Modeling of non-uniform heat dissipation and prediction of hot spots in power transistors

L. ZHU and K. VAFAI
Department of Mechanical Engineering

and

L. XU
Department of Electrical Engineering, The Ohio State University, Columbus, OH 43210, U.S.A.

(Received 9 May 1997 and in final form 24 September 1997)

Abstract—In this work a heat generation model induced by current crowding is proposed. This model, in turn, is applied to the thermal calculation of typical multi-finger emitter structures of BJTs. The internal temperature distribution obtained from this model provides clear indication of the existence of hot spots in base-emitter contact region. The characteristics of the location and the magnitude of hot spots are also presented in this paper. Various pertinent effects due to the non-uniform heat dissipation are discussed. The result of this study serves to locate potential hot spots, which must be considered to prevent thermal breakdown of power BJTs as well as other types of power devices. © 1998 Elsevier Science Ltd. All rights reserved

1. INTRODUCTION

Contemporary power semiconductor devices are known to generate high heat flux up to 100 W cm⁻². This heat dissipation increases substantially with the current/voltage handling capability, operation frequency and the scaling-down of the device size. As a consequence of the intense heat dissipation, modern power devices operate at a high internal temperature. In most cases, heat distributes unevenly, causing localized high temperature regions, known as hot spot, within the power semiconductor devices. The maximum temperature in hot spots is likely to exceed an intrinsic temperature, at which the intrinsic carrier density equals to the dopant carrier density, thus causing transistor devices to fail its function and leading to possible thermal runaway. In order to understand this problem, the prediction of hot spots becomes crucial in the thermal design of integrated circuits (IC).

Due to the emitter current crowding in discrete and monolithic power transistor devices such as bipolar junction transistors (BJTs), Joule heating occurs mainly at the base-emitter contact rather than distributing uniformly within transistor devices. This non-uniform heat dissipation causes hot spots and leads to thermal runaway phenomena. Previous works in thermal modeling of these types of power transistors have not addressed this problem. In power devices such as BJTs, current crowding is an important cause of hot spots. Current crowding is caused by the lateral base-emitter (BE) current flowing from the base to the emitter. Due to the two-dimensional structure of power BJT devices and the sizable base resistance, the base current creates a lateral voltage drop between base-emitter contact region and the central emitter region. The voltage at the BE contact is higher than the voltage at the central emitter. As the current density increases exponentially with an increase of the applied voltage on a pn junction, the current density at the BE contact becomes drastically larger than the current density at the central emitter. This mechanism is known as the emitter current crowding, a commonly observed problem in power BJT devices. Corresponding to the non-uniform distribution of current density, the heat generation rate at the BE contact is also much larger than that at the central emitter, since heat dissipation rate is proportional to the current density. This non-uniform distribution of Joule heating induced by the emitter current crowding can produce potentially damaging hot spots.

Electrical effects of the current crowding have been investigated extensively in the solid state literature for semiconductors since the early years of transistor technology [1] and through the modern monolithic devices [2]. The classic model of Hauser [1] has been widely used as an analytical model which describes the distributed nature of emitter base junction voltage by introducing the concept of an equivalent base resistance. A recent extension of Hauser's model to the transient current crowding is given by Jin and Fossum in [3]. Thermal effects of current crowding,
however, have been generally ignored in previous thermal modeling of power semiconductor devices, although the important issue of current crowding has been recognized and briefly discussed [4].

Two examples of the pertinent works about thermal modeling of transistor devices include works of Higgins [5] and Webb [6]. Higgins simulated the thermal behavior of Al/GaAs/As power Heterojunction Bipolar Transistors (HBTs) and established some quantitative tradeoff between power density, chip layout and junction temperature on multi-finger structures that is similar to the one studied in this paper. Likewise, Webb used the similar multi-finger structure for the thermal simulation of Gallium Arsenide HBTs and Metal-Semiconductor Field Effect Transistors (MESFETs). In both of these works a uniformly distributed heat generation model was utilized and the current crowding effects was not considered. The exclusion of the current crowding effect in previous thermal studies stems from the inherent difficulty of modeling a complex coupled thermal-electrical phenomena.

