# Compact Modeling of Tunneling Breakdown in PN Junctions for Computer Aided ESD Design (CAD-for-ESD)

Yeshwant Subramanian and R. Bruce Darling
University of Washington, Seattle, WA, USA, yeshwant@ee.washington.edu

## ABSTRACT

This paper presents compact, physically-based electrothermal models for direct as well as indirect bandgap tunneling processes in pn-junctions for use in network simulators (e.g. Saber or VHDL-A). The model for indirect tunneling has been validated using a 3.3V Si Zener diode (1N4728). Self-heating effects have also been included. The above tunneling breakdown models, together with the compact models for avalanche breakdown presented previously[1] constitute a complete, compact representation of breakdown in ESD zener diodes. Their utility lies in the simulation of large systems of interconnected ESD structures, without detailed device analysis, permitting a 'CAD-for-ESD' approach in commercial ESD design.

***Keywords***: Tunneling, Compact Modeling, ESD, CAD.

## 1 INTRODUCTION

Zener diodes used as ESD protection structures may undergo breakdown due to the avalanche process or tunneling or a combination of both, depending on process parameters that set their nominal breakdown voltage. In the existing literature [4], analytical equations for direct as well as indirect tunneling are derived at $0^{o}$K. In this paper, compact temperature-dependent I-V models for direct as well as indirect tunneling are derived using piecewise linear approximations of the Fermi distribution functions. In the context of ESD zener breakdown, they complement the compact avalanche breakdown models developed previously [1]. The model for the indirect tunneling process has been validated using a 1N4728 3.3V Si zener diode. Self-heating has also been included by using a thermal equivalent circuit for the bulk of the diode. The utility of the models in a 'CAD-for-ESD' environment is demonstrated through Saber simulations of ESD events in example protection circuits.

## 2 ELECTRICAL MODEL

In this section, the temperature-dependent electrical models for direct as well as indirect bandgap tunneling processes are derived. The model for indirect tunneling is validated using a 3.3V Si zener diode (1N4728).

### 2.1 Direct Tunneling

Consider a planar PN junction under reverse bias. A sketch of its band diagram is shown in Fig. 1.

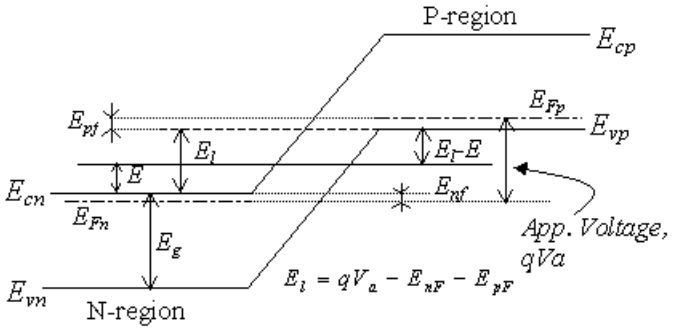


Figure 1: Band diagram sketch of a PN junction.

Existing treatments of direct tunneling [2,3,4] assume a parabolic variation of the (imaginary) kinetic energy of electrons tunneling through a barrier. The corresponding tunneling current density under reverse bias in an isotropic and planar PN junction is given by[2]:

$$J = \frac{qm_{eff}}{18\hbar^3}\exp\left(\frac{-\pi m_{eff}^{\frac{1}{2}} E_g^{\frac{3}{2}}}{2\sqrt{2}qF\hbar}\right) \times \int_0^{E_t}\left[f_p(E) - f_n(E)\right]\left(1-\exp\left(\frac{-E_\perp}{\overline{E}}\right)\right)dE \quad (1)$$

where:

$E_\perp$ is the smaller of $E$, $E_t$-$E$;
$E$, $E_{cn}$ and $E_{vp}$ are as shown in Fig. 1;
$F$ is the maximum electric field in the junction;
$f_p(E)$ and $f_n(E)$ are the Fermi distribution functions for electrons in the P and N regions respectively;
$m_{eff}$ is the electron effective mass;
$\overline{E}$ is given by:

$$\overline{E} = \frac{\sqrt{2}qF\hbar}{2\pi m_{eff}^{\frac{1}{2}} E_g^{\frac{1}{2}}} \;;$$

q is the electronic charge; and
$\hbar$ is the reduced Planck's constant.

In [4], Eq. (1) is analytically evaluated at T=0K (i.e. assuming 'step profiles' for $f_p$ as well as $f_n$). In this paper, temperature dependence is included in the analytical evaluation of Eq. (1) by using piecewise linear approximations for $f_p(E)$ and $f_n(E)$, as shown in Fig. 2.

