State University of New York at Stony Brook

College of Engineering Technical Report #344

SIMULATION OF A BIPOLAR TRANSISTOR BY AN

INFINITE NETWORK

Prasad Subramaniam and Armen H. Zemanian

Report No. 2
to the
Air Force Office of Scientific Research AFSC
for

Grant F49620 - 79 - C - 0172, AFOSR-80-205

September 1st, 1980

Simulation of a Bipolar Transistor by an Infinite Network[*]

Prasad Subramaniam and Armen H. Zemanian

Department of Electrical Engineering

State University of New York at Stony Brook

Abstract

In order to analyze a semiconductor device accurately and realistically, we need to study the two-dimensional behavior of the electrical properties of the device. There are methods available now which solve the current transport equations of the device using finite-difference or finite-element approximations. These consume much computer time. In a bipolar transistor it is the minority carriers in the base that determine the behaviour of the device. In this report, we present a method to solve for the two-dimensional distribution of the electron density in the base region of an n-p-n transistor. The current density and current continuity equations are reduced at steady state to a partial-differential equation by assuming low injection levels and neglecting carrier generation. The base region is divided into a square mesh and the partial-differential equation discretized on this mesh by using finite differences. The resulting problem is converted into one of finding the voltages in an infinite electrical network. The characteristic resistance approach is used to arrive at the solution with a considerable saving in computer time.

1. Introduction

A currently active area of research in semiconductor devices is the numerical computation of the electrical behavior of the device for various geometries and doping configurations. Methods of exhaustive two-dimensional analysis of the device are available which use finite-difference approximations of the basic semiconductor equations.[5] Other methods exist which use the finite element approach.[1] However there are no methods which provide for a simplified two-dimensional analysis. In this report we present a method of analysis for a specific geometry of a bipolar transistor as shown in Figure 1. In a bipolar transistor it is the minority carriers in the base region that determine the behavior of the device. A two-dimensional analysis of these carriers in the base would be useful in the understanding of the properties of the device. Our method solves for the two-dimensional minority-carrier density in the base with a considerable saving in computer time. In Section 2, we discuss the basic equations that govern the transistor. We make certain assumptions to simplify the equations and finally reduce the problem to one of solving a partial-differential equation with given boundary conditions. We then divide the base region into a fine square mesh and discretize this equation on this mesh using a finite-difference approximation. In Section 3 we discuss the characteristic resistance method for semi-infinite grids and see how we can exploit this method to solve our problem with a substantial reduction in computer time. In Section 4 we discuss the numerical aspects involved in the computation of the minority-carrier densities. In Section 5 we discuss a numerical example with typical values for the doping parameters.

2. Basic Semiconductor Device Equations

The geometric configuration of the device under consideration is shown in Figure 1. B_1 and B_2 denote the two edges of the depletion regions
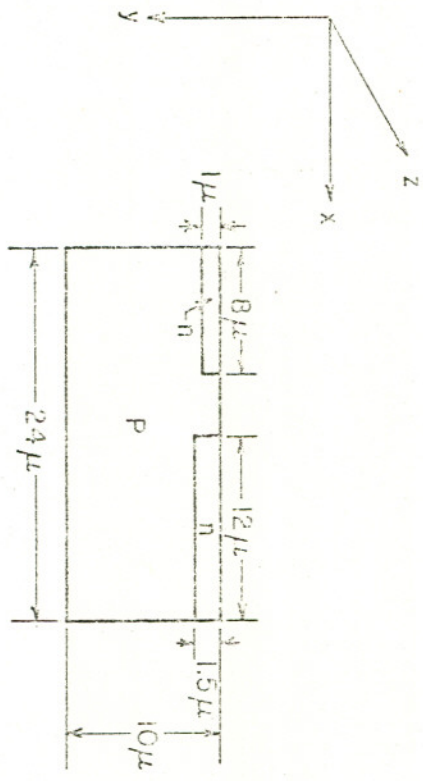
Figure . 1

bordering the p-type base region of the transistor. In the present exposition we have taken B_1 and B_2 with square corners; nevertheless our method allows rounded corners or more generally curvilinear B_1 and B_2. We use the simple theory of p-n junctions [2]. This includes a number of approximations chief among them being (a) The junction is abrupt and the doping is uniform in each region of the transistor. This assumption helps in simplifying the calculations needed to obtain the location of the depletion-layer edge. (b) The Boltzman boundary condition is used to calculate the carrier densities at the depletion layer edge. (c) The effects due to carrier generation and high-level injection are neglected.

Then assuming that all the potential drops occur across the junction depletion region, the current density and continuity equations are given by

$$J_n = - q\mu_n n_p \varepsilon + qD_n \nabla n_p \tag{1}$$

where J_n is the electron-current density

    q the electronic charge

    μ_n the mobility of the electron

    n_p the electron density in the base

    ε the electric field, and

    D_n the electron-diffusion constant

Also, $\dfrac{\partial n_p}{\partial t} = G_n - U_n + \dfrac{1}{q} \nabla \cdot J_n \tag{2}$

where G_n is the electron-generation rate, and

    U_n the electron-recombination rate.

Under low-injection conditions, U_n can be approximated by the expression

$$\frac{n_p - n_{po}}{\tau_n}$$

where n_{po} is the electron density in the base at thermal equilibrium and τ_n is the electron lifetime in the base. Equation (2) now reduces to

$$\frac{\partial n_p}{\partial t} = G_n - \frac{n_p - n_{po}}{\tau_n} + n_p \mu_n \nabla \cdot \varepsilon + \mu_n \varepsilon \cdot \nabla n_p + D_n \nabla^2 n_p \tag{3}$$

Under low-injection levels the term $n_p \mu_n \nabla \cdot \varepsilon$ can be neglected. By neglecting the generation rate and taking the steady-state condition, we get

$$0 = \frac{n_p - n_{po}}{\tau_n} + \mu_n \varepsilon \cdot \nabla n_p + D_n \nabla^2 n_p \tag{4}$$

In the neutral base region there is no electric field. Hence, only the diffusion term remains, yielding [6]

$$\nabla^2 n_p = \frac{n_p - n_{po}}{D_n \tau_n} \tag{5}$$

Let $n = n_p - n_{po}$ represent the excess carriers. Then (5) can be rewritten as

$$\nabla^2 n = \frac{n}{D_n \tau_n} \tag{6}$$

In general, n varies with the spatial coordinates. With known biases applied to the emitter and collector the boundary values of n at the emitter and collector depletion layer edge denoted by n_e and n_c respectively are given by

and
$$\begin{aligned} n_e &= n_{po}(\exp(V_{BE}/V_T) - 1) \\ n_c &= n_{po}(\exp(V_{CB}/V_T) - 1) \end{aligned} \tag{7}$$

where V_{BE} is the base-emitter forward voltage

V_{CB} is the collector-base reverse voltage, and

V_T is the Boltzmann voltage, given by

$$V_T = \frac{kT}{q}$$

where k is the Boltzmann constant, and

T is the absolute temperature.

