Circuits and Systems, 2016, 7, 2286-2295
Published Online July 2016 in SciRes. http://www.scirp.org/journal/cs
http://dx.doi.org/10.4236/cs.2016.79199

On a Parasitic Bipolar Transistor Action in a Diode ESD Protection Device

Jin Young Choi

Department of Electronic & Electrical Engineering, Hongik University, Sejong, Korea
Email: jychoi@hongik.ac.kr

Received 19 April 2016; accepted 10 May 2016; published 19 July 2016

Abstract

In this work, we show that an excessive lattice heating problem can occur in the diode electrostatic discharge (ESD) protection device connected to a V_{DD} bus in the popular diode input protection scheme, which is favorably used in CMOS RF ICs. To figure out the reason for the excessive lattice heating, we construct an equivalent circuit for input human-body model (HBM) test environment of a CMOS chip equipped with the diode protection circuit, and execute mixed-mode transient simulations utilizing a 2-D device simulator. We analyze the simulation results in detail to show out that a parasitic pnp bipolar transistor action relating nearby p^+-substrate contacts is responsible for the excessive lattice heating in the diode protection device, which has never been focused before anywhere.

Keywords

ESD Protection, Diode Protection Device, Bipolar Transistor, Mixed-Mode Simulation, RF IC

1. Introduction

CMOS chips are vulnerable to electrostatic discharge (ESD) due to the thin gate oxides used, and therefore protection devices are required at input pads. A large size for the protection devices is needed to reduce discharge current density and thereby to protect them against thermal-related problems. However, adopting a large size tends to increase parasitic capacitances added to the input nodes generating other problems such as gain reduction and poor noise characteristics in RF ICs [1].

To reduce the added parasitics, various techniques have been suggested. However, basic approaches should be to reduce size of protection devices connected to input nodes, and protection schemes utilizing diode or thyristor protection devices were suggested [2] [3]. Among them, the diode input protection scheme is most popular in RF ICs [4]-[6].

How to cite this paper: Choi, J.Y. (2016) On a Parasitic Bipolar Transistor Action in a Diode ESD Protection Device. *Circuits and Systems*, **7**, 2286-2295. http://dx.doi.org/10.4236/cs.2016.79199

In the diode input protection scheme, it is essential to include a V_{DD}-V_{SS} clamp NMOS device in an input pad structure to provide discharge paths for all possible human-body model (HBM) test modes. Although a large size for the clamp NMOS device is essential to prevent thermal device failure, it does not add parasitics to an input node since it is not directly connected to it. A pn diode structure with a small holding voltage is utilized as the diode protection device connected to an input pad.

In RF ICs, to reduce parasitics added to input nodes, it is also important to minimize the diode size by optimizing device structure. Some efforts have been made to optimize diode structures [6], but the efforts for optimization seem to be insufficient.

In [6], geometric parameters of CMOS shallow-trench-isolation (STI) diodes were investigated to optimize diode structures for high-speed RF applications based on 2-D simulations. The parameters for optimization included diode length, anode & cathode width, anode-to-cathode spacing, and STI depth to suggest guidelines in designing the diode protection device structure. In this study, however, the authors made a mistake not to include nearby p-type substrate contacts in the target device structure. Our study shows that device heating can be most severe at these missing contacts, which has never been focused before anywhere.

In this work, based on 2-D mixed-mode (device and circuit) simulation, we figure out that the mechanism leading to diode protection device failure is mostly related to a trigger of parasitic pnp bipolar transistor involving the p-type substrate contacts.

A 2-D device simulator, together with a circuit simulator, is utilized as a tool for the analysis. The analysis methodology utilizing a device simulator has been widely adopted with credibility [7] [8] since it can provide valuable information relating the mechanisms leading to device failure, which may not be obtained by measurements.

In Section 2, we introduce the popular diode input protection scheme, which can be implemented into input pad structures of CMOS RF ICs to provide protection against HBM and MM (Machine mode) ESD events, and briefly explain discharge modes in HBM tests. We show conventional NMOS and diode protection device structures assuming usage of standard CMOS processes, and show device characteristics based on DC device simulations.

In Section 3, we construct an equivalent circuit model for CMOS chips equipped with the input protection devices to simulate various input HBM test situations, and execute mixed-mode transient simulations of circuits. Based on the simulation results, we explain the discharge characteristics in detail. By analyzing the simulation results, we show that severe lattice heating in the diode device can occur due to a trigger of parasitic pnp bipolar transistor action. Conclusions are made in Section 4.

2. Diode ESD Protection Scheme and Protection Device Structures

Figure 1 shows the popular diode input protection scheme, which minimizes added parasitics to input nodes in CMOS RF ICs. In **Figure 1**, a CMOS inverter is assumed as an input buffer.

As an example of a HBM ESD test, when a positive ESD voltage is applied to an input pin, which is connected to the pad in **Figure 1**, with a V_{SS} pin, which is connected to the V_{SS} bus, grounded, the forward-biased diode D_2 and the V_{DD}-V_{SS} clamp NMOS device M_1 in series forms a main discharge path.

