3D TCAD studies of Snapback Driven Failure in Punch-through TVS Diodes under System Level ESD Stress Conditions

Jhnanesh Somayaji B, Monishmurali, Ajay Singh, N Kranthi K and Mayank Shrivastava

Advanced Nanoelectronic Devices and Circuits Research Group, Department of Electronic Systems Engineering, Indian Institute of Science, Bangalore- 560012, India; email: mayank@iisc.ac.in

Abstract **–**This paper explains the ESD failure dynamics of TVS diode, reported for the first time. The failure is universal to TVS diode structure designed as ESD protecting element. This work gives the detailed physical insights into failure mechanisms, using 2D & 3D TCAD simulations. Failure at very low currents due to non-uniform bipolar turn on, for longer pulse duration, is explained. 2D simulation shows the non -uniformity due to current crowding while 3D simulation supports the failure dependency with thermally driven filamentation. Finally, three device engineering schemes are proposed to improve the phenomena of low current ESD failure under longer pulse duration, through which non-uniform bipolar turn–on is limited.

I. Introduction

According to IEC 61000-4-2 standard, System level ESD robustness is a standard mandatory requirement for mobile phones compliant with the CE certification standard. Board designs of mobile manufacturer include ESD protection along with other EMC/EMI requirements. While there is a wide range of fails due to IEC pulses, most critical for hardware design are physical damage and latch-up.

ESD damage to ICs can occur at any stage, from assembly, through board-level soldering, to end-user interactions. Today, more than 60% of the electronic system failures are related to over voltage and electrostatic discharge Phenomena [1]. Very Large Scale Integrated (VLSI) circuits are very much sensitive and should be protected from voltage transients originating from lightning, inductive load switching, and ESD events. High voltage modules, which are the primary components for building system on chips (SoC), are often prone to ESD events. The levels of ESD strikes from both voltages and currents can be much greater in the end-user environment. When an electrical system experiences an ESD event, the system can continue to work with a soft failure. While device-level modules are usually sufficient for the controlled ESD environment of the factory floor, they are completely inadequate for system-level testing. With the evolution of power supply voltage scaling, better off-chip discrete low-voltage Transient Voltage Suppressors (TVS) are required for circuits/systems operating at 1 to 3V [1,2]. Conventional silicon TVS devices consist of p-n junction that breaks down at a well-controlled voltage, and operate at the avalanche condition [2]. However, stand-off voltage or low breakdown needs high doping concentrations, which cause both the capacitance and the leakage current to increase dramatically. Several concepts have been explored in the literature to reduce the leakage and capacitance. One of the concepts is to use the breakdown voltage between the emitter and collector of the BJT, which is much lower than the avalanche breakdown voltage of the base-collector junction [2]. Other way is to simply stack several P-N junctions in series so that the standoff voltage is approximately 0.5V times the number of junctions. Finally, a punch-through TVS structure, which may be thought of as a BJT, is based on NPN or PNP open-base bipolar transistors. The punch-through effect takes place when the two existent depletion regions of the P-N junctions meet each other. Beyond this point, a small rise of the applied bias reduces the collector-base potential barrier and the

Authorized licensed use limited to: J.R.D. Tata Memorial Library Indian Institute of Science Bengaluru. Downloaded on November 25,2022 at 10:13:26 UTC from IEEE Xplore. Restrictions apply.

electrons diffused from the emitter can be driven towards the collector. A punch-through diode having low base doping is considered so that punch-through occurs at a voltage lower than the avalanche breakdown voltage. The punch-through diode shows much lower leakage current and capacitance comparing to Zener diode. However, when a p^+-n-p^+ punch-through diode conducts large punch-through currents, the mobile carriers result in large field and voltage drop, i.e., large dynamic resistance and clamping voltage. Thus four-layer structure punch-through TVS was proposed [2-5]. The device is comprised of a n^+-p^+-p^--n^+ structure with relatively light doping in the p^+ and p- layers so that the reverse-biased p^+-n^+ junction will not break down by avalanche. In the n^+-p^+-p^--n^+ four-layer structure, it is found that as doping concentration in the p^+ layer increases, the device exhibits a weak snapback effect. By using this snapback effect caused by impact ionization, we could significantly reduce the clamping voltage of the device. With appropriately selected doping levels in the base layer, the stand-off voltage of the punch-through TVS device, can cover a wide range of standoff voltage from 4 V to as low as 1 V.

Although few papers [1-5] have explored characteristics of 3-layered and 4-layered punch through TVS structures in the vertical or lateral dimensions, no literature provides a comprehensive study of TVS failures and its physical insights into failure dynamics. Another concern about a TVS device, is its transient response. Spikes causing circuit damage when TVS diodes turn on, are not acceptable. This work attempts to bridge this gap.

Under the transient conditions, the device failure does depend on the pulse duration and the fact that failure current is the direct function of TLP pulse duration is explained systematically using detailed 2D and 3D TCAD simulations.