The position of the depletion layer is obtained by solving the charge-neutrality equations of the device. If the doping is N_{DE}, N_A, and N_{DC} in the

emitter, base, and collector respectively and if W_{EB} and W_{CB} are the depletion layer widths at the emitter-base junction and at the collector base junction respectively, we have [2;p.126]

$$W_{EB} = \sqrt{\frac{2\varepsilon_S}{q} \left(\frac{N_A + N_{DE}}{N_A N_{DE}}\right) V_{BE}} \qquad (8)$$

and

$$W_{CB} = \sqrt{\frac{2\varepsilon_S}{q} \left(\frac{N_A + N_{DC}}{N_A N_{DC}}\right) V_{CB}}$$

If W_E and W_B are the depletion layer widths in the emitter and base respectively, making up the emitter-base depletion layer we can solve for W_E and W_B individually by using the following simultaneous equations.

$$W_{EB} = W_E + W_B \qquad (9)$$

and the charge neutrality equation

$$N_A W_B = N_{DE} W_E \qquad (10)$$

Combining (9) and (10), we get

$$W_E = \frac{N_A}{N_{DE} + N_A} W_{EB} \qquad (11)$$

Thus, we know the location of the edge of the emitter-base depletion layer in the base region. Similarly, we can find the edge of the collector-base depletion layer in the base region. Now, we are all set to solve equation (6) since we have the boundary values and the location of the boundaries where these values hold.

A closed-form solution to equation (6) is practically impossible to obtain due to the nature of the boundary. However, approximate numerical solutions can be obtained. We divide the base region of the n-p-n transistor into a fine square mesh and discretize (6) on this mesh by means of a finite difference approximation. We get the following equation

$$(4 + \frac{h^2}{D_n \tau_n}) \, n_{i,j} - n_{i,j-1} - n_{i,j+1} - n_{i+1,j} - n_{i-1,j} = 0 \qquad (12)$$

where h is the distance between adjacent nodes of the mesh, and $n_{i,j}$ is the excess-carrer density evaluated at the node (i,j) of the mesh.

Equation (12) can also be interpreted as Kirchoff's node equation at the node (i,j) of an electrical network as in Figure 2, where $n_{i,j}$ is taken to be the voltage at the node (i,j). The values for the voltage sources along B_1 and B_2 are taken to be E_1 and E_2. These values represent the discrete boundary values of the density of excess carriers, which is obtained from (7). We now have a resistive grounded grid network with a conductance $y_c = \frac{h^2}{D_n \tau_n}$ connecting every node to ground and unit conductance connecting every node to its adjacent nodes, on the plane of the grid.

In the transistors of integrated circuitry, the bottom surface of the semiconductor wafer has negligible effect on the variation of n in the region near the top surface where transistor behavior is taking place. For all practical purposes, we can assume that the bottom edge is effectively at infinity. However, ordinary circuit analysis does not permit us to compute the node voltage in a grid that extends infinitely downward. On the other hand, truncating the grid at some finite distance downwards leads us to analyses requiring impractically long computer times if the truncation occurs at realistic distances from the top surface. Truncations that are sufficiently close to the top sur- face to allow practicable computer times are two close to be representative of the actual configuration. This close proximity of the truncation distorts the variations in n_p from their actual values. Our solution to this dilemma is to assume that the semiconductor does extend infinitely downward and exploit a method for analysing an infinite uniform resistive grounded grid that occupies the half-plane [8]. This method is based on an extension of the characteristic resistance concept to uniform ladder networks of Hilbert ports. Since we have

assumed that doping is constant the diffusion constant and minority-carrier lifetime, which depend on the doping, may also be assumed to be constant. This leads us to a uniform grid structure. In a later report we will allow variations in these quantities.

3. The Characteristic Resistance Method for Infinite Grids

The characteristic impedance concept of analysing a uniform ladder network can be extended to infinite half-plane resistive grids to obtain a unique solution for the voltage and currents with a finite total power dissipation [8]. Normally there are infinite possible solutions to such a network since power can be fed in from infinity [7]. However, for practical purposes we are interested in the solution that obtains power only from sources within the network. It can be proved that the characteristic resistance method does indeed give this solution [8].

The first step in the analysis is to formulate an operator-valued ladder network representation of the grid of Figure 2. To this end, we modify that grid by converting the voltage sources into current sources. We start by shifting all the voltage sources at the boundary into voltage sources in adjacent branches in a standard fashion [3; p.131]. This is shown in Figure 3. In the process we delete all self-loops that arise under these alterations. We then convert the voltage sources to current sources by Norton's theorem. This gives us the network in Figure 4. Our objective at this point is to convert the network of Figure 4 into one that can be decomposed into a ladder network whose voltages and currents are members of Hilbert's space ℓ_{2r} of two-sided infinite vectors of quadratically summable real-valued elements, i.e., vectors,