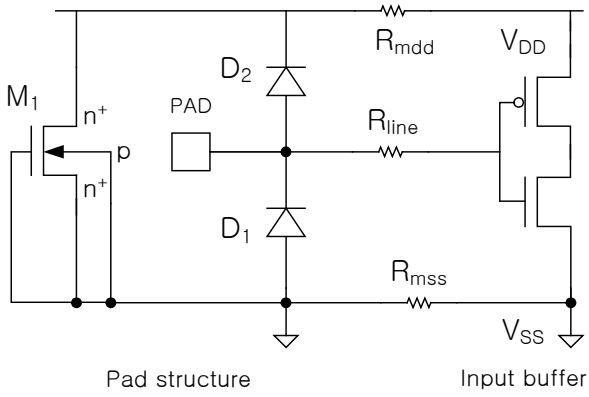

Figure 1. Diode input protection scheme.

Figure 2 shows the NMOS protection device structure assumed in this work, which is utilized as the V_{DD}-V_{SS} clamp device M_1 in the protection circuit in **Figure 1**. The structure represents a conventional protection device incorporating n^+ source and drain ESD implants, which is implied by the relatively deep junctions. In order to alleviate drain-contact melting problems caused by lattice heating, the gate-drain contact spacing is chosen as 3.5 μm, which can be regarded as a minimal. The principal structure parameters are same with those previously described in [9].

The p^+ junctions located at the upper left/right corners represent diffusions for substrate ground contacts. A series resistor of 1 MΩ·μm, which is not shown in **Figure 1**, is connected at the bottom substrate node considering distributed resistances reaching to substrate contacts located farther away.

DC simulations were performed using a 2-D device simulator ATLAS [10]. All necessary physical models including an impact ionization model were included in the simulations. The lattice-heating model included joule heat, generation-recombination heat, and Peltier-Thomson heat. The source, the gate, and the substrate are grounded, and the drain bias is varied for simulation.

From the DC simulation results, it was confirmed that the breakdown voltage, the snapback voltage, and the holding voltage are 9.3 V, 9.4 V, and 4.6 V, respectively. A 2nd breakdown [11] occurs when the drain current is about 1.3 mA/μm, which is the current value per 1μm of device width. More explanation for the DC characteristics are described in [9].

Figure 3 shows the reference diode structure assumed in this work. The p^+ anode and the n^+ cathode forms the diode. The p^+ anode and the n^+ cathode do not incorporate ESD implant steps, which is implied by the relatively shallow junctions, to reduce added parasitics to the input node.

The n well is assumed to have a Gaussian doping profile with 10^{17} cm^{-3} of peak concentration, and the doping profiles of the p^+ and n^+ junctions are similar to those of the p^+ junctions in **Figure 2**. The contact spacing between the n^+ cathode and the p^+ anode was chosen as 2.4 μm. A series resistor is also connected at the bottom substrate node as in the NMOS device.

We note that in the previous study in [9], we made a mistake not to consider nearby p-type substrate contacts, which can cause a severe lattice heating problem leading to device failure. In the reference diode structure, we added two p^+ substrate contacts (p^+ subl & p^+ subr shown in **Figure 4**) at the upper left/right corners in **Figure 3**.

Table 1 summarizes the principal structure parameters.

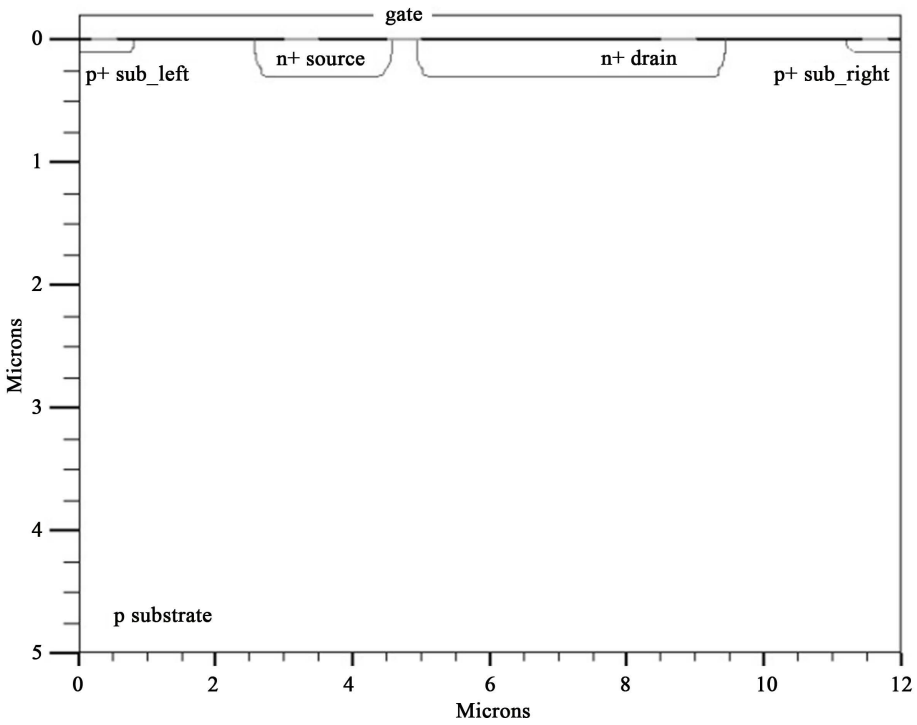

Figure 2. Cross section of the NMOS device.

