JOURNAL OF COMPUTATIONAL PHYSICS 68, 420–461 (1987)

Application of LBI Techniques to the Solution of the Transient, Multidimensional Semiconductor Equations

J. P. KRESKOVSKY AND H. L. GRUBIN

Scientific Research Associates, Inc.,
P. O. Box, 498, Glastonbury, Connecticut 06033

Received May 7, 1985; revised April 11, 1986

An analysis and computational procedure are developed for the solution of the coupled system of transient, multidimensional equations governing current continuity for electrons and holes, and the electrostatic potential in semiconductors. The analysis is unique in that through it the 3×3 coupled system is reduced to 2×2 coupled, parabolic system, the continuity equations, and a scalar, elliptic equation, Poisson's equation, which are solved sequentially without introducing stability restrictions or iteration. The coupled continuity equations are integrated through implementation of an LBI algorithm followed by solution for the electrostatic potential using a scalar ADI procedure. The stability and accuracy of the method are explored in one-dimension followed by application to a three-dimensional transient charge collection simulation.

INTRODUCTION

A major concern in the multidimensional, transient simulation of semiconductor devices is the large amount of computational effort required to obtain numerical solutions to the governing current continuity and Poisson's equations. These equations form a nonlinear, time-dependent, coupled system of PDEs. When attempting to solve these equations numerically, implicit methods appear attractive due to their favorable stability properties. However, since the equations are nonlinear and coupled, some form of linearization must be introduced. This may be accomplished by treating each equation as a scalar equation and iterating to account for coupling and nonlinearity, or by expressing the system as a linearized block 3×3 coupled system which is solved iteratively using, for example, Newton's method. Corttell and Buturla [1] have developed procedures following both approaches and discuss the advantages and drawbacks of each. In either case, large systems of linear algebraic equations must be solved repeatedly at each time step. These linear systems are frequently solved using direct or iterative matrix inversion procedures which may result in excessive computer run times. Two areas can thus be identified which contribute significantly to the computational effort required in obtaining solutions to the semiconductor equations; the use of nonlinear iteration, and inefficiency of the methods used to solve the linearized system of equations.

0021-9991/87 $3.00

The approach presented herein reduces the effort in both these areas through application of consistently split, linearized, block implicit (LBI) methods to obtain solutions to the governing semiconductor equations. LBI and related methods were originally developed by Briley and McDonald [2, 3], and Lindemuth and Killeen [4] in independent investigations, and have previously been applied to obtain solutions of the Navier–Stokes equations in the field of computational fluid dynamics (CFD). The methods center about a formal linearization technique which allows solutions of nonlinear, coupled PDEs in one space dimension to be computed efficiently without iteration. The efficiency of the procedure is retained for multidimensional problems through consistent splitting of the LBI operators using ADI techniques which reduce the broad-banded multidimensional system of equations to systems of one-dimensional equations having narrow block-banded matrix structures. The term consistent splitting is introduced since, in the ADI technique employed, the intermediate level solutions are consistent with the original system of equations. Briley and McDonald discuss the possible problems which may arise if inconsistent splitting schemes are used in [2].

LBI techniques may be applied directly to solve the coupled block 3×3 system consisting of the continuity and Poisson's equations in one space dimension. However, when considering multidimensional problems, the block 3×3 system cannot be ADI split due to the absence of a time derivative in Poisson's equation. In the present procedure, this problem is circumvented by reformulating the continuity equations in a manner which allows the coupled block 3×3 system to be reduced to a coupled block 2×2 system, which is solved efficiently, without iteration, using an LBI method, and a scalar, elliptic equation (Poisson's equation) solved using an iterative ADI procedure. Consequently, iteration is limited to the solution of Poisson's equation and nonlinear iteration is eliminated completely. The favorable stability properties of the fully coupled implicit method are retained, as will be demonstrated.

Analysis

The analysis which follows is founded on the need to obtain the governing equations in a form suitable for rapid and efficient numerical solution. Thus, in contrast to other analysis and computational procedures where numerical methods are simply employed to generate solutions to given sets of partial differential equations, the concept of numerical solution and the methods to be employed are an integral part of the present analysis. In some instances these considerations dictate the direction which the formulation takes.

The Governing Equations

The motion of electrons and holes in the semiconductor material in question, and the electrostatic potential, are assumed to be governed by the classical drift and diffusion equations and Poisson's equation

$$\frac{\partial N}{\partial t} = \frac{\partial}{\partial x_i}(N\mu_n E_i) + \frac{\partial}{\partial x_i}\left(D_n \frac{\partial N}{\partial x_i}\right) + R \tag{1}$$

$$\frac{\partial P}{\partial t} = -\frac{\partial}{\partial x_i}(P\mu_p E_i) + \frac{\partial}{\partial x_i}\left(D_p \frac{\partial P}{\partial x_i}\right) + R \tag{2}$$

$$\begin{aligned} \frac{\partial}{\partial x_i}\left(\varepsilon \frac{\partial \psi}{\partial x_i}\right) &= e(N - P - C) \\ &= -e\rho. \end{aligned} \tag{3}$$

In Eqs. (1)–(3), N is the electron concentration, P the hole concentration, μ_k and D_k are mobilities and diffusivites for electrons and holes ($k = N, P$), E_i is the electric field in the ith direction, R is any suitable recombination/generation mechanism, ψ is the electrostatic potential, ε is the permittivity of the material in question, e is the electron charge, C is the net impurity concentration of electrons and holes, and ρ is the space charge. The electric field is related to the electrostatic potential as

$$E_i \equiv -\frac{\partial \psi}{\partial x_i} \tag{4}$$

and the net impurity concentration is given as

$$C \equiv (N_o - P_o). \tag{5}$$

Here N_o and P_o are imposed doping distributions of donor and accepter ions, respectively.

The mobilities and diffusivities appearing in Eqs. (1) (2) are dependent upon the material under consideration, and may reflect field and/or doping dependence, as well as carrier-to-carrier scatering effects. Furthermore, although the Einstein relation between mobility and diffusivity may be implemented for specific problems, it is not implicitly assumed here.

Upon specification of the functional dependence of the mobilities and diffusivities on the carrier concentrations and electric field, and specification of C, Eqs. (1)–(3) may be solved as given. However, the characteristics of this coupled system, aside from its nonlinearity, are dictated by the elliptic nature of Poisson's equation. While this does not introduce any significant sources of computational inefficiency when considering one-dimensional solutions, solution of this system as a block 3×3 coupled system in multidimensions would necessitate the use of methods for coupled elliptic systems. Such methods typically involve the direct or iterative solution of a large system of linear equations generated by the discretization of the governing PDEs. When the nonlinearity of the governing equations considered, it is apparent that these large systems may need to be solved repeatedly at each time step. Alternatively, if the appearance of the potential could be eliminated from the continuity equations the block 3×3 coupled system would be reduced to a block 2×2 coupled parabolic system and a scalar, elliptic equation. This coupled

parabolic system would be ideally suited for solution using a noniterative, split LBI procedure [2, 3], and the scalar, elliptic equation need then be solved only once per time step. The noniterative solution of the continuity equations and the need to solve Poisson's equation only once would result in a substantial reduction in the computational effort required to solve the system at each time step.

In order to implement such a solution procedure, the system of equations must be reformulated through manipulation of the continuity equations to allow sequential solution without affecting stability. The drift terms in Eqs. (1) and (2) are first expanded as

$$\frac{\partial}{\partial x_i}(N\mu_n E_i) = N\mu_n \frac{\partial E_i}{\partial x_i} + E_i \frac{\partial}{\partial x_i}(N\mu_n) \tag{6}$$

and

$$\frac{\partial}{\partial x_i}(P\mu_p E_i) = P\mu_p \frac{\partial E_i}{\partial x_i} + E_i \frac{\partial}{\partial x_i}(P\mu_p). \tag{7}$$

Poisson's equation may also be expanded yielding

$$\frac{\partial^2 \psi}{\partial x_i^2} = \frac{e}{\varepsilon}(N - P - C) - \frac{1}{\varepsilon}\frac{\partial \varepsilon}{\partial x_i}\frac{\partial \psi}{\partial x_i}. \tag{8}$$

Using the definition of electric field (Eq. (4)), Eq. (8) may be substituted into Eqs. (6) and (7) to obtain

$$\frac{\partial}{\partial x_i}(N\mu_n E_i) = -\frac{e}{\varepsilon}\mu_n N(N - P - C) + E_i\left(\frac{\partial N\mu_n}{\partial x_i} - \frac{N\mu_n}{\varepsilon}\frac{\partial \varepsilon}{\partial x_i}\right) \tag{9}$$

and

$$\frac{\partial}{\partial x_i}(P\mu_p E_i) = -\frac{e}{\varepsilon}\mu_p P(N - P - C) + E_i\left(\frac{\partial P\mu_p}{\partial x_i} - \frac{P\mu_p}{\varepsilon}\frac{\partial \varepsilon}{\partial x_i}\right). \tag{10}$$