$$x = [\ldots \ldots x_{-1} \ x_0 \ x_1 \ldots \ldots]^T \tag{13}$$

such that the x_j are real numbers and
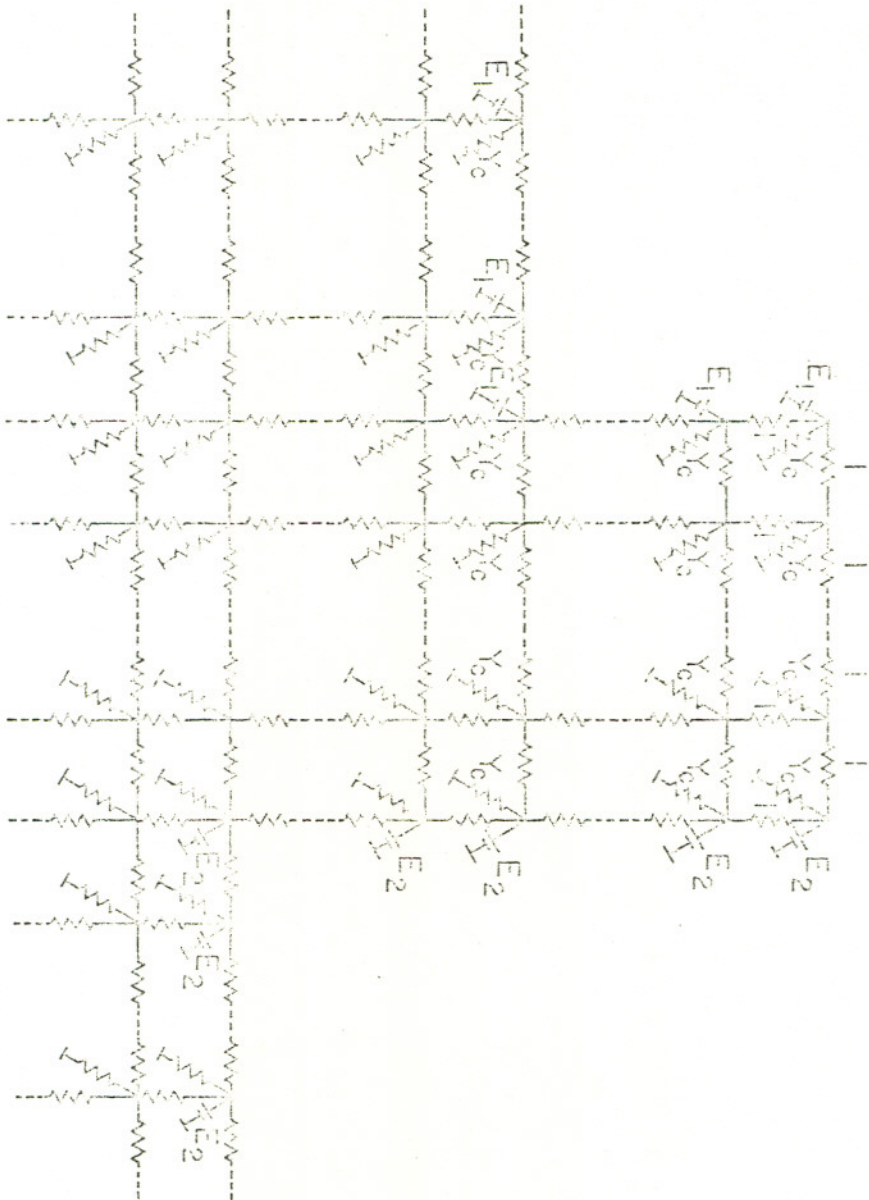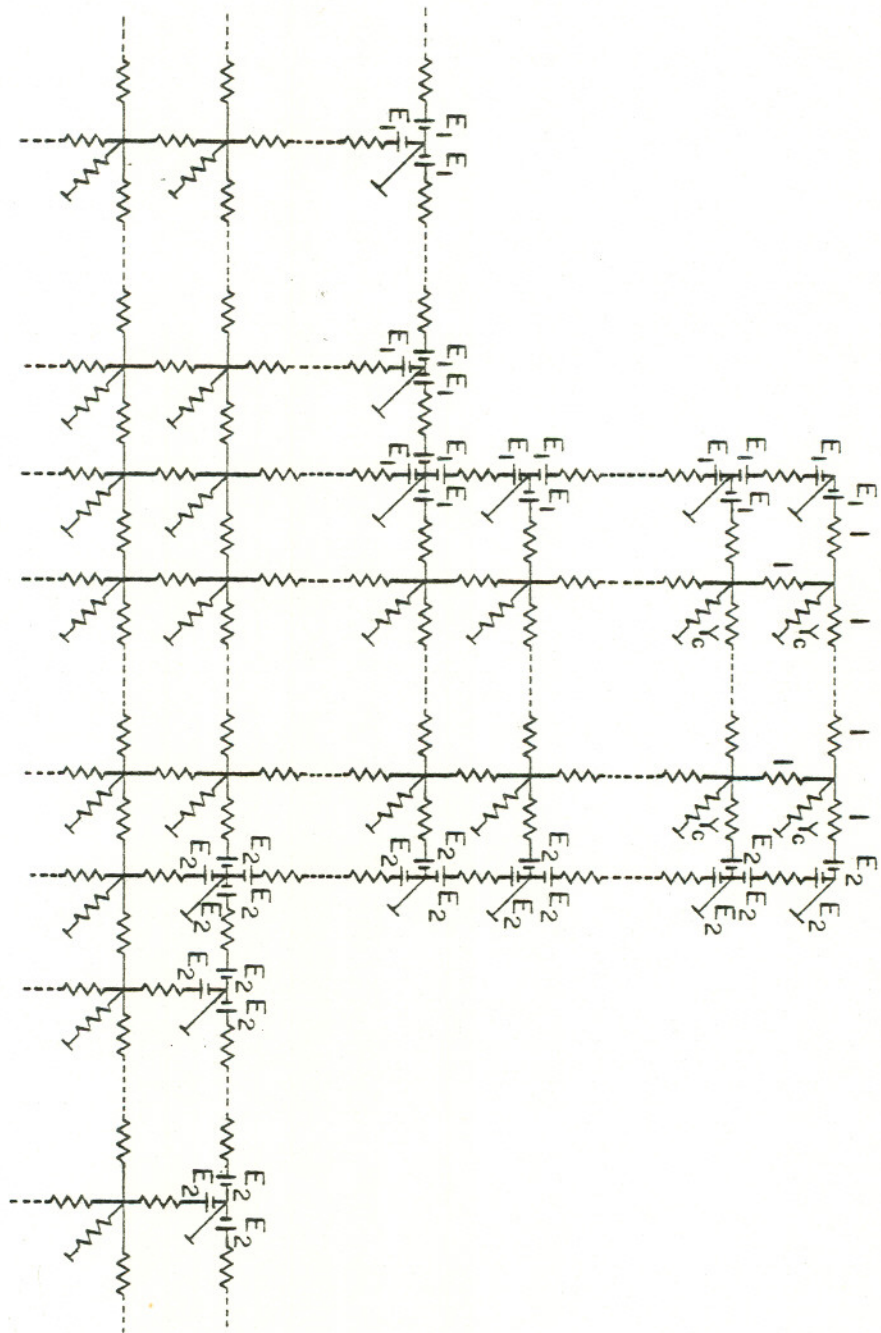
Figure 2