II. TVS Punch through diode

Fig. 1 illustrates the cross section of 4-layered $n^+P^+P^-n^+$ vertical punch-through TVS diode. The device has Cathode which acts as collector and Anode at the bottom which is the emitter of an open base BJT. Buffer and Epi layers correspond

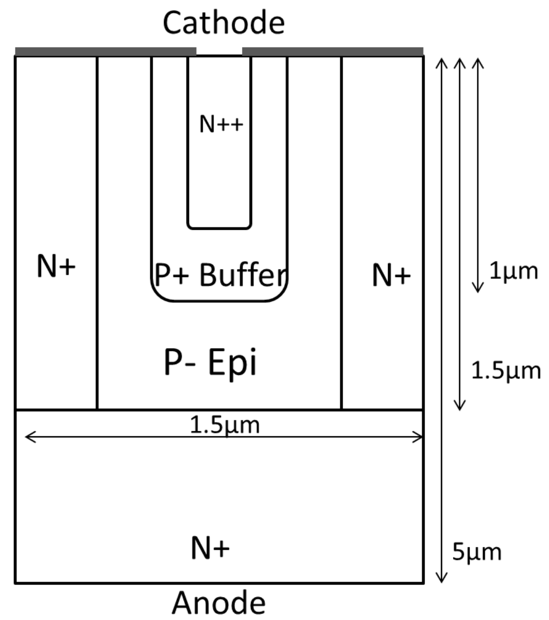


Figure 1: 2D Cross-sectional view of 4-layered vertical TVS structure. The width of the device, in Z plane, is into the paper.

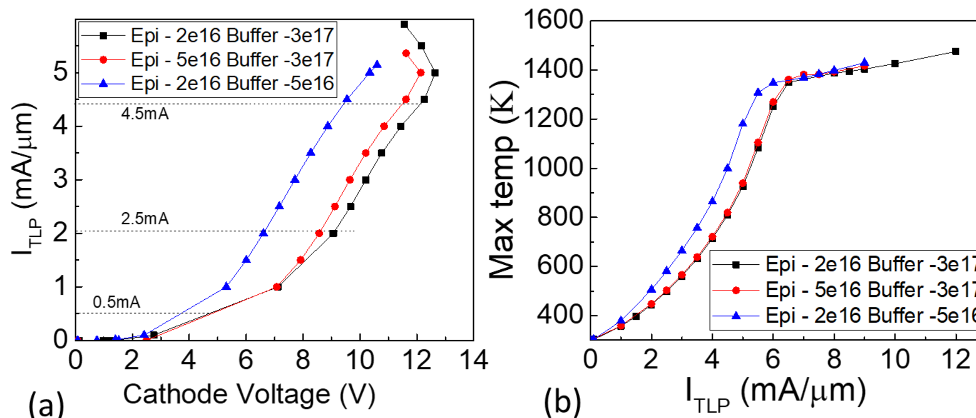


Figure 2: Simulated TLP characteristics of TVS structure for different buffer and Epi dopings (a) I-V characteristics (b) Lattice temperature.

Authorized licensed use limited to: J.R.D. Tata Memorial Library Indian Institute of Science Bengaluru. Downloaded on November 25,2022 at 10:13:26 UTC from IEEE Xplore. Restrictions apply.

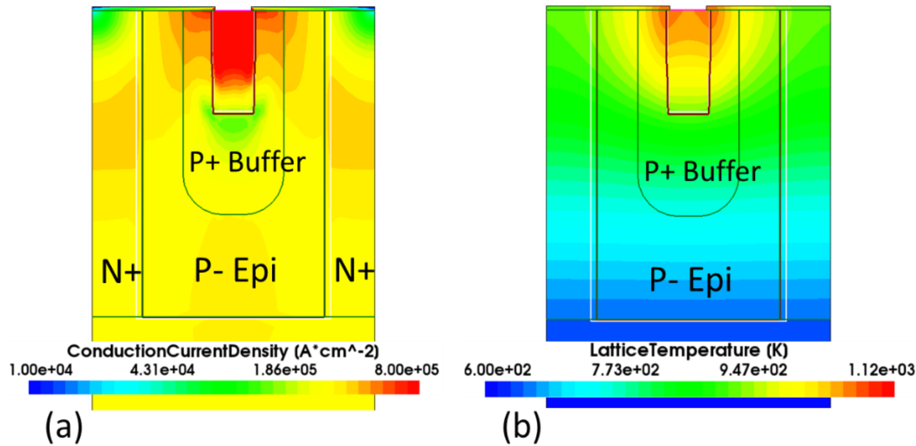


Figure 3: Contours of (a) Conduction current density (b) Lattice temperature taken at 80% of It2 after the bipolar is triggered.