The current continuity equations may now be expressed as

$$\frac{\partial N}{\partial t} = -\frac{e}{\varepsilon}\mu_n N(N - P - C) + E_i\left(\frac{\partial N\mu_n}{\partial x_i} - \frac{N\mu_n}{\varepsilon}\frac{\partial \varepsilon}{\partial x_i}\right) + \frac{\partial}{\partial x_i}\left(D_n \frac{\partial N}{\partial x_i}\right) + R \tag{11}$$

and

$$\frac{\partial P}{\partial t} = \frac{e}{\varepsilon}\mu_p P(N - P - C) - E_i\left(\frac{\partial P\mu_p}{\partial x_i} - \frac{P\mu_p}{\varepsilon}\frac{\partial \varepsilon}{\partial x_i}\right) + \frac{\partial}{\partial x_i}\left(D_p \frac{\partial P}{\partial x_i}\right) + R. \tag{12}$$

The importance of this reformulation lies in the substitution of the space charge for the Laplacian of the electrostatic potential. This substitution allows a strong coupling between the continuity and Poisson's equations to be retained while removing its dependence on the explicit appearance of the potential. A weaker coupling, due to the appearance of the electric field as a coefficient, may be eliminated without affecting the stability of the overall solution algorithm by simply lagging this coefficient. As will become apparent when the numerical solution technique is discussed, the reformulation allows the block 3×3 coupled system to be reduced to a block 2×2 coupled system plus a decoupled, scalar, elliptic equation. The decoupling of the elliptic equation results in a highly efficient algorithm in which the coupled 2×2 system is solved noniteratively, using LBI techniques, followed by the ADI solution of the scalar elliptic equation. Since the space charge is treated fully implicitly in both systems, stability is retained without introducing iteration between the 2×2 coupled system and the scalar elliptic system, as will be demonstrated subsequently. For this reason, this is referred to as a space charge coupled algorithm (SCC).

When applied to transient simulations, using a backward Euler time discretization, both the fully coupled and the decoupled SCC algorithms have a time truncation error which is $\mathcal{O}(\Delta t)$. However, the SCC algorithm contains an additional $\mathcal{O}(\Delta t)$ error, relative to the fully coupled approach, due to the lagging of the electric field, which facilitates the decoupling. This is a transient error and appears as an error in the total current which is proportional to the electrical displacement. However, the Poisson residual

$$\omega \equiv \frac{\partial}{\partial x_i}\left(\varepsilon \frac{\partial \psi}{\partial x_i}\right) - e(N - P - C) \tag{13}$$

will remain zero throughout the transient and the steady state solution obtained will have no error relative to the fully coupled approach.

Conservation of Total Current

Although the SCC algorithm discussed above allows a highly efficient solution procedure to be implemented, the errors in the total current make it unattractive, and it is not recommended, since conservation of total current during transients is usually a primary concern. This error can be eliminated, however, by recasting Poisson's equation in a statement of total current conservation through temporal differentiation:

$$\frac{\partial}{\partial t}\left[\frac{\partial}{\partial x_i}\left(\varepsilon \frac{\partial \psi}{\partial x_i}\right)\right] = \frac{\partial}{\partial x_i}\left[E_i(\mu_n N + \mu_p P)\right] + \frac{\partial}{\partial x_i}\left[D_n \frac{\partial N}{\partial x_i} - D_p \frac{\partial P}{\partial x_i}\right]. \tag{14}$$

This total current constraint is solved in lieu of Eq. (3), and the resulting algorithm is referred to as the space charge coupled, current conserving algorithm (SC^4).

While this reformulation allows rigorous conservation of total current to be maintained during transients using a decoupled, sequential solution technique, the $\mathcal{O}(\Delta t)$ error which appeared in the total current in the SCC algorithm now appears as a related $\mathcal{O}(\Delta t)$ error in the Poisson residual. An equation governing this residual may be derived from the continuity equations, Eqs. (11) and (12), and the total current constraint, Eq. (14), with the result

$$\left.\frac{\partial \omega}{\partial t}\right|^{n+1} = -\left[\frac{(\mu_n N + \mu_p P)}{\varepsilon}\omega\right]^{n+1} + \nabla(\psi^{n+1} - \psi^n)\cdot\nabla(\mu_n N + \mu_p P)^{n+1}, \quad (15)$$

where the superscripts n and $n+1$ identify the time levels to be used in the temporal discritization. The first term on the RHS of Eq. (15) gives a temporal decay of any non-zero Poisson residual. The second term is $\mathcal{O}(\psi^{n+1} - \psi^n) = \mathcal{O}(\Delta t)$ and arises since the electric field is lagged in the continuity equations and treated fully implicitly in the total current constraint. This term may generate localized Poisson residuals in regions where there are substantial gradients of the mobility-carrier density product, however, it is easily shown that such residuals are bound by

$$|\omega_{\max}| \leqslant |\varepsilon\nabla(\Delta\psi_{\max})\cdot\nabla\ln(\mu_n N + \mu_p P)|, \quad (16)$$

where $\Delta\psi_{\max}$ is the maximum change in potential across any time step. This suggests that the maximum value of the Poisson residual during a transient can be limited by selecting time steps such that $\Delta\psi^{n+1} = (\psi^{n+1} - \psi^n)$ never exceeds a prespecified maximum value. As steady state is approached, $\Delta\psi^{n+1}$ will go to zero independent of Δt and, in accordance with Eq. (15), any local Poisson residual will decay to zero as well. A unique steady solution will thus be obtained with no additional error relative to a fully coupled approach, even if the initial Poisson residual is non-zero. It should be noted that in the calculations to be presented here, as well as in a variety of other practical calculations [5–7], consideration of Poisson residuals has not had any impact on the selection of time steps.

An approach similar to the present SC[4] approach has also been followed by Mock [8] in the development of a stable, fully decoupled algorithm. However, there are important differences between Mock's approach and that followed here. Mock [8] solves the continuity equations in the form given by Eqs. (1) and (2). The system of Eqs. (1), (2), and (14) does not contain information regarding the space charge. Thus, as Mock points out, if the Poisson residual is not initially zero, or if the temporal integration of Eq. (14) is not sufficiently accurate, the residual may grow unacceptably, and a unique steady state solution need not be obtained.

Mock also indicates potential problems regarding conservation of total current in his decoupled algorithm. This can be traced to the order in which the equations are solved. The electrostatic potential is advanced first by Mock using an equation analogous to Eq. (14). Since the carrier concentrations are not yet available at the advanced time level, they are lagged. As a result the total current constraint is not rigorously satisfied at the advanced time level. To alleviate this problem, iteration

may be introduced negating to some extent the advantages obtained by decoupling. By contrast, in the algorithm considered here, the carrier concentrations are advanced first, and the total current constraint is satisfied rigorously on a fully implicit basis without iteration. It is emphasized that in the present formulation Eq. (14) is used only to insure conservation of total current, and not to provide stability.

Nondimensionalization

Before attempting to solve the system of Eqs. (11), (12), and (14) numerically, it is advantageous to recast these equations in dimensionless form. Historically [6], such equations have been nondimensionalized using intrinsic carrier density n_i, Debye length, L_D, a reference potential given as $V_D = kT/e$, etc. Such quantities may be desirable from a physical point of view; however, they are not generally well-suited for computational purposes. If such quantities are used, the computational variables may exhibit large exponents which can lead to machine overflows. The approach taken here is to use the following arbitrary reference quantities. Carrier concentrations are normalized by N_r, potentials by ψ_r, lengths by X_r, velocity (μE_i) by V_r, time by $t_r \equiv X_r/V_r$, mobility by μ_r, diffusivity by D_r, and permittivity by ε_r. Equations (11), (12), and (14), expressed in nondimensional form using these reference quantities, are

$$\frac{\partial N}{\partial t} = -SnCn\frac{\mu_n N}{\varepsilon}(N - P - C) + CnE_i\left(\frac{\partial N\mu_n}{\partial x_i} - \frac{N\mu_n}{\varepsilon}\frac{\partial \varepsilon}{\partial x_i}\right) + \frac{1}{Re}\frac{\partial}{\partial x_i}\left(D_n\frac{\partial N}{\partial x_i}\right) + R \tag{17}$$

$$\frac{\partial P}{\partial t} = SnCn\frac{\mu_p P}{\varepsilon}(N - P - C) - CnE_i\left(\frac{\partial P\mu_p}{\partial x_i} - \frac{P\mu_p}{\varepsilon}\frac{\partial \varepsilon}{\partial x_i}\right) + \frac{1}{Re}\frac{\partial}{\partial x_i}\left(D_p\frac{\partial P}{\partial x_i}\right) + R \tag{18}$$

$$\frac{\partial}{\partial t}\left[\frac{\partial}{\partial x_i}\left(\varepsilon\frac{\partial \psi}{\partial x_i}\right)\right] = SnCn\left\{\frac{\partial}{\partial x_i}[E_i(N\mu_n + P\mu_p)]\right\} + \frac{Sn}{Re}\left\{\frac{\partial}{\partial x_i}\left(D_n\frac{\partial N}{\partial x_i} - D_p\frac{\partial P}{\partial x_i}\right)\right\}, \tag{19}$$

where the dimensionless parameters *Cn*, *Re*, and *Sn* are given by

$$Cn = \frac{\mu_r \psi_r}{x_r V_r}, \qquad Re = \frac{x_r V_r}{D_r}, \qquad Sn = \frac{e x_r^2 N_r}{\varepsilon_r \psi_r}. \tag{20}$$

As stated previously, the reference quantities are arbitrary; however, when properly chosen, the computational variables will typically vary between zero and unity. The dimensionless parameters given by Eq. (20) then indicate relative magnitude of each

term in the governing equations. The reference quantities are usually chosen as follows:

N_r—the maximum carrier density specified in the doping profile

ψ_r—the maximum boundary value specified for potential

X_r—the maximum dimension of the device

V_r—the saturation velocity of the fastest carrier

μ_r—the maximum mobility of either carrier

D_r—the maximum diffusivity of either carrier.