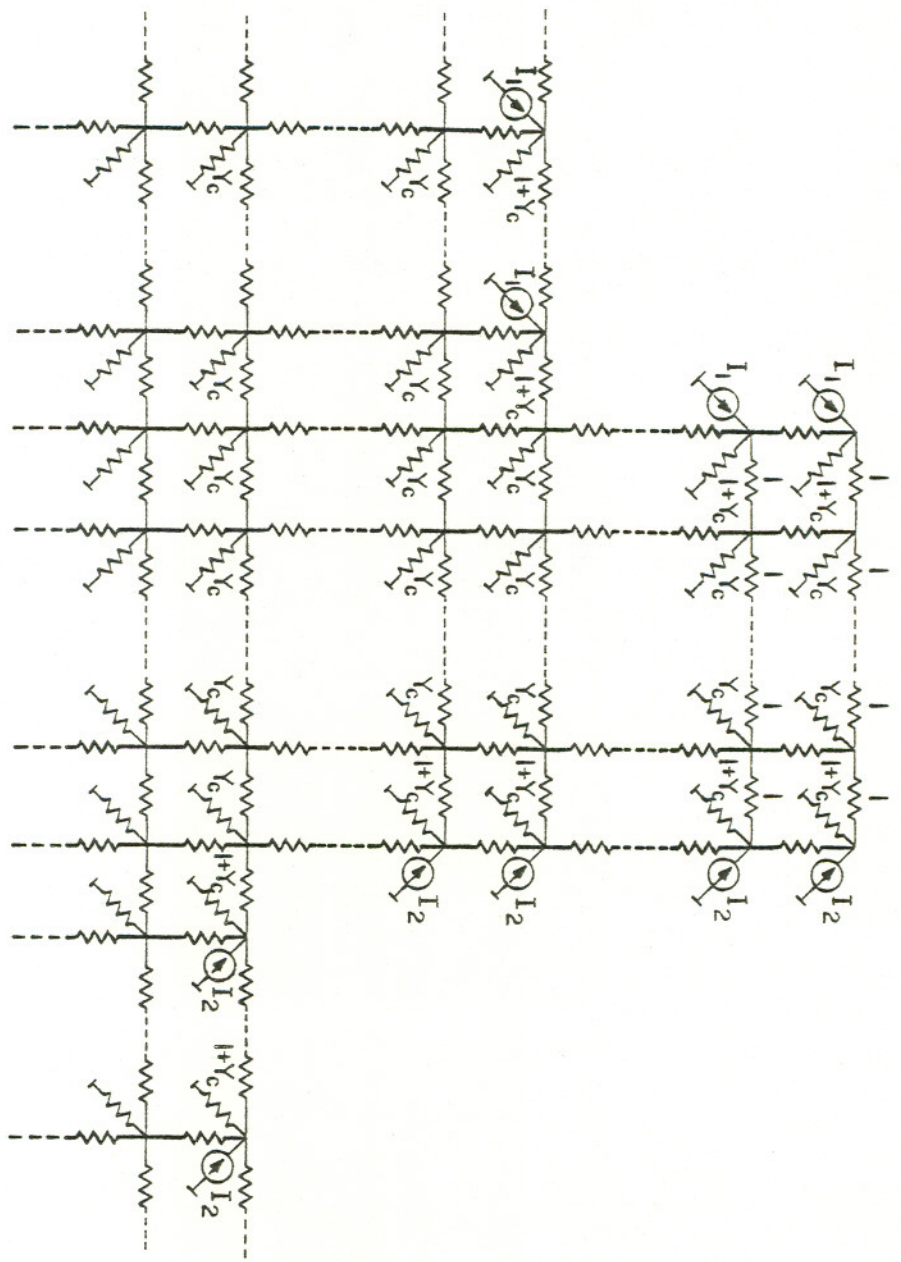Figure 3

$$\sum_{j=-\infty}^{\infty} x_j^2 < \infty .$$
(14)

Here, the superscript T represents matrix transpose. The admittances and impedances in the ladder network will be bounded linear operators on ℓ_{2r} To obtain such a decomposition we augment the network of Figure 4 with series connections of unit resistances as is indicated in Figure 5. Since the added circuitry does not introduce additional loops into the network, this augmentation does not alter the electrical properties of the network. The result is a rectangular pattern of nodes that fully occupies the half-plane.

We now treat each horizontal row of nodes with the branches between them as well as their branches to ground as a Hilbert port. Connecting every horizontal row of nodes to its succeeding horizontal row of nodes are vertical branches of unit resistance. Each such horizontal row of vertical branches is taken to comprise another Hilbert port. The resulting ladder network of Hilbert ports is shown in Figure 6. H, J, K and L are all members of ℓ_{2r} as follows. We denote the entries of H, J, K and L by h_ν, j_ν k_ν and ℓ_ν respectively, where $\nu = \ldots\ -1,o,1\ldots$ Then

$$h_\nu = \begin{cases} I_1 & \nu = 1 \\ I_2 & \nu = a \\ 0 & \text{otherwise} \end{cases}$$

$$j_\nu = \begin{cases} I_1 & \nu = 0, -1, -2\ldots \\ I_2 & \nu = a \\ 0 & \text{otherwise} \end{cases}$$
(15)