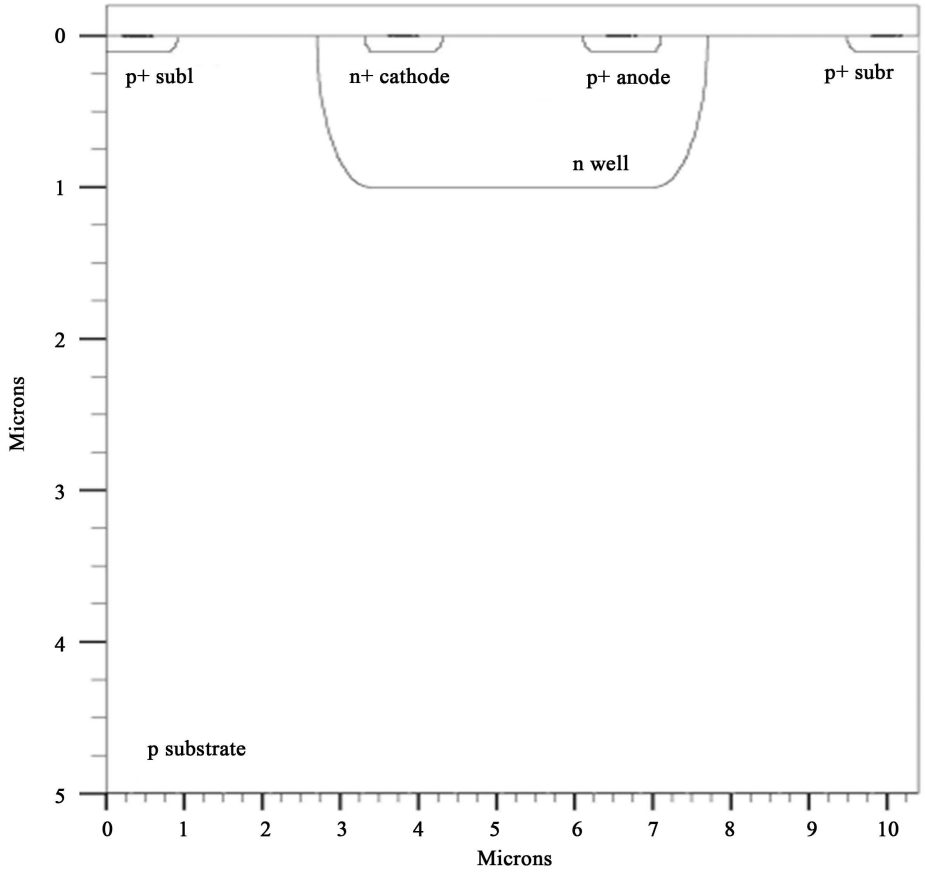

Figure 3. Cross section of the reference diode device.

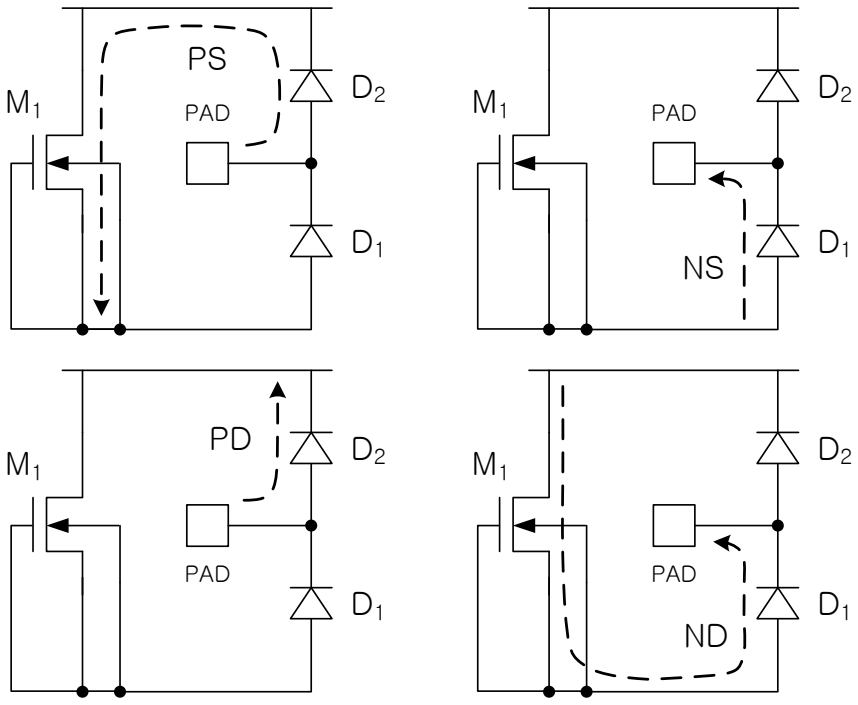

Figure 4. Main discharge paths for each test mode.

Table 1. Principal parameters of the diode protection device.