to the base region. The Base region has graded doping profile with Buffer region highly P^+ doped compared to Epi layer. The Buffer layer has a depth of 1μm while Epi layer thickness is about 1.5μm. An open base BJT always provides a trade-off between achieving low leakage current and low clamping voltage (low dynamic resistance, Rdyn). Punch-through graded base region with P^+ buffer and P^- epi offers better design window by achieving both (leakage & Capacitance) with better Rdyn [3]. The device is stressed with a TLP pulse of 100ns with a rise time of 10ns. Figures 2a and 2b show simulated TLP characteristics for devices having different dopings in the buffer and Epi regions. The simulations were carried out using Sentaurus TCAD. In addition to Poisson equation and carrier continuity equations, high field saturation mobility model, avalanche UniBo2 model, and thermodynamic models were included to understand the device physics. The net p-base region doping determines the punch-through breakdown before reverse biased junction breakdown occurs. Furthermore, as the buffer doping increases, the failure threshold hikes (It2/unit width). But the attempt here is to study the detailed dynamics of the failure inside the TVS structure. The initial current (0.5mA) through the device originates from the punch-through breakdown. At moderate currents (~2mA), the parasitic bipolar turns on. Fig. 3(a) illustrates the current density contours (taken at 80% of It2) for the case of 100ns pulse width, after the parasitic bipolar is triggered. At higher currents (>4.5mA), most of the current flows through the emitter whereas the base current is defined by β of the n-p-n. At lower TLP current (<1mA) there are high fields only at the n-p junction, whereas at moderate currents, (>2.5mA) the onset of bipolar turn on leads to development electric fields, under the entire cathode. At the onset of bipolar turn on, the device suffers from high current densities, leading to thermal failure [6] (Fig. 3(b)). Though, an interesting observation comes when the device is investigated under transient conditions. When the device was stressed for the pulse duration of 100ns, max. lattice temperature was found to be unsettled or it was still in the state of rising. Also, this was observed for many of the injected currents. This concern led us to increase the pulse duration of the ESD stress to 1000ns.

III. Low Current Failure: Physical Insights

Fig. 4 describes the device transient characteristics when stressed for period of 1000ns with a rise time of 10ns. It is clearly visible that as the pulse duration increases towards 1000ns, the device fails at lower currents. (Fig. 4(a)). Device failure condition was considered for max. lattice temperature of 1200K. In other words, the failure current (It2) is lowered as a function of ESD pulse duration (Fig. 5(a)). It was found that the device fails due to non-uniform bipolar turn-on, which eventually results in localized current crowding, and device failure (Fig. 3(a-b)). The non-uniformity originates from electro thermal instability [7-8]. As the pulse duration increases, the time to turn on the device bipolar raises. This can be observed in Fig. 4(b). The turn on time decreases for higher injected current. We can learn from Fig. 5(b) that as the

Authorized licensed use limited to: J.R.D. Tata Memorial Library Indian Institute of Science Bengaluru. Downloaded on November 25,2022 at 10:13:26 UTC from IEEE Xplore. Restrictions apply.

injected current raises, the non-uniformity in turning on the bipolar increases. For the purpose of investigation, lowest current at which device fails under 1000ns stress pulse

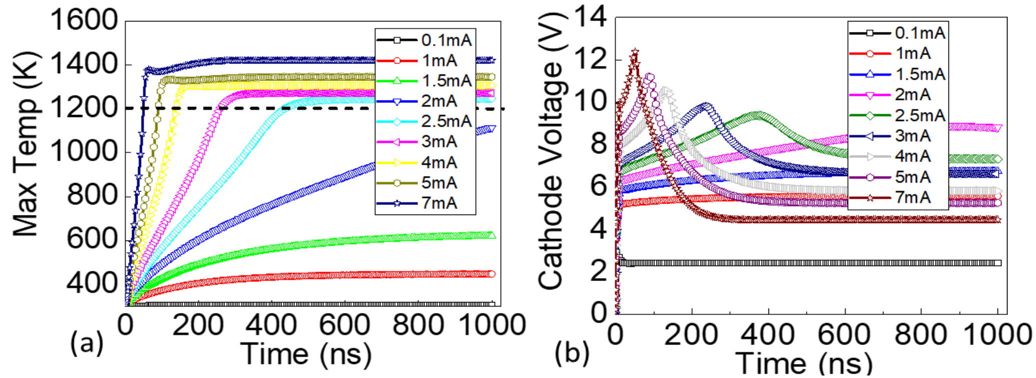


Figure 4: Transient simulation at different injected currents, as a function of time. (a) Max. lattice temperature (b) Cathode voltage

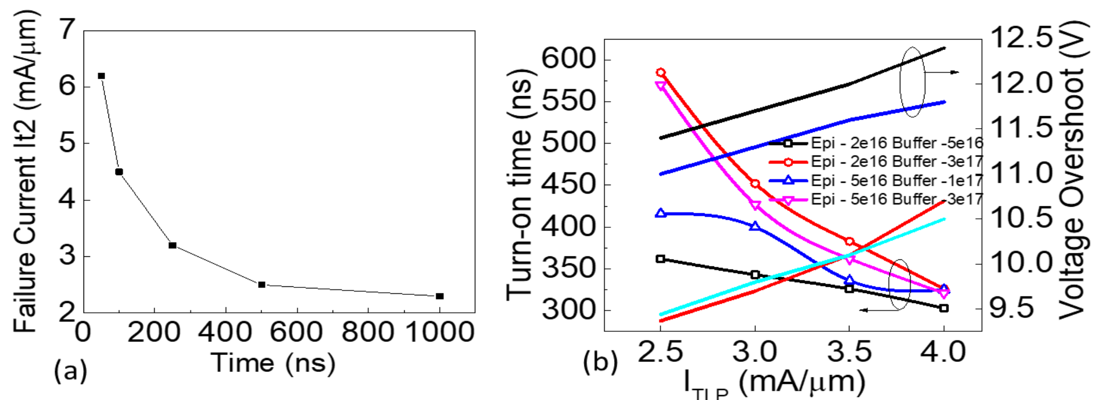


Figure 5: (a) Failure current as a function of time (b) Turn-on time and peak voltage overshoot for different injected currents. Low current ESD failures can be observed, decreasing with increase of the pulse duration.

