LUMP PARTITIONING OF IC BIPOLAR TRANSISTOR MODELS FOR HIGH-FREQUENCY APPLICATIONS

Nelson Chan* and Robert Dutton

Integrated Circuits Laboratory
Stanford University
Stanford, CA 94305 U. S. A.

ABSTRACT

A physical basis for parameter extraction of a two lump bipolar model for high-frequency applications is presented. S-parameter data taken from devices is compared to that generated from a two lump model predicted by physical theory using 2-D finite element device simulation. Erros within 5% in magnitude and 5^{o} in phase are obtained for all four S-parameters for a frequency range up to f_T and across the full current range.

Based on the finite lement analysis and the Dynamic Grid Method, the sidewall base charge is shown to plav a significant role in the formulation of the two lump model when the base diffusion current becomes dominant. This paper describes the numerical approach to lump partitioning based on physical principles. A methodology for formulating the two lump model is also presented.

*Nelson Chan is now with Intel Corp., Santa Clara, CA., U.S.A.

The integrated bipolar transistor is a distributed device, with distributed base resistance, base-emitter and base-collector capacitance (Figure 1). At high frequencies, due to the distributed effects of resistance and capacitance, a one lump model is not adequate to represent the device--frequently more lumps are needed. The multi-lump models have been studied extensively [1-6] and results show an improved fit of model parameters to measured data in high-frequency applications. However, in these studies, the criteria for lump partitioning are not discussed and no physical basis given. The general approach is empirical.

Richardson [7] derived a two lump model based on a uniform transmission line representing the active base. AC crowding and diode nonlinearity effects were taken into account. These two effects together produced a net DC effect on the emitter current. Based on this net DC effect the transistor is partitioned into two sections. The model is valid for low crrent operation of a device in which the active base can be considered as a uniform transmission line. However, at high currents, the effect of sidewall base charge becomes significant and this model cannot explain the observed lump partitioning ratios.

The present work incorporates Richardson's theory for lump partitioning in the low current region. In addition, the sidewall contribution in high current/high frequency modeling is analyzed using a two-dimensional finite element technique (the Dynamic Grid Method) to map current flow lines and calculate the current/charge distribution in the sidewall base [8].

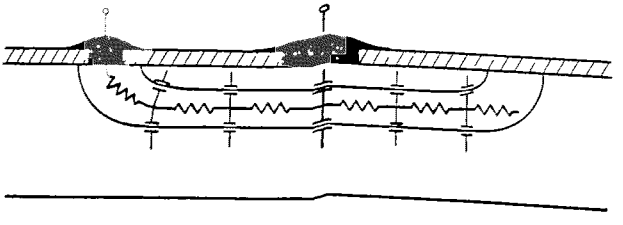

Figure 1 The integrated bipolar transistor

This study has led to a more systematic basis for lump partitioning from physical principles. The predicted two lump model has been compared to an empirical model derived from S-parameter fit of standard bipolar transistors [8]. Errors less than 5% in magnitude and 5^{o} in phase are obtained for all four S-parameters for frequencies up to f_T. The predicted partitioning ratios track the empirical ratios closely for the full current range. These results confirm the physical basis for lump partitioning of high-frequency bipolar transistors.

Section II shows the inadequacy of the one lump model and the improvement using a two lump model based on S-parameter fit to measured data. A physical basis is given to the lump partitioning in this section. In Section III the numerical approach to lump partitioning, based on the finite element analysis and the dynamic grid method, is described. A methodology for formulating the two lump model is presented in Section IV. The conclusion is given in Section V.

II. LUMP PARTITIONING

A. One Lump Model

The one lump hybrid-pi model shown in Figure 2 has been extensively used as a high-frequency model for the integrated bipolar transistors. This a a one-transistor (one-lump) small signal model that describes the intrinsic transistor action and incorporates the extrinsic resistance and parasitic capacitance. Though simple and physical, this model is not accurate for high-frequency applications.

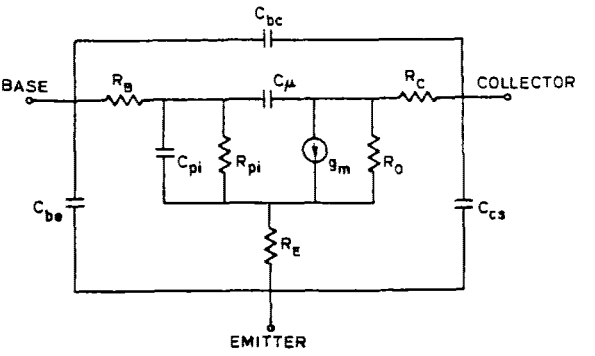

Figure 2 One-lump hybrid-pi model

When compared to measuredS-parameters of real devices, large errors are observed at frequencies near f_T, the cutoff frequency. Figure 3a shows the extent of the discrepancies at three current levels -- low (100 uA), intermediate (1 mA), and high (5 mA) for standard bipolar devices (Figure 4) [8] with an emitter area of 20x20 μ^2, a base area of 30x50 μ^2 and a total area of 80x80 μ^2, with a one-sided base contact. The depth of the epitaxial layer is 5 μm. The base-emitter junction depth is one micron and the base-collector junction depth is 1.7 micron. As shown in Figure 5a the average dc beta is 60-80 over the current range of one microampere to ten milliamperes. The maximum f_T is about 450 MHz peaking at 3 mA collector current (Figure 5b). Both beta and f_T fall off very rapidly near 10 mA of collector.

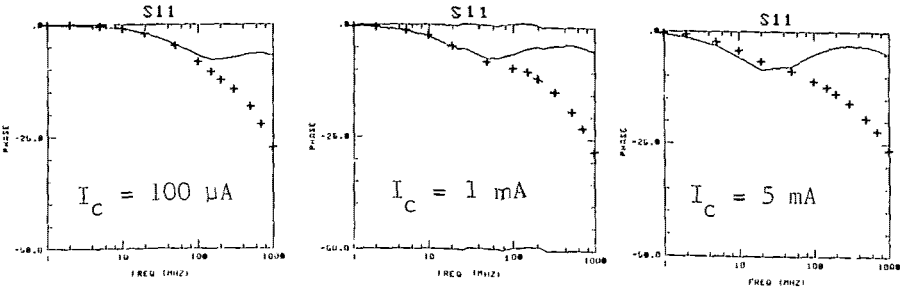

Figure 3a Comparison between one lump model and data (S11 phase)