As a result of these choices, t_r will be the time required for the fastest carrier to traverse the longest dimension of the device when moving at its saturated drift velocity.

The Numerical Method

LBI Procedures

The numerical method used in the present algorithm is based on application of consistently split, linearized, block implicit (LBI) methods, as developed by Briley and McDonald [2, 3] to the system given by Eqs. (17)–(19). LBI methods have been applied to a wide variety of problems in the field of CFD (cf. [10–12]) with a high degree of success. Thus, application of such methods to semiconductor problems can draw on a vast amount of related experience generated using LBI techniques.

LBI techniques center about the use of a formal linearization procedure in which systems of coupled nonlinear PDEs in one space dimension are reduced to a system of linear equations, which upon application of spatial differencing, may be expressed as a block coupled matrix system. The resulting system may then be solved efficiently, without iteration, to advance the solution. The benefits of the procedure are retained for multidimensional problems through application of ADI schemes in their natural extension to block coupled systems. The ADI procedures reduce the multidimensional system of equations, having broad-banded matrix structures to systems of one-dimensional equations with narrow block-banded structures which are solved efficiently using fundamental block-elimination methods.

Briley and McDonald [2] considered the coupled system of l nonlinear, time-dependent, multidimensional equations given by

$$\frac{\partial H(\phi_l)}{\partial t} = D(\phi_l) + S(\phi_l). \tag{21}$$

In Eq. (21), ϕ_l represent the vector of dependent variables, $H(\phi_l)$ and $S(\phi_l)$ are nonlinear functions of ϕ_l which contain no spatial operators, and $D(\phi_l)$ is a general,

nonlinear, multidimensional, partial differential operator. Equation (2) is first time differenced about $t^n + \beta\,\Delta t$ as

$$\frac{H(\phi)^{n+1} - H(\phi)^n}{\Delta t} = \beta[D(\phi)^{n+1} + S(\phi)^{n+1}] + (1-\beta)[D(\phi)^n + S(\phi)^n], \quad (22)$$

where $\Delta t = t^{n+1} - t^n$, and the subscript l has been dropped for clarity. The parameter $\beta = 1$ for a fully implicit scheme or $\beta = 0.5$ for the Cranck–Nicolson formulation. The implicit level nonlinear operators H, D, and S are then formally linearized using a Taylor series expansion about the explicit time level

$$G(\phi)^{n+1} = G(\phi)^n + \left.\frac{\partial G(\phi)}{\partial \phi}\right|^n \Delta\phi^{n+1} + O(\Delta t^2). \quad (23)$$

Equation (22) may then be expressed at each grid point in the solution domain as a matrix equation of the form

$$(A + \Delta t L)\,\Delta\phi^{n+1} = \Delta t[D(\phi^n) + S(\phi^n)], \quad (24)$$

where

$$A \equiv \left(\frac{\partial H}{\partial \phi}\right)^n - \beta\,\Delta t\left(\frac{\partial S}{\partial \phi}\right)^n \quad (25)$$

and

$$L \equiv -\beta \left.\frac{\partial D}{\partial \phi}\right|^n. \quad (26)$$

As a result, the nonlinear, coupled system of PDEs given by Eq. (21) has been reduced to a block $l \times l$ coupled, linear system of temporal difference equations (Eq. (24)) which, upon spatial differencing, need only be solved once per time step to obtain a solution. Additionally, since the linearization error is at worst of the same order as the temporal discretization error, the linearization is not expected to introduce significant inaccuracies.

Application of Eq. (24) to second order PDEs in one space dimension, using standard three-point spatial difference approximations requires the solution of one block $l \times l$ tridiagonal system per time step. Such a system can be solved efficiently using standar block-tridiagonal elimination procedures.

The application of the LBI algorithms given by Eqs. (24)–(26) to multidimensional problems results in the loss of the narrow, block-banded matrix structure obtained in one space dimension. The discretization of the multidimensional spatial operator results in a broad-banded structure, which, if solved by direct or iterative methods, can be extremely inefficient. Typically, the computational effort required to solve such a system using a direct method varies as N^2 to N^3 (where N is the total number of grid points) [1]. Iterative matrix solver effort may vary as N per

iteration, but may exhibit slow convergence as the number of grid points rises, or if highly unequally spaced mesh are used. Such observations led Briley and McDonald [2, 3] to develop consistently split LBI algorithms for multidimensional problems. The splitting is accomplished by dividing the multidimensional spatial operator, L, into one-dimensional operators associated with each coordinate direction

$$L = L_x + L_y + L_z. \tag{27}$$

Equation (24) is then split following the scalar ADI development of Douglas and Gunn [13]:

$$(A + \Delta t L_x)\,\Delta\phi^* = \Delta t[D(\phi^n) + S(\phi^n)] \tag{28a}$$

$$(A + \Delta t L_y)\,\Delta\phi^{**} = A\,\Delta\phi^* \tag{28b}$$

$$(A + \Delta t L_z)\,\Delta\phi^{***} = A\,\Delta\phi^{**}. \tag{28c}$$

Here $\Delta\phi^*$, $\Delta\phi^{**}$, and $\Delta\phi^{***}$ are intermediate solutions of Eqs. (28a)–(28c). Again, if three point operators are used to approximate the spatial operators, L_i, each of Eqs. (28a)–(28c) will be block-tridiagonal, and can be efficiently solved. The block size and band width are independent of the number of grid points; hence, the computational effort required to solve the sequence varies linearly with the total number of grid points regardless of the number of space dimensions considered. For two-dimensional problems, Eq. (28c) is omitted. Elimination of the intermediate steps in Eqs. (28a)–(28c) yields

$$\begin{aligned}(A + \Delta t L_x)\,A^{-1}(A + \Delta t L_y)\,A^{-1}(A + \Delta t L_z)\,\Delta\phi^{***} \\ = \Delta t[D(\phi^n) + S(\phi^n)] + O(\Delta t^2).\end{aligned} \tag{29}$$

Thus

$$\Delta\phi^{n+1} = \Delta\phi^{***} + O(\Delta t^2). \tag{30}$$

The development given above presents a brief outline of the LBI method used in the present investigation. A more detailed development, as well as in-depth discussion of LBI methods, the linearization procedure, and related topics may be found in the article by Briley and McDonald [2].

Application to the Semiconductor Equations

The semiconductor equations given by Eqs. (17)–(19) may be expressed in the form of Eq. (21), where

$$H(\phi) = \begin{Bmatrix} N \\ P \\ 0 \end{Bmatrix} \tag{31}$$

$$D(\phi)=\begin{Bmatrix} C_n E_i\left(\dfrac{\partial N\mu_n}{\partial x_i}-\dfrac{N\mu_n}{\varepsilon}\dfrac{\partial \varepsilon}{\partial x_i}\right)+\dfrac{1}{Re}\dfrac{\partial}{\partial x_i}\left(D_n\dfrac{\partial N}{\partial x_i}\right) \\ -C_n E_i\left(\dfrac{\partial P\mu_p}{\partial x_i}-\dfrac{P\mu_p}{\varepsilon}\dfrac{\partial \varepsilon}{\partial x_i}\right)+\dfrac{1}{Re}\dfrac{\partial}{\partial x_i}\left(D_p\dfrac{\partial P}{\partial x_i}\right) \\ -\dfrac{\partial}{\partial t}\left[\dfrac{\partial}{\partial x_i}\left(\varepsilon\dfrac{\partial \psi}{\partial x_i}\right)\right]+Sn\,Cn\left[\dfrac{\partial}{\partial x_i}E_i(\mu_n N+\mu_p P)\right] \\ +\dfrac{Sn}{Re}\left[\dfrac{\partial}{\partial x_i}\left(\dfrac{D_n}{\partial x_i}-D_p\dfrac{\partial P}{\partial x_i}\right)\right] \end{Bmatrix} \tag{32}$$

$$S(\phi)=\begin{Bmatrix} -Sn\,Cn\dfrac{\mu_n N}{\varepsilon}(N-P-C)+R \\ Sn\,Cn\dfrac{\mu_p P}{\varepsilon}(N-P-C)+R \\ 0 \end{Bmatrix} \tag{33}$$

for $\phi=(N, P, \psi)^{\mathrm{T}}$. The permittivity, mobilities, and diffusivities appearing in Eqs. (31)–(33) are evaluated at the explicit, or t^n level, and are assumed known. Equations (31)–(33) may then be linearized following the example of Eq. (23), using Eq. (4) to express E_i in terms of ψ. The resulting linearized equations can then be expressed as a block 3×3 coupled system which is of the form of Eq. (24), and which can be solved efficiently by block tridiagonal elimination in one space dimension. For multidimensional problems, however, this block 3×3 coupled system cannot be split following Eq. (28) since the A matrix has the form

$$A=\left|\begin{array}{cc:c} x & x & 0 \\ x & x & 0 \\ \hdashline 0 & 0 & 0 \end{array}\right| \tag{34}$$

and is singular.