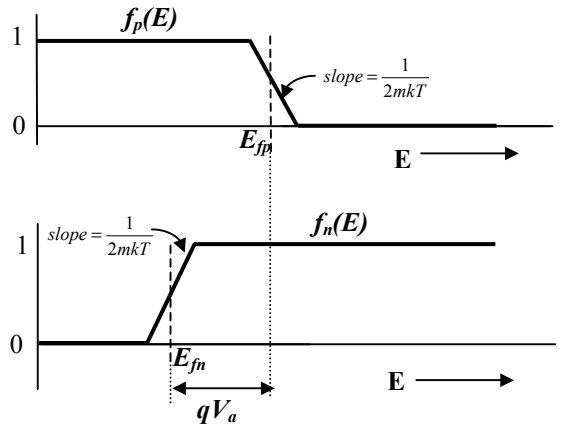


Figure 2: Piecewise linear approximations for the Fermi distribution functions, $f_p(E)$ and $f_n(E)$. A parameter $m$ is used to vary the slope of the two PWL functions.

Depending on the applied bias voltage, the term $f_p(E)$-$f_n(E)$ in Eq. (1) can have either a trapezoidal or triangular shape, as shown in Fig. 3.

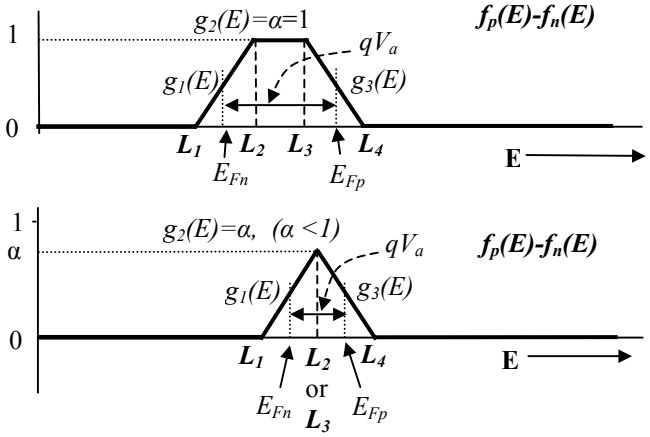


Figure 3: Two possible shapes for the difference function $f_p(E)$-$f_n(E)$. The functions $g_1(E)$, $g_2(E)$ and $g_3(E)$ correspond to the analytical representations of each individual PWL segment (or point). $L_1$, $L_2$, $L_3$ and $L_4$ denote the x-coordinates of transition points between segments.

Using the above PWL form for $f_p(E)$ - $f_n(E)$, Eq. (1) can be rewritten as:

$$J = \frac{qm_{eff}}{18\hbar^3}\exp\left(\frac{-\pi\sqrt{m_{eff}}E_g^{\frac{3}{2}}}{2\sqrt{2}qF\hbar}\right)\int_{E_a}^{E_b} g_i(E)\left(1-\exp(-\frac{E_j}{\overline{E}}\right)dE \quad (2)$$

where:
$i = 1$ if $L_1 \le E \le L_2$;
$i = 2$ if $L_2 \le E \le L_3$; and $E_j = E_1 = E$ if $E \le E_l/2$; *(i.e., j=1)*
$i = 3$ if $L_3 \le E \le L_4$ $E_j = E_2 = E$ if $E \le E_l/2$; *(i.e., j=1)*

Consider the integral in Eq. (2) written in the following form:

$$F_{ij}(E_a, E_b) = \int_{E_a}^{E_b} g_i(E)h_j(E)dE \quad (3)$$

where

$$h_j(E) = 1-\exp\left(\frac{-E_j}{\overline{E}}\right)$$

Let $A$, $B$, $C$, $D_1$ and $D_3$ be defined by the following expressions:

$$A = \exp\left(\frac{-E_a}{\overline{E}}\right);\quad B = \exp\left(\frac{-E_b}{\overline{E}}\right);\quad C = \exp\left(\frac{-E_l}{\overline{E}}\right)$$

$D_1 = E_{nf} + mkT$ and

$D_3 = E_{nf} + mkT - qV_a$.

where $E_{nf}$, $qV_a$ and $E_l$ are as defined in Fig. 1.