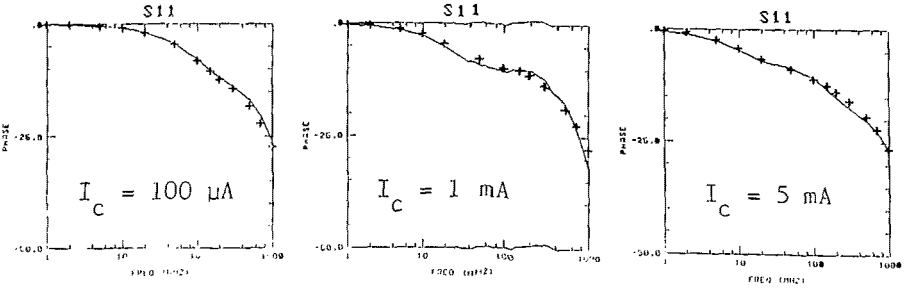

Figure 3b Comparison between two lump model and data (S11 phase)

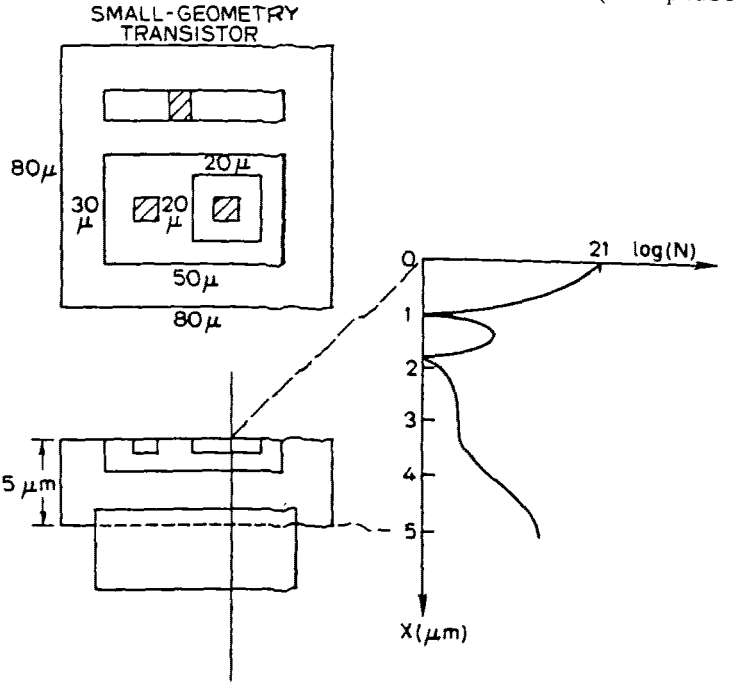

Figure 4 Surface geometry, cross-sectional view and doping profile of the bipolar transistor.

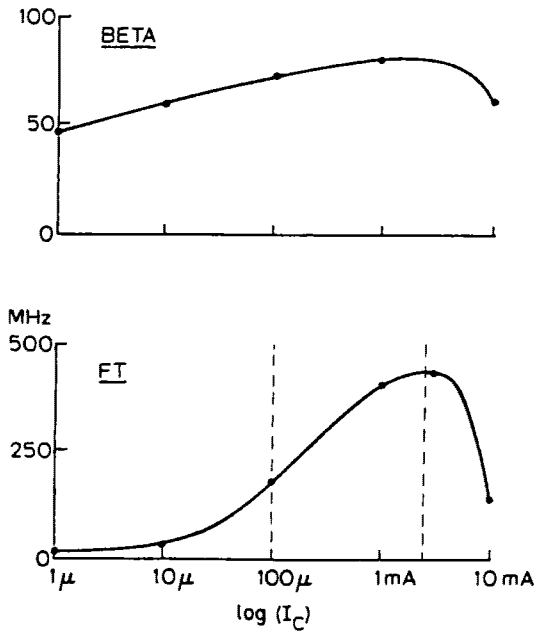

Figure 5 (a) Beta vs I_C plot
(b) f_T vs I_C plot of the bipolar transistor

B. Two Lump Model

The inadequacy of the one lump model to predict the high-frequency behavior is attributed to the transmission line characteristics of the active base (Figure 1). To improve the fit a multi-lump approach is generally taken [1-6]. For this study an empirical approach has been taken to fit S-parameter data using a two lump model with a split base (Figure 6).

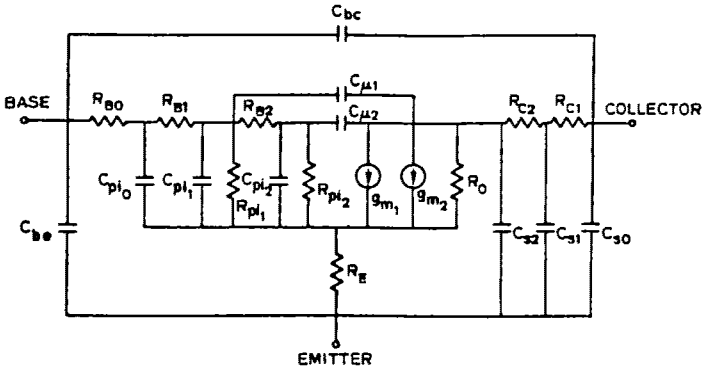

Figure 6 Two lump model

The two lumps were obtained by partitioning the one lump model. A scaling factor is employed to give the proper ratios of the two lumps. The factor, designated as F, can be regarded as an area ratio of the first lump to the total emitter area. Assuming that the current injection is uniform across the emitter area, the following parameters of the two lump model can be obtained from the one-lump element values:

$$g_{m1} = F \cdot g_m$$
$$g_{m2} = (1-F) \cdot g_m$$
$$R_{B1} = F \cdot R_B$$
$$R_{B2} = (1-F) \cdot R_B$$
$$Rpi_1 = Rpi/F$$
$$Rpi_2 = Rpi/(1-F)$$
$$Cpi_1 = F \cdot Cpi$$
$$Cpi_2 = (1-F) \cdot Cpi$$
$$C_{\mu 1} = F \cdot C_\mu$$
$$C_{\mu 2} = (1-F) \cdot C_\mu$$

and the other model elements unchanged.

However, if the current injection across the emitter area is not uniform, as in the case of high-level injection, F is no longer an area scaling factor. Instead, it is regarded as a current ratio, dependent on the bias condition. A more general definition of F is simply the ratio of the diffusion capacitance of the first lump Cpi_1 to the total diffusion capacitance Cpi_T.