$$k_\nu = \begin{cases} I_2 & \nu = a \\ 0 & \text{otherwise} \end{cases}$$

$$\ell_\nu = \begin{cases} I_\nu & \nu = a+1, a+2,\ldots \\ 0 & \text{otherwise} \end{cases}$$
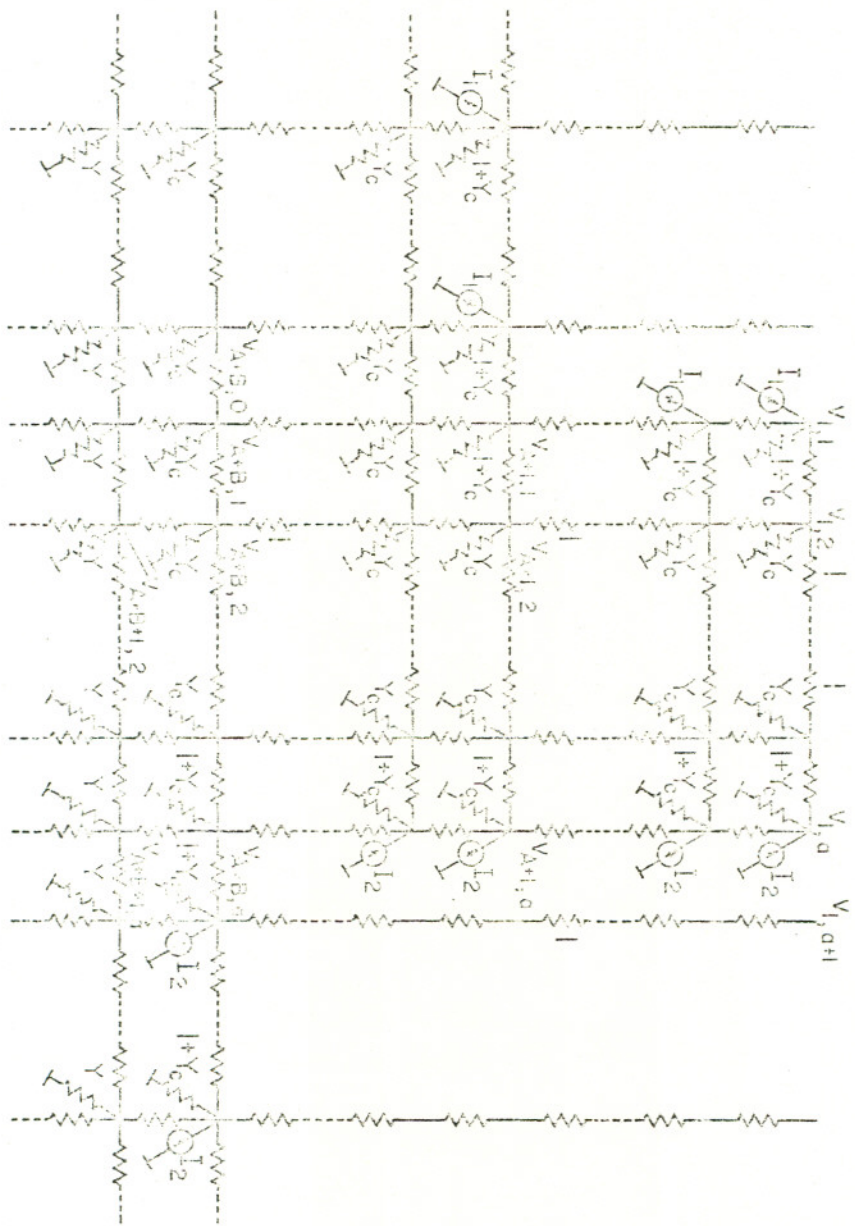
Figure 5

W, X, P and Q are admittance operators whose infinite matrix representations have entries as follows. We denote the entries of these matrices by $w_{\nu\mu}$, $p_{\nu\mu}$ and $q_{\nu\mu}$ respectively where $\nu,\mu = \ldots.. -1,0,1\ldots.$ Then,

$$w_{\nu\mu} = \begin{cases} 2+y_c & \nu=\mu= 1,2\ldots\ a \\ -1 & \mu=\nu+1;\ \nu=1,2\ldots.\ a-1. \\ -1 & \mu=\nu-1;\ \nu=2,3\ldots.\ a \\ 0 & \text{otherwise.} \end{cases}$$

$$x_{\nu\mu} = \begin{cases} 3+y_c & \nu=\mu= \ldots\ldots\ -2,\ -1,0. \\ 2+y_c & \nu=\mu= 1,2\ldots.\ a \\ -1 & \mu=\nu+1;\ \nu= \ldots\ldots.\ -1,0,1\ \ldots..\ a-1. \\ 1 & \mu=\nu-1;\ \nu= \ldots\ldots.\ -1,0,1\ \ldots..\ a \\ 0 & \text{otherwise.} \end{cases} \qquad (16)$$

$$p_{\nu\mu} = \begin{cases} 2+y_c & \nu=\mu= \ldots\ldots\ -1,0,1\ \ldots..\ a \\ -1 & \mu=\nu+1;\ \nu= \ldots..\ -1,0,1\ \ldots.\ a-1. \\ -1 & \mu=\nu-1;\ \nu= \ldots.\ -1,0,1\ \ldots.a \\ 0 & \text{otherwise.} \end{cases}$$

$$q_{\nu\mu} = \begin{cases} 2+y_c & \nu=\mu= \ldots..\ -1,0,1\ \ldots\ a \\ 3+y_c & \nu=\mu= a+1,\ a+2\ \ldots.. \\ -1 & \mu=\nu+1;\ \nu= \ldots.\ -1,0,1\ \ldots. \\ -1 & \mu=\nu-1;\ \nu= \ldots.\ -1,0,1\ \ldots. \\ 0 & \text{otherwise} \end{cases}$$

Finally, the impedance operator Z is the identity operator on ℓ_{2r}, i.e., $Z=1$. Y is defined as a Laurent matrix with elements $y_{\nu\mu}$ given by

$$y_{\nu\mu} = \begin{cases} y_c+2 & \nu=\mu = \ldots..\ -1,0,1\ \ldots. \\ -1 & \mu=\nu-1;\ \nu= \ldots..\ -1,0,1\ldots. \\ -1 & \mu=\nu+1;\ \nu= \ldots..\ -1,0,1\ldots \\ 0 & \text{otherwise.} \end{cases} \qquad (17)$$

The ladder network beyond the node (A+B+1) can be replaced by its characteristic impedance Z_o, since we make the assumption that the repetitive structures that follow continue indefinitely. We evaluate this characteristic impedance by pursuing the method adapted in [8]. The operators Z,W,X,P,Q and Y represent the Hilbert ports, and, since the interconnection of the ports preserves the port conditions, the solution of the ladder network of Figure 6 is the same as that of the grid of Figure 2. A formal mathematical proof is provided in [8].

4. Computational Aspects

In order to be able to compute the voltages and currents of the grid network of Figure 2 on a computer, certain modifications have to be made. It is no longer possible to work with Hilbert ports since computers have only a finite memory. Moreover the transistor we are modelling is anyway only finite. We thus truncate all operators into finite matrices in such a way that the grid encompasses the entire region of the transistor. Thus all matrices are finite and of order N where N is number of interior points on the grid of Figure 2. The characteristic impedance Z_o is now a truncated version of the Hilbert operator Z_o.