As discussed earlier, the crowded current density at BE contact generates high heat dissipation rates, leading to hot spots. Hot spots substantially alter the semiconductor properties, increasing the intrinsic carrier density, which in turn increases the current density. This forms a positive feedback loop between current density and hot spots and makes the problem thermo-electrically coupled. One approach to this coupled problem is to simultaneously solve both the carrier transport equations as well as the energy equation [7] which usually involves a complex and a highly time-consuming simulation. Another approach which is adopted in this work is to obtain a semi-analytical formulation and to establish a heat generation model in which non-uniform distribution of Joule heating caused by current crowding is taken into consideration through an analytical model of lateral distributions of current density and junction voltage in the transistor.

2. MODELING OF THE HEAT GENERATION

In order to reduce current crowding, the emitters and the bases of practical power transistors are interleaved as narrow fingers, a part of which is shown in Fig. 1. The schematics of current crowding in one emitter strip is illustrated in Fig. 2. Figure 2(a) shows the lateral distribution of base current in a npn BJT. Usually the base region is moderately doped and is made relatively thin in order to obtain a large current gain. Consequently, there is a considerable base resistance which generates a lateral potential difference underneath the emitter region. In the npn BJT, this potential decreases from the BE contact towards the center of the emitter. The number of electrons injected

![Fig. 1. Multi-finger structure of emitters and bases in power BJT transistors.](image-url)
Modeling of non-uniform heat dissipation

from the emitter is exponentially dependent on the local BE voltage, so that more electrons are injected near the emitter boundary than at the center of the emitter, causing the crowding of the emitter current near the BE boundary. This results in current-crowding at the boundary as shown in Fig. 2(b). In addition, the substantial current density near the emitter boundary causes localized heating effects. To model this process, first, it should be noted that only one half of the BE finger structure needs to be analyzed (Fig. 3) since the base current distributes symmetrically with respect to the center of the emitter. The heat generation (power loss) in power transistors occurs when the device is at on-state, off-state and during switching transitions. Only the on-state Joule heating is accounted here as the off-state power loss due to the leakage current is usually negligible. The heat dissipation during switching transition, which is proportional to the switching frequency, is also not significant since BJTs normally are not high frequency devices [4]. The heat generation (Joule heating) in general is given by

\[ H = J \cdot E \]

\[ = J_c \frac{V_{BE}(x)}{W_b} \]  

(1)

where \( H \) is the Joule heat generation, \( J \) is the total current density vector and \( E \) is the electric field and \( J_c, V_{BE}(x) \) and \( W_b \) define collector current density, BE junction voltage, and base thickness, respectively. The Joule heating dissipates mainly within the base region directly underneath the emitter (shaded area in Fig. 3). This is because the heavily doped emitter and the collector conduction layers have substantially smaller resistance compared to that of the base. Furthermore the resistance of the collector depletion region is also negligible when the BJT is at the on-state (forward active mode operation). The total current in the base includes the base current flowing laterally from the BE contact to the central emitter and the collector current flowing vertically from BE junction to the base-collector (BC) junction. The total current density is approximated as the collector current density \( J_c \) only, as indicated in the second part of equation (1). This is due to the fact that the magnitude of the lateral base current is much smaller than the collector current by a fraction of the common emitter gain. Note that the BE junction voltage, \( V_{BE} \), is a function of position \( x \) due to the lateral distribution of base current.