The possible analytical forms for the integral in Eq. (3) are then given by:

$$F_{21}(E_a, E_b) = \overline{E}\left(E_b - E_a + \overline{E}(B-A)\right) \quad (4)$$

$$F_{22}(E_a, E_b) = \overline{E}\left(E_b - E_a + \overline{E}\left(\frac{C}{B}-\frac{C}{A}\right)\right) \quad (5)$$

For i=1 and i=3:

$$F_{i1}(E_a, E_b) = \frac{(-1)^{\frac{i-1}{2}}}{2mkT}\left(\begin{array}{l} D_i F_{21}(E_a, E_b) + \frac{E_b^{\,2} - E_a^{\,2}}{2} \\ -\overline{E}^2\left(A\left(\frac{E_a}{\overline{E}}+1\right) - B\left(\frac{E_b}{\overline{E}}+1\right)\right)\end{array}\right) \quad (6)$$

$$F_{i2}(E_a, E_b) = \frac{(-1)^{\frac{i-1}{2}}}{2mkT}\left(\begin{array}{l} D_i F_{22}(E_a, E_b) + \frac{E_b^{\,2} - E_a^{\,2}}{2} \\ -\overline{E}^2\left(\left(\frac{C}{B}\right)\left(\frac{E_b}{\overline{E}}-1\right) - \left(\frac{C}{A}\right)\left(\frac{E_a}{\overline{E}}-1\right)\right)\end{array}\right) \quad (7)$$

Eqs. (4)-(7) represent analytical expressions for all the possible components $F_{ij}$ that could contribute to the integral in Eq. (2). For a given PN-junction, the number of contributing components (as well as each component's corresponding limits $E_a$ and $E_b$) is dependent upon the location of the quasi-Fermi levels with respect to the band edges in the N and P regions (see Fig. 1). This is illustrated in Fig. 4 using two examples of junctions with differing $E_{pf}$. A compact model for direct tunneling must therefore determine the contributing components $F_{ij}$ and their corresponding limits $E_a$ and $E_b$ based on band information derived from its process-related parameters.

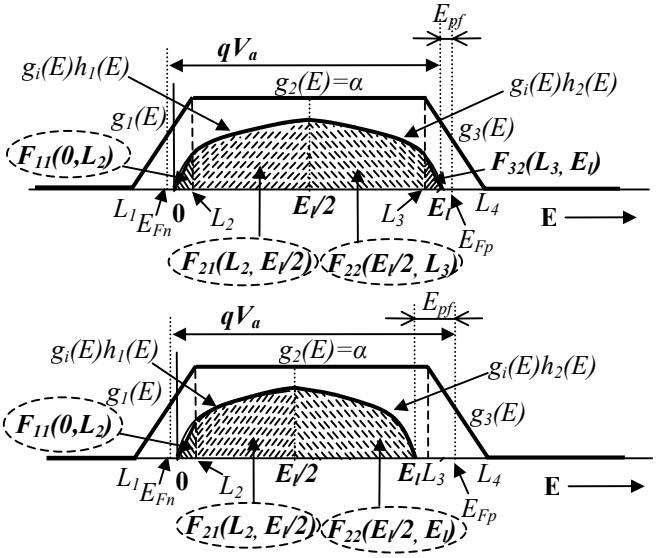


Figure 4: Examples of PN-Junctions with differing contributing components $F_{ij}$ (encircled) to direct tunneling current due to differing $E_{pf}$.

## 2.2 Indirect Tunneling

The basic equation for indirect tunneling has been considered in [2] using results by Keldysh [5] and Price [6]:

$$J = \frac{q\overline{E}^2\left(m_{x1}m_{y1}m_{z1}m_{x2}m_{y2}m_{z2}\right)^{\frac{1}{2}}M^2V}{4\left(2^{\frac{3}{4}}\right)\pi^{5/2}\hbar^{13/2}m_{rx}{}^{\frac{1}{4}}E_G{}^{\frac{3}{4}}F^{\frac{1}{2}}}$$

$$\times\left\{\begin{array}{l}(n+1)\exp\left(\dfrac{-4(2m_{rx})^{\frac{1}{2}}}{3\hbar F}\left(E_G+\hbar\omega\right)\right)\\ +n\exp\left(\dfrac{-4(2m_{rx})^{\frac{1}{2}}}{3\hbar F}\left(E_G-\hbar\omega\right)\right)\end{array}\right\}$$

$$\times\int_0^{E_l}\left[f_p(E)-f_n(E)\right]\left(1-\exp\left(-\frac{E}{\overline{E}}\right)\right)\left(1-\exp\left(-\frac{E_l-E}{\overline{E}}\right)\right)dE \qquad (10)$$

where, in addition to symbols common with the direct tunneling case (that have meanings indicated in Fig.1 or against Eq. (1)):