Next, we write down Kirchoff's node equations at nodes 1 through (A+B+1) of the multiport ladder network. where I_k, V_k, H, J, K and L are now vectors in R^N and W, X, Y, P, Q, Z and Z_o are matrices in R^N. The current equations are

$$
\begin{aligned}
H &= WV_1 + I_1 \\
H &= WV_2 + I_2 - I_1 \\
H &= WV_3 + I_3 - I_2 \\
&\quad\bullet \quad \bullet \quad \bullet \quad \bullet \quad \bullet \quad \bullet \quad \bullet \quad \bullet \quad \bullet \\
H &= WV_A + I_A - I_{A-1} \\
J &= XV_{A+1} + I_{A+1} - I_A \\
K &= PV_{A+2} + I_{A+2} - I_{A+1} \\
&\quad\bullet \quad \bullet \quad \bullet \quad \bullet \quad \bullet \quad \bullet \\
K &= PV_{A+B} + I_{A+B} - I_{A+B-1} \\
L &= QV_{A+B+1} + I_{A+B+1} - I_{A+B}
\end{aligned}
\tag{18}
$$

Figure 6

The voltage equations are:

$$V_1 - V_2 \quad = \quad Z I_1$$
$$V_2 - V_3 \quad = \quad Z I_2$$
$$- - - - - - - - - - - -$$
$$V_{A+B} - V_{A+B+1} = Z I_{A+B}$$
$$V_{A+B+1} \quad = \quad (Z + Z_0) I_{A+B+1}$$

(19)

These equations can be combined into the following matrix equations in partitioned form:

$$
\begin{bmatrix}
W & & & & & & \\
 & W & & & & & \\
 & & \ddots & & & & \\
 & & & W & & & \\
 & & & & P & & \\
 & & & & & P & \\
 & & & & & & P \\
 & & & & & & \quad Q
\end{bmatrix}
\begin{bmatrix}
V_1 \\
V_2 \\
\vdots \\
\vdots \\
\vdots \\
\vdots \\
V_{A+B+1}
\end{bmatrix}
+
\begin{bmatrix}
1 & & & & \\
-1 & 1 & & & \\
 & -1 & 1 & & \\
 & & \ddots & & \\
 & & & -1 & 1 \\
 & & & & -1 & 1
\end{bmatrix}
\begin{bmatrix}
I_1 \\
I_2 \\
\vdots \\
\vdots \\
\vdots \\
I_{A+B+1}
\end{bmatrix}
=
\begin{bmatrix}
H \\
H \\
\vdots \\
H \\
J \\
K \\
\vdots \\
K \\
L
\end{bmatrix}
$$

(20)

and

$$
\begin{bmatrix}
1 & -1 & & & \\
 & 1 & -1 & & \\
 & & \ddots & & \\
 & & & 1 & -1 \\
 & & & & 1
\end{bmatrix}
\begin{bmatrix}
V_1 \\
V_2 \\
\vdots \\
\vdots \\
V_{A+B+1}
\end{bmatrix}
=
\begin{bmatrix}
Z & & & \\
 & Z & & \\
 & & \ddots & \\
 & & & Z \\
 & & & & (Z + Z_0)
\end{bmatrix}
\begin{bmatrix}
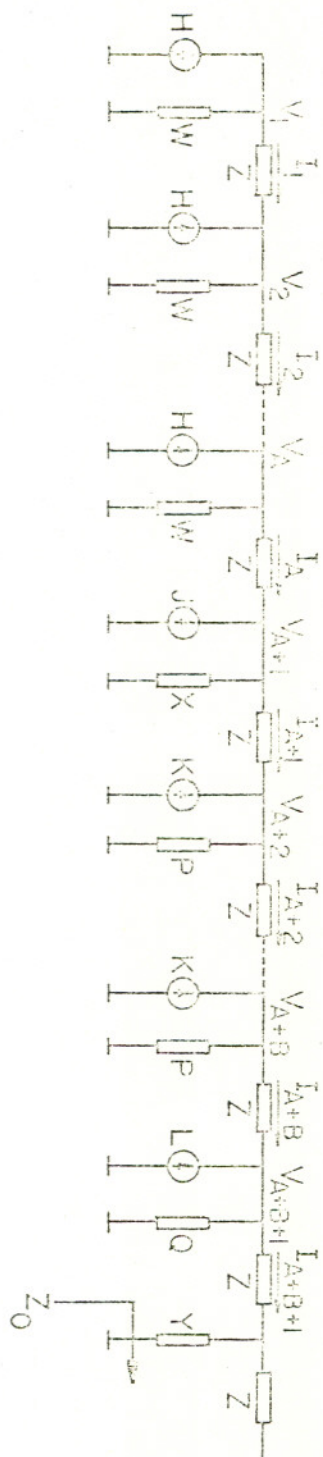I_1 \\
I_2 \\
\vdots \\
\vdots \\
I_{A+B+1}
\end{bmatrix}
$$

(21)

where 1 represents the identity matrix of order N. In this formulation
and   are the vectors that represent the voltage and current of the multiport
ladder network. Equation (20) and (21) can be written as

$$AV + \vec{B}I = C$$

(22)

and

$$BV = DI$$

where V, I and C are vectors of dimension $N(A+B+1)$ and A, B and D
are matrices of the same dimension. The voltage V whose elements yield the
minority-carrier densities along the interior nodes of Figure 2 can be ob-
tained by

$$\left(A + B^T D^{-1} B\right) V = C$$

(23)

Z and Z_o are truncations of positive Hilbert operators. D is hence
invertible, and $\left(A + B^T D^{-1} B\right)$ is also nonsingular, being a positive-definite
symmetric of the form

$$\begin{bmatrix}
W+z^{-1} & -z^{-1} & & & & & & \\
-z^{-1} & W+2z^{-1} & -z^{-1} & & & & & \\
 & \cdots & \cdots & \cdots & & & & \\
 & & -z^{-1} & W+2z^{-1} & -z^{-1} & & & \\
 & & & -z^{-1} & X+2z^{-1} & -z^{-1} & & \\
 & & & & -z^{-1} & P+2z^{-1} & -z^{-1} & \\
 & & & & & \cdots & \cdots & \cdots \\
 & & & & & -z^{-1} & P+2z^{-1} & -z^{-1} \\
 & & & & & & -z^{-1} & Q+z^{-1}+(Z+Z_o)^{-1}
\end{bmatrix}$$

(24)

All the components of $(A + B^T D^{-1} B)$ can be evaluated easily except for $(Z + Z_o)^{-1}$
A simple method of evaluating $(Z + Z_o)^{-1}$ is by truncating the Hilbert operator
$(Z+Z_o)^{-1}$ To do this, we exploit the natural isomorphism that exists between
the Hilbert coordinate space ℓ_{2r} and the space of functions quadratically
integrable on the unit circle. We can then show that the Laurent operators
Y and Z correspond to multiplication by

$$\tilde{y}(x) = 2 + y_c - 2\cos x$$
$$\tilde{z}(x) = 1. \tag{25}$$

The characteristic impedance Z_o is then mapped under the isomorphism into
multiplication by

$$\tilde{Z}_o(x) = \frac{1}{2} \left[\left(1 + \frac{4}{\tilde{y}(x)}\right)^{1/2} - 1 \right] \tag{26}$$