The collector current density, when BC junction is forward biased, is given by the ambipolar diffusion theory as

\[ J_c = J_S \left[ \exp \left( \frac{V_{BE}(x)}{V_T} \right) - 1 \right] \]

\[ \approx J_S \exp \left( \frac{V_{BE}(x)}{V_T} \right) \]  

(2)

where \( J_S \) is the reverse saturation current density of the base region. The thermal voltage, \( V_T \), is defined as \( V_T = (kT/e) \), with \( k \) being the Boltzmann constant, \( T \), the temperature and \( e \), the electron charge. The BE junction voltage, \( V_{BE} \), can be written as

\[ V_{BE} = V_{BE}^0 - \int_0^x J_S(x) \rho \, dx \]  

(3)

Fig. 2. Schematics of current crowding. (a) The lateral voltage drop induced by the base current. (b) The emitter current crowding.

Fig. 3. One half of the base and emitter finger structure in BJTs.
where $V_{BE}^0$ is the BE voltage at BE contact, $x = 0$, $I_B$ the base current and $\rho$ is the effective specific resistance of the base which can be evaluated as
\[ \rho = \frac{1}{2pe\mu_p W_E L_E} \] (4)
in which $p$ is the hole concentration in the $p$-doped base region which can be approximated as the base acceptor doping density $N_a$. The hole's mobility, $\mu_p$, is assumed to be constant. The emitter length and the half emitter width, $W_E$ and $L_E$, respectively, are shown in Fig. 3.

The majority part of $I_B$ is the hole current that is back-injected from the base to the emitter, which is due to the recombination of excess minority carrier electrons with majority carrier holes at BE junction. $I_B$ can thus be written as [1]
\[ I_B(x) = I_B^0 - \int_0^x 2J_E W_E \exp \left( \frac{V_{BE}(x)}{V_T} \right) - 1 \right] dx \] (5)
where $V_{BE}$ in equation (5) can be eliminated by differentiating equation (5) twice and utilizing equation (3). After some manipulation, we obtain the governing equation for $I_B$ as
\[ \frac{d^2 I_B}{dx^2} + \frac{\rho}{V_T} \frac{dI_B}{dx} = 0. \] (6)

For a $n p n$ BJT at the on-state, $I_B \approx 0$ and $(dI_B/dx) < 0$. In this case the general solution for equation (6) takes the following form
\[ I_B(x) = C_1 \tan \left[ \frac{C_1 \rho}{2V_T} (C_2 - x) \right] \] (7)
where $C_1$ and $C_2$ are unknown constants to be determined from boundary conditions. One such boundary condition is the symmetry of $I_B$ about the center of the emitter,
\[ I_B(L_E) = 0. \] (8)

Using equation (8), the two unknown constants $C_1$ and $C_2$ in equation (7) can be reduced to one parameter and equation (7) may be re-written as
\[ I_B(x) = \frac{2V_B Z}{\rho L_E} \tan \left[ Z \left( 1 - \frac{x}{L_E} \right) \right] \] (9)
where $Z$ is the current crowding parameter. Another boundary condition is that at the BE contact, namely $x = 0$, we arrive at
\[ I_B(0) = I_B^0 = \frac{2V_B Z}{\rho L_E} \tan (Z). \] (10)

Substituting equation (9) in equation (5) we obtain an expression for $V_{BE}$ in terms of $Z$
\[ V_{BE}(x) = V_{BE}^0 - 2V_T \ln \left[ \frac{\cos Z \left( 1 - \frac{x}{L_E} \right)}{\cos Z} \right] \] (11)

In addition to equation (10), another expression for $I_B^0$ can be obtained by substituting equations (8) and (11) into equation (3). This results in
\[ I_B^0 = W_E L_E J_S \exp \left( \frac{V_{BE}^0}{V_T} \right) \sin \frac{2Z}{Z} \] (12)

Combining the two expressions, i.e. equations (10) and (12) for $I_B^0$. The constraint equation for the current crowding parameter $Z$ is found to be
\[ \cos^2 \frac{Z}{Z^2} = 2 \pi \rho e \mu_p W_E \exp \left( - \frac{V_{BE}^0}{V_T} \right). \] (13)