Parameter	Values
n well depth	1.0 μm
peak doping concentration in p^+ & n^+ junctions	1.2×10^{20} cm^{-3}
p^+ & n^+ junction depth/length	0.1 μm/0.8 μm
n^+ cathode-p^+ anode contact spacing	2.4 μm
n^+ cathode-p^+ anode junction spacing	1.8 μm
p^+ sub-n well spacing	1.9 μm

The p^+ substrate contacts and the p^+ anode were grounded and the n^+ cathode was biased positively or negatively to simulate DC reverse-bias or forward-bias characteristics, respectively.

From the DC simulation results, it was confirmed that the forward diode drop is 0.67 V when the diode current is 0.2 μA/μm, and the reverse breakdown voltage is about 11.3 V.

Table 2 summarizes the principal DC characteristics of the two protection devices.

Since HBM tests for input pins should include all possible discharge modes, tests are performed for 5 modes defined below.

1) PS mode: $+V_{ESD}$ at an input pin with a V_{SS} pin grounded
2) NS mode: $-V_{ESD}$ at an input pin with a V_{SS} pin grounded
3) PD mode: $+V_{ESD}$ at an input pin with a V_{DD} pin grounded
4) ND mode: $-V_{ESD}$ at an input pin with a V_{DD} pin grounded
5) PTP mode: $+V_{ESD}$ at one input pin with another input pin grounded

Main discharge paths for each test mode in the diode protection scheme are shown in **Figure 4** and **Figure 5**.

Figure 4 shows main discharge paths in the diode protection scheme. In a PS mode, the forward-biased D_2 and an npn bipolar transistor in M_1 in series provides a main discharge path, and in an NS mode, the forward-biased D_1 provides it. In a PD mode, the forward-biased D_2 provides a main discharge path, and in an ND mode, an npn bipolar transistor in M_1 and the forward-biased D_1 in series provides it.

Figure 5 shows a main discharge path for a PTP mode. As shown in **Figure 5**, two forward-biased pn diodes D_2 and D_3 in series with an npn bipolar transistor in M_2 provides a main discharge path.

Since local lattice heating is proportional to a product of current density and electric field intensity, temperature-related problems in the protection devices can occur in the NMOS device rather than in the diode device since the holding voltage of the npn bipolar transistor in the NMOS device is much larger than the diode drop in the forward-biased diode. Therefore width of the diode devices can be small, but we should assign sufficient device width for the NMOS devices considering PS, ND, and PTP modes.

3. Mixed-Mode Transient Simulations

Figure 6 shows an equivalent circuit of an input HBM test situation in case of a PS mode.

The portion indicated as 'Test environment' in **Figure 6** is an equivalent circuit for test equipment connection. M_1, D_1, and D_2 form a protection circuit at the input pad. A CMOS inverter is assumed as an input buffer inside a chip, which is modeled by a capacitive network. C_{ngate} and C_{pgate} represent gate-oxide capacitances of an NMOS transistor and a PMOS transistor, respectively. More explanation about all the circuit elements are described in detail in [9].

Using ATLAS, we performed mixed-mode transient simulations utilizing the equivalent circuit in **Figure 6**. We note that the NMOS device in **Figure 2** is used as the protection device M_1, and the diode device in **Figure 3** is used as the protection device D_2 and D_1.

When a mixed-mode simulation is performed, the active protection devices are solved by device and circuit simulators simultaneously. Notice that up to six active protection devices are included in a mixed-mode simulation, which correspond to the PTP-mode simulations in **Figure 5**.

For all the mixed-mode transient simulations performed for each test mode, V_{ESD} = 2000 V was assumed. Widths of the protection devices were adjusted to maintain peak lattice temperature inside all the protection

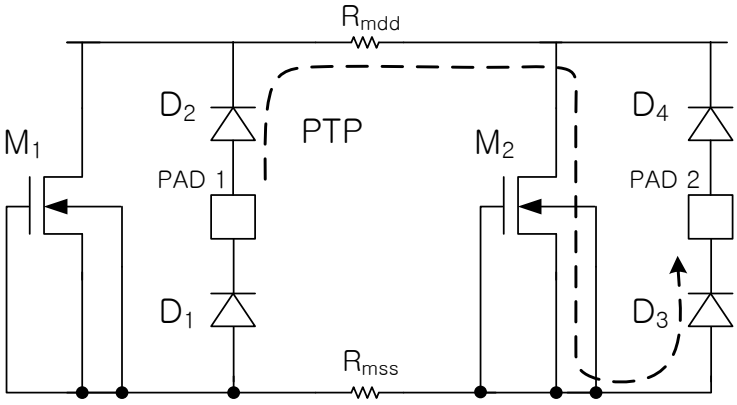

Figure 5. Main discharge path for a PTP mode.