Using a circuit analysis program [9] to represent the two lump model, S-parameters were generated and compared with measured data. The circuit elements in the model were adjusted to optimize the fit to all four S-parameters (S11,S12,S21,S22) in both magnitude and phase to within 5% and 5^o respectively. The results of the fit to S11 phase are illustrated in Figure 3b (compare this to Figure 3a).

C. Physical Basis for the Two Lump Model

The optimized two lump model has been shown to fit measured S-parameters, both magnitude and phase, at frequencies up to f_T for the full current range. The results of this study show that two mechanisms form the physical basis of this two lump model. One is AC crowding with a diode nonlinearity [7]. Based on this effect the device can be separated into two regions, one in the periphery of the emitter which sustains most of the AC current flow and the other in the center region with negligible AC response. This is valid for low current operation. At high currents, however, this effect becomes les important as most of the DC current flows along the periphery due

to DC current crowding [10]. In this high current regime it is the sidewall base charge as a result of current spreading [11, 12] which controls the high frequency behavior of the device. Thus the combined effect of AC crowding-diode nonlinearity and sidewall charge form the physical basis for lump partitioning.

1. AC Crowding and Diode Nonlinearity

The distributed base of an integrated bipolar transistor can be represented by distributed capacitances and resistances (Figure 7). Thus for AC analysis it can be treated as a transmission line. If a small AC base-emitter voltage signal is impressed at the start of the line, the voltage will attenuate along the line due to the complex impedance. As a result, the AC emitter current that responds to the base input voltage also attenuates along the line. This phenomenon is known as AC crowding [3], resulting in more AC current flowing near the edge than in the center where the input signal is weakest due to attenuation.

If a time average of the AC emitter current is evaluated, a net shift in the DC emitter current results due to the diode nonlinearity [7]. Based on this DC shift the transistor can be divided into two regions, one with a full AC response and the other with negligible response.

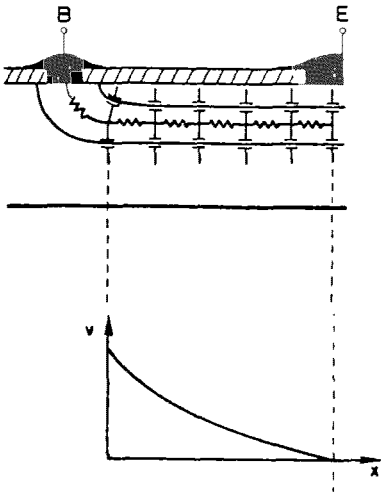

Figure 7 AC crowding

2. Sidewall Effect

At high current levels there is an additional effect to be considered. Due to lateral current spreading [11,12], there is stored charge in the sidewall base of the transistor (Figure 8). To assess the extent of the high-frequency effect due to these sidewall charges, a two-dimensional analysis is necessary to calculate the amount of minority charge (Cpi) stored in the sidewall. For this purpose a finite element numerical analysis was applied to the sidewall base transport of the bipolar transistors and a dynamic grid method developed to map current flow lines as well as orthogonal minority density contours. Figure 9 shows the output of the dynamic grid method using a 14x14 grid. The amount of emitter current, hence Cpi, in the sidewall is then calculated.

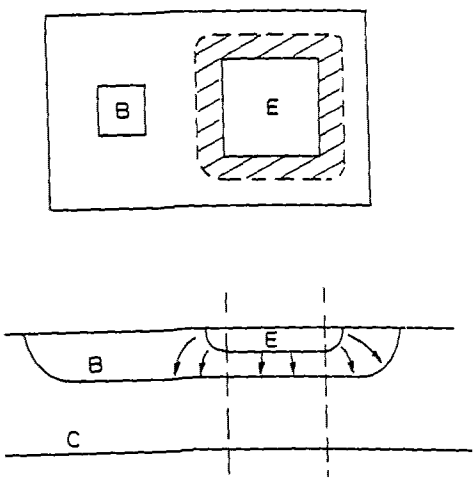

Figure 8 Sidewall transistor action

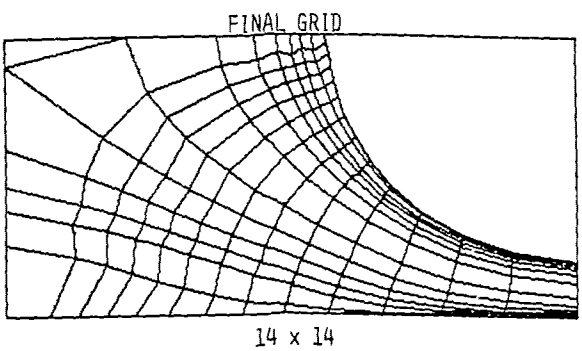

Figure 9 Dynamic grid output

3. Comparison with Measured Ratios

Experimental ratios F over the full current range have been obtained by optimizing the two lump model to fit measured S-parameters as described in Section II-B. The empirical results are shown as open circles in Figure 10, which is a plot of Cpi ratios of the sidewall to the total charge as a function of the collector current. The ratios calculated using the AC crowding/diode nonlinearity theory are depicted as open triangles. At low currents this effect is significant, as exhibited by a large ratio. However, for high currents, the ratio decreases to a negligible level, suggesting that the first lump is no longer important, which is contrary to what the data show. Thus the effect of AC crowding/diode nonlinearity is seen mainly in low current operation.

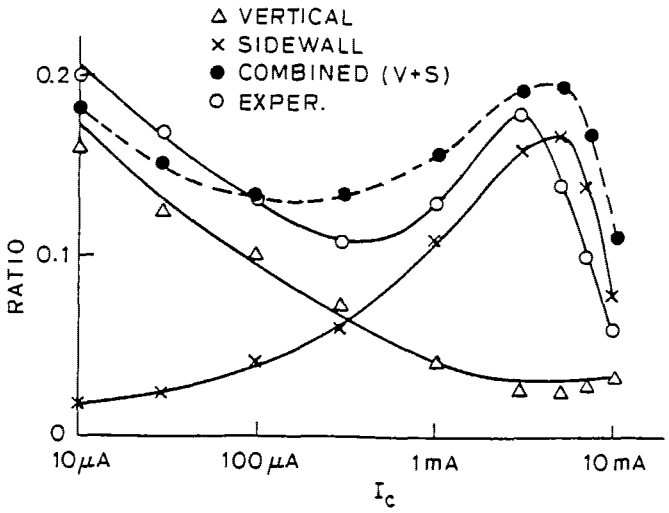

Figure 10 Sidewall effects on base minority charge ratio as a function of collector current.

