Thermal Breakdown Modeling and Simulation of GGNMOS under ESD Stress

Yao Liu, Yingjun Gao
College of Physical Science and Engineering Technology
Guangxi University
Nanning, China
E-mail: malena326@126.com

***Abstract*—Based on the thermal breakdown behavior of grounded-gated n-channel metal-oxide-semiconductor *(GGNMOS)* under Electrostatic Discharge*(ESD)* conditions, the electrothermal models are built and optimized by modeling thermal breakdown temperature, heat source and temperature-dependent parameters, which are coupled with the electrical characteristics of GGNMOS. Thus, the thermal breakdown current I_{t2}, which is also the failure threshold of GGNMOS, can be simulated by the models. A comparison between the transmission line pulsing *(TLP)* testing results and simulation results shows that the maximum error value of I_{t2} is only 8.95%, which confirms the accuracy of the models. The models can be provided for the thermal breakdown simulation of GGNMOS with various process parameters and different ESD pulse duration owing to the physical-level modeling. The studies have great significance for the reliability design of microelectronic device.**

Keywords- ESD; GGNMOS; Electrothermal effects; Modeling; Reliability design

I. INTRODUCTION

As one of the most widely used ESD (Electrostatic Discharge) protection device, silicon-substrate GGNMOS (grounded-gated n-channel metal-oxide-semiconductor) discharges the high ESD current by turning on the parasitic lateral bipolar transistor (LNPN), which protects the innel circuits from the high current damage. However, as the current increases, thermal breakdown(alse called second breakdown) can occur in GGNMOS, then the device will be melt and thermal failure is incurred. Hence, I_{t2} is a measure of the maximum high current capability of GGNMOS, which is commonly used as the figure of merit for process-dependent ESD capability. The current simulation tools such as SPICE, MEDICI, IETSIM are all not special for ESD protection devices, they can not provide the complete models on thermal breakdown operation region[1][2][3][4]. In this case, based on electrothermal effects, a novel approach of modeling thermal breakdown special for GGNMOS under ESD stress is discussed in this paper.(The models before thermal breakdown were presented in our previous paper[5].)

II. PHYSICAL-LEVEL MODELING

The relationship between the drain voltage V_D and the drain current I_D of GGNMOS is shown in Fig. 1, where V_{t1} is the turn-on voltage of the LNPN , V_{sp} is the holding

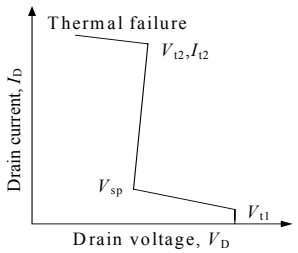

Figure 1. Typical I-V curve of GGNMOS.

voltage, and (V_{t2}, I_{t2}) are the thermal breakdown voltage and current[6]. Once the LNPN turns on, V_D decreases rapidly to a minimum voltage V_{sp}. As the drain current continue to increase, the power dissipation at the reverse-biased collector-base junction of LNPN, which can be approximated by $V_D \times I_D$, causes the temperature in the device to rise. Then the thermal current increases due to the rising temperature, which can result in thermal breakdown if the heat can not be dissipated quickly[7]. Once thermal breakdown is initiated, the transistor enters a unstable negative resistance region, current filamentation and local hot spots begin to occur and the temperature in these regions very quickly reaches the semiconductor melt temperature and permanent damage is incurred. The above analysis indicate that thermal breakdown modeling can be based on the breakdown temperature and heat source models firstly.

A. Thermal breakdown temperature model

The conditions for thermal breakdown are defined by the internal temperature of the device at which the voltage begins to collapse. Accurate measurements of the temperature rise during the ESD event are still not possible due to technical restrictions, so the breakdown temperature can be obtained by modeling. Based on the analysis of the npn transistor, the breakdown is triggered when the collector-base junction of LNPN reaches its intrinsic temperature T_i. T_i is defined as the temperature at which the intrinsic carrier concentration n_i is equal to the background doping concentration given by N_d for n-type material[8]. n_i is given by Selberher [9], coupled with the dependence of E-quality m_0 and energy gap E_g on T[9], the thermal breakdown temperature model for GGNMOS can be approximatively given by

$$3.88\times10^{16}T_i^{1.5}\exp(-7.0\times10^3/T_i)=n_i=N_{BC} \tag{1}$$

where N_{BC} is the substrate doping concentration, for GGNMOS, $N_d = N_{BC}$. The simulation results show that the breakdown temperature T_i varis from 460K to 1000K when N_{BC} is in the usual range $10^{14}\sim10^{18}$cm^{-3}.