The partitioning of the A matrix indicated by the dotted lines in Eq. (34) suggests that if the L matrix (the linearized D-operator) could be similarly partitioned Poisson's equation could be decoupled from the continuity equations. From examination of the D operator, it is apparent that this can be accomplished if the electric field, E_i, appearing in the continuity equations, is evaluated explicitly. It will be demonstrated that this does not affect stability, and as a result, the coupled, block 3×3 system is reduced to a coupled, block 2×2 system and a scalar elliptic equation. This decoupled system is solved using the algorithm shown in Fig. 1. The carrier concentrations are first advanced through application of the noniterative, consistently split LBI method (Eqs. (28a)–(28(c)) to the block 2×2 system. Then, with the solution for the carrier concentrations at the advanced time level known,

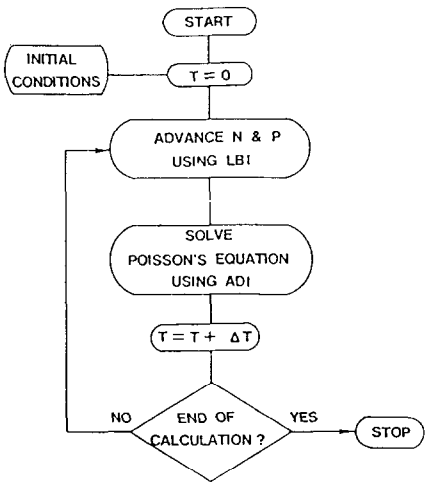

FIG. 1. Computational algorithm used to advance the solution in time.

the decoupled, scalar elliptic equation for the electrostatic potential is solved, completing a time step. The elliptic equation is solved using an iterative ADI procedure with cycled acceleration parameters [2, 13, 14]. Consequently, the use of iteration is limited to solution of Poisson's equation, and nonlinear iteration is eliminated completely. It should be noted that if the continuity equations were not reformulated, and the potential was simply lagged, sequential solution of Eqs. (1)–(3) in the order followed here introduces stability limits [15].

Spatial Differencing

The spatial difference approximations applied to the L and D operators must be developed in such a manner that the various components of the current vectors as well as the total current are rigorously conserved on a finite difference basis. To achieve this, rather than difference equations Eqs. (17)–(19) directly, the equations expressed in terms of current vectors are differenced. These difference approximations, which are based on the central difference scheme, are manipulated into a form analogous to Eqs. (17)–(19). The carrier concentrations, mobilities, diffusivities, permittivities, and potentials are defined at grid points, i, $i \pm 1$, etc. The electric field, E, and the current density, J, are defined at half points, $i \pm \frac{1}{2}$, etc. Values of point functions required at half points are computed by simple averages

$$\phi_{i+1/2} = \tfrac{1}{2}(\phi_i + \phi_{i+1}). \tag{35}$$

The overall result is similar to that which is obtained from a control volume approach [16].

To obtain the difference approximations to the continuity equations, consider first Poisson's equation, expressed in terms of the electric field

$$-\frac{\partial}{\partial x}(\varepsilon E) = e(N - P - C). \tag{36}$$

This equation is differenced as

$$\frac{\varepsilon_{i+1/2}E_{i+1/2} - \varepsilon_{i-1/2}E_{i-1/2}}{\Delta_A} = -e(N_i - P_i - C_i), \tag{37}$$

where

$$\Delta_A = \tfrac{1}{2}(\Delta_m + \Delta_p) \tag{38}$$

and

$$\Delta_p = x_{i+1} - x_i \tag{39}$$

$$\Delta_m = x_i - x_{i-1}. \tag{40}$$

The permittivities in Eq. (37) are obtained from Eq. (35), thus Eq. (37) may be expressed as

$$\frac{1}{2}\left[\frac{\varepsilon_{i+1}E_{i+1/2} - \varepsilon_{i-1}E_{i-1/2}}{\Delta_A}\right] + \frac{\varepsilon_i}{2}\left[\frac{E_{i+1/2} - E_{i-1/2}}{\Delta_A}\right] = -e(N_i - P_i - C_i). \tag{41}$$

The second term on the LHS of Eq. (41) may be added and subtracted to yield, with some rearrangement,

$$-\left(\frac{E_{i+1/2} - E_{i-1/2}}{\Delta_A}\right) = \frac{e}{\varepsilon_i}(N_i - P_i - C_i) + \frac{1}{2\varepsilon_i}\left[E_{i+1/2}\left(\frac{\varepsilon_{i+1} - \varepsilon_i}{\Delta_A}\right) + E_{i-1/2}\left(\frac{\varepsilon_i - \varepsilon_{i-1}}{\Delta_A}\right)\right]. \tag{42}$$

Equation (42) is recognized as a difference analog to Eq. (8).

The electron curent

$$J_N = N\mu_n E + D_n\frac{\partial N}{\partial x} \tag{43}$$

is considered next. The divergence of the drift component of current is approximated as

$$\frac{\partial(N\mu_n E)}{\partial x} = \frac{(N\mu_n)_{i+1/2}E_{i+1/2} - (N\mu_n)_{i-1/2}E_{i-1/2}}{\Delta_A}. \tag{44}$$

Expressing the $N\mu_n$ product at the half points in terms of Eq. (35)

$$\frac{\partial(N\mu_n E)}{\partial x} = \frac{(N\mu_n)_i}{2}\left[\frac{E_{i+1/2} - E_{i-1/2}}{\Delta_A}\right] + \frac{1}{2}\left[\frac{(N\mu_n)_{i+1}E_{i+1/2} - (N\mu_n)_{i-1}E_{i-1/2}}{\Delta_A}\right]. \quad (45)$$

Adding and subtracting the first term on the RHS of Eq. (45)

$$\frac{\partial(N\mu_n E)}{\partial x} = (\mu_n N)_i\left[\frac{E_{i+1/2} - E_{i-1/2}}{\Delta_A}\right] + \frac{1}{2}\left[E_{i+1/2}\left(\frac{(N\mu_n)_{i+1} - (N\mu_n)_i}{\Delta_A}\right) + E_{i-1/2}\left(\frac{(N\mu_n)_i - (N\mu_n)_{i-1}}{\Delta_A}\right)\right], \quad (46)$$

Eq. (42) may then be substituted in Eq. (46) to yield

$$\frac{\partial(N\mu_n E)}{\partial x} = -\frac{(\mu_n N)_i}{\varepsilon_i}(N_i - P_i - C_i) + \frac{1}{2}\left[E_{i+1/2}\left(\frac{(N\mu_n)_{i+1} - (N\mu_n)_i}{\Delta_A}\right) + E_{i-1/2}\left(\frac{(N\mu_n)_i - (N\mu_n)_{i-1}}{\Delta_A}\right)\right] - \frac{(\mu_n N)_i}{2\varepsilon_i}\left[E_{i+1/2}\left(\frac{\varepsilon_{i+1} - \varepsilon_i}{\Delta_A}\right) + E_{i-1/2}\left(\frac{\varepsilon_i - \varepsilon_{i-1}}{\Delta_A}\right)\right], \quad (47)$$

which is recognized as the difference analog to Eq. (9). Finally, the diffusive contribution to the current is approximated as

$$\frac{\partial(D(\partial N/\partial x))}{\partial x} = \left[\frac{D_{i+1/2}((N_{i+1} - N_i)/\Delta_p) - D_{i-1/2}((N_i - N_{i-1})/\Delta_m)}{\Delta_A}\right]. \quad (48)$$

Similar approximations are obtained for the divergence of the hole current. The difference analog for the total current constraint, Eq. (19) is obtained using the LHS of Eq. (37) for the displacement, and Eqs. (44) and (48) (and their hole counterparts) for the drift and diffusion currents.

The approximation given above are all formally $\mathcal{O}(\Delta^2)$ if the mesh spacing is uniform. For unequally spaced meshes the formal accuracy of these approximations reduces to $\mathcal{O}(\Delta)$; however, the leading terms in the truncation error are typically of the form

$$TE \doteq (1 - AR)\,\Delta\{\cdots\} \pm \frac{(1 + AR^3)}{(1 + AR)}\Delta^2\{\cdots\} + \cdots, \quad (49)$$

where $AR = \Delta_p/\Delta_m$. Thus, provided the mesh spacing does not change too rapidly, the accuracy of the scheme may approach second order for unequally meshes.

Artificial Diffusion

It is well known that the central difference approximations used in the present solution procedure may give rise to spatial oscillations in the solution, and may even result in the failure of iterative methods to converge, if the potential difference between grid points exceeds twice the thermal voltage, $2kT/e$. As a result, many workers in the solid state community have abandoned the use of the central difference scheme in favor of the exponential scheme of Scharfetter and Gummel [17], which suppresses such oscillations. The Scharfetter–Gummel scheme was not used here since the difference approximations based on it cannot be manipulated into a form analogous to Eqs. (16)–(18) as required.

The problems with the central difference scheme alluded to above are not unique to the field of solid state device simulation, but may occur whenever central differences are applied to a convection-diffusion equation. Numerous investigations in the field of CFD have examined the problem, and Roache [18] discusses the problem in detail for Burgers equation. From a physical point of view, such spatial oscillations are typically an indication that the spatial gradients in the solution are locally too steep to be resolved on the specified grid. Alternatively, it is possible to interpret such oscillations as being the result of Fourier components in the solution with wave lengths less than the Nyquist frequency, $2\Delta x$, which cannot be supported on the mesh.