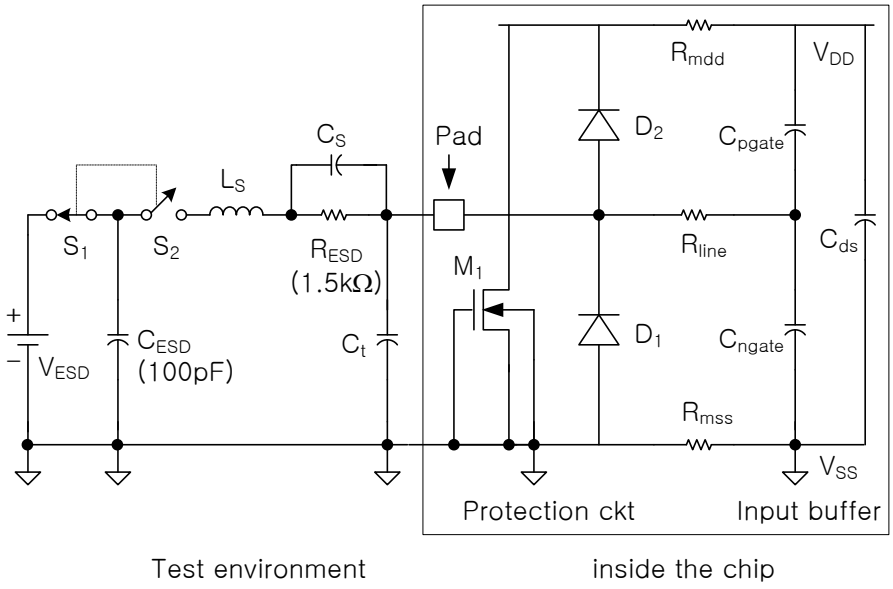

Figure 6. Equivalent circuit of an input-pin HBM test situation.

Table 2. Principal DC characteristics of the two protection devices.

Protection device	Holding voltage (V)	Breakdown voltage (V)	Snapback voltage (V)	Vbe (V) (2 μA/μm)
NMOS	4.6	9.3	9.4	
Diode		11.3		0.67

devices below 500˚K for all the mixed-mode simulations, resulting 250 μm, 15 μm, and 15 μm for the NMOS device, the diode device connected to a V_{DD} bus, and the diode device connected to a V_{SS} bus, respectively.

As an example of the mixed-mode simulation results, **Figure 7** shows variations of voltages developed on C_{ngate} and C_{pgate} in the input buffer as a function of time in a PS mode. We note that pad voltage, which is not shown in **Figure 7**, is almost same with the voltage developed on C_{ngate}.

Before explaining the result shown in **Figure 7**, we note that the simulated anode current of D_2 in **Figure 6**, which lies in the main discharge path, as a function of time is similar to that in [9]. The anode current peaks up to 1.37A with a rise time of about 7 ns, and shows decaying characteristics with a time constant of roughly $R_{ESD}C_{ESD}$ = 1.5 kΩ × 100 pF = 0.15 μs, which can be expected from the equivalent circuit in **Figure 6**.

By monitoring the simulation results in detail, we confirmed that the forward-biased diode in D_2 is triggered

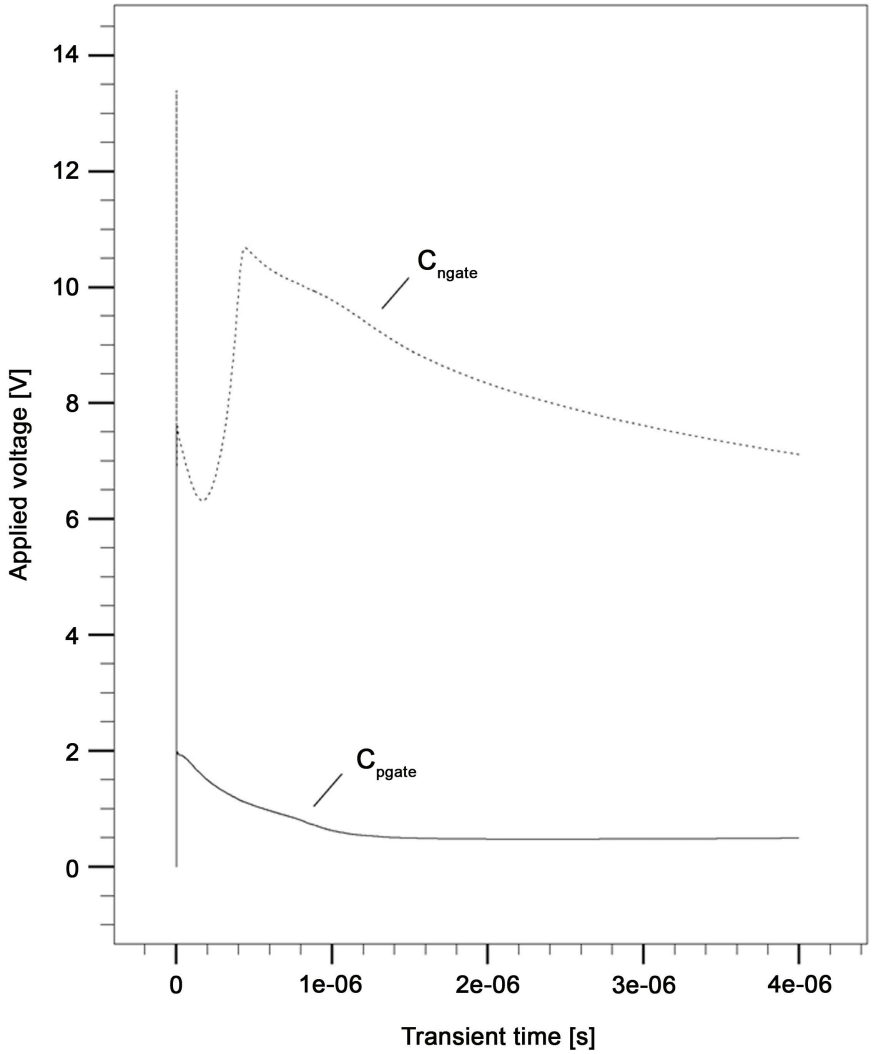

Figure 7. Variations of voltages on C_{ngate} and C_{pgate} in a PS mode.