(which is the worst-case scenario), is considered here. Fig. 6 illustrates the voltage and lattice temperature plot, for 2.5mA/µm stressed for 1000ns. Following the plot, the contours are extracted at different time instants marked as (i), (ii), (iii), and (iv) taken at 220, 370, 500, and 850ns respectively. Fig. 7 depicts the contours taken at these instants. It is evident that after breakdown occurs, the electric field rises with time at the cathode (Fig. 7(a)) due to which the voltage increases. This is attributed to lowered bipolar efficiency (instance (i)) and reaches peak (instance (ii)). A similar observation is seen in the contour of the e-density (Fig. 7(b)). Initially, before the bipolar triggers, the current density is uniform. Once the

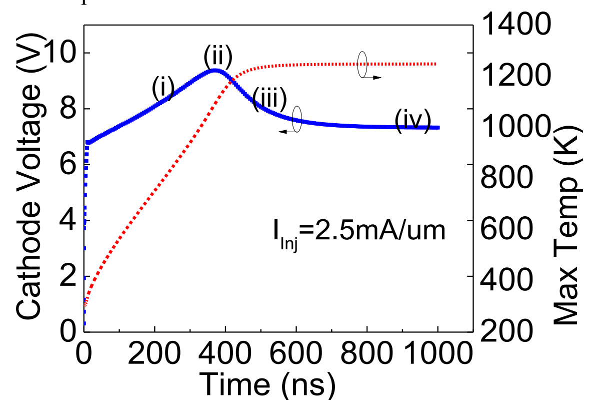


Figure 6: Cathode voltage and max. lattice temperature as a function of time, for injected current of 2.5mA/µm. Four instances are marked at different time scales, for which the device contours are pulled out.

bipolar starts triggering, the current is conduction gets crowded at the emitter-base junction as marked in Fig. 7(c). Similarly, the e-density increases with time (as temperature starts

Authorized licensed use limited to: J.R.D. Tata Memorial Library Indian Institute of Science Bengaluru. Downloaded on November 25,2022 at 10:13:26 UTC from IEEE Xplore. Restrictions apply.

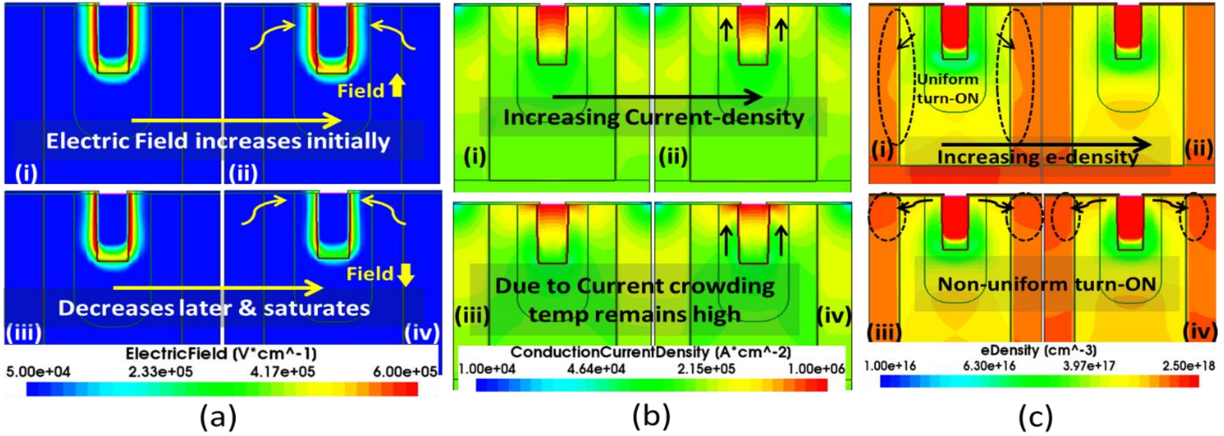


Figure 7: Contour explaining taken at four instants of time w.r.t Fig. 4. (a) Electric field (b) Conduction current density (c) e-density

raising. In addition, with time, the non-uniformity bipolar turn-on dominates. One important observation to be noted, is that as a result of weak bipolar, the triggering occurs after longer duration (instance (iii)). This results in non-uniformity of impact ionization since the e-density increases no-evenly. The reason being that the lateral base emitter junction becomes dominant comparing to the vertical base emitter junction. This results in increment of conduction current density (Fig. 7(c)). It is worth mentioning at this point, that due to higher current density the temperature keeps increasing constantly, reaching a failure point. This becomes clearer in Fig. 8, where the temperature rises with injected current stressed with longer pulse duration.

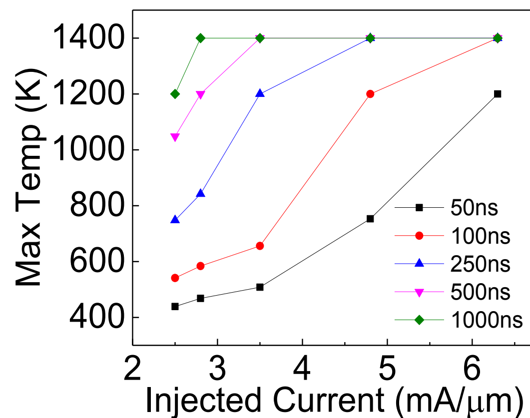


Figure 8: Maximum lattice temperature as a function of injected currents stressed for different time pulses from 50ns to 1000ns.