B. Heat source model

An analytical approach to heat source modeling starts with the solution of the heat conduction equation, which is given by[10]

$$\frac{\partial T}{\partial t}-D\cdot\nabla^2(T)=\frac{P(t)}{\rho\cdot C_P} \tag{2}$$

where ρ is the density of the semiconductor, C_p is the specific heat, $D=K/(\rho C_p)$ is the thermal diffusivity, K is the thermal conductivity, $P(t)$ is the power dissipation of the heat source.

It's very complicated to deduce the heat source model from (2), therefore, two assumptions to the model are proposed: *a)* Assuming the heat source as a ideal rectangular; *b)* Assuming that the heat is dissipated in a semi-infinite medium, because the thermal diffusion length $(DT)^{0.5}$ is on the order of 10μm for a pulse duration of 1μs, which is much smaller than the distance between the protection circuit and the edge of the chip. With these assumptions, for a rectangular heat source $-a/2<x<a/2$, $-b/2<y<b/2$ and $0<z<c/2$, using the Green's function method, the time t and temperature T dependence of the heat source model is given by

$$\begin{cases} P(t)=\dfrac{ab\sqrt{\pi K\rho C_p}\cdot(T-T_0)}{\sqrt{t}-\sqrt{t_c/2}}, & \text{if } 0<t<t_c \\ P(t)=\dfrac{4\pi Ka(T-T_0)}{\ln(t/t_b)-2-c/b}, & \text{if } t_b<t<t_a \\ P(t)=\dfrac{4\pi Ka(T-T_0)}{\ln(t/t_b)-2-c/b}, & \text{if } t_b<t<t_a \\ P(t)=\dfrac{2\pi Ka(T-T_0)}{\ln(a/b)+2-c/(2b)-\sqrt{t_a/t}}, & \text{if } t_a<t \end{cases} \tag{3}$$

where T_0=300K is the room-temperature, the diffusion times t_a, t_b and t_c are defined as

$$t_a=a^2/(4\pi D),\quad t_b=b^2/(4\pi D),\quad t_c=c^2/(4\pi D) \tag{4}$$

C. Electrothermal models

The above analysis indicates that the heat source model has four time-dependent regions, while the dividing time point t_a, t_b and t_c are related to a, b, c and D. Hence, electrothermal modeling should start with the solution of the heat source dimensions of GGNMOS under ESD stress.

Fig. 2 shows the three-dimensional and sectional views of a single finger MOS device. For an MOS device, the heat source dimensions can be approximated by

$$\begin{cases} a=W, c=x_j \\ b=b_0\ln\dfrac{V_{sp}\mu_{eff}}{b_0 v_{sat}}, \text{where } b_0=\sqrt{\dfrac{\varepsilon_{si}t_{ox}x_j}{\varepsilon_{ox}}} \end{cases} \tag{5}$$

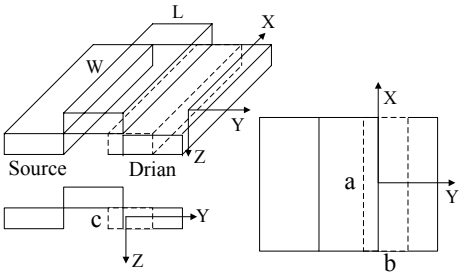

Figure 2. Three-dimensional and sectional views of a single finger NMOSFET.

Where W is the gate width, t_{ox} is the gate oxide thickness, x_j is the junction depth, v_{sat} is the electron saturation velocity, μ_{eff} is the effective carrier mobility and ε_{si}, ε_{ox} are the dielectric constants of silicon and silicon dioxide.

The drain voltage is nearly unchanged after decreasing to V_{sp}, so $V_{t2}\approx V_{sp}$, and I_{t2} can be given by the equation

$$P(T_i)=I_{t2}V_{sp} \tag{6}$$

Where $P(T_i)$ is the power dissipation on breakdown point. Hence, the electrothermal models can be obtained by combinations of (1) and (3)–(6), where the time t dependence of the power dissipation is determined by the ESD pulse duration t_i.