It is not possible to measure $V_{BE}^0$ directly. However, it is possible to describe $I_B^0$ in terms of the collector current for the entire BJT device which is controlled by the external circuit. The total collector current in the device can be written as
\[ I_C = 2N_E W_E \int_0^x J_C(x) dx \] (14)
where $N_E$ is the effective number of emitter strips. The factor 2 in equation (14) accounts for the fact that only one half of the emitter strip is considered. Using equations (11) and (2) we can explicitly integrate equation (14). This results in
\[ I_C = 2W_E N_E \int_0^{V_{BE}^0} \frac{\cos^2 Z}{Z^2} \left( \frac{V_{BE}^0}{V_T} \right) \sin \frac{2Z}{Z} \left( \frac{1 - \frac{x}{L_E}}{Z} \right) dx \] (15)

Combining equations (15) and (13) to eliminate the exponential term in equations (15) and (13), resulting in a constraint equation for $Z$,
\[ \frac{Z \sin Z}{\cos^2 Z} = \frac{L_E I_C}{4kpe \mu_p W_S E N_E \frac{1}{T}} \] (16)
and $V_{BE}^0$ can be given as
\[ V_{BE}^0 = -V_T \ln \left( \frac{\cos^2 Z}{Z^2} \right) \exp \left( \frac{L_E^2}{2kpe \mu_p W_S T} \right). \] (17)

Finally substituting equations (15), (11), and (7) into equation (1), the heat generation model of BJTs which incorporates current crowding is attained as
\[ H(x) = \frac{J_S}{W_b} V_{BE}(x) \frac{\cos^2 Z}{Z^2} \exp \left( \frac{V_{BE}^0}{V_T} \right). \] (18)
This heat generation model is applied to the energy equation to simulate the temperature distribution in the one-half base-emitter strip shown in Fig. 3. The relevant energy equation for the BJT is

\[ K(T) \left( \frac{\partial^2 T}{\partial x^2} + \frac{\partial^2 T}{\partial y^2} \right) + H(T) = 0 \]  

(19)

where \( \kappa \) is the silicon thermal conductivity that varies considerably with temperature. A temperature dependent model for \( \kappa \) which is given by

\[ \kappa(T) = \frac{1}{3 \times 10^{-4} + 1.56 \times 10^{-5} T + 1.65 \times 10^{-8} T^2} \]  

(20)

is adopted in this work to account for this variation.

3. NUMERICAL METHODOLOGY

The governing equations are discretized using a finite element procedure. Temperature within each element is interpolated by linear functions, expressed in terms of temperatures to be determined at the four nodes as

\[ T = \theta^T T^{(e)} \]  

(21)

where \( \theta \) is the shape function and \( T^{(e)} \) is the nodal temperature array. The Galerkin form of the Method of Weighted Residuals is used to reduce the error introduced during the discretization process to zero, by making the residuals orthogonal to the interpolation function. The resulting discretization equation is

\[ K(T)T = Q + H(T) \]  

(22)

where \( K \), \( Q \) and \( H \) are stiffness matrix, load array from heat flux boundary condition and load array from heat source, respectively. They are defined as

\[ K(T) = \sum_{i=1}^{n} \int_{V_e} \kappa(T) \frac{\partial \theta}{\partial x_j} \frac{\partial \theta}{\partial x_i} dV \]

\[ Q = \int_{V_e} q \theta ds \]  

(23)

\[ H(T) = \sum_{i=1}^{n} \int_{V_e} -H(T)\theta dV. \]

It should be noted that \( K \) and \( H \) are temperature dependent. A successive substitution procedure is used to solve equation (22). At every iteration step \( n \), \( \kappa \) is updated with \( T^{(n)} \) through the use of equation (20). The value of \( Z \) is obtained by solving equation (16). The values for \( V_{BE} \) and \( V_{BE} \) are obtained through the use of equations (11) and (17), respectively, while \( H \) is calculated from equation (18). The updated value, \( T^{(n+1)} \), is then obtained from

\[ K(T^{(n)})T^{(n+1)} = Q + H(T^{(n)}). \]  

(24)

This iterative procedure continues until the iteration criteria is satisfied. An advantage of this method is its fairly large convergent radius, which is essential for this nonlinear problem. The application of the Galerkin-based FEM is described by Taylor and Hood [8], and its application in the finite element program used in the present work is well documented [9].