IV. 2D vs 3D Observations

Ensuring device width (W) of 10μm, the 3D structure was simulated using the simulation setup/framework borrowed from [8-9]. 2D TCAD enforces uniform electrical and thermal response along the device width. Variations in electrical and thermal response of the device in width plane is, however, naturally captured when 3D geometry is studied using 3D TCAD simulations. Fig. 9 compares the 2D and 3D simulated TLP I-V characteristics, extracted from 100ns pulses of ESD stressed using TCAD simulation. It shows a distinct behavior which is attributed to the charge modulation within the device. Although the TLP I-V characteristics show slight deviation, the failure current is about 30% lower for 3D device compared to 2D device. The failure dynamics in 3D device can be explained in a similar way as the 2D device, but with some more important observations. Fig. 10 shows the transient behavior of the 3D device for the same current (2.5mA) as simulated for 2D device. It can be explained by the following arguments. (a). Increase in the electric field and voltage (b) Initial Impact Ionization triggering a non-uniform bipolar turn on (c) Non uniform current conduction. (d) Predominant rise of the temperature. (e) Increase in the impact ionization which leads to rise of the electric field resulting in a thermal run away. As 3D simulation captures the non-uniformity very well, sudden rise in the lattice temperature is well observed. This leads to onset of filamentation and after certain high current, to localized current crowding ending with

Authorized licensed use limited to: J.R.D. Tata Memorial Library Indian Institute of Science Bengaluru. Downloaded on November 25,2022 at 10:13:26 UTC from IEEE Xplore. Restrictions apply.

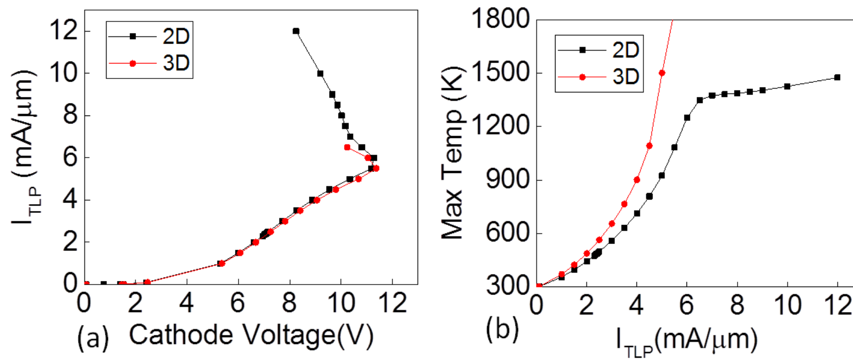


Figure 9: TLP I-V characteristics and max. lattice temperature vs. I_{TLP} for 2D and 3D device. The 3D shows lower It2 compared to 2D devices which indicate a severe current crowding resulting in a filamentation.

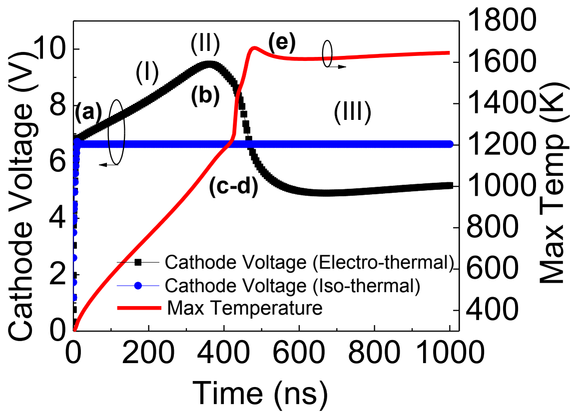


Figure 10: 3D simulated characteristics for the injected current of 2.5mA/um for electro-thermal and isothermal setups. (I), (II), (III) shows the different time instants at which the contours are extracted for failure study.

filamentation failure [8-10]. To confirm that it is a thermal failure, an iso-thermal 3D simulation was conducted, injecting the same current. It was very clear that the voltage remains constant throughout the pulse duration and that there was no filamentation which proves that the filament is purely due to thermal instability. 3D contours were extracted for better understanding of the device failure mechanism. Fig. 11 explains the contours extracted at three-time instants (refer Fig. 10) (I)-230ns, (II)-370ns & (III)-700ns. Fig. 11(a) shows the contours of the electric field, indicating the increase of the transient voltage (in Fig. 10) and decreases gradually (at (III)).