when the pad voltage in the early stage of discharge increases to 13.4 V, which is 0.82 ns after S_2 in **Figure 6** is closed. At this moment, the voltage developed across D_2 in **Figure 7** corresponds to 7.6 V. We also confirmed that the parasitic bipolar transistor in M_1 is triggered when the pad voltage in the early stage of discharge increases to 11.1 V, which is 0.75 ns after S_2 is closed. Due to this trigger time difference, the voltage (13.4 V) smaller than the sum of the bipolar breakdown voltage (9.4 V) and the diode trigger voltage (7.6 V) appears across C_{ngate} in the early stage of discharge in **Figure 7**.

After that, main discharge through the forward-biased diode in D_2 and the parasitic npn bipolar transistor in M_1 in series proceeds when the pad voltage drops to a sum of the forward diode drop and the bipolar holding voltage, which is about 7 V as shown in **Figure 7**.

We can see that the pad voltage increases again and reaches to 10.7 V at about 0.45 μs when M_1 conducts in a breakdown mode since the drain current of M_1 is reduced below the holding current for the bipolar transistor action. From the simulation result, it is confirmed that the peak voltage 10.7 V corresponds to a sum of the forward diode drop in D_2 (1.2 V) and the breakdown voltage of M_1 (9.5 V), which are somewhat different from the values in **Table 2** due to the difference in the current value and the lattice temperature.

In **Figure 7**, the voltage developed on C_{pgate} is maintained low all the time after the forward diode is triggered since it is equal to the forward diode drop in D_2.

We note that the variations of voltages developed on C_{ngate} and C_{pgate} in **Figure 7** are very similar to those

previously presented in Figure 16 in [9], and the peak voltages across gate oxides in the early stage and later stage of discharge are very similar to those previously presented in **Table 3** and **Table 4** in [9].

Figure 8 shows variation of overall peak temperature inside the protection devices including M_1, D_1, and D_2 from the same simulation result. The overall peak temperature increases up to 893˚K at about 36 ns, and shows a second peak (655˚K) at 430 ns.

In the previous study in [9], we confirmed that the overall peak temperature increases up to 495˚K at about 30ns, when the bipolar transistor current in M_1 still dominates the discharge, and decreases slowly. We also confirmed that the overall peak temperature appears at the gate-side n^+ drain junction in M_1. We also confirmed that the peak temperature inside D_2, whose device width is set to 15 μm, increases up to only 485˚K at about 45 ns, and the peak temperature appears at the n^+-cathode junction.