While researchers in the field of CFD have developed exponential schemes paralleling that of Scharfetter and Gummel [19, 20], they have also developed a means of stabilizing the central difference scheme through the introduction of artificial diffusion [3]. Artificial diffusion damps the high frequency components of the solution which cannot be supported on the local mesh and suppresses the oscillations which would otherwise be present. This approach is followed in the present solution procedure. Artificial diffusion is introduced locally by augmenting the physical diffusion coefficient. A computational diffusion coefficient is thus defined as

$$D_{\text{comp}} = D_{\text{phys}} \left(\frac{Re_{\Delta x_i}}{Re_c} \right), \qquad Re_{\Delta x_i} > Re_c, \tag{50a}$$

$$D_{\text{comp}} = D_{\text{phys}}, \qquad Re_{\Delta x_i} < Re_c, \tag{50b}$$

where

$$Re_{\Delta x_i} = \frac{|CnE_i\mu|\, \Delta x_i Re}{D_{\text{phys}}}. \tag{51}$$

If the Einstein relation is implemented, Eq. (51) reduces to

$$Re_{\Delta x_i} = \frac{\Delta\psi_i e}{kT}, \tag{52}$$

or artificial diffusion is introduced if the potential difference between grid points exceeds $Re_c\, kT/e$. If Re_c is taken as 2.0, artificial diffusion is added when the potential difference between grid points exceeds the familiar value, $2kT/e$.

Obviously, the use of artificial diffusion must be implemented with extreme care. If, for example, it is introduced in a region where the carrier concentrations are relatively uniform, the effect will be insignificant regardless of the potential variation between grid points. However, if it is introduced in a region where significant gradients of the carrier densities exist, the diffusion component of the current vectors may be in error. Under such circumstances, the results must be evaluated carefully to determine if the calculation should be performed using local mesh refinement.

It is of interest to note that the present hybrid scheme using central differences with artificial diffusion and the Scharfetter–Gummel scheme have the same limiting forms as the electric field approaches zero and infinity [21]. At the limit of zero field, both schemes reduce to the standard three-point central difference approximation to the diffusion term. At high field, both schemes reduce to a two-point upwind difference approximation to the drift term. In this sense, the Scharfetter–Gummel scheme can be thought of as containing an "implicit" artificial diffusion. An advantage of the present formulation is that the artificial diffusion can be directly controlled, and the magnitude of its effect explicitly evaluated. Additionally, it can be shown [21] that Re_c can be chosen as a function of $Re_{\Delta x}$ in such a manner that the present hybrid scheme is equivalent to the Scharfetter–Gummel scheme.

Estimation of Computational Effort

The computational effort required to solve the linearized system of difference equations is considered here. Since only scalar or block 2×2 tridiagonal matrices must be solved as a result of ADI splitting, the overall solution algorithm is highly efficient. Solution of a block-tridiagonal matrix requires approximately $3M\ (l^2 + l^3)$ operations, where M is the number of diagonal blocks and l is the block size [22]. Solution of a scalar tridiagonal system requires approximately $5M$ operation [22]. For a three-dimensional problem with a grid size by NX by NY by NZ each of Eqs. (28a)–(28c) must be solved $NY * NZ$, $NX * NZ$, and $NX * NY$ times, respectively. For the block 2×2 structure considered here the total number of operations required to solve this sequence is thus approximately

$$\begin{aligned}(NY * NZ)(3 * NX)(12) + (NX * NZ)(3 * NY)(12) \\ + (NX * NY)(3 * NZ)(12) = 108N_T,\end{aligned} \tag{53}$$

where $N_T = NX * NY * NZ$, the total number of grid points. The operations required for the scalar tridiagonal system which must be solved is, similarly,

$$(NY * NZ)(5 * NX) + (NX * NZ)(5 * NY) + (NX * NY)(5 * NZ) = 15N_T. \tag{54}$$

Thus, the total number of operations required to solve the linear systems at each time step is of the order

$$OPT = 108N_T + (15N_T)I, \tag{55}$$

where I is the number of iterations required in the ADI solution of the scalar elliptic equation. For two-dimensional problems these numbers reduce to $72N_T$ and $10N_T I$, respectively. The elliptic ADI solver typically converges to the required degree of accuracy to insure conservation of total current in from 50 to 100 iterations ($I = 50$ to 100).

The estimates given above are for implementation on a scalar machine. On a vector machine such as the CRAY-1, it is possible to achieve a substantial reduction in the apparent effort by performing the solution of each of Eqs. (28a)–(28c) in parallel. For example, consider the x-implicit sweep in which Eq. (28a) must be solved along each of the $NY * NZ$ x-implicit lines. If a vector length VL is defined, then VL of the x-implicit lines can be solved in parallel with an effort equal to that of a single x-line plus the overhead associated with the vector operations. The number of vector operations for the x sweep would be proportional to $(NY * NZ)/VL$ rather than $(NY * NZ)$. The equivalent number of operations for the complete algorithm is then of the order of

$$OPT_V = 108N_T/VL + (15N_T/VL)I.$$

On a square mesh with $NX = NY = NZ = N$, $N_T = N^3$, and if sufficient storage and vector length are available for $VL = N^2$,

$$OPT_V = 108N_T^{1/3} + (15N_T^{1/3})I.$$

Computed Results

The results to be discussed subsequently were all computed assuming silicon as the semiconductor material in question. The mobilities were determined from the relationship of Caughey and Thomas [23] as given by Kurata [24]:

$$\mu = \left[\frac{\mu_{max} - \mu_{min}}{1 + (|C|/N_{ref})^{\alpha} + [(NP)^{1/2}/2.04N_{ref}]^{\alpha}} + \mu_{min}\right]\left[\frac{1}{[1 + (|E|/E_{ref})^{\beta}]^{1/\beta}}\right]. \tag{56}$$

The constants appearing in Eq. (56) are given in Table I. This relationship was used for all calculations reported here. The diffusivities were obtained assuming the Einstein relationship

$$D = \frac{kT}{e}\mu. \tag{57}$$

TABLE I

Constant	Value for hole mobility	Value for electron mobility
N_{ref}	6.3×10^{16}	8.5×10^{16}
α	0.76	0.72
μ_{max}	495.0	1330.0
μ_{min}	47.7	65.0
E_c	1.95×10^4	8.0×10^3
β	1.0	2.0

The boundary conditions applied are for either ohmic contacts or free surfaces. At ohmic contacts the contact potential is set to the applied potential plus the built-in potential

$$\psi = \frac{kT}{e} \ln\left(\frac{N}{N_i}\right) + V_{APP}. \tag{58}$$

The carrier densities at such contacts are specified assuming zero space charge under equilibrium conditions

$$N - P = C \tag{59}$$

$$NP = N_i^2. \tag{60}$$

At free surfaces, the electric field and the particle currents are set to zero.

Finally, the recombination model used in the present computations is taken as a combination of the Shockely–Read–Hall and Auger recombination models

$$R = \frac{NP - N_i^2}{\tau_P(N + N_i) + \tau_N(P + N_i)} + r(N + P)(NP - N_i^2). \tag{61}$$

Stability, Convergence, and Accuracy Tests

The stability of the present algorithm has not been analyzed on a theoretical basis. However, one of the authors [10, 25] has had extensive experience with related algorithms in the field of CFD in which the coefficients of derivatives in parabolic equations, which are lagged, are determined through the solution of a Poisson equation. The resulting algorithms have been extremely stable and accurate. The present algorithm is expected to exhibit similar qualities. To verify the stability characteristics of the present solution algorithm, several one-dimensional calculations were performed. Under the one-dimensional constraint direct comparisons can be easily made between the fully coupled LBI approach in which Eqs. (3), (11), and (12) are solved as a block-coupled system, the SCC approach in

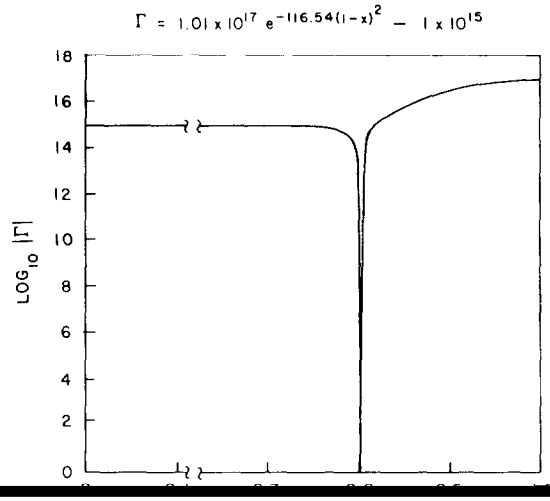

FIG. 2. Doping profile for one-dimensional test calculation. $\Gamma = 1.01 \times 10^{17} e^{-116.54(1-x)^2} - 1 \times 10^{15}$.

which Eqs. (3), (11), and (12) are solved with the electric field, E_i, appearing in Eqs. (11) and (12) evaluated at the explicit time level, and the SC[4] current conserving approach in which Eqs. (11), (12), and (14) are solved. The characteristics of this last algorithm are of primary concern since it is this algorithm which will be applied to multidimensional problems. The test problem considered is a 15-μm N^+P silicon device with a $10^{17}/cm^3$ N^+ region on a $10^{15}/cm^3$ P-type substrate. The doping distribution is given as

$$C = 1.01 \times 10^{17} e^{-116.54(1-x)^2} - 1 \times 10^{15} \tag{62}$$

and is shown in Fig. 2. The reference quantities and resulting dimensionless constants are given in Table II.