The Hilbert operator $(Z+Z_o)^{-1}$ is thus mapped under the isomorphism into multi-
plication by

$$\tilde{A}(x) = 2 \left\{ \left(1 + \frac{4}{\tilde{y}(x)}\right)^{1/2} + 1 \right\}^{-1} \tag{27}$$

If a_m are elements of the main row of the Laurent matrix $(Z+Z_0)^{-1}$ with $a_m = a_{-m}$ we have

$$\tilde{A}(x) = \sum_{m=0}^{\infty} a_m \cos mx$$

and
$$a_m = \frac{1}{\pi} \int_0^{\pi} \tilde{A}(x) \cos mx \, dx \qquad m = 0,1,2.... \tag{28}$$

Now we need to compute only as many a_m as are needed for the desired accuracy.
The truncated matrix $(Z + Z_o)^{-1}$ is of order N and has 2M+1 non-zero diagonal
elements, M+1 being the number of a_m that are to be evaluated. We set $a_m = 0$
for $M+1 < m \leq N$.

In order to solve equation (23) more efficiently, we can reduce the
order of the equation. Since many of the components of the   and
for k≤A+B fall outside the base region of the transistor, they have no physi-
cal significance as far as the base minority-carrier density is considered.
They were earlier introduced merely to facilitate the mathematical theory
of the Hilbert-space operators, of which they formed a part. As far as compu-
tational purposes are concerned, they can be eliminated by truncating the
matrices appropriately. This leaves us with a matrix $(A+B^T D^{-1} B)$ of much
smaller dimension. It is still of the same form, but the individual blocks
are no longer of the same dimension N. The first A blocks are of dimension
N_L where N_L is the number of grid points at the top edge of the base region.
The next B blocks are of dimension N_{TL} which is the number of grid points
along the horizontal line passing through the boundary B_1. The last block is
of dimension N. Thus, the order of the system has been considerably reduced.
Every component of the voltage vector   now represents a node in the base
region of the transistor.

Since the matrix $(A + B^T D^{-1} B)$ is symmetric, positive-definite, and block
tridiagonal, it is easier to solve (23) iteratively rather than by a direct
inversion. A direct inversion would be expensive on the computer both in terms
of computer time and memory. All the elements of the matrix would have to be
stored and despite the fact that the matrix is symmetric and contains many
nonzero elements, the size of storage required for reasonable values of A, B
and N is large. A direct-inversion process does not take advantage of the
small number of non-zero elements, since the inverse is rarely a sparse matrix.
For a large system the number of computations involved to obtain the components
of the inverse matrix and then calculate the solution is of the order of the
cube of the system dimension. On the other hand, an iterative method requires
minimal storage. There are only a handful of distinct elements in the matrix

which have to be stored. Convergence of most iteration methods is guaranteed due to the positive-definite nature of the system. [4. p.50]

The successive overrelaxation method was used to solve equation (23). A problem arises in the choice of the optimum relaxation factor. Before we proceed any further, let us define a few preliminaries.

Definition 1.   A matrix M is said to be diagonally block tridiagonal if it is of the following form

$$M = \begin{bmatrix} A_1 & B_1 & & & \\ C_2 & A_2 & B_2 & & \\ & \cdot & \cdot & \cdot & \\ & & & C_n & A_n \end{bmatrix} \tag{29}$$

where A_i are square diagonal matrices and C_i and B_i are rectangular matrices.

Definition 2.   A matrix M is said to have property A if there exists a linear transformation T such that $T^{-1}MT$ is diagonally block tridiagonal.

The successive overrelaxation method has been studied in detail for matrices that possess property A and the optimum overrelaxation factor for such matrices is given by [4; p.56]

$$w_{opt} = \frac{2}{1 + \sqrt{1-\lambda_1^2}} \tag{30}$$

where λ_1 is the largest eigenvalue of the iteration matrix which can be obtained by decomposition of the system matrix M. Now, most system matrices that arise out of finite-difference approximations of boundary-value problems do possess the property A. The matrix in question $\left(A + B^T D^{-1} B\right)$ however does not possess this property. The anamoly is due to the presence of the characteristic impedance terms Z_o in the last block. However, it is reasonable to assume that the matrix $\left(A + B^T D^{-1} B\right)$ almost has the property A since the last block forms only a small portion of the matrix and the matrix is mainly dominated

by the other blocks. Thus (30) would give a near-optimum choice for the para-
meter w. We now need to get an approximate estimation of λ_1. We start the
iteration with an initial value of w = 1. After a few iterations, the quotients
of the solution norms converge to λ_1^2. We can save some computer time by
choosing some arbitrary component of the solution vector and taking the
quotient of two such successive components to arrive at λ_1^2. In our method the
first component v_1 was used. Thus

$$w = \frac{2}{1 + \sqrt{1 - v_1^{k+1}/v_1^k}} \tag{31}$$

was chosen, where v_1^k is the value of v_1 after the kth iteration. This approx-
imation is justified since it is expected that individual components would vary
in the same fashion as the norm.

In order to simplify computations, the whole system was scaled down by a
factor of n_{po}. Furthermore, since the boundary conditions were incompatible,
E_1 being of the order of 10^9 and E_2 being of the order of 1, two different
sets of solutions were obtained, one with the second boundary condition set
equal to zero, the other with the first boundary condition set equal to zero.
Then two solutions were superposed. This procedure is valid since we are
dealing with a linear system where the principle of superposition holds good.
The solution was then appropriately scaled by the factor n_{po}.