$m_{x1}$,$m_{x2}$,$m_{y1}$,$m_{y2}$, $m_{z1}$ and $m_{z2}$ are the components of the effective mass tensor;

$m_{rx}$ is reduced effective mass, given by:

$$m_{rx} = \frac{m_{x1}m_{x2}}{m_{x1}+m_{x2}};$$

$M^2V$ is a parameter for phonon scattering;

ω is the phonon frequency;

n is the phonon occupation factor, given by:

$$n = \left(\exp\left(\frac{\hbar\omega}{kT}\right)-1\right)^{-1}$$

Using the PWL form in Fig. 3 for $f_p(E)$-$f_n(E)$, the integral in Eq. (10) may be rewritten as:

$$F_i(E_a,E_b) = \int_{E_b}^{E_b} g_i(E)h(E)dE \qquad (11)$$

where:

$$h(E) = \left(1-\exp\left(\frac{-E}{\overline{E}}\right)\right)\left(1-\exp\left(\frac{-(E_l-E)}{\overline{E}}\right)\right);\ \text{and}$$

$i = 1$ if $L_1 \le E \le L_2$;
$i = 2$ if $L_2 \le E \le L_3$; and
$i = 3$ if $L_3 \le E \le L_4$

As in the direct tunneling case, let

$$A = \exp\left(\frac{-E_a}{\overline{E}}\right);\quad B = \exp\left(\frac{-E_b}{\overline{E}}\right);\quad C = \exp\left(\frac{-E_l}{\overline{E}}\right)$$

$D_1 = E_{nf} + mkT$ and $D_3 = E_{nf} + mkT - qV_a$.

The possible analytical forms for the integral in Eq. (11) are:

$$F_2(E_a,E_b) = \left((E_b-E_a)(1+C)+\overline{E}\left(B-A-\frac{C}{B}+\frac{C}{A}\right)\right) \qquad (12)$$

If i=1 or i=3:

$$F_i(E_a,E_b) = \frac{(-1)^{\frac{i-1}{2}}}{2mkT}\left(\begin{array}{l}D_iF_2(E_a,E_b)+\dfrac{E_b{}^2-E_a{}^2}{2}\\ -\overline{E}^2\left(\left(1+\dfrac{E_a}{\overline{E}}\right)A\right.\\ \left.-\left(1+\dfrac{E_b}{\overline{E}}\right)B\right)-\overline{E}^2C\left(\begin{array}{l}\left(1-\dfrac{E_a}{\overline{E}}\right)\dfrac{1}{A}\\ -\dfrac{1}{B}\left(1-\dfrac{E_b}{\overline{E}}\right)\end{array}\right)\end{array}\right) \qquad (13)$$

Similar to the direct tunneling case, the indirect tunneling current for a given bias voltage is evaluated using one or more of the components $F_i$ in Eqs. (12) and (13) for the integral in Eq. (10). The compact model for indirect tunneling must determine the choice of the contributing components $F_I$ (as well as the limits $E_a$ and $E_b$ in each case) from process parameters in a manner similar to that shown in Fig. 4.

## 2.3 Validation

The direct as well as indirect models are implementable in the standard modeling languages for network simulators (e.g. Saber or VHDL-A). Both models have been implemented and tested in Saber. The validation of the indirect tunneling model using a 3.3V 1N4728 Si zener diode is shown in Figs. 5 and 6. Contact resistance was modeled with a 10-ohm series resistance. At a global level, the model compares well with 1N4728 data.

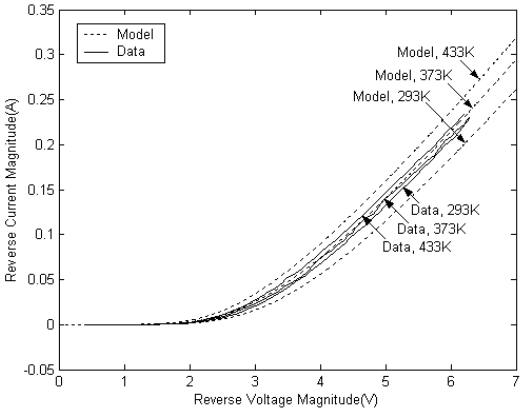


Figure 5: Indirect tunneling model performance (dotted curves) vs. 1N4728 data (solid curves).]