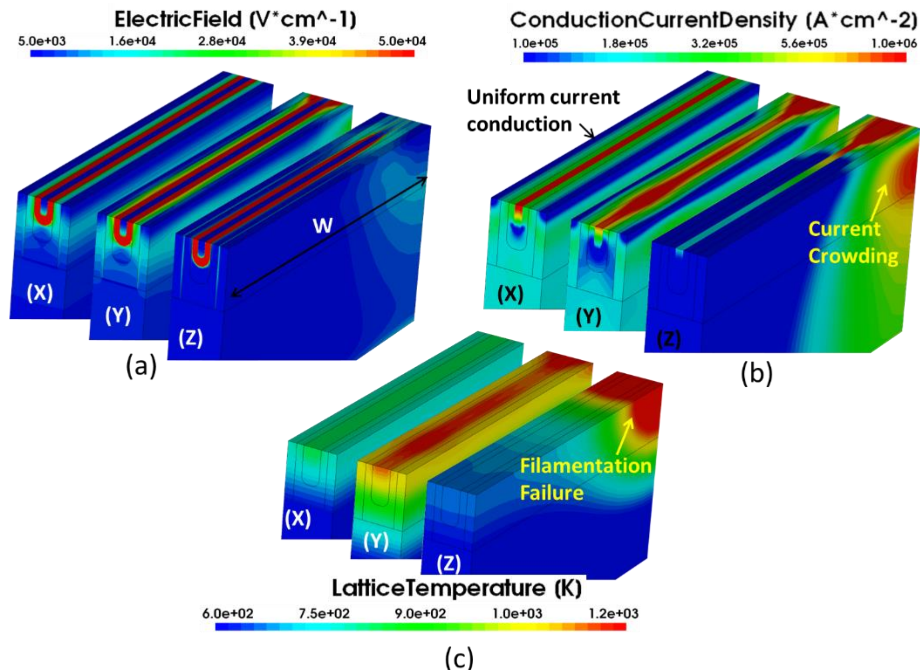


Figure 11: 3D contours of (a) Electric field (b) Conduction current density and (c) Lattice temperature, captured at three instants of time (I), (II), (III) w.r.t Fig. 10.

Authorized licensed use limited to: J.R.D. Tata Memorial Library Indian Institute of Science Bengaluru. Downloaded on November 25,2022 at 10:13:26 UTC from IEEE Xplore. Restrictions apply.

Similarly, the contours of the current density demonstrate uniform conduction initially and then switch to non-uniformity leading to current crowding and formation of the filament at the other end of the device along the width, as shown in Fig. 11(b) (at (III)). The lattice heating (Fig. 11(c)) during the filamentation failure, can be observed under the contour of lattice temperature (at (III)), eventually leading to device failure. This can also be observed in Fig. 12. The 3D device failure current is a function of the pulse duration. The device survives only if the pulse duration is shorter. Fig. 13 summarizes the entire phenomena using a flow-chart.

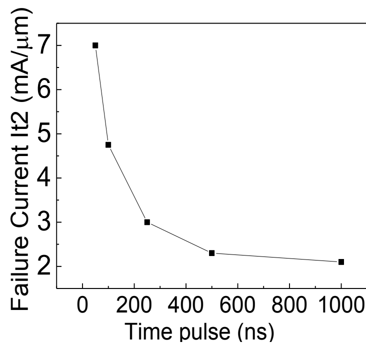


Figure 12: 3D Failure current as a function of pulse duration.

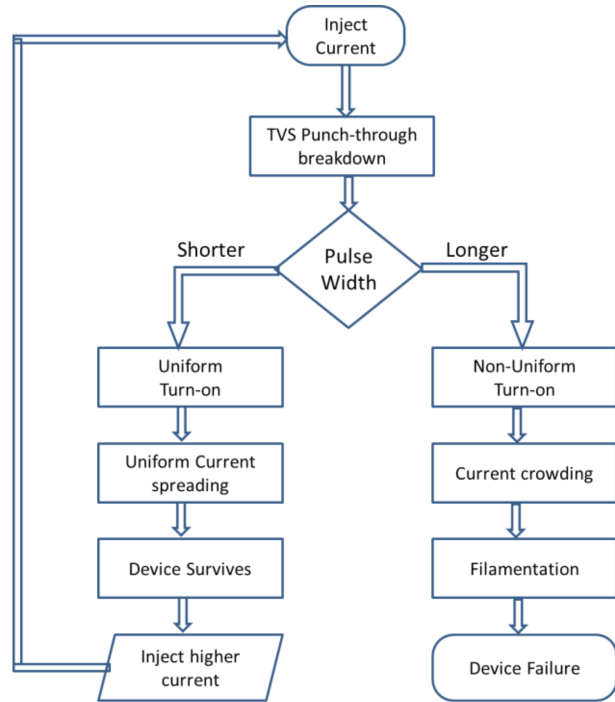


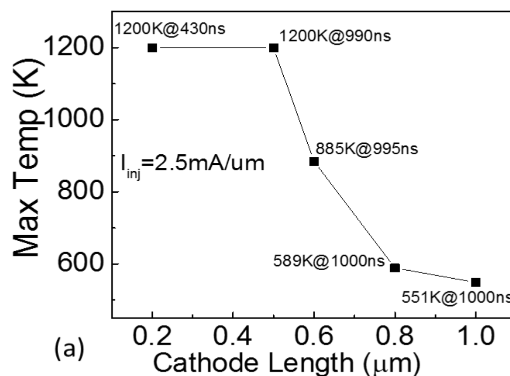
Figure 13: Flow chart summarizing the series of discussed physical events

VI. Parametric Optimization

This section will focus on providing the basic design schemes/solutions to improve the low current ESD failures in TVS diodes, using the insights developed in this work. It is a well-known fact that the efficient heat distribution determines the failure current. Better the distribution of the bipolar turn-on, better the uniformity and hence lesser the current crowding. Keeping this concept in mind, following are the engineering schemes that are implemented for TVS diode.