D. Parameters models

The unknown parameters in the above electrothermal models are μ_{eff}, v_{sat}, D and K, which are all temperature dependent, so the parameters modeling should be based on the physical-level relationship.

1) μ_{eff} is defined to describe the mobility degradation due to high fields under ESD stress, which can be modeled by[8]

$$\mu_{eff}=\frac{\mu_0}{[1+(E/E_{sat})^n]^{1/n}} \tag{7}$$

Where E_{sat} is the saturation electric field ($E_{sat}\approx10^4$V/cm); n is a fixed constant ($n=2$ for electrons), μ_0 is the zero-bias mobility. Considering the temperature effects, μ_0 can be described by Caughey-Thomas empirical mobility model[9] for electrons in drain

$$\mu_0=\mu^{\min}+\frac{\mu^L-\mu^{\min}}{1+[(T/300)^{-3.8}/N_E/N_{ref}]^{\beta}} \tag{8}$$

Where μ^L is the lattice mobility, μ^{min} is the coefficient for ionized impurity scattering, N_E is the drain impurity concentration(collector doping concentration for LNPN), N_{ref} is a constant and β is an empirically determined constant. Considering the typical value of $N_E=10^{18}$cm^{-3}, the parameters in (8) can be obtained as $\mu^{min}=55.24$ cm^2/V·s, $\mu^L=1448\times(T/300)^{-2.33}$, and $\beta=0.733$.

2) The temperature-dependent v_{sat} is given by

$$v_{sat}=\frac{2.4\times10^7}{1+0.8\exp(T/600)} \tag{9}$$

3) For a rectangular box with volume $\Delta=abc$, the thermal conductivity K is given by

$$K = 1.5486(T/300)^{-4/3} \tag{10}$$

4) For the above rectangular heat source, $\rho C_p=1.574\times(T/300)^{0.1}$, combined with $D=K/\rho C_p$ and (10), the thermal diffusivity D can be written as

$$D = 0.984(T/300)^{-1.433} \tag{11}$$

E. Optimization of the electrothermal models

It's necessary to reduce the complexity of the electrothermal models, an optimized approach is to simplify (3) by determining which time region the various ESD pulse duration t_i belong to.

For NMOS, there is always $a>>b>c$, so the ranges of t_b and b should be determined firstly. Equation (5) shows that b is a function of v_{sat}、V_{sp}、μ_{eff}、x_j and t_{ox}, while v_{sat} and μ_{eff} are temperature dependent; t_b is a function of b and D. Hence, the dependence of μ_{eff}/v_{sat} and D on breakdown temperature T_i can be simulated by combinations of (7) ,(8) ,(9) and (11), as shown in Fig. 3 and Fig. 4. We can see that in the range $460 < T_i < 1000$, μ_{eff}/v_{sat} varis from 0.18×10^{-5}cm/V to 2.75×10^{-5}cm/V on high fields($E=10^4\sim10^5$V/cm) when N_E is 10^{18}cm^{-3} , and D varis from 0.175 cm^2/s to 0.533cm^2/s.

Table I gives the typical value ranges of x_j and t_{ox} and the empirical value range of V_{sp} for usual shallow-junction GGNMOS. Inputting the value range of these parameters and the simulated μ_{eff}/v_{sat} and D into (4) and (5), the value ranges of t_b and b can be simulated as shown in Table I, which indicate that the dividing time point t_b is much smaller than the ESD pulse duration t_i(50～200ns). Hence, even N_E is not 10^{18}cm^{-3} and the other input parameters are out of the typical value ranges, Equation (3) can be limited on the region $t_b<t_i$ for GGNMOS.

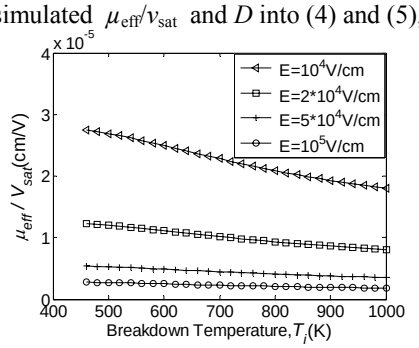

Figure 3. Dependence of μ_{eff}/v_{sat}(the ratio of effective carrier mobility and electron saturation velocity) on the thermal breakdown temperature.