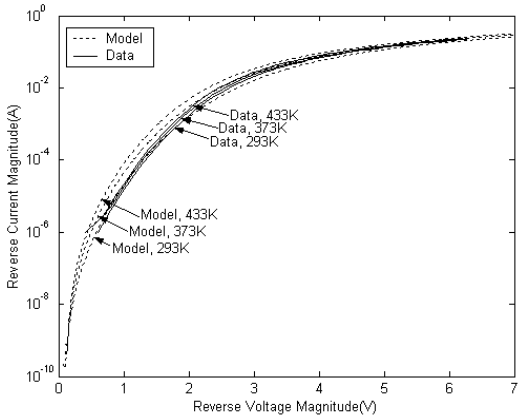


Figure 6: Indirect tunneling model performance (dotted) vs. 1N4728 data (solid) on a semilog scale.

## 2.4 System Simulation

Fig. 7 shows an ESD-zener protected CMOS gate being subjected to a 6kV HBM ESD pulse. Fig. 7 shows Saber simulation (using the 1N4728 model) results showing the key output characteristics.

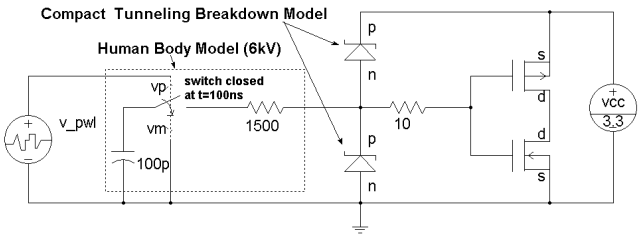


Figure 7: Schematic of dual-diode ESD protection circuit for CMOS gates being subjected to a 6kV HBM ESD pulse.

Self-heating, though included in the model (as in [1]) has a negligible effect on the performance of the protection diodes in the circuit of Fig. 7 (due to the low temperature coefficient and high lead thermal mass).

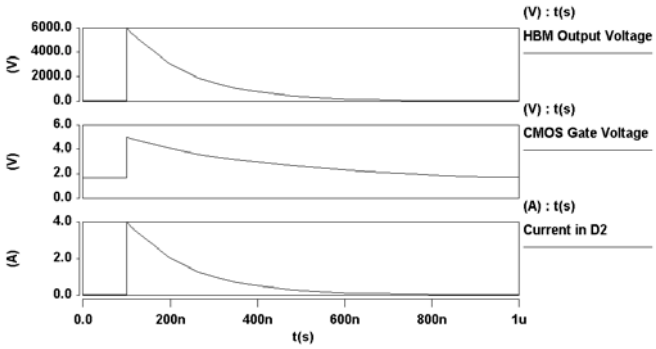


Figure 8: Saber simulation of the circuit in Fig. 7. The diode D2 breaks down upon application of the pulse, protecting the CMOS gate input. Self-heating effects were negligible.

# 3 CONCLUSIONS

Compact, physically based electrothermal models for direct as well as indirect bandgap tunneling processes have been developed and implemented in a network simulator (Saber). The model for indirect tunneling has been validated using a 1N4728 3.3V Si zener diode. The models are intended for use in system-level simulations involving large networks of interconnected ESD structures (e.g. in a 'CAD-for-ESD' environment). An example simulation of the 1N4728 model in a dual diode ESD protection circuit for a CMOS gate input is used to demonstrate their utility. The compact tunneling models developed in this paper, coupled with the avalanche breakdown models developed previously [1] constitute a complete, compact representation of breakdown in ESD zener diodes.

# REFERENCES

[1] Y. Subramanian and R.B. Darling, "Compact Modeling of Avalanche Breakdown in PN-Junctions for Computer-Aided ESD Design", *Proc. MSM 2001*, Hilton Head Island, SC, March 2001, pp. 48-51.

[2] E. O. Kane, "Zener Tuneling in Semiconductors", *J. Phys. Chem. Solids*, vol. 12, 1959, pp. 181-188.

[3] E. O. Kane, "Theory of Tunneling", *J. Applied Physics*, vol. 32, no. 1, January 1961, pp. 83-91.

[4] J. L. Moll, *Physics of Semiconductors*. McGraw Hill, 1964, pp. 240-259.

[5] L.V. Keldysh, *Soviet Phys. –JETP*, vol. 6(33), no. 4, 1958, pp. 763-770.

[6] P.J. Price and J. M. Radcliffe, *IBM J. Research Develop.*, vol. 3, no. 3, 1959.