Design-I: Cathode Length

Cathode lateral length is one such parameter that allows us to decrease the lateral current crowding.



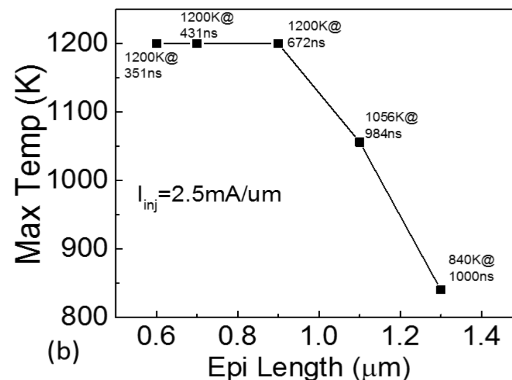


Figure 14: Maximum lattice temperature for (a) different Cathode lengths varied. After 0.5μm the temperature falls down to 551K for 1000ns. (b) Variation in Epi-lengths shows impact on Max temperature. Simulations were carried out for injected current of 2.5mA/μm.

Authorized licensed use limited to: J.R.D. Tata Memorial Library Indian Institute of Science Bengaluru. Downloaded on November 25,2022 at 10:13:26 UTC from IEEE Xplore. Restrictions apply.

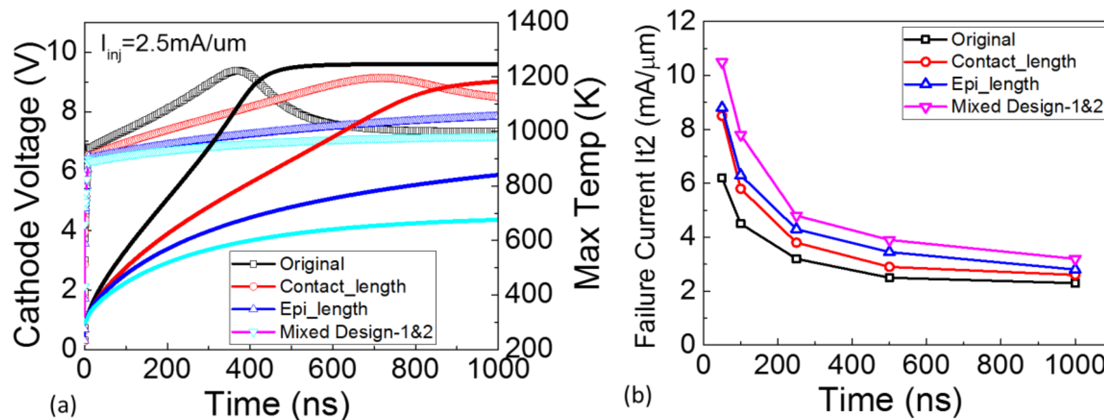


Figure 15: Cathode voltage and maximum lattice temperature (b) Failure current as a function of pulse duration for original device and engineered device namely contact length, Epi length, and mixed design, extracted from injected current of 2.5mA/µm.

Fig. 14(a) shows device maximum lattice temperature when subjected to injected current of 2.5mA/µm for different varied cathode lengths. It can be observed that as the cathode length increases, the maximum lattice temperature falls down to 551K, even at longer pulse duration of 1000ns. On the other hand, with lower cathode length, the injected current rises with the temperature of 1200K at a time instant of 430ns. This is attributed to the fact that increasing the cathode length boosts up the relaxation in the current crowding and this helps to withstand more heat within the device and the device survives for longer duration.

Design-II: Epi Length

P- Epi layer corresponds to lightly doped base region of an open base BJT. As soon as the current crowding occurs, the bipolar turn-on is the immediate result of the non-uniformity in impact ionization. The idea here is to weaken the efficiency of the bipolar triggering in order to diminish this non-uniformity. Fig. 14(b) illustrates the max. lattice temperature as a function of the Epi-length. Similar to cathode length variation, a trend can be observed where the temperature falls down with the increase in the Epi-length. The temperature does not reach 1200K at even 1000ns of pulse duration. The reason being that increasing of the base length helps in weakening the bipolar turn-on. In other words, the bipolar turn-on is distributed throughout the epi-region rather than concentrating at the lateral junctions.

Design-III: Mixed Design

Mixed Design includes increasing both the lateral lengths of the cathode and the epi layer. This approach helps in achieving the uniform turn-on. Fig. 15(a) shows the cathode voltage and the max. lattice temperature as a function of time. The case is taken once again considering the worst-case scenario for injected current of 2.5mA/um for the pulse duration of 100ns. As Fig. 16 depicts, engineering schemes help in delaying the turn-on and the voltage overshoot. This directs to the observation that the max. temperature is lesser for engineered schemes comparing to original device, and much relaxed in the case of mixed-design. This can be appreciated from Fig. 15(b). The mixed design has the highest It2 in all the time scale. Fig. 16 depicts the current density and the electron density contours, comparing it with the original device. It is clearly visible that mixed design approach helps in achieving the uniform bipolar turn-on. The current density is much more distributed unlike the original case. Furthermore, the e-density shows how well the lateral bipolar is weakened and the vertical bipolar becomes stronger. These engineering schemes do not completely solve the problem of low current ESD failures at long pulse duration, but reduces the severity by spreading the heat even at 1000ns.