The calculated ratios based on the sidewall analysis are shown as crosses in Figure 10. Indeed, the peak ratio occurs in the high current regime. This means that the effect of the first lump cannot be neglected due to the sidewall charge even though the effect of the AC crowding/diode nonlinearity is minimal at this current level. The high current effects such as base pushout and conductivity modulation have been studied and found to be negligible [8]. The slight increase of the ratios at very high currents (shown by the triangles in the figure) is due to these effects.

The combined effect of the AC crowding/diode nonlinearity and sidewall charge is shown as solid circles in Figure 10. This constitutes a set of predicted ratios based on physical principles. This set tracks the experimental ratios closely. A more detailed explanation of the qualitative fit is given

2. Sidewall Effect

At high current levels there is an additional effect to be considered. Due to lateral current spreading [11,12], there is stored charge in the sidewall base of the transistor (Figure 8). To assess the extent of the high-frequency effect due to these sidewall charges, a two-dimensional analysis is necessary to calculate the amount of minority charge (Cpi) stored in the sidewall. For this purpose a finite element numerical analysis was applied to the sidewall base transport of the bipolar transistors and a dynamic grid method developed to map current flow lines as well as orthogonal minority density contours. Figure 9 shows the output of the dynamic grid method using a 14x14 grid. The amount of emitter current, hence Cpi, in the sidewall is then calculated.

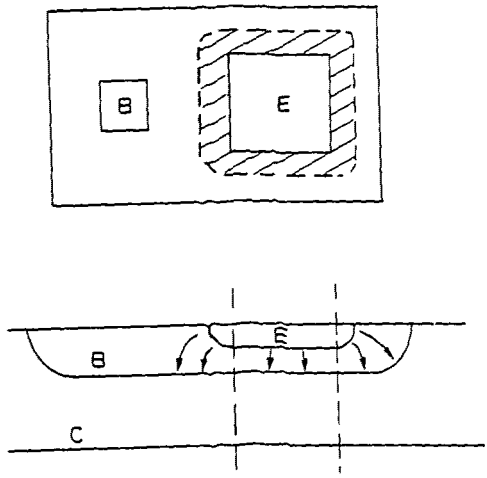

Figure 8 Sidewall transistor action

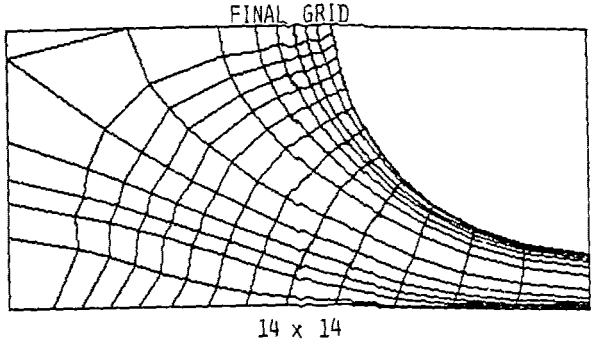

Figure 9 Dynamic grid output

3. Comparison with Measured Ratios

Experimental ratios F over the full current range have been obtained by optimizing the two lump model to fit measured S-parameters as described in Section II-B. The empirical results are shown as open circles in Figure 10, which is a plot of Cpi ratios of the sidewall to the total charge as a function of the collector current. The ratios calculated using the AC crowding/diode nonlinearity theory are depicted as open triangles. At low currents this effect is significant, as exhibited by a large ratio. However, for high currents, the ratio decreases to a negligible level, suggesting that the first lump is no longer important, which is contrary to what the data show. Thus the effect of AC crowding/diode nonlinearity is seen mainly in low current operation.

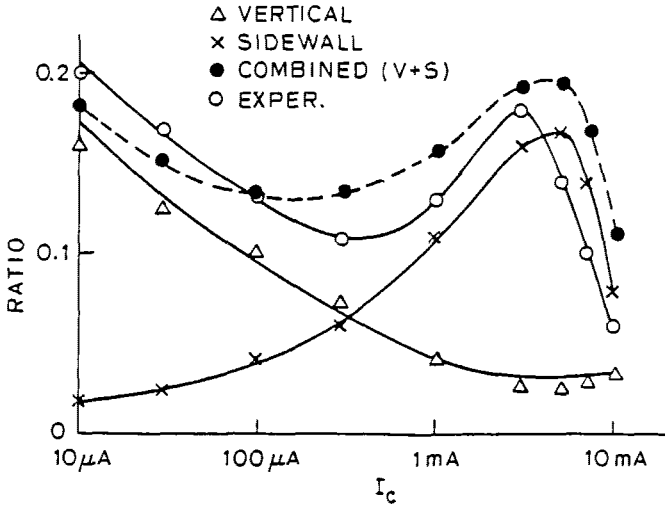

Figure 10 Sidewall effects on base minority charge ratio as a function of collector current.

The calculated ratios based on the sidewall analysis are shown as crosses in Figure 10. Indeed, the peak ratio occurs in the high current regime. This means that the effect of the first lump cannot be neglected due to the sidewall charge even though the effect of the AC crowding/diode nonlinearity is minimal at this current level. The high current effects such as base pushout and conductivity modulation have been studied and found to be negligible [8]. The slight increase of the ratios at very high currents (shown by the triangles in the figure) is due to these effects.

The combined effect of the AC crowding/diode nonlinearity and sidewall charge is shown as solid circles in Figure 10. This constitutes a set of predicted ratios based on physical principles. This set tracks the experimental ratios closely. A more detailed explanation of the qualitative fit is given

in Ref.[13]. In the next section, the numerical approach and the methodology of calculating the sidewall charge is described.

III. NUMERICAL APPROACH TO SIDEWALL ANALYSIS

A. Basic Semiconductor Device Equations

The current and charge distribution in the sidewall base of an integrated bipolar transistor is analyzed by solving a boundary-valued partial differential equation which describes the sidewall (Figure 11). The Poisson and continuity equations are solved in this region having both Neumann and Dirchlet boundary conditions.

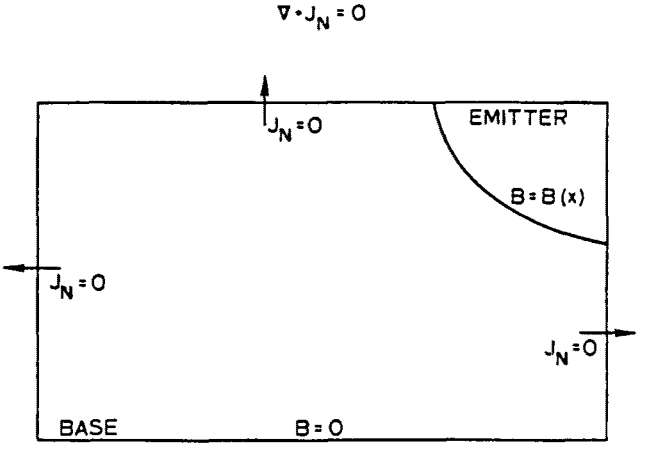

Figure 11 Boundary value problem of the continuity equation for the sidewall.

