Dynamic Voltage Overshoot During Triggering of an SCR-Type ESD Protection

Guido Notermans®, Senior Member, IEEE, Hans-Martin Ritter, Steffen Holland®, and Dionyz Pogany, Member, IEEE

Abstract—During triggering of an on-board ESD protection, e.g., an SCR, its voltage may temporarily rise to a large value. This work identifies and distinguishes the two major factors that contribute to this voltage overshoot: (1) the conductivity modulation in the silicon and (2) the inductance of the metal traces. The overshoot is shown to have a major impact on the residual current into the IC during a system level discharge (ESD gun test), in particular in high-speed interfaces, such as USB3. This work presents a physics-based behavioral model which accurately describes both contributions to the voltage overshoot. The conductivity modulation model is based on the premise that the doping concentration of a region in the SCR is insufficient to conduct the trigger current. It follows that a minimum charge must be injected into the low-doped region to enable conduction of the trigger current. SCR triggering is delayed until this minimum charge has been injected. During the delay time, the SCR voltage is determined by the high (∼1 kΩ) off-resistance of the device. After triggering, the voltage is determined by the low-impedance (∼0.25 Ω) on-resistance. The net effect is a voltage overshoot. This first overshoot is followed by a second inductive voltage overshoot at the time when L·dI/dt is maximum. The paper demonstrates how the resulting model can be used to improve system level robustness for a typical USB3 interface.

Index Terms—ESD, protection, TVS, conductivity modulation, USB3.

I. INTRODUCTION

HIGH-SPEED data connections use high data rates up to 10 or even 20 Gb/s. Driver ICs for high-speed interfaces are very sensitive to electrostatic discharges. Therefore, system vendors use additional on-board protection devices, often called Transient Voltage Suppressors (TVS), with an ESD robustness of 15 kV contact discharge according to IEC 61000-4-2 [1], [2].

Several approaches to model the transient voltage overshoot during triggering of such protections have been reported in recent years [3]–[7]. The inductance of on-board and on-chip protection have a significant effect [7], demonstrated by measurements and SEED [8] simulations.

Fig. 1 depicts the current distribution between internal (IC) and external (TVS) protection during an IEC discharge. In the ideal case, almost all of the gun current I_{gun} flows into the TVS.

This current is denoted I_c. A non-zero residual current I_s will flow into the IC, however, because it is connected to the same node. The impedances Z_c and Z_s include the contribution from the ohmic resistance and inductance of the protections. Note that Z_c contains on-board parasitics in addition to the on-chip passive components.

If we ignore the voltage drop on the protections V_c and V_s for the moment, the residual current I_s can be expressed as [7]:

$$I_s \approx \frac{Z_c}{Z_c + Z_s} \cdot I_{gun}$$

Fig. 2 illustrates the impact of the impedances Z_c and Z_s in either protection [7]. During the relatively slow second gun peak, Z_c and Z_s are dominated by the ohmic resistance and the residual current I_s is determined by $\frac{R_c}{R_c + R_s}$. Because this ratio is small, the second peak is all but eliminated in the residual current (red curve). During the fast first peak, the impedance of the inductance $|Z| = \omega L$ is dominant, and the current distribution is determined by the ratio of the inductances $\frac{L_c}{L_c + L_s}$. This ratio is about 0.3 for a typical USB3 application [7], which means that a significant portion of the gun current still flows into the IC during the first IEC peak, although the TVS has triggered in time.

In real life, the impact of V_c and V_s on the current distribution is not negligible. In particular, V_c may initially be quite high during triggering of the TVS. The conductivity
Fig. 2. Total gun current (I\textsubscript{gun}) in black and residual current (I\textsubscript{s}) in red [7] during an IEC 61000-4-2 discharge on a USB3 interface board.

Fig. 3. a) Cross-section of SCR; b) metal stack only.

modulation (CM) [3], [4] in the device causes a switching delay during which the voltage rises temporarily. This voltage overshoot in the protection increases the residual current into the IC. In this paper, which is an extension of [9], we will show how to take the dynamic V\textsubscript{c} into account. This way, we may extend the static model [7], [10] to obtain a complete dynamic model for the current distribution during a gun discharge.