V. Conclusion

Punch-through TVS diodes were found to be suffering from low current failures under longer pulses. 2D and 3D simulations depict the clear picture and help in understanding the failure

Authorized licensed use limited to: J.R.D. Tata Memorial Library Indian Institute of Science Bengaluru. Downloaded on November 25,2022 at 10:13:26 UTC from IEEE Xplore. Restrictions apply.

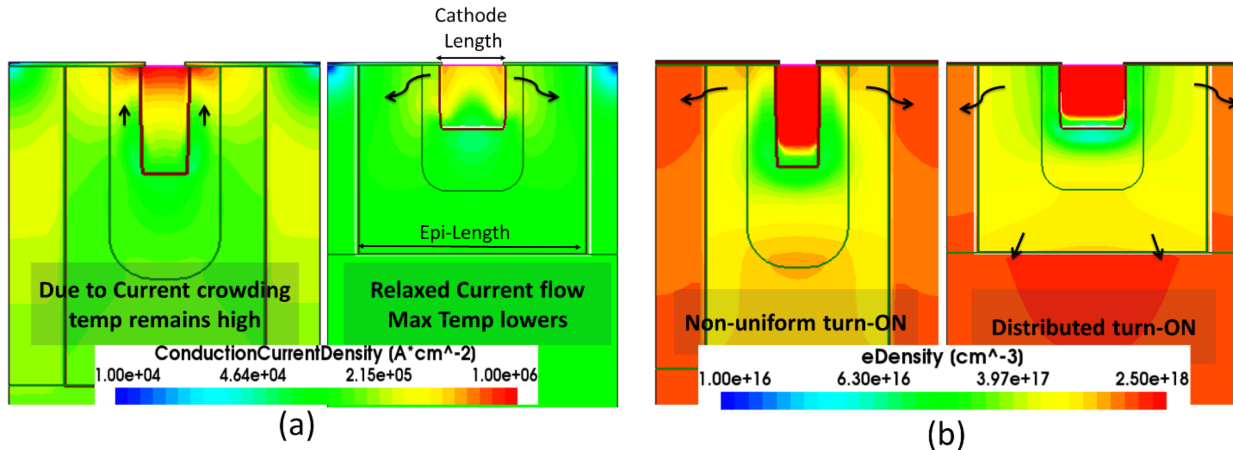


Figure 16: Contours of (a) conduction current density (b) electron density. The contours show the difference between the original device and the modified device design in having uniform and relaxed current conduction spread laterally and vertically.

dynamics. 2D simulation explains the reason behind the failure mechanism as current crowding, while 3D simulation supports the reason with thermally driven instability. The behavior was found to be dominated by the early lateral n-p-n turn on leading to non-uniformity in conduction, resulting in an early failure when the device is stressed for longer pulse duration. Engineering schemes like increasing the cathode length and epi-length are proposed which reduce the severity of low current ESD failures under long pulse duration stress.

VI. References

[1] J. Urresti., et al. “Low voltage TVS devices: design and fabrication." Proceedings. International semiconductor Conference. IEEE, Vol. 2. pp. 257-260

[2] King, Ya-Chin, et al. "Punch-through diode as the transient voltage suppressor for low-voltage electronics." IEEE Transactions on Electron Devices 43.11 (1996): 2037-2040.

[3] Yu, Bin, et al. "Punch-through transient voltage suppressor for EOS/ESD protection of low-voltage IC's." Electrical Overstress/Electrostatic Discharge Symposium Proceedings. IEEE, 1995.

[4] Urresti, Jesús, et al. "Lateral punch-through TVS devices for on-chip protection in low-voltage applications." Microelectronics Reliability 45.7-8 (2005): 1181-1186.

[5] Bobde, Madhur, et al. "A novel ESD super-clamp structure for TVS applications." 2008 Twenty-Third Annual IEEE Applied Power Electronics Conference and Exposition. IEEE, 2008.

[6] Shrivastava, Mayank, et al. "Part I: On the behavior of STI-type DeNMOS device under ESD conditions." IEEE transactions on electron devices 57.9 (2010): 2235-2242.

[7] Shrivastava, Mayank, et al. "Highly resistive body STI NDeMOS: An optimized DeMOS device to achieve moving current filaments for robust ESD protection." 2009 IEEE International Reliability Physics Symposium. IEEE, 2009.

[8] Shrivastava, Mayank, et al. "Part II: On the three-dimensional filamentation and failure modeling of STI type DeNMOS device under various ESD conditions." IEEE transactions on electron devices 57.9 (2010): 2243-2250.

[9] Paul, Milova, et al. "Physics of current filamentation in ggNMOS devices under ESD condition revisited." IEEE Transactions on Electron Devices 65.7 (2018): 2981-2989.

[10] Pogány, Dionyz, et al. "Thermally-driven motion of current filaments in ESD protection devices." Solid-state electronics 49.3 (2005): 421-42

Authorized licensed use limited to: J.R.D. Tata Memorial Library Indian Institute of Science Bengaluru. Downloaded on November 25,2022 at 10:13:26 UTC from IEEE Xplore. Restrictions apply.