Electrostatic Discharge Symp.*, 1993, pp. 233–237.

[67] F.-C. Hsu, P.-K. Ko, S. Tam, C. Hu, and R. S. Müller, "An analytical breakdown model for short-channel MOSFET's," *IEEE Trans. Electron Devices*, vol. 29, no. 11, pp. 1735–1740, Nov. 1982, doi: 10.1109/T-ED.1982.21018.

[68] M. Mergens, W. Wilkening, S. Mettler, H. Wolf, and W. Fichtner, "Modular approach of a high current MOS compact model for circuit-level ESD simulation including transient gate coupling behavior," in *Proc. IEEE Int. Rel. Phys. Symp. 37th Annu.*, San Diego, CA, USA, 1999, pp. 167–178, doi: 10.1109/RELPHY.1999.761609.

[69] J. Li, S. Joshi, R. Barnes, and E. Rosenbaum, "Compact modeling of on-chip ESD protection devices using Verilog—A," *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst.*, vol. 25, no. 6, pp. 1047–1063, Jun. 2006, doi: 10.1109/TCAD.2005.855948.

[70] J. Li et al., "Predictive full circuit ESD simulation and analysis using extended ESD compact models: Methodology and tool implementation," in *Proc. Elect. Overstress/Electrostatic Discharge Symp.*, Reno, NV, USA, 2010, pp. 1–8.

[71] S. Joshi, R. Ida, P. Givelin, and E. Rosenbaum, "An analysis of bipolar breakdown and its application to the design of ESD protection circuits," in *Proc. IEEE Int. Rel. Phys. Symp. 39th Annu.*, Orlando, FL, USA, Apr./May 2001, pp. 240–245, doi: 10.1109/RELPHY.2001.922908.

[72] K.-H. Meng and E. Rosenbaum, "The need for transient I-V measurement of device ESD response," in *Proc. Elect. Overstress/Electrostatic Discharge Symp.*, Tucson, AZ, USA, 2012, pp. 1–8.

[73] K.-H. Meng and E. Rosenbaum, "Verification of snapback model by transient I-V measurement for circuit simulation of ESD response," *IEEE Trans. Device Mater. Rel.*, vol. 13, no. 2, pp. 371–378, Jun. 2013, doi: 10.1109/TDMR.2013.2258672.

[74] N. Thomson, N. Jack, and E. Rosenbaum, "Exponential-edge transmission line pulsing for snap-back device characterization," in *Proc. IEEE Int. Rel. Phys. Symp. (IRPS)*, Anaheim, CA, USA, Apr. 2012, pp. 3E.2.1–3E.2.6, doi: 10.1109/IRPS.2012.6241820.

[75] A. Amerasekera, M.-C. Chang, J. A. Seitchik, A. Chatterjee, K. Mayaram, and J.-H. Chern, "Self-heating effects in basic semiconductor structures," *IEEE Trans. Electron Devices*, vol. 40, no. 10, pp. 1836–1844, Oct. 1993, doi: 10.1109/16.277342.

[76] J. Wu and E. Rosenbaum, "Gate oxide reliability under ESD-like pulse stress," *IEEE Trans. Electron Devices*, vol. 51, no. 9, pp. 1528–1532, Sep. 2004, doi: 10.1109/TED.2004.834683.

[77] A. Ille et al., "Ultra-thin gate oxide reliability in the ESD time domain," in *Proc. Elect. Overstress/Electrostatic Discharge Symp.*, Tucson, AZ, USA, 2006, pp. 285–294.

[78] Y. Fong and C. Hu, "The effects of high electric field transients on thin gate oxide MOSFETS," in *Proc. EOS/ESD Symp.*, 1987, pp. 252–257.

[79] D. F. Ellis, Y. Zhou, J. A. Salcedo, J.-J. Hajjar, and J. J. Liou, "Prediction and modeling of thin gate oxide breakdown subject to arbitrary transient stresses," *IEEE Trans. Electron Devices*, vol. 57, no. 9, pp. 2296–2305, Sep. 2010, doi: 10.1109/TED.2010.2053864.

[80] C. H. Diaz, S.-M. Kang, and C. Duvvury, "Circuit-level electrothermal simulation of electrical overstress failures in advanced MOS I/O protection devices," *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst.*, vol. 13, no. 4, pp. 482–493, Apr. 1994, doi: 10.1109/43.275358.

[81] A. Righter, J. Salcedo, A. Olney, and T. Weyl, "CDM ESD current characterization—Package variability effects and comparison to die-level CDM," in *Proc. 31st EOS/ESD Symp.*, Anaheim, CA, USA, 2009, pp. 1–10.

[82] V. Kireev, J. Karp, M. Hart, S. Jeong, and P. Upadhyaya, "Effect of delay in package traces on CDM stress and peak current," in *Proc. 31st EOS/ESD Symp.*, Anaheim, CA, USA, 2009, pp. 1–10.

[83] J. P. Di Sarro, B. Reynolds, and R. Gauthier, "Influence of package trace properties on CDM stress," *IEEE Trans. Device Mater. Rel.*, vol. 14, no. 3, pp. 810–817, Sep. 2014, doi: 10.1109/TDMR.2014.2329552.

[84] M. Etherton et al., "A new full-chip verification methodology to prevent CDM oxide failures," in *Proc. 37th Electr. Overstress/Electrostatic Discharge Symp. (EOS/ESD)*, Reno, NV, USA, Sep. 2015, pp. 1–10, doi: 10.1109/EOSESD.2015.7314811.

[85] D. Abessolo-Bidzo, R. Derikx, P. Cappon, and S. Zhao, "A study of HBM and CDM layout simulations tools," in *Proc. 40th Electr. Overstress/Electrostatic Discharge Symp. (EOS/ESD)*, Reno, NV, USA, Sep. 2018, pp. 1–10, doi: 10.23919/EOS/ESD.2018.8509764.

[86] K.-H. Meng and E. Rosenbaum, "Fast circuit simulator for transient analysis of CDM ESD," in *Proc. 37th Electr. Overstress/Electrostatic Discharge Symp. (EOS/ESD)*, Reno, NV, USA, Sep. 2015, pp. 1–10, doi: 10.1109/EOSESD.2015.7314765.

[87] E. R. Worley, "Distributed gate ESD network architecture for inter-power domain signals," in *Proc. Electr. Overstress/Electrostatic Discharge Symp.*, Grapevine, TX, USA, Sep. 2004, pp. 1–10, doi: 10.1109/EOSESD.2004.5272602.