Section II will present the measurements of test structures which were used to distinguish between inductive and CM voltage overshoot. In Section III an enhanced behavioral clamp model will be presented which implements both overshoots and, thus, provides a complete dynamic clamp model. In Section IV it will be shown how to use the new insight to design on-board protections with significantly reduced overshoot. The discussions in Section V will be followed by the conclusions in Section VI.

II. MEASUREMENTS OF TEST STRUCTURES

Several versions of SCR protection devices (Fig. 3a) and a metal stack test structure (Fig. 3b) were designed in a dedicated bipolar technology with a critical dimension of 1 \( \mu \)m [11] with an additional triple well (dpw).

The SCRs have a floating p-substrate and were laid out on either bulk substrates (“reference”) or with identical layout and cross-section on SOI substrates (“improved”). The SCR devices are multi-finger devices with a total width of 700 \( \mu \)m. The SCR test structures are triggered by an external trigger between “signal” and p-sub, indicated by “T” in Fig. 3a (“triggered”). For comparison, identical devices without external trigger (“untriggered”) were implemented as well.

Because the metal test structures (Fig. 3b) have an identical metal stack layout as in the SCR devices (except for the contacts), they have approximately the same metal inductance as the SCR and, thus, exhibit the same inductive voltage drop during the first IEC peak, but none of the CM overshoot of the SCR. From a comparison of the voltage waveforms of both test structures, we will be able to distinguish which part of the overshoot in a vf-TLP pulse is due to the inductance and which part is caused by the CM overshoot.

A. Very-Fast TLP Measurements of Test Structures

An HPPI TLP-8010C system [13] with a 33 GHz Tektronix DPO73304DX oscilloscope and 20 GHz TCA292D adapters were used to record the 5 ns vf-TLP [12] waveforms. The effective bandwidth of the system is 20 GHz and the resolution 100 GS/s. The current risetime is set to 1 ns, unless otherwise noted.

The HPPI system uses well-matched 50 W Kelvin probes. We applied the inverse filtering option to de-embed the Pico-Probe parasitics [14]. No additional voltage de-embedding [5] was deemed necessary.

Fig. 4 shows the measured voltage waveforms for a combined SCR and metal stack (Fig. 3a) and for a metal stack alone (Fig. 3b) as well as the current waveform.

The high time resolution allows separation of the inductive and the CM overshoot, which could not be distinguished previously [5], [7]. The inductive voltage drop:

\[ V = L \cdot \frac{dl}{dt} \]  

for both SCR and metal stack occurs at the inflection point in the current waveform (circle), where \( \frac{dl}{dt} \) is maximum. The CM overshoot occurs only in the SCR waveform and
Fig. 5. Transient $v_f$-TLP voltages for a triggered SCR on SOI for several currents. The first peak voltage saturates at high current, only the second peak increases with current.

Fig. 6. Transient voltage (solid lines) and current (dotted) of a triggered SCR on SOI. The trigger delay, indicated by the arrows, depends on current amplitude.

on a shorter timescale than the current risetime. The model developed in the next section will explain this in detail.

B. High Current

Fig. 5 shows that for high currents, the inductive overshoot increases with current, as expected according to (2). The CM overshoot appears to saturate at about 12 V, however. The reason for this apparent saturation will be explained in Section V.

C. Low Current

Fig. 6 shows current (dotted lines) and voltage (solid lines) waveforms for low current. In this case, there is no second peak, because $L\frac{dl}{dt}$ is too small. We observe a delayed triggering which appears to be smaller when the initial current before the device triggers is higher. We hypothesize that before the device can trigger a minimum charge must be injected into the low-doped substrate to enable conduction of the trigger current. Let us denote this charge $Q_{t1}$. The hypothesis explains why the delay is smaller for higher currents. At higher currents, $Q_{t1}$ is, obviously, reached earlier.