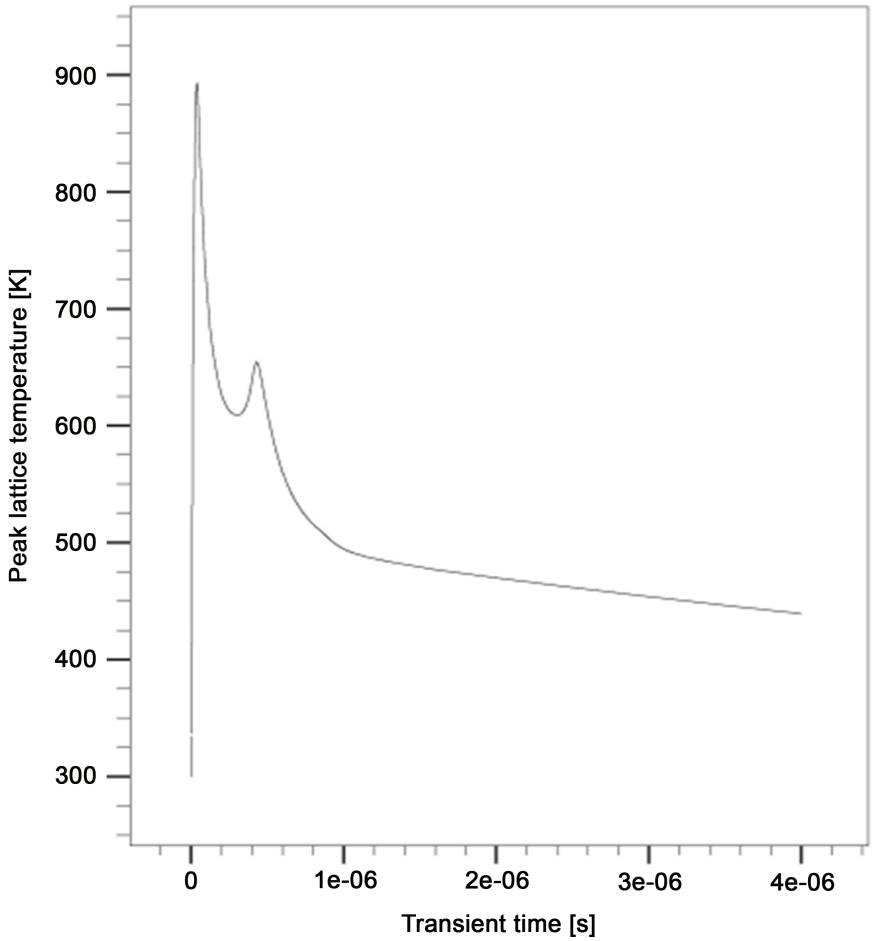

Figure 8. Overall peak temperature variation in a PS mode.

Table 3. Terminal voltages and currents of D_2 at 36 ns.

Terminal	Voltage (V)	Current (mA)	Percentage of Current (%)
p^+ anode	7.203	1070	100
n^+ cathode	5.287	970.6	90.71
p^+ sub_right	0	77.8	7.27
p^+ sub_left	0	21.5	2.01
M_1 drain	5.286	970.6	90.71

Table 4. Terminal voltages and currents of D_2 at 430 ns.

Terminal	Voltage (V)	Current (mA)	Percentage of Current (%)
p^+ anode	10.635	73.11	100
n^+ cathode	9.514	50.72	69.37
p^+ sub_right	0	15.73	21.52
p^+ sub_left	0	6.52	8.92
M_1 drain	9.514	50.72	69.37

However in this new simulation with the modified diode device structure in **Figure 3**, we confirmed that the peak temperature inside D_2 increases up to 893˚K at about 36 ns, and it peaks again up to 655˚K at 430 ns, and both peak temperatures appear at the p^+-subr junction, which is a result quite different from that in the previous study. We also confirmed that peak temperature inside M_1 increases up to 468˚K at about 37 ns, which is not much different from that in the previous study.

Notice that only the difference made in this simulation is that we added two upper left/right corner p^+-substrate contacts in the diode device considering nearby p-substrate contacts.

To figure out the reason for lattice heating at the p^+-subr junction, we first monitored terminal currents of D_2 at 36 ns when the 1st temperature peaking occurs. Terminal voltage and current values at 36 ns are summarized in **Table 3**. The drain current of M_1 is also given for explanation. Percentage of each current regarding the p^+-anode current are also given in **Table 3**.

We note here again that the forward-biased D_2 and the npn bipolar transistor in M_1 in series provides a main discharge path in a PS mode.

We can see that 90.71% of the p^+-anode current flows to the drain of M_1 through the n^+ cathode. The rest (9.28%) of the p^+-anode current flows to the p^+-subr and p^+-subl junctions, which must be related to lateral pnp bipolar actions. The p^+-anode, the n^+-cathode (n-well), and the p^+-subr contacts act as an emitter, a base, and a collector, respectively. Notice that the p^+-anode/n^+-cathode junction is forward-biased by 1.916 V, and the n^+-cathode/p^+-subr junction is reverse-biased by 5.287 V based on the data shown in **Table 3**. Severe joule heating at the p^+-subr junction is possible since the current density is high and the electric field is high with the large reverse bias.

We can see that the lateral pnp bipolar action occurs more easily to the right-hand side since the p^+-subr junction is closer to the p^+-anode junction. 3.6 (77.8 mA/21.5 mA) times larger pnp bipolar current flows to the p^+-subr junction compared to that to the p^+-subl junction.

We also confirmed that the 2nd hottest spot in D_2 at this moment is the n^+-cathode junction, whose temperature increases up to about 460˚K.

To figure out the reason for the 2nd temperature peaking at 430 ns, we summarize again terminal voltage and current values of D_2 at 430 ns in **Table 4**.

We can see that the lateral pnp bipolar action is still a reason for the lattice heating at 430 ns from the data given in **Table 4**. 69.37% of the p^+-anode current flows to the drain of M_1 through the n^+ cathode. The rest (30.44%) of the p^+-anode current flows to the p^+-subr and p^+-subl junctions. 2.4 (15.73 mA and 6.52 mA) times larger pnp bipolar current flows to the p^+-subr junction compared to that to the p^+-subl junction.

Compared to the 1st temperature peaking situation, the current levels are lower with the decreased discharging current (73.11 mA vs. 1.07 A), but the reverse bias is a lot larger, resulting the 2nd peaking at the p^+-subr junction. We also confirmed that the 2nd hottest spot in D_2 at this moment is the p^+-subl junction, whose temperature increase up to about 490˚K.

From the analysis above, we can conclude that the excessive lattice heating at the p^+-subr junction in D_2 is caused by a trigger of parasitic pnp bipolar transistor action. Notice that in a PS mode simulation, the p^+-subr and p^+-subl contacts are grounded, the n well is connected to the V_{DD} bus, and the p^+ anode is connected to the pad, where a positive ESD voltage is applied.