TABLE II

Quantity	Value
N_r	$1.0 \times 10^{17}/cm^3$
ψ_r	3.0 V
X_r	15.0×10^{-4} cm
V_r	1.0×10^7 cm/sec
μ_r	1300.00 cm^2/V-sec
D_r	34.4 cm^2/sec
t_r	150.0 psec
Cn	0.0266
Re	435.8
Sn	1.22×10^4

For the first series of calculations, a forward bias of three volts was applied across the device. The initial carrier distributions were set equal to the doping distribution and the electrostatic potential was initialized consistently (zero space charge). A mesh of 51 equally spaced grid points was used with $Re_c = 2$. An initial dimensionless time step of 0.001 was chosen. The time step was increased by a factor of 1.5 per time step to a maximum of 100. This maximum time step is 100 times the transit time across the device for a carrier traveling at its saturated drift velocity; thus it represents a time step of a magnitude which would typically be significantly larger than that used for an accurate transient calculation. It is also greater than the convective and diffusive limits associated with explicit integration of the continuity equations. The minimum time step is approximately 3.5 times the dielectric relaxation time. The dielectric relaxation time typically represents a stability limit for decoupled algorithms in which Poisson's equation is solved sequentially in the order followed here [15]. Thus if such a stability limit were present in either the SCC or SC^4 algorithms, the present choice of time steps should cause the solution to diverge or exhibit instability.

The convergence behavior for the fully coupled, SCC and SC^4 algorithms is shown in Fig. 3, where the maximum residual for the continuity equations, defined as

$$\mathrm{Res}_{\max} = \mathrm{MAX}(|\nabla \cdot J_n + R|, |\nabla \cdot J_p + R|), \tag{63}$$

and the magnitude of the time step are plotted as a function of time step number. The maximum residual for all three algorithms is reduced from approximately 1.0 to 1.0×10^{-7} in about 50 time steps and no signs of instability are evident. Thus, this comparison demonstrates that the time step restrictions associated with the dielectric relaxation time have been eliminated by the reformulation of the continuity equations in the SCC and SC^4 algorithms. The steady state distributions of potential and electron and hole concentration for this problem are shown in Figs. 4–6. In Fig. 5, the carrier concentrations are plotted on a log scale, whereas in Fig. 6 they are shown on a linear scale. All three algorithms converged to the same solution, as they should in the absence of any transient behavior.

It may also be observed from Fig. 3 that the manner in which the residual decays is similar for all three algorithms. This is particularly true of the fully coupled and the SC^4 algorithms. Such behavior suggests, at least for this test problem, that the $\mathcal{O}(\Delta t)$ error associated with the Poisson residual in the SC^4 algorithm does not adversely affect the transient accuracy of the solution.

Previously, it was stated that the current conserving algorithm would converge to a unique solution even if the initial Poisson residual was not zero. This property of the SC^4 algorithm was attributed to the manipulation of the continuity equations into a form which retains information about the doping distribution and which implicitly contains the relationship between the potential and the space charge given by Eq. (3). To demonstrate this property, a calculation was performed in which the initial potential distribution was at variance with the equilibrium dis-

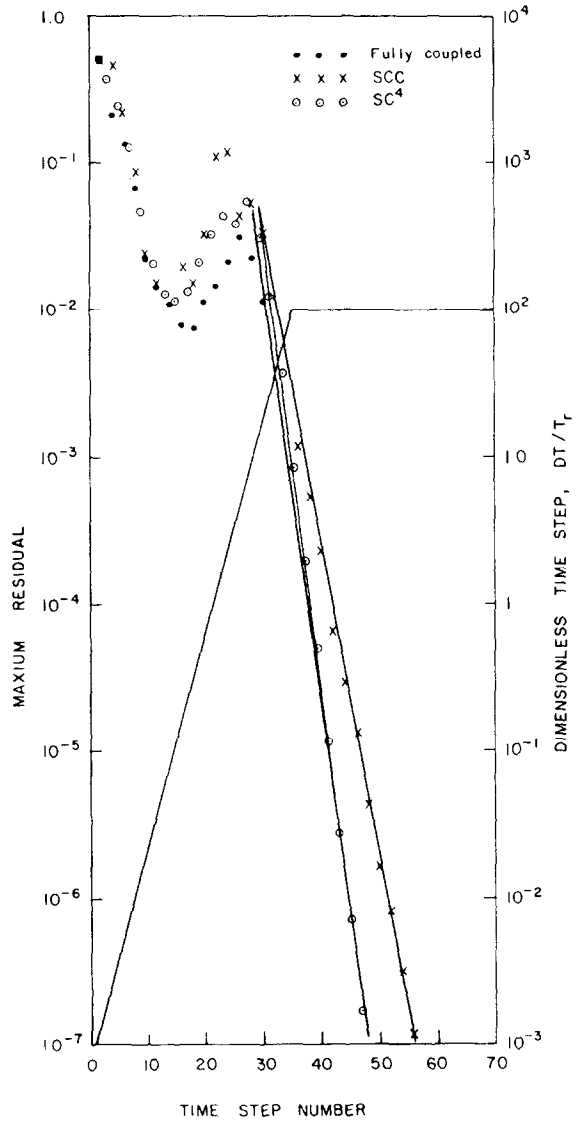

FIG. 3. Comparison of residual decay for fully coupled, SCC, and SC[4] Algorithms.

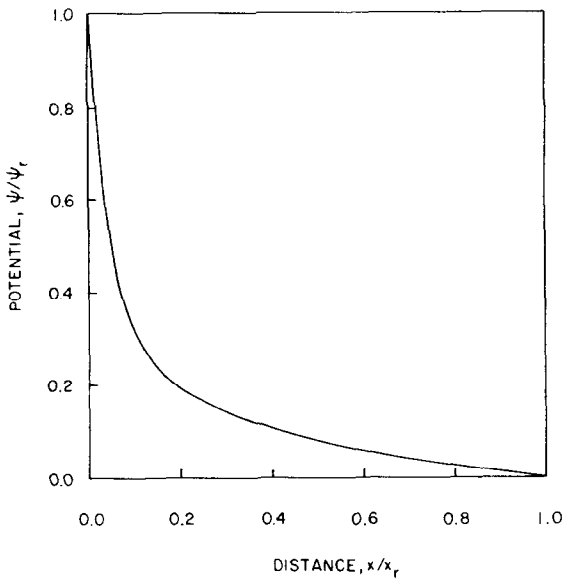

FIG. 4. Steady state potential distribution for forward bias junction (all algorithms).

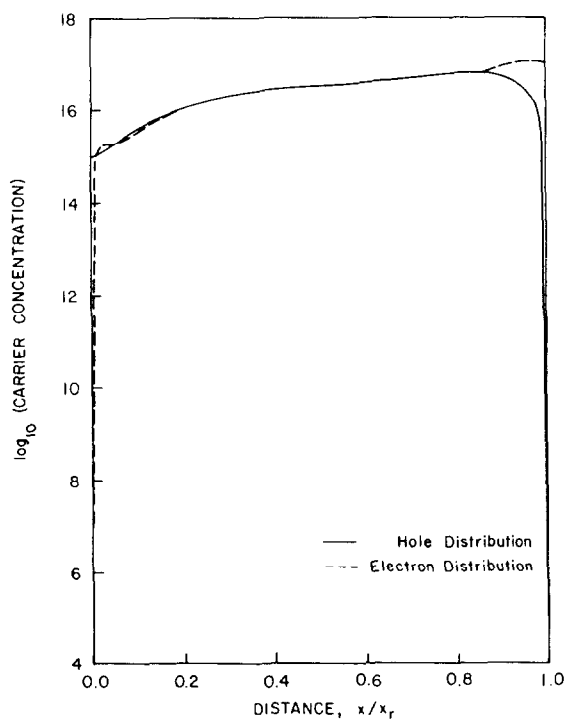

FIG. 5. Steady state distribution of electrons and holes for forward bias junction (all algorithms).