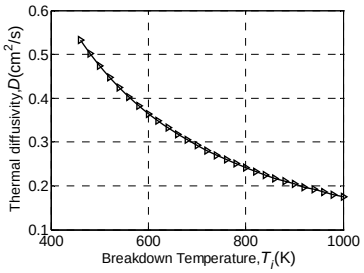

Figure 4. Dependence of D on the thermal breakdown temperature.

The combinations of (4), (5) and (11) can simulate the dependence of t_a on W for different breakdown temperatures as shown in Fig. 5.

Fig. 5 shows that for a constant gate width W, the higher T_i is, the larger t_a is; for a constant T_i, t_a increases as W increases. For various W and T_i, t_a may be or not be in the range 50～200ns of t_i. Using these relationships, (3) can be simplified as

$$\begin{cases} P(t)=\dfrac{4\pi Ka(T-300_0)}{\ln(t_i/t_b)-2-c/b} & \text{if } W^2(\dfrac{T_i}{300})^{1.433}\ge 12.36\cdot t_i \\ P(t)=\dfrac{2\pi Ka(T-300)}{\ln(a/b)+2-c/(2b)-\sqrt{t_a/t_i}} & \text{if } W^2(\dfrac{T_i}{300})^{1.433}\le 12.36\cdot t_i \end{cases} \tag{12}$$

Hence, the optimized electrothermal models can be obtained by combinations of (1), (12)and (4) - (6).

III. SIMULATION AND TLP TESTING RESULTS

I_{t2} can be simulated by combinations of the optimized electrothermal models and parameters models ((7)-(11)). Two silicon-substrate GGNMOS samples, with the process parameters shown in Table II, are used in the simulation, where ε_{si}=11.9F/cm and ε_{ox}=3.96F/cm for Si- substrate material. Considering the ESD pulse duration t_i is 200ns, input these process parameters into the optimized electrothermal models and parameters models, which are coupled with the models before thermal breakdown[6], then the I_D-V_D curves on the safe operation region of two samples can be simulated by MATLAB, as shown in Fig. 7. Fig. 7 also shows the corresponding thermal breakdown current I_{t2}.

TABLE I. THE VALUE RANGES OF b AND T_b ON TYPICAL VALUE RANGES OF PARAMETERS.

x_j(μm)	t_{ox}(nm)	V_{sp}(V)	b(μm)	t_b (ns)
0.1-0.15	2～10	2～10	0.051～0.104	0.004～0.049

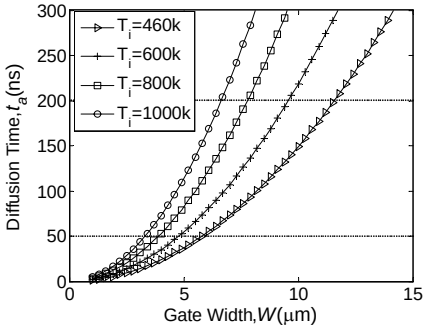

Figure 5. Variation of t_a with W on different T_i .

TABLE II. PROCESS PARAMETERS OF TWO SAMPLES

Parameters	sample 1	sample 2
W	10μm	7μm
gate length L	0.6μm	0.36μm
N_{BC}	1.6×10^{17}cm^{-3}	1.8×10^{17}cm^{-3}
N_E	1.0×10^{18}cm^{-3}	1.1×10^{18}cm^{-3}
sheet resistance $R_{\square}$	5000Ω	4000Ω
x_j	0.1μm	0.1μm
t_{ox}	4.12nm	3nm

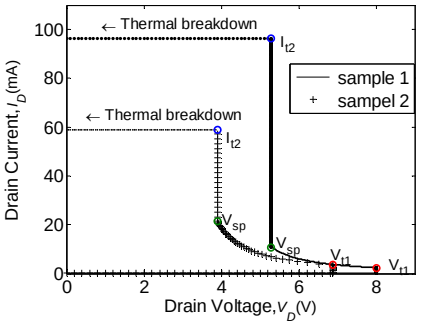

Figure 6. The simulated I_D-V_D curves of two samples.

Note that the gate width in sample 1 is 10μm, the simulated T_i=811.06K, t_a=336.5ns, so $t_a > t_i$; while the gate width in sample 2 is 7μm, the simulated T_i=820.58K, t_a=167.36ns, so $t_a < t_i$. Hence, the time regions of (12) chosen by the two samples in the simulation are different.