A current of 25 mA is lower than the SCR trigger current $I_{t1}$. Therefore, all current flows through the trigger circuit (with a static trigger voltage $V_{t1} \approx 7$ V). At 30 mA of initial current, the required conductivity modulation is reached 3 ns after the pulse rising edge, the SCR triggers and the voltage drops to the SCR holding voltage $V_h \approx 1.2$ V. At that time, the current rises to 150 mA. For a still larger initial current of about 60 mA, the SCR triggers already after 1.5 ns (200 mA after triggering).

D. Peak Voltage vs. Current

Fig. 7 summarizes the peak voltage in each vf-TLP pulse as a function of TLP current, both for the complete SCR (blue) and for the metal stack only (green). Since the risetime of the pulse is set to 1 ns in each of the pulses, the time derivative of the current $\frac{dl}{dt}$ is proportional to $I$. This means that a pure inductive load with a finite ohmic resistance, such as the metal stack, will show up as a straight line. Its slope includes an ohmic part which scales with $I$ and an inductive part which scales with $\frac{dl}{dt}$.

The curve for the SCR comprises two distinct peaks (see Fig. 7 inset). For currents below 25 A the CM peak dominates and above 25 A the inductive peak is larger. Below 25 A, the CM peak increases rapidly for small currents and then levels off around 12 V (see Fig. 5). Beyond 25 A, the curve approaches the inductive dependence, i.e., the peak voltage increases linearly with current (and, thus, with $\frac{dl}{dt}$ as well).

III. OVERSHOOT MODELING

The observations in the previous section indicate a way to implement a dynamical SEED model for an SCR.

A. Static Behavioral Model

The basic static model, used in [7], is a piecewise linear (PWL) Verilog-A model [10] in series with an inductance.
in which \( I_{\text{inj}} \) denotes the injected electron current into the substrate, \( I_0 \approx 1 \) pA the diode leakage current, \( n_{\text{inj}} \) the injected electron concentration, and \( n_0 \approx 5 \times 10^5 \text{ cm}^{-3} \) the minority carrier concentration in the p-sub. In order to pass a trigger current of \( I_1 \approx 50 \) mA, we need to inject \( n_{\text{inj}} \approx 6 \times 10^{15} \text{ cm}^{-3} \), according to (3). By multiplying \( n_{\text{inj}} \) by the volume into which the electrons diffuse (“Vol” in Fig. 8) and the electron charge \( q \), we arrive at the necessary charge to trigger the SCR:

\[
Q_{\text{inj}} = n_{\text{inj}} \cdot q \cdot \text{Vol}
\]  

We estimate Vol \( \approx 2 \times 10^{-8} \text{ cm}^{-3} \), which puts \( Q_{\text{inj}} \approx 40 \) pC. The total injected charge will be used in the dynamic clamp model as criterion to switch on the SCR. The SCR switches from its high-impedance off-state to its low-impedance on-state once \( Q_{\text{inj}} \) is reached:

\[
Q_{\text{inj}}(\tau) = \int_0^\tau I \cdot dt > Q_{\text{inj}} \approx I_1 \cdot \tau
\]  

which happens after a trigger delay time \( \tau \). Note that in [16] good results were obtained by integrating the voltage beyond the static trigger voltage, which lacks a physical interpretation. The model presented in the present work offers the straightforward physical interpretation as the charge needed to enable sufficient conductivity modulation for the trigger current to flow.

3) Modulated Resistance: The dynamic model includes a model for the modulated resistance of the low-doped part of the silicon. Analogous to [3], [4], we have modeled the modulation of resistance \( R_{\text{cm}} \) by taking the total injected charge into account:

\[
R_{\text{cm}}(t) = R_0 \cdot \frac{1}{1 + \frac{Q_{\text{inj}}}{Q_0}}
\]  

where \( R_0 \) is the off-resistance before triggering, mainly determined by the doping of the intrinsic region, \( Q_{\text{inj}} \) the total injected charge according to (5) and \( Q_0 = n_0 \cdot q \cdot \text{Vol} \) represents the charge due to the doping concentration. Implementation of (6) in Verilog-A is straightforward. The correction for reduced mobility at higher electric fields [4] was omitted, because it was found to have a negligible effect on the resistance.