The basic equations for semiconductor device analysis are:

$$\text{Poisson Equation :} \quad \nabla^2 V = \frac{q}{\varepsilon}(n - p + N) \tag{1}$$

$$\text{Continuity Equation:} \quad q\frac{\partial p}{\partial t} = -\nabla \cdot J_p - qR \tag{2a}$$

$$q\frac{\partial n}{\partial t} = \nabla \cdot J_n - qR \tag{2b}$$

with the following auxiliary equations:

$$\text{Transport Equations:} \quad J_p = -qD_p \nabla p + q\mu_p pE \tag{3a}$$

$$J_n = qD_n \nabla n + q\mu_n nE \tag{3b}$$

$$\text{Electric Field Equation:} \quad E = -\nabla V \tag{4}$$

Recombination Equation:

$$R = \frac{pn - p_o n_o}{\tau_n(p + p_1) + \tau_p(n + n_1)} \qquad (5)$$

where

ε = semiconductor permittivity

V = voltage

q = electron charge

p = hole density

n = electron density

N = net doping density

J_p = hole current density

J_n = electron current density

R = net generation rate

D_p = hole diffusion coefficient

D_n = electron diffusion coefficient

μ_p = hole mobility

μ_n = electron mobility

E = electric field

p_o = equilibrium hole density

n_o = equilibrium electron density

τ_p = hole lifetime

τ_n = electron lifetime

p_1, n_1 = hole and electron concentrations that would exist if the Fermi level were at the trap level.

The present analysis is confined tothe sidewall base region of an integrated NPN bipolar transistor. Low-level injection in the base is assumed, so the electric field is not modified by excess carriers but fixed by the doping concentraion. Although the present analysis does not model high current injection, this is not a severe limitation to our modeling effort. Calculation of collector current for the vertical transistor by the dynamic grid method tracks the SEDAN [14] results (which include high level effects) for the full current range, with a maximum deviation of 35% at 10 mA [8], where the onset of high level effect occurs near 3 mA. With this assumption of low level injection, the electric field is

$$E = \frac{kT}{q}\frac{\nabla p}{p} = \frac{kT}{q}\frac{\nabla N}{N} \qquad (6)$$

Defining a variable B as the relative excess minority density

$$B = \frac{n_e}{n_o} \tag{7}$$

where n_e is the excess density and n_o the equilibrium density, and using the Einstein relation

$$\frac{D_n}{\mu_n} = \frac{kT}{q} \tag{8}$$

the continuity equation (2b) becomes

$$\nabla \cdot J_n = \nabla \cdot \left(\frac{qD_n n_i}{N} \nabla B \right) = 0 \tag{9}$$

for steady state condition with no recombination. The above equation is then solved using finite element method with boundary conditions shown in Figure 11.

B. Finite Element Formulation

Using the Galerkin's method of weighted residuals the divergence equation (9) is solved by:

$$\int_\Omega \phi_i (\nabla \cdot J_n - qR) ds = 0 \tag{10}$$

where ϕ_i is the interpolation function (triangular elements in this case) and Ω is the solution region (sidewall base).

To derive the set of finite element equations for Eq.(10), it is also first desirable to apply the divergence theorem to the equation. Since the integration is performed over the entire solution region, it is equivalent to integrating element by element throughout the region. In this way Eq.(10) can be expressed in the following form:

$$\sum_{e(i)} \int_{\Omega_e} \phi_i (\nabla \cdot J_n - qR) ds = 0 \tag{11}$$

where e(i) represents all the elements connected to node i and $\int_\Omega$ denotes integration over the area of the contributing element. Only those elements conncected to node i are included in the above equation as a direct consequence of the definition of ϕ_i.

$$\phi_i(x_j, y_j) = \begin{cases} 0, & i \neq j \\ 1, & i = j \end{cases} \tag{12}$$

Applying the divergence theorem [23] element by element results in

$$\sum_{e(i)} \int_{\Gamma_e} \phi_i J_n \bar{n} dL = \sum_{e(i)} \int_{\Omega_e} (\nabla\phi_i J_n - q\phi_i R) ds \quad (13)$$

where Γ_e is the boundary of the element and $\bar{n}$ is the unit vector normal to the boundary. For the nodes which are not "contact nodes" (boundary nodes that contain external current flow), the line integral in Eq.(13) is reduced to zero for the following reasons:

1. Across element interfaces there is a cancellation of line integral along sides shared by two elements.
2. On sides of elements which do not contain node i, ϕ_i is zero by definition.
3. When node i is not an interier point but a point on the surface of the interior domain where a Neumann condition prevails, there is no current flow through the node and thus

$$J_n \cdot \bar{n} = 0$$

Therefore

$$\sum_{e(i)} \int_{\Omega_e} (\nabla\phi_i J_n - q\phi_i R) ds = 0 \quad (14)$$

for i = 1, ..., M, where M is the number of nodes which are not contact nodes. This is equivalent to

$$\sum_{e(i)} \int_{\Omega_e} (\nabla\phi_i K \nabla B - q\phi_i R) ds = 0 \quad (15)$$

Using the finite representation of B

$$B(X,Y) = \sum_{j=1}^{N} \phi_j B_j(X,Y) \quad (16)$$

Eq.(15) becomes

$$\sum_{e(i)} \left\{ K\nabla\phi_i \nabla \{ \sum_{j=1}^{N} \phi_j B_j \} - q\phi_i R \right\} ds = 0 \quad (17)$$

for i = 1 to M. Using the property of the interpolation function as defined in Eq.(12) and rearranging terms

$$\sum_{e(i)} \left\{ \sum B_j \int_{\Omega} K\nabla\phi_i \nabla\phi_j ds - \int_{\Omega_e} q\phi_i R ds \right\} = 0 \quad (18)$$

where i = 1, ..., M and j runs through the indices which designate neighbors of node "i". We use the triangular elements, so there will be three such neighbors. The above equation is the finite element representation of Eq.(9). When the interpolation

function ϕ is specified, Eq.(18) will be reduced to a set of simultaneous algebraic equations with B_j as the unknown.