The main discharge current flows through the p^+-anode/n^+-cathode forward biased diode in D_2, whose voltage drop is small. At the same time, the pnp bipolar current flows from the p^+ anode to the grounded p^+-subr junction via the n well, whose potential gets high by the applied ESD voltage to the input pad. The current density at the

reverse-biased p^+-subr/n-well junction is large enough to result in the severe temperature peaking by joule heating.

The pnp bipolar current also flows from the p^+ anode to the p^+-subl junction via the n well. However p^+-subl junction is located farther from the p^+ anode when compared to the p^+-subr junction, and therefore the pnp bipolar action reaching to the p^+-subl junction is much weaker.

We also confirmed that the lattice heating problem in D_2 is severer in a PTP mode test. Lattice temperature at the p^+-subr junction in D_2 peaks up to 1500˚K in the PTP mode simulation, which is the maximum temperature limit set in this simulation set. The PTP simulation stops at 5.2 ns due to temperature limit exceeding.

We note here that the lattice heating problem caused by the parasitic pnp bipolar action relating the p^+- sub contacts in the diode protection device has never been focused in prior studies.

4. Conclusions

We encountered with an excessive lattice heating problem in the diode ESD protection device connected to a V_{DD} bus in the popular diode input protection scheme, which is favorably used in CMOS RF ICs.

To figure out the reason for the excessive lattice heating, we constructed an equivalent circuit for input HBM test environment of a CMOS chip equipped with the diode protection circuit, and executed mixed-mode transient simulations utilizing a 2-D device simulator. We analyzed the simulation results in detail to figure out that a parasitic pnp bipolar transistor action relating nearby p^+-substrate contacts is responsible for the severe lattice heating in the diode protection device, which has never been focused before anywhere.

Acknowledgements

This work was supported by 2014 Hongik University Research Fund, and in part by the International Science and Business Belt Program through the Ministry of Science, ICT and Future Planning (2015K000270).

References

[1] Leroux, P. and Steyaert, M. (2001) High-Performance 5.2GHz LNA with On-Chip Inductor to Provide ESD Protection. *Electronics Letters*, **37**, 467-469. http://dx.doi.org/10.1049/el:20010271

[2] Chatterjee, A. and Polgreen, T. (1991) A Low-Voltage Triggering SCR for On-Chip ESD Protection at Output and Input Pads. *IEEE Electron Device Letters*, **12**, 21-22. http://dx.doi.org/10.1109/55.75685

[3] Worley, E.R., Gupta, R., Jones, B., Kjar, R., Nguyen, C. and Tennyson, M. (1995) Sub-Micron Chip ESD Protection Schemes Which Avoid Avalanching Junctions. *Electrical Overstress/Electrostatic Discharge Symposium Proceedings*, Phoenix, 12-14 September 1995, 13-20. http://dx.doi.org/10.1109/eosesd.1995.478263

[4] Yeh, C.-T., Ker, M.-D. and Liang, Y.C. (2010) Optimization of Layout Style of ESD Protection Diode for Radio-Frequency Front-End and High-Speed I/O Interface Circuits. *IEEE Transactions on Device and Materials Reliability*, **10**, 238-246. http://dx.doi.org/10.1109/TDMR.2010.2043433

[5] Yang, M.-T., Du, Y., Teng, C., Chang, T., Worley, E., Liao, K., Yau, Y.-W. and Yeap, G. (2010) BSIM4-Based Lateral Diode Model for LNA Co-Designed with ESD Protection Circuit. 11*th International Symposium on Quality Electronic Design* (*ISQED*), San Jose, 22-24 March 2010, 87-91.

[6] Au, T. and Syrzycki, M. (2013) Investigation of STI Diodes as Electrostatic Discharge (ESD) Protection Devices in Deep Submicron (DSM) CMOS Process. 26*th Annual IEEE Canadian Conference on Electrical and Computer Engineering* (*CCECE*), Regina, SK, 5-8 May 2013, 1-5.

[7] Feng, H., Chen, G., Zhan, R., Wu, Q., Guan, X., Xie, H. and Wang, A.Z.H. (2003) A Mixed-Mode ESD Protection Circuit Simulation-Design Methodology. *IEEE Journal of Solid-State Circuits*, **38**, 995-1006. http://dx.doi.org/10.1109/JSSC.2003.811978

[8] Fankhauser, B. and Deutschmann, B. (2004) Using Device Simulations to Optimize ESD Protection Circuits. *International Symposium on Electromagnetic Compatibility*, 9-13 August 2004, Vol. 3, 963-968. http://dx.doi.org/10.1109/isemc.2004.1349956

[9] Choi, J.Y. (2010) A Comparison Study of Input ESD Protection Schemes Utilizing NMOS, Thyristor, and Diode. *Communications and Network*, **2**, 11-25. http://dx.doi.org/10.4236/cn.2010.21002

[10] ATLAS II Framework, Version 5.10.2.R, Silvaco International, 2005.

[11] Amerasekera, A., van Roozendaal, L., Bruines, J. and Kuper, F. (1991) Characterization and Modeling of Second Breakdown in nMOST's for Extraction and ESD-Related Process and Design Parameters. *IEEE Transactions on Electron Devices*, **38**, 2161-2168. http://dx.doi.org/10.1109/16.83744