The TLP testing I_D-V_D curves of the two samples are shown in Fig. 7, where the stress pulse duration of the testing is 200ns. Table III shows the comparison between the simulated and measured I_{t2} of TLP testing in two samples. We can see that the maximum error is 8.95%, which is in the range of permissible error.

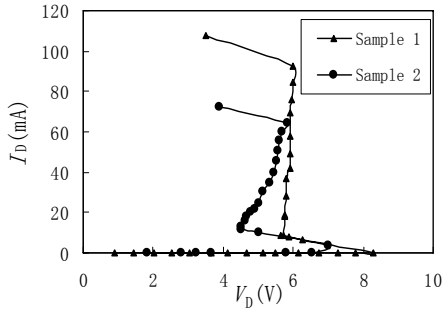

Figure 7. The TLP testing I_D-V_D curves of two samples.

TABLE III. THE COMPARED RESULTS BETWEEN TLP TESTING AND SIMULATION

	TLP TESTING	SIMULATION	ERROR
I_{T2} IN SAMPLE 1	**92.6MA**	**96.42MA**	**4.13%**
I_{T2} IN SAMPLE 2	**64.6MA**	**58.82MA**	**8.95%**

IV. CONCLUSION

This paper presents a novel method of thermal breakdown modeling of GGNMOS based on the physical characteristics of the device and the effects of temperature on the parameters, which avoid the complexity and large error of parameters extraction in other simulation methods. The models can simulate the failure threshold I_{t2} of GGNMOS, and the simulation results are in good agreement with TLP measurements. Combining the models in this paper with the models before thermal breakdown, which were built in previous studies, the I_D-V_D curves on the safe operation region can be simulated. This simple modeling method allows ESD protection circuit designers to rapidly evaluate the figure of merit for ESD capability in their device designs and to reduce the iterations and cycle times involved in achieving successful ESD performance.

ACKNOWLEDGMENT

The work is supported by National Nature Science Foundation of China (Nos. 50661001 and 51161003), and Natural Science Foundation of the Guangxi University (No. DD052064).

REFERENCES

[1] A. Ramanujan, M. Kadi, J. Tremenbert, F. Lafon, B.Mazari, "Modeling IC snapback characteristics under electrostatics discharge stress," IEEE Transactions on Electromagnetic Compatibility, vol. 51, pp. 901–908, November 2009.

[2] Alto.Palo, MEDICI Two-Dimensional device simulation program. Fremont,CA: Avant! Corp., 2001

[3] Haolu Xie, Rouying Zhan, Albert Wang, and R.Gafiteanu, "3D electro-thermal modeling of GGNMOS ESD protection structure," The 2004 IEEE Asia-Pacific Conference on Circuits and Systems, vol. 1, pp. 61-64, December 2004.

[4] Li-Fei Zhang, Chin-Chang Liao, Wei Liu, and Waisum Wong, "A new failure mechanism of gate-grounded MOSFET ESD device in 90nm technology," IEEE Conference on Electron Devices and Solid-State Circuits, pp. 61-64, December 2005.

[5] Yao Liu, Ruohe Yao, and Yingjun Gao, "Modeling of ESD protection device GGNMOS before second breakdown," Microelectronics, vol. 38, 2008, pp. 647-651.

[6] Yuxi Jiang, Jiao Li, Feng Ran, and Yang Dian, "Layout optimization and modeling of an ESD-protection n-MOSFET in 0.13um silicide CMOS technology," The International Conference on ICEPT-HDP , Shanghai, pp. 1-6, July 2008.

[7] U. Lindefelt, "Heat generation in semiconductor device," J. Appl. Physics, vol. 75, pp.942-957, 1994.

[8] A. Amerasekera, and C. Duvvury, ESD in silicon integrated circuits, 2rd ed., New York: John Wiley & Sons, Ltd, 2002, pp. 330-333.
[9] S. Selberher, Analysis and simulation of semiconductor device, Berlin: Springer-Verlag, 1984, pp. 23-39.
[10] S. Ramaswamy, E. Rosenbaum, and Sung-Mo Kang, "Circuit-level electrothermal simulation techniques for designing output protection devices," IRWS, Lake Tahoe, CA, pp. 79-82, Octorber 1994.