4. RESULTS AND DISCUSSIONS

The simulations in this work are based on an actual BJT transistor used in power electronics applications. Three monolithical BJT structures are connected serially to form a Darlington structure for a multiplied current gain, as shown in Fig. 4. Most of the heat dissipation occurs in the third BJT (the main amplifier) as the current in that BJT is the largest. The effective number of emitter strips of the third BJT is set at 25. The base is usually moderately doped and \( N_a \) is approximated as \( 10^{16} \text{ cm}^{-3} \) as a typical doping level. The length of the emitter is 500 \( \mu \text{m} \) and the thickness of the base is 8 \( \mu \text{m} \). The rated collector current of a typical BJT device is set at 100 \( \text{A} \). Various \( I_c \) values have been used in this simulation to study corresponding effect on the heat generation model as well as on the temperature distribution. The hole mobility \( \mu_h \) is taken to be 480 \( \text{cm}^2 \text{V}^{-1} \text{s} \) which is a typical value for BJTs. The leakage current \( J_g \) in the model includes the reverse saturation current and the reverse recombination current. It is taken as \( 1.8 \times 10^{-8} \text{ A cm}^{-2} \) again a typical value. The distribution of \( V_{BE} \) is shown in Fig. 5. From the center of the emitter to the BE contact, \( V_{BE} \) increases gradually from 0.58 \( \text{V} \) to 0.64 \( \text{V} \) across most of the base region and then rises rapidly from 0.64 \( \text{V} \) to 0.89 \( \text{V} \) within a distance of about 10 \( \mu \text{m} \) near BE boundary. Because of the exponential dependence of the injected electrons on \( V_{BE} \), the rapid increase of \( V_{BE} \) in the thin layer near BE contact corresponds to a drastic increase in the current density and hence results in significant heat dissipation in this region.

A parametric study of the heat generation model given in equation (18) has been made. The parameters whose influence have been studied include the total collector current of the device, \( I_c \), the emitter stripe width \( 2L_e \), and the internal temperature. Figure 5 also shows the lateral distribution of the heat generation for different collector currents. The internal temperature is assumed to be 80°C, a typical operation temperature of actual power transistors. Two observations of the heat generation rate distribution shown in Fig. 6 are worthy to be presented. First, we find that the heat dissipated in the BJTs is to a large degree concentrated on the BE boundary, contrasting the uniform distribution models that previous thermal investigations have assumed. In fact, the maximum heat dissipation rate near the BE boundary is about \( 3 \sim 5 \times 10^4 \) times larger than the minimum heat dissipation rate occurring at the central emitter region.
Fig. 4. Three Darlington-connected BJTs in a typical device. (a) Structure of three Darlington-connected BJTs. (b) Simplified circuit of three Darlington-connected BJTs.

Fig. 5. Distribution of the heat dissipation rate $H$ and the base-emitter voltage $V_{BE}(x)$. 
Second, we observe from Fig. 5 that the maximum heat dissipation rate increases linearly with \( I_c \) at BE boundary but remains almost constant at the emitter center.

Another important parameter is \( 2L_E \), the half width of emitter strip. Figure 6 shows the change of maximum heat dissipation rate as a function of \( 2L_E \). The collector current and the temperature used in the calculation of Fig. 6 are 100 A and 80°C, respectively. As \( 2L_E \) varies, the number of the total emitter strips, \( N_E \), varies accordingly so that \( N_EL_E \) remains a constant and the total \( I_c \) remains unchanged.