In order to arrive at the total device resistance, we need to add the remaining resistance of the metallization, etc. (\( R_{\text{on}} \approx 0.25 \Omega \)):

\[
R(t) = R_{\text{cm}}(t) + R_{\text{on}}
\]  

Before triggering, \( R(t) \) is mainly determined by the high resistance of the p-sub. After triggering, \( R_{\text{cm}}(t) \ll R_{\text{on}} \) and \( R(t) \) reduces to the on-resistance \( R_{\text{on}} \).

4) Protection On-State and Capacitance: In previous work [3]–[5], [9], the impact of the diffusion capacitance of the protection on the voltage waveform was ignored. In this work we will show how to make the simulation of the voltage waveform more realistic by taking this capacitance into account.

First of all, it is important to realize that in the ESD case, when the protection is conducting, the device capacitance is not small. A high-speed protection is designed to

![Diagram of a triggered SCR highlighting triggering.](image-url)
have a very small off-state capacitance to maintain application data integrity during normal operation. During an ESD discharge, however, the current flowing through the protection increases the capacitance dramatically. There are two junction capacitances in series in the path of the electron trigger current (step 2 in the triggering mechanism; Fig. 8): the relatively high nw/dpw and the low n+/p-sub junction capacitance. Let us denote these capacitances \( C_{\text{rev}} \) and \( C_{\text{fw}} \), respectively.

The overall off-state capacitance is low because the low \( C_{\text{fw}} \) (∼ 0.25 pF) ‘hides’ the large \( C_{\text{rev}} \) (estimated to be ∼20 pF). When current flows through the protection, this junction becomes forward biased and its diffusion capacitance \( C_{\text{diff}} \) increases linearly with current [15]:

\[
C_{\text{fw}} = C_{\text{diff}}(t) = C_0 \cdot I(t)
\]

in which \( C_0 \) is a scaling constant and \( I(t) \) the total current through the protection. When the ESD current reaches the mA range, the total capacitance has increased by 2 orders of magnitude.

Fig. 9 shows both current (red) and voltage waveform (blue) at a low current before triggering of the SCR. \( C_{\text{fw}} \) can be estimated from the current peak according to \( I = C_{\text{fw}} \cdot \frac{dV}{dt} \) to be about \( C_{\text{fw}} \approx 10 \text{ pF} \) and it will increase linearly with current according to (8). This means that during and after triggering, \( C_{\text{fw}} \) will be too large to hide \( C_{\text{rev}} \).

The large \( C_{\text{rev}} \) affects the voltage waveform, because when the device snaps back from the trigger voltage \( V_{t1} \) to the holding voltage \( V_h \), capacitor \( C_{\text{rev}} \) discharges via the dynamic device resistance \( R_{\text{cm}}(t) \) (6). The discharge current effectively slows down the voltage drop from \( V_{t1} \) to \( V_h \), which can be modeled by adding a voltage \( V_{\text{cm}} \) to the model:

\[
V_{\text{cm}}(t) \approx -R_{\text{cm}}(t) \cdot C_{\text{rev}} \cdot \frac{dV(t)}{dt}
\]

in which \( V(t) \) represents the total device voltage. Capacitance \( C_{\text{rev}} \) can be used as a free parameter, within reasonable limits, to fit the voltage waveform to the measurements. Note that (9) basically adds an RC-integrator to the voltage in the model which could be used to model switching delays from other physical causes, e.g., on-state spreading (OSS) [17].

Fig. 10 illustrates that the consideration of the device capacitance allows to predict the actual voltage waveform (red), in particular the voltage drop following the first peak, very accurately. The current waveform was adjusted to match the measured waveform (blue).

Without (9) the voltage drop after the CM peak is instantaneous [9].

C. Testing the Model in a Circuit Simulation

Circuit simulations were performed using Keysight’s Advanced Design System (ADS) to assess the accuracy of the model’s predictions. Fig. 11 shows simulated current and voltage waveforms at high current: 5 A, 15 A, and 25 A. The voltage waveform exhibits a double peak as in the measurements (Fig. 5). The second peak is the inductive peak and the maximum occurs at the inflection point of the current, as in the measurements.