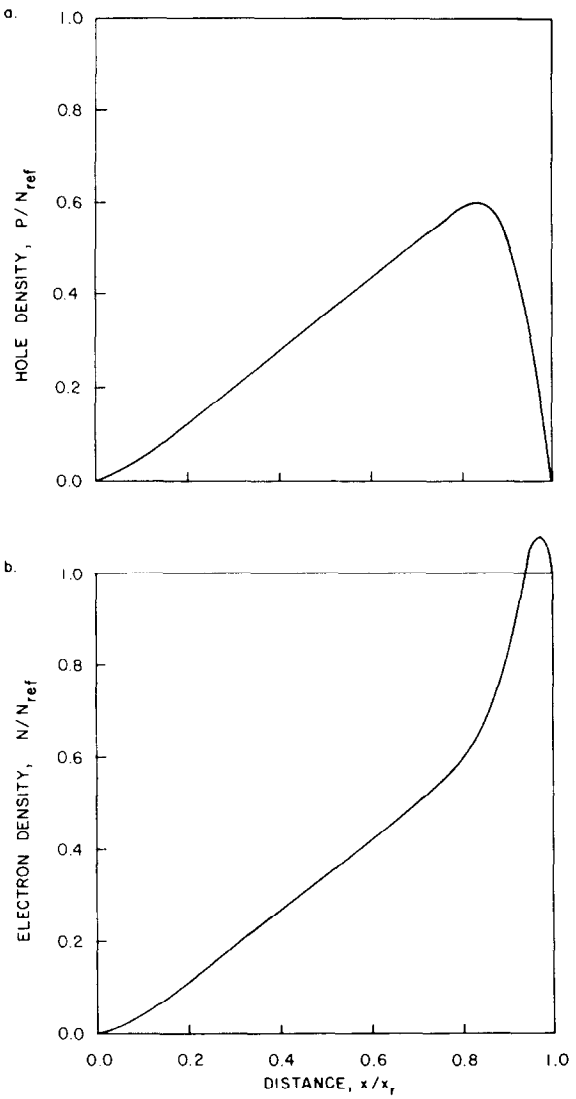

FIG. 6. Steady state hole (a) and electron (b) density distribution for forward junction (all algorithms).

tribution given by Eq. (3). The initial potential distribution *used* in this calculation varied linearly across the device as shown by the broken line in Fig. 7. This is the correct distribution only if the initial space charge, $\rho_i = 0$. However, the initial carrier distributions were not set equal to the doping distribution, but specified in a manner such that $\rho_i \neq 0$. The consistent potential distribution for the assumed carrier distribution is shown by the solid line in Fig. 7. The solution was integrated in time, using the inconsistent initial potential, to a steady result as discussed

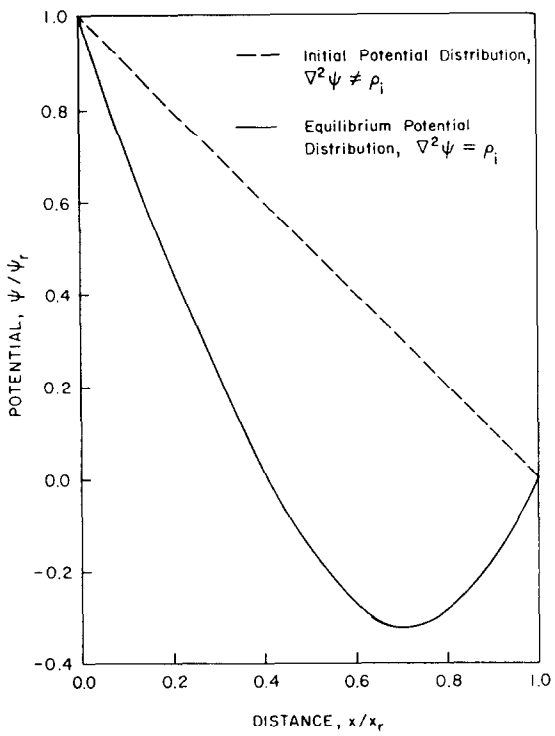

FIG. 7. Comparison of initial nonequilibrium and equilibrium potential distributions used in uniqueness test. Forward bias.

previously. The steady result was found to be identical to the solutions generated from consistent initial conditions, shown in Figs. 4–6, thus showing that a unique solution is obtained independent of the initial conditions, and that solution is identical to the solution obtained using a fully coupled algorithm.

The computations previously described serve to demonstrate the stability of the present method for large time steps as well as its convergence behavior. However, it is also important to understand the effects of mesh refinement and artificial diffusion on the computed solutions. To this end, the forward bias solution was recomputed varying the number of grid points and Re_c. Figure 8 shows the variation of potential across the device for $Re_c = 2$, using 14, 51, and 101 grid points. The effects of varying the number of grid points while maintaining the same value of Re_c is twofold. First, as the number of grid points is reduced the spatial truncation error increases. Second, since the grid spacing increases, the voltage drop between adjacent points also increases. Thus, a larger percentage of the total number of grid points will have $Re_{\Delta x} > Re_c$, resulting in artificial diffusion being introduced over more of the solution domain with a further degradation of the formal solution accuracy. The results shown in Fig. 8 demonstrate this effect, and as can be seen the solution obtained using only 14 points, while qualitatively correct,

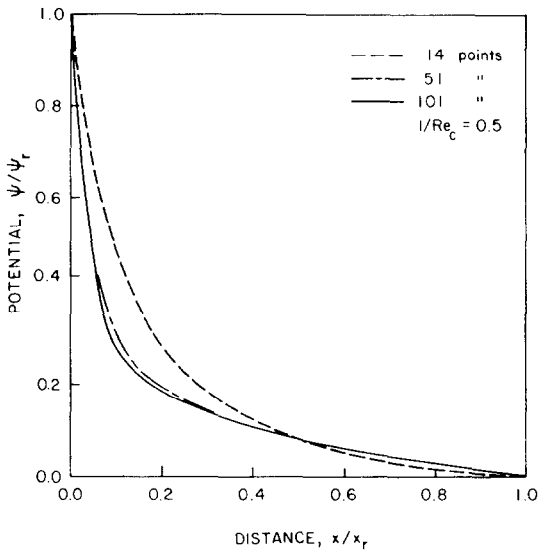

FIG. 8. Effect of grid spacing on potential distribution.

shows significant quantitative differences when compared to the solutions obtained using 51 and 101 points. The latter two solutions tend to collapse onto each other, indicating sufficient resolution. This is varified further in Fig. 9, where the current density is shown as a function of the number of grid points with $Re_c = 2$. In Fig. 9, the results are also compared to the solution obtained using 100 unequally spaced points; the grid constructed such that $Re_{\Delta x} < Re_c$ at all points (no artificial diffusion). As can be seen, the current asymptotically approaches a limiting value as the number of grid points is increased, and the result with 101 equally spaced points with artificial diffusion is very close to that obtained using the unequally spaced grid with no artificial diffusion.

The effect of artificial diffusion alone on the solution can be examined by varying

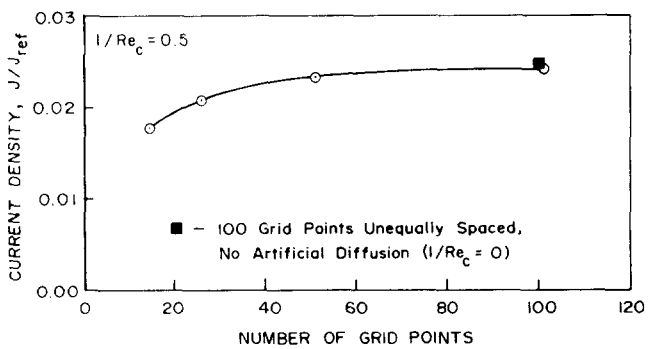

FIG. 9. Effect of mesh spacing on current flux.

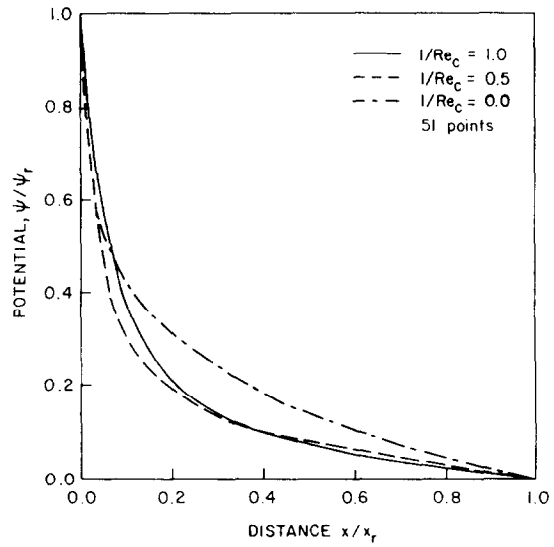

FIG. 10. Effect of artificial diffusion on potential distribution.

Re_c while keeping the number of grid points constant. The results for such solutions are shown in Figs. 10 and 11 for $Re_c = 1.0$, 2.0, and ∞ using 51 grid points. $Re_c = \infty$ gives no artificial diffusion. Figure 10 shows the solution for the potential, and when compared with Fig. 8 it is apparent that the solution for $Re_c = \infty$ is actually the most inaccurate on this mesh even though only physical diffusion is present. The voltage drop between some grid points is in excess of $2kT/e$, and spatial oscillations of the carrier densities were present in the converged solution. Such oscillatory solutions, although numerically exact, are of course nonphysical and may be grossly inaccurate, depending on the magnitude of the oscillations. The introduction of artificial diffusion when the potential difference between grid points exceeds $2kT/e$ damps these oscillations. There is, as a result, a trade-off between the inaccuracy caused by such oscillations in the solution and inaccuracy caused by artificial diffusion. This is demonstrated further in Fig. 11, where the current density for these solutions is shown as a functions of Re_c. It is apparent that without artificial diffusion the current is overpredicted by more than 100%, while introduction of artificial diffusion with $Re_c = 2$, which suppresses the oscillations in the carrier densities, results in less than 10% error. Values of $Re_c > 2$ generally reduce the magnitude of the oscillations, but do not suppress them entirely. $Re_c < 2$ introduces excessive artificial diffusion.