Figure 5 shows the heat generation corresponding to collector currents of 50, 100, and 150 A. The heat generation rate increases with current level linearly at the BE contact and remains almost constant at the emitter center. The variation of maximum heat dissipation rate corresponding to a range of emitter width is shown in Fig. 6. For \( L_E > 100 \mu m \), the current crowding has become well-developed and further increase of \( L_E \) does not increase heat generation significantly. However, for \( L_E \) smaller than 55 \( \mu m \), the maximum heating decreases substantially with a decrease in \( L_E \). This indicates the advantages of using narrow emitter strips to reduce heating dissipation and hot spots, as well as to increase current gain and electrical performance. The development of new technologies such as the self-aligning polysilicon technology has made it possible to fabricate emitter strips with width considerably smaller than the minimum strip width given by conventional lithography technology. Thus the quantitative correlation between heat generation and emitter strip width is quite pertinent in the design of bipolar junction transistor devices. Temperature dependency of the heat generation is also illustrated in Fig. 6, in which the collector current used is 100 A and the \( L_E \) is taken as 50 \( \mu m \).

As can be seen in Fig. 6 the maximum heat dissipation rate increases linearly with an increase in device temperature. This linear dependency agrees with previous works [7, 10]. This correlation between the heat dissipation rate and the device temperature may result in a potential failure, usually termed secondary or thermal breakdown. In BJTs as well as other minority transistors, the current is carried by minority carriers and is proportional to the density of minority carriers, which in turn is proportional to \( T^{1.5} \). Therefore if the temperature at one location in the transistor become higher than its surrounding, more current will flow to that location, generating more heating and thus might further increase the temperature in that area if the heat dissipation rate is greater than the heat removal rate. This results in the thermal stability problem. Temperature dependency results in Fig. 6 along with other results presented in this work can be used as an important guideline in the thermal design of BJT transistor devices.

Figure 7 displays the results of a two-dimensional thermal simulation of the BE stripe shown in Fig. 3. The half width of emitter used in the simulation is 50 \( \mu m \), as measured from the device. Boundary condition used in the simulation are described below. At the symmetry lines (the base and the emitter) an adiabatic condition prevails. The heat released from the top of the device is negligible and thus the an adiabatic condition also prevails on the top. The boundary at the bottom attached to a heat sink is taken as a constant temperature condition at 80°C which represents a typical junction temperature in power transistor devices.
devices. Since the objectives of this study is to locate the hot spot, the relative rise of temperature from the bottom boundary temperature does not change substantially for different boundary temperatures. Figure 7 shows the temperature distribution for two different collector currents, i.e., 50 A and 100 A. The hot spot can be clearly seen in Fig. 7, localized at the BE contact region with maximum temperature rise (from boundary temperature) of 5.8 and 8.0°C for \( I_c \) values of 50 and 100 A, respectively. These local temperature rises are considerably higher than 2–3°C, previously reported by Higgins [5], where the heat generation due to emitter current crowding was not taken into consideration. It should also be noted as seen in Fig. 7 that the hot spot region is rather constrained (within 10 \( \mu m \)) in the x direction, implying a very large heat flux flowing in that direction.

5. CONCLUSIONS

In the study a heat generation model incorporating current crowding effects has been proposed. It is established that the heat dissipation is concentrated at the base emitter contact instead of distributing uniformly as assumed in previous models. It is found that the maximum heat generation rate drops rapidly with small emitter width whenever the half emitter width is smaller than 55 \( \mu m \). As expected it is seen that the heat generation rate increases with temperature, indicating a potential thermal runaway mechanism caused by the loss of thermal stability. Hot spots associated with the localized heat generation rate are clearly observed and it is shown that the maximum temperature rise is higher than what has been reported previously. It is found that the lateral heat flux at the BE contact is very high and as such quite crucial in establishing the operating characteristics of a power device. The result of this study serves to locate potential hot spots, which must be considered to prevent thermal breakdown of power BJTs as well as other types of power devices.

Acknowledgement—The support from NASA Grant no. NCC3-324 is acknowledged and appreciated.

REFERENCES

1. Hauser, J. R., The effects of distributed base potential on emitter-current injection density and effective base
Modeling of non-uniform heat dissipation