Note that the CM peak levels off at higher current, and the inductive peak continues to increase with current, which agrees with the measurements (Fig. 5) as well. At 25 A both peaks have approximately the same height.

The results for low currents are shown in Fig. 12. The simulated trigger delay compares well with the measurements (Fig. 6). After triggering, the device enters the on-state (see the blue lines with on-state currents of 150 and 200 mA).

Note that the voltage drop after the first (CM) peak in Fig. 11 and Fig. 12 is less abrupt than in the previous work [9], because the device capacitance has now been taken into account in the model, as explained above.
Fig. 12. Simulated voltage (red) for a triggered SCR on SOI for low currents (blue).

Fig. 13. Simulated peak voltage vs. current (black dotted line and measurements (blue symbols) for a triggered SCR on SOI.

Fig. 14. Current vs. dynamic (peak) voltages for the “reference” (blue circles) and “improved” device (red squares and grey triangles). A significant reduction of the overshoot voltage can be observed in the improved device. Inset: Current vs. static voltage for triggered (red squares) and untriggered (grey triangles) SCR.

Fig. 15. Current vs. dynamic (peak) voltages for the “reference” (blue circles) and “improved” device (red squares and grey triangles). A significant reduction of the overshoot voltage can be observed in the improved device.

IV. IMPROVED SCR PROTECTION

In this section, we will deploy the new dynamic model to explore ways to improve system level protection for a high-speed system (see [7]). Let us start by extending the expression for the residual current (1) to account for the device voltages $V_c$ and $V_s$ (Fig. 1) as follows:

$$I_s \approx \frac{Z_e}{Z_e + Z_s} \cdot I_{\text{gun}} + \frac{V_c - V_s}{Z_e + Z_s}$$

in which $V_c$ and $V_s$ represent the (dynamic) voltage drop on external and internal protection, respectively.

Equation (10) suggests two ways to reduce $I_s$: (1) decreasing the impedance ratio (the approach followed in [7]), or (2) decreasing $V_c$, the option which we are exploring in this work (the option to increase $V_s$ is not available to us, as it is determined by the IC design). We may write:

$$V_c(t) = V_{t1} + V_{cm}(t)$$

in which $V_{t1}$ is the static trigger voltage of the external trigger ($T$ in Fig. 8), and $V_{cm}(t)$ the CM overshoot voltage. Therefore, we can reduce $V_c$ by either reducing $V_{t1}$ or $V_{cm}$.

Since $V_{cm}$ depends directly on the charge $Q_{t1}$ via (6) which needs to be injected into the low-doped substrate to provide the necessary conductivity modulation, it is clear from (4) that decreasing the volume to be filled should lead to a lower overshoot voltage in the first peak. A very efficient way to reduce the volume is to create the SCR on an SOI substrate which limits the depth to a few $\mu$m instead of typically 100 $\mu$m on a bulk substrate. Fig. 14 shows the peak voltages in a vTLP measurement of an improved SCR on an SOI substrate (“triggered” or “untriggered”) compared to a reference device (“reference”) with identical layout and cross-section on a bulk substrate. The improved SCR has a significantly reduced overshoot in the first peak from about 30 V to 20 V at 40 A of current. This reduction has a huge impact on system level robustness.

When using this protection, for instance, in a typical USB3 system, as described in [7], the protection level increases from 6 kV to over 15 kV and is essentially limited only by the pass level of the TVS and not by the residual current at lower gun levels. Both CM peak and inductive peak are lower for the improved version. The latter because the CM peak ‘spills over’ into the inductive peak and because of minor improvements to the metal stack.

V. DISCUSSION

Let us now return to the question of what determines the apparent saturation of $V_{cm}$ at higher currents, as illustrated in Fig. 14 (“triggered” and “untriggered”), which compares the overshoot for a triggered vs. an untriggered SCR. In both cases the overshoot voltage saturates at the same value, whereas the static trigger voltage is very different (inset). This illustrates that, although the trigger is needed to get a low static triggering voltage $V_{t1}$, it does not lower the dynamic overshoot voltage $V_{cm}$.