C. Current Formulation and Conservation Properties

In the application of the finite element method to semiconductor analysis, much attention has been focused on the problem of current continuity and conservation. This problem arises due to the inherent nature of the linear elements, that there is slope discontinuity at element interfaces. Thus current is not continuous across element boundaries, resulting in the violation of the conservation law. This problem can be overcome by using higher-order elements with a large number of internal nodes or simply by refining the element to an infnitestimal size. However, this approach is not feasible in computer analysis due to its complexity.

The approach by Barnes and Lomax [16-19] has been adopted for this study for the following reasons. First, the definition of current and the resulting conservation properties are consistent with the finite element theory. Secondly, it is a simple and direct approach. Both local and global current conservation are preserved, even for the simple, linear triangular elements.

The following is a direct application of Barnes and Lomax's definition of current to the sidewall base transport problem in this study. After applying the divergence theorem to Galerkin's formulation of the continuity equation for minority transport, Eqs.(11) and (13) result. If node i is not a contact point, the line integral is zero. If node i is a contact node with flux across the boundary, the line integral has a finite value and can be evaluated as

$$\int_{\Gamma} \phi_i J_n \bar{n} \, d\ell = \sum_{e(i)} \int_{\Omega_e} (\nabla \phi_i J_n - q\phi_i R) \, ds \qquad (19)$$

Now define the electron current (per unit length in the Z-direction) associated with (flowing into) node i as

$$I_{n_i} \triangleq \int_{\Gamma} \phi_i J_n \bar{n} \, d\ell \qquad (20)$$

As a result of applying the divergence theorem, this current density is also

$$I_{n_i} = \sum_{e(i)} \int_{\Omega_e} (\nabla \phi_i J_n - q\phi_i R) \, ds \qquad (21)$$

In a more general form

$$I_{n_i} = \int_{\Omega} (\nabla \phi_i J_n - q\phi_i R) \, ds \qquad (22)$$

where Ω covers the entire region. Equations (21) and (22) are actually equivalent as a result of the definition of the ϕ function. Equation (22) represents a general curent formulation, ascribing the current contributions from the elements to node i.

With this kind of current definition, when the currents are summed over all the contact nodes, the current conservation properties are preserved and the continuity equation is satisfied globally. When the currents are summed over the peripheral nodes of an element, the same conservation law holds. Thus current is also conserved locally, or on an element-by-element basis. This property of current conservation does not require inter-element slope continuity. For a proof of this global and local current conservation using the Barnes-Lomax approach, see Ref.(20).

D. The Dynamic Grid Method

The finite element analysis was applied to the sidewall base to calculate the distribution of current and charges. However, this analysis by itself is not sufficient to predict the charge ratios F because it does not generate current flow paths which are required in calculating Cpi. Hence, a dynamic grid method is developed to map the current flow lines starting from an initial grid. With such a map, Cpi is simply the line integral of the product of the transconductance (g_m) and the forward transit time (τ_F) along the current path.

$$C_T = \sum_{i=1}^{m-1} g_{m_i} \cdot \tau_{F_i} \qquad (23)$$

where m-1 is the number of current flow paths in the sidewall, g_m the transconductance and τ_F the transit time [21]

$$\tau_F = \int_0^W \left(\frac{1}{D_n N_B(x)} \int_x^W N_B(x)\, dx \right) dx \qquad (24)$$

where W is the total length of the flow path and D_n, N_B are respectively the minority diffusion coefficent and the doping concentration in the base.

The flow chart for the dynamic grid method (DGM) is shown in Figure 12. First, an initial grid is generated (Figure 13 shows a 14x14 initial grid). The requirement for the grid configuration is not too critical to the results. The primary constraint is that the lines generated should correspond symbolically to the approximate pattern of the current flow lines (see Figure 9) and rules for boundary conditions need to be observed.

Having established an initial grid, the next step is to set up equations to solve for DX and DY, the amount of change in grid positions. When DX and DY are solved for all the grid points, a new grid is formed by moving the old grid points by

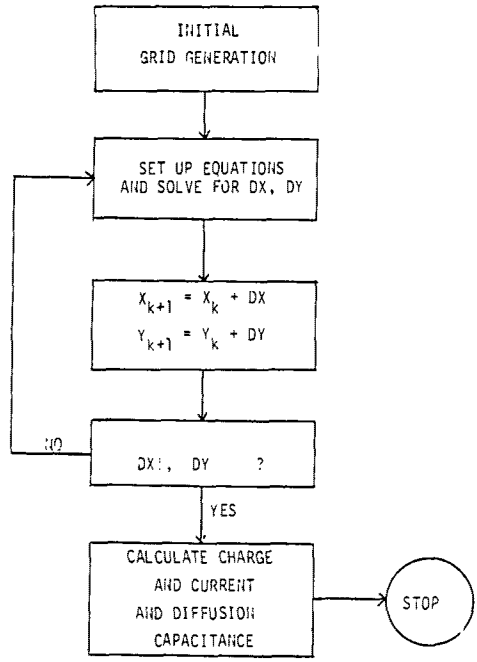

Figure 12 Flow chart for the DGM

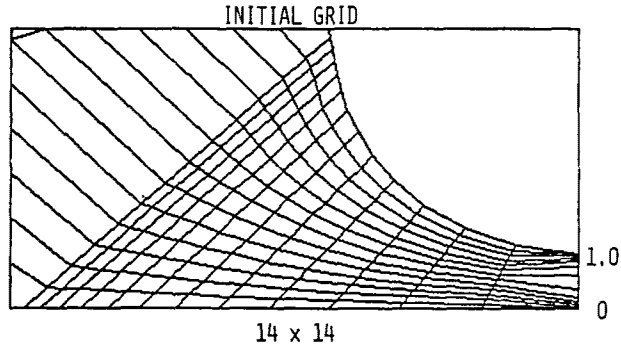

Figure 13 An initial grid for the DGM

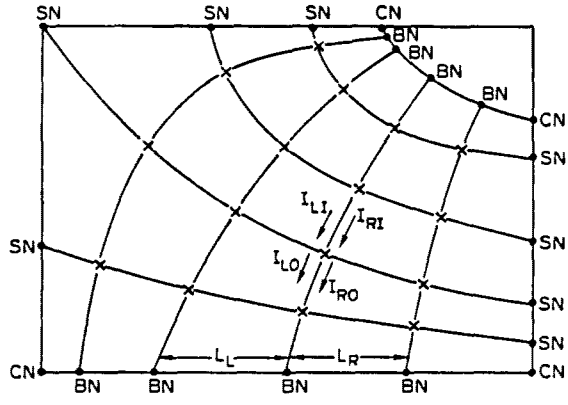

Figure 14 Designation of grid nodes and current flow

DX and DY displacement. Then a test is made to determine if further iterations are necessary. Depending on the outcome, the procedure is either repeated or stopped after calculating the currents and charges in the final grid. In this way, the DGM iterations continue until the solution coverges (see Figure 9 for the final grid).