5. Example.

As an example, we will consider a transistor with a uniformly doped
base, collector, and emitter. The base doping is $N_A = 4.428851 \times 10^{13}/cm^3$, the
emitter doping is $N_{DE} = 10^{16}/cm^3$, and the collector doping is $N_{DC} = 10^{15}/cm^3$. We
assume that all impurities are ionized. Let the collector-base reverse-bias
voltage be $V_{CB} = 500mV$ and base-emitter forward bias voltage be $V_{BE} = 560mV$. The
dimensions of the transistor are indicated in Figure 1. The depletion widths

at the collector and emitter junctions are evaluated. The location of the depletion layer edge is found to be approximately 0.1 μ from the junction edge. The base region is then divided into a mesh with h = 0.1μ. This gives us $N_L = 39$ points at the top edge of the base, N_{TL} = 120 points along the horizontal line through the boundary B_1, and finally N = 241 points. The value of Y_c is calculated to be 3.0959752×10^{-6} The scaled boundary values along B_1 and B_2 are then evaluated to be E_1 = 2.2596178×10^9 and E_2 = -1 in units of n_{po}. The other parameters are A = 11 and B = 5 which are dictated by the geometry. D_n and τ_n are assumed to be constants with values of 32.3 cm^3/sec and 1μsec respectively. w was chosen to be 1 for the first 50 iterations. At the end of 50 iterations an approximate value of w was calculated from (31). w's of 1.93 and 1.74 were obtained for the two solutions which were then superposed and scaled. The iterations took 273 and 320 steps respectively to converge. The total computer time on the UNIVAC 1100 multiprocessing system was about 1 minute for the entire computation. Figure 7 gives a two-dimensional profile of the electron minority-carrier distribution as obtained by this method.

6. Conclusion

In this report a new method has been presented to solve for the two-dimensional variation of the minority-carrier densities in the base region of a bipolar transistor. The characteristic resistance approach reduces the size of the linear system to be solved and hence saves considerable computation. Further research is presently being carried out with non-uniform doping densities and hence variable diffusion constant and carrier lifetime.

References

1. J.J. Barnes and R.J. Lomax, "Finite Element Methods in Semiconductor Device Simulation" IEEE Transactions on Electron Devices, Vol. ED-24 No. 8, August 1977. p. 1082-1089.

2. J.L. Moll, "Physics of Semiconductors", McGraw-Hill Book Company 1964.

3. B. Peikari, "Fundamentals of Network Analysis and Synthesis", Prentice-Hall, Englewood Cliffs 1974.

4. H.R. Schwartz, H. Rutishauser, R. Stiefel and P. Hertelendy, "Numerical Analysis of Symmetric Matrices", Prentice Hall, Englewood Cliffs 1973.

5. J.W. Slotboom, "Computer-aided two-dimensional analysis of bipolar transistors", IEEE Transactions on Electron Devices, Vol. ED-20 No. 8, August 1973. pp. 669-679.

6. S.M. Sze, "Physics of semiconductor devices", Wiley Interscience.

7. A.H. Zemanian, "Infinite Electrical Networks", Proceedings of the IEEE Vol. 64 (1976) pp. 6-17.

8. A.H. Zemanian, "The Characteristic Resistance Method for Grounded Semi-infinite Grids". College of Engineering Technical Report No. 330. State University of New York at Stony Brook, New York 11794, August 1979.
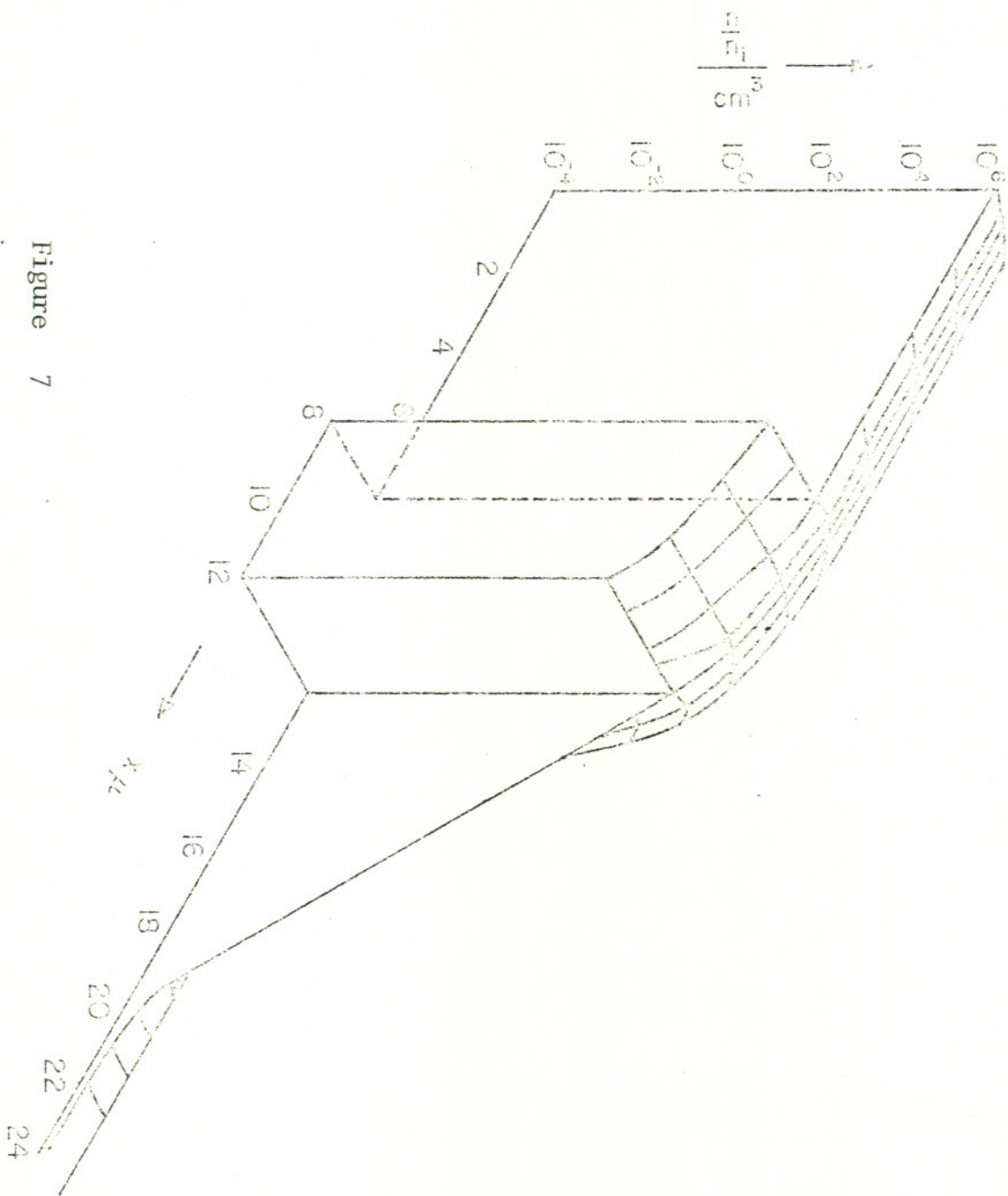
Figure 7

$\dfrac{n}{n_i} \cdot \dfrac{1}{cm^3}$