What causes the CM overshoot to saturate? In recent literature [6] it was proposed that the saturation voltage is equal to the breakdown voltage of the reverse-biased junction
Fig. 15. TCAD simulations at a current of 5 A in the improved device. Total overshoot voltage (solid line) for a pulse risetime of 1 ns is mainly determined by the nw/dpw junction breakdown voltage (green dashed line) and not by the voltage drop in the substrate (dashed red line).

in the SCR plus two forward-biased diodes. In the present work (see Fig. 3), the nw/dpw junction with a breakdown voltage of about $BV \approx 11$ V forms the reverse biased junction and the p+/nw and psub/n+ junctions are both forward biased. This adds up to roughly 13 V which fits the saturation voltage shown in Fig. 14. When the voltage reaches the breakdown voltage, additional carriers are created by avalanche generation, which limits additional voltage increase.

In contrast to [6], the devices in our investigation contain a low-doped region ("p-sub" in Fig. 3b), over which an additional voltage drop might occur. It can be shown, however, using TCAD simulations [18] that the total overshoot voltage for a risetime of 1 ns is mainly determined by the breakdown voltage of the nw/dpw junction and only for a small part by the ohmic voltage drop over low-doped p-sub region (Fig. 15). Therefore, the explanation from [6] applies equally well to our case, despite the low-doped region.

Another question to be considered is the effect of the vf-TLP current risetime. The essential question is whether the current ramp is slow enough to accommodate the necessary current injection during the ramp or not. For a risetime of 1 ns, this requirement appears to be fulfilled (Fig. 5). But for shorter risetimes this may not be the case. Then, the required conductivity modulation cannot be achieved during the current ramp and the voltage keeps rising beyond the saturation value. This effect is shown in Fig. 16, which compares overshoot voltage vs. current curves for several risetimes.

The 1 ns risetime curve is shown as reference. For shorter risetimes, such as 300 ps and 600 ps, the conductivity modulation cannot keep up with the current ramp and the CM voltage overshoot increases beyond the saturation voltage. For this reason, the 300 ps and 600 ps curves do not show the kink, at the current where the I-V curve switches to its inductive dependence [18], [19]. The black dashed lines in Fig. 16 represent ADS simulation results using the dynamic model. There is good agreement between model and measurements.

Note that for a risetime of 0.6 ns and shorter, the peaks caused by conductivity modulation and inductance can no longer be resolved (even with a 33 GHz scope) because they occur at approximately the same time. Note, furthermore, that for the device in literature [6] with n-well and p-well abutting, the doping level is high enough to ensure a limited voltage drop at the forward biased junction at any risetime, because there is no low-doped p-sub in the structure.

VI. CONCLUSION

We have shown that the voltage waveform during switching of a protection comprises a double peak:

1) a first fast peak caused by the trigger delay due to conductivity modulation in the SCR, and

2) a second peak caused by the inductance of the protection.

At currents below a certain value, 25 A in our case, the first peak due to conductivity modulation (CM peak), is dominant. The conductivity modulation, furthermore, causes a trigger delay at low currents. This is due to the fact that a minimum charge $Q_{11}$ needs to be injected into the SCR to achieve sufficient conductivity modulation. At currents above 25 A the second voltage peak due to metal parasitic inductance dominates.

Both peaks and the trigger delay can be simulated through an extension of the static clamp model [10] by adding a critical CM charge $Q_{11}$ and by adding the diffusion capacitance to the model, the smooth snapback may be simulated accurately.

The charge $Q_{11}$ determines the delay during switching as well as the height of the CM overshoot. By reducing the critical charge $Q_{11}$ through the shrinking of the active device volume, it is possible to reduce the CM voltage overshoot during the first IEC peak to a level which does not limit the system level ESD performance of an application. System performance is then essentially boosted to the pass level of the TVS and has been shown to exceed 15 kV for a USB3 application.

REFERENCES