The DGM makes use of the physical properties of current flow boundaries to formulate the necessary equations. At any node which is postulated to lie on a flow boundary, there should be no lateral current. In other words, current flux is conserved in each flow path. This property, when expressed in terms of current equations, becomes (see Figure 14)

$$I_{(RI)} = I_{(RO)} \tag{25a}$$

and

$$I_{(LI)} = I_{(LO)} \tag{25b}$$

where $I_{(RI)}$ and $I_{(RO)}$ are currents entering (IN) and leaving (OUT) a node due to elements associated with this node on the (R) flow path. $I_{(LI)}$ and $I_{(LO)}$ in the figure have similar interpretation. For boundary nodes, only one of the above equations is necessary.

The following is a description of the solution method of DGM. Starting from the basic current formulation, the result has been shown to be Eq.(13)

$$I_i = \sum_{e(i)} \int_{\Omega_e} F_i \, ds \tag{26}$$

where

$e(i)$ = elements containing node i

$$F_i = K \nabla \phi_i \cdot \nabla B - q \phi_i R \tag{27}$$

and

$K = qDn_i^2/n$

ϕ_i = interpolation function

B = relative excess minority density

R = net generation rate

Together with the application of the physical property of current along a flow boundary, for XN, SN nodes (Figure 14), the equations for the right and left sides are

$$\sum_{e_i(RI)} \int_{\Omega_e} F_i \, ds = \sum_{e_i(RO)} \int_{\Omega_e} F_i \, ds \tag{28a}$$

$$\sum_{e_i(LI)} \int_{\Omega_e} F_i \, ds = \sum_{e_i(LO)} \int_{\Omega_e} F_i \, ds \qquad (28b)$$

where $e_i(RI)$ and $e_i(RO)$ represent the elements contributing to currents entering or leaving node i on the rightof the node ($e_i(LI)$ and $e_i(LO)$ have similar meaning). For BN nodes:

$$L_L \cdot \sum_{e_i(RI)} \int_{\Omega_e} F_i \, ds = L_R \cdot \sum_{e_i(LO)} \int F_i \, ds \qquad (28c)$$

where L_L and L_R are left and right boundary segments at node i. Equation (28) can be written as:

$$H_i = \sum_{e(in)} \int F_i \, ds - \sum_{e(out)} \int F_i \, ds \qquad (29)$$

When Eq.(29) is assembled using Eq.(27) for all the nodes, a matrix equation results:

$$[H] \begin{Bmatrix} DX \\ DY \end{Bmatrix} = 0 \qquad (30)$$

The solution for this matrix equation is the set of displacements DX, DY to the grid nodes at which the current flux conservation requirement holds. This set of new positions represents the final grid.

Since current is a nonlinear function of the grid position, the matrix equation (30) is essentially nonlinear. Hence a direct solution is not possible, and a nonlinear equation solver is necessary. For this reason, the Newton-Ralphson iterative technique [15] has been selected for this study, for it has been extensively used and is a reliable method when the initial guess is suitably close to the final solution.

Using the Newton method, the following equation is solved

$$H'(\underline{Z}^k)(\underline{Z}^{k+1} - \underline{Z}^k) + H(\underline{Z}^k) = 0 \qquad (31)$$

where $\underline{Z}^k$ is the unknown vector of grid positions at k^{th} iteration and H' the Jacobian matrix. Alternatively,

$$D\underline{Z}^k = -(H')^{-1}(\underline{Z}^k)H(\underline{Z}^k) \qquad (32)$$

where $\underline{Z}^{k+1} = \underline{Z}^k + D\underline{Z}^k$.

At the k^{th} iteration, the Jacobian as well as the functional value of H are evaluated in terms of the known grid positions. The new grid positions will be the sum of the present positions and their respective changes. The entire process is then repeated until DZ converges.

E. Error Analysis of the Dynamic Grid Method

The error estimates for the finite element method can be found in the literature (for example, Strang and Fix [22]) and will not be included. Only the error analysis of the DGM is presented here. Since there is no analytical solution to the dynamic grid problem, the only way to check its accuracy is by a grid refinement technique. Within reasonable storage and cost of the IBM 360/168 used in this study, a maximum gridsize of 28x28 (obtained by using the Interpolative DGM [8]) was taken as the reference. The following grids were tested: 5x5, 7x7, 10x10, 14x14, 20x20. This series of grid stepsize was selected for its doubling of nodes of each step. For each grid size the DGM was allowed to run until the change in grid positions amounted to less than a millionth of the grid dimensions.

The total base charging capacitance (Cpi_T) for each grid was calculated and compared to that of the reference grid after the dynamic grid converges. For the grids listed above, the following errors were obtained: 200, 70, 25, 1.5, 0.66 % respectively. This error trend shows that the dynamic grid method is self-consistent and accurate.

IV. FORMULATION OF THE TWO LUMP MODEL FROM PHYSICAL PRINCIPLES

To formulate the two lump model suitable for circuit simulation, (e.g. SPICE analysis), first a vertical analysis is performed and the AC crowding/diode nonlinearity theory is applied to predict the Cpi ratio (F) as a function of collector current. Then a sidewall analysis using the Dynamic Grid Method follows to generate another set of Cpi ratio. The separate Cpi ratios obtained from the vertical and sidewall analyses are combined to produced a total ratio, based on which a two lump model can be formulated at each collector current level.

To generate the two lump circuit the following parameters can be split from the one lump value according to the calculated Cpi ratio: Cpi, Rpi, g_m, R_B, C_μ (see Section II-B). The major features of the two lump model (Figure 6) are the two base sections, which represent the two parallel transistors. R_{BO} is the base resistance in the inactive base and Cpi_O is the sidewall junction capacitance. These two elements model the diode action of the sidewall. The first transistor lump consists of R_{B1}, Rpi_1, Cpi_1, $C_{\mu 1}$ and g_{m1}. The second transistor lump consists of R_{B2}, Rpi_2, Cpi_2, $C_{\mu 2}$ and g_{m2}. In general, the first lump models the sidewall transistor action, and the second lump models the vertical transistor action. In low-current operation the first lump may also include a small part of the vertical transistor. In high-current operation the first lump represents only the sidewall, as the AC response is purely

an "edge" effect. In the overall modeling scheme, the first lump is the dominant part of the model which influences the high-frequency response of the device. The second lump is important in the intermediate frequency range. The two lumps together, acting as one lump, influence the DC-low frequency response.

V. DISCUSSION

A one lump model has been found inadequate to describe the high-frequency behavior of the bipolar transistor, particularly near the f_T range. A two lump circuit model is derived from an empirical fit to measured S-parameters. The results show marked improvement in fitting data with an accuracy of 5% in magnitude and 5^o in phase for frequencies up to f_T and over the full current range.

A physical basis, with the aid of a two-dimensional finite element analysis, has been presented to account for the lump partitioning. Close tracking of the predicted partitioning ratios to empirical ratios over the full current range confirms the validity of the physical theory.

The Dynamic Grid Method, based on the finite element theory as applied to the sidewall base, gives an estimate of the DC current/charge distribution of the bipolar transistors operated under different bias conditions. Although the fit to empirical data is approximate, showing the correct tracking, the sidewall analysis results in a physical understanding of how DC current flow influences high-frequency behavior of these devices.

Based on this study, the importance of the sidewall charge contribution to high-frequency device operation, particularly in the high current regime, cannot be neglected.

REFERENCES

1. PRITCHARD, R.L. and COFFEY, W.N.
"Small Signal Parameters of Grown Junction Transistors at High Frequencies"
IRE Convention Record (1954), Part 3, pp. 89-98.

2. PRITCHARD, R.L.
"Two-Dimensional Current Flow in Junction Transistors at High Frequencies"
Proc. IRE, pp. 1152-1160.

3. GHOSH, H.N.
"A Distributed Model of Junction Transistor and its Application in the Prediction of the Emitter-Base Diode Characteristics, Base Impedance and Pulse Response of the Device"
IEEE Trans. on Electron Devices, Vol. ED-12, No. 10, pp. 513-531, Oct. 1965.

4. LANGE, J. and CARR, W.N.
"An Application of Device Modeling to Microwave Power Transistors"
IEEE J. Solid State Circuits, Vol. SC-7, No. 1, pp. 71-80, February 1972.

5. MACNEE, A.B. and TALSKY, R.J.
"High-Frequency Transistor Modeling For Circuit Design"
IEEE J. Solid State Circuits, Vol. SC-7, No. 4, pp. 320-322, August 1972.

6. OLLINS, R.J. and RATNER, S.J.
"Computer-Aided Design and Optimization of a Broad-Band High-Frequency Monolithic Amplifier"
IEEE J. Solid State Circuits, Vol. SC-7, No.6, pp. 487-492, Dec. 1972.

7. RICHARDSON, R.E., Jr.
"Quiescent Operating Point Shift in Bipolar Transistor with AC Excitation"
IEEE J. Solid State Circuits, Vol. SC-14, No. 6, pp. 1087-1094, Dec. 1979.

8. CHAN, N.N.
"Lump Partitioning of Integrated Circuit Bipolar Transistor Models for High Frequency Applications"
Stanford Electronics Laboratories Technical Report, No. G201-4, August 1979.

9. NAGEL, L.W. and PEDERSON, D.O.
"Simulation Program with Integrated Circuit Emphasis"
16th Midwest Symposium on Circuit Theory, Waterloo, Ontario, April 12, 1973.

10. HAUSER, J.R.
"The Effects of Distributed Base Potential on Emitter-Current Injection Density and Effective Base Resistance for Stripe Transistor Geometries"
IEEE Trans. on Electron Devices, ED-11, No. 5, pp. 238-242, May 1964.

11. VAN DER ZIEL, A. and AGOURIDIS, A.
"The Cutoff Frequency Falloff in UHF Transistors at High Currents"
Proc. IEEE (Letter), Vol. 54, No. 3, pp. 411-412, Mar. 1966.

12. WHITTIER, R.J. and TREMERE, D.A.
"Current Gain and Cutoff Frequency Falloff at High Currents"
IEEE Trans. on Electron Devices, Vol. ED-16, No. 1, pp. 39-57, Jan. 1969.

13. CHAN, N. and DUTTON, R.
"Lump Partitioning of IC Bipolar Transistor Models for High-Frequency Applications"
Submitted to IEEE Trans. on Computer-Aided Design of Integrated Circuits and Systems.

14. D'AVANZO, D.C.
"One-Dimensional Semiconductor Device Analysis (SEDAN)"
Computer Aided Engineering of Semiconductor IC, J. D. Meindl, et al, Stanford University, Feb. 1978.

15. BJORCK, A. (translated by N. Anderson)
"Numerical Method"
Prentice Hall, Englewood Cliffs, New Jersey, 1974.

16. BARNES, J.J. and LOMAX, R.J.
"Two-Dimensional Finite Element Simulation of Semiconductor Devices"
Elect. Lett., Vol. 10, pp. 341-343, Aug. 8, 1974.

17. BARNES, J.J. and LOAMX, R.J.
"Transient Two-Dimensional Simulation of Submicrometer Gate-Length M.E.S.F.E.T."
Elect. Lett., Vol. II, pp. 519-521, Oct. 16, 1975.

18. BARNES, J.J. and LOMAX, R.J.
"Finite-Element Simulation of GaAs MESFETs with Lateral Doping Profiles and Submicron Gates"
IEEE Trans. Electron Devices, Vol. ED-23, pp. 1042-1084, September 1976.

19. BARNES, J.J.
"A Two-Dimensional Simulation of MESFETs"
Technical Report No. RADC-TR-76-153, Rome Air Development Center, Griffiss Air Force Base, New York, NY, May 1976.

20. BARNES, J.J. and LOMAX, R.J.
"Finite Element Methods inSemiconductor Device Simulation
IEEE Trans. on Electron Devices, Vol. ED-24, No. 8, pp. 1082-1089, August 1977.

21. MOLL, J.L. and ROSS, I.M.
"The Dependence of Transistor Parameter on The Distribution of Base Layer Resistivity"
Proc. IRE, Vol. 41, No. 1, pp. 72-78, Jan. 1956.

22. STRANG, G. and FIX, G.J.
"An Analysis of the Finite Element Method"
Prentice Hall, Inc., Englewood Cliffs, New Jersey, 1973.

23. SCHEY, H.M.
"Div, Grad, Curl, and All That"
W. W. Norton and Co., Inc., New YOrk, p. 48, 1973.