In addition to the forward bias solutions discussed above, a similar series of computations were performed for the same device subject to a reverse bias of three volts. All three algorithms again converged to the same solution, which is shown in Figs. 12 and 13. The effect of mesh resolution on the potential distribution is shown

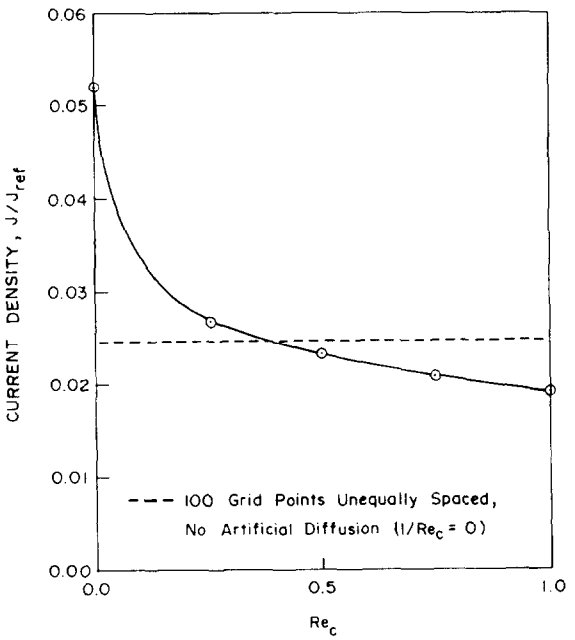

FIG. 11. Effect of artificial diffusion on current flux.

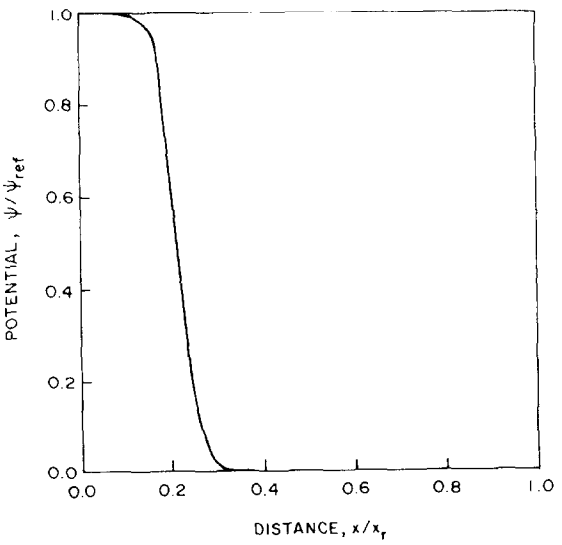

FIG. 12. Steady state potential distribution for reverse bias junction (all algorithm).

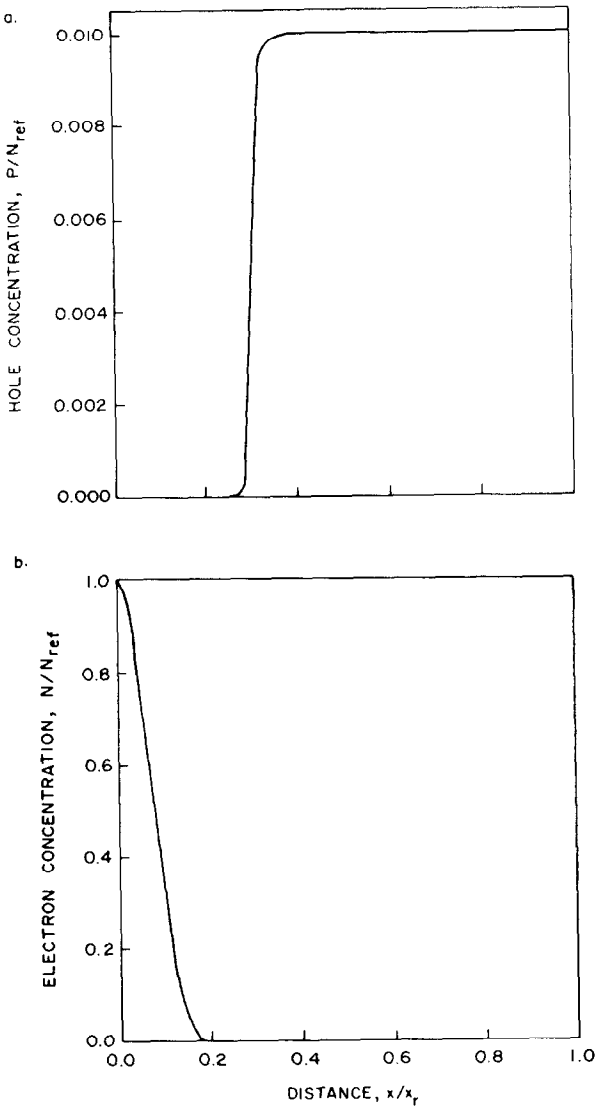

FIG. 13. Steady state hole (a) and electron (b) density distribution for reverse bias junction (all algorithms).

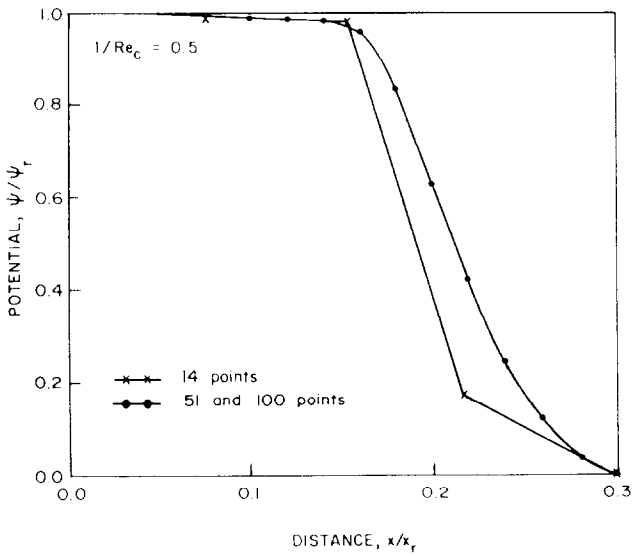

FIG. 14. Effect of grid spacing on potential distribution.

tribution is shown in Fig. 15. Here it is observed that without artificial diffusion, oscillations appear in the potential distribution. Whereas in the forward bias calculation, artificial diffusion was found to have a significant effect on the solution accuracy, particularly in the predicted current level, this is not the case for reverse bias. Under reverse bias, the majority of the potential drop occurs across the junction and artificial diffusion is introduced only in this region, depending on the

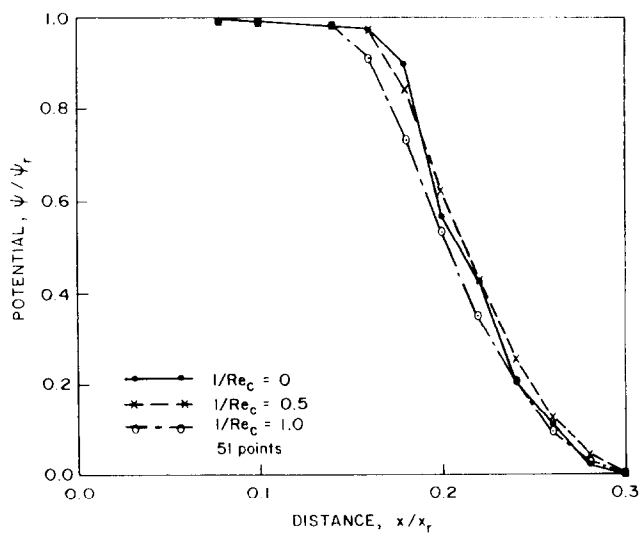

FIG. 15. Effect of artificial diffusion on potential distribution.

mesh spacing. Since under reverse bias the net current flux is zero, and since the drift and diffusive components of the current oppose each other other in the junction region, the effect of augmenting the physical diffusion coefficient is to smear, or broaden, the carrier profiles in this region. This is similar to shock smearing in fluid dynamics and only affects the solution in the immediate region of the junction. Thus, as is true in many simulation, if the detailed structure of the carrier concentration profiles at the edge of the depletion region is not of interest, artificial diffusivity can be effective in reducing the total number of grid points (within reasonable limits) required in the region of reverse bias junctions. Although this may not be of particular concern in one-dimensional calculations, the reduction in the total number of grid points in two- or three-dimensional simulations can be significant and result in a substantial reduction in the computational effort required to obtain a solution.

Three-Dimensional Transient Simulation

The full potential of the present technique is demonstrated here through the computation of the three-dimensional, transient response of a reverse bias silicon diode junction subjected to incident radiation using the SC[4] algorithm. The incident radiation leaves in its path a wake of excess electron-hole pairs. These excess carriers respond initially by diffusion and to the equilibrium field, separating slightly, and upsetting the charge balance within the device. The equilibrium field distorts and the excess carriers are collected over a finite length of time at the device contacts. When the majority of the excess charge is collected, the electric field is restored to its initial, equilibrium state.

The use of this problem as an example and test for the present algorithm is attractive because it involves large, localized disturbances within the device and requires that each components of current (electron, hole, and displacement) be rigorously conserved over the entire device volume. Additionally, the solution returns to the initial conditions once the charge is collected. Should any portion of the procedure fail, the results would immediately show evidence of such a failure.

The specific device structure considered is shown in Fig. 16. The $10^{18}/\text{cm}^3$ N^+ region is 2.0 μm in width, penetrates 0.5 μm into the device, and extends from the top to the bottom of the device. The P-type substrate is doped to $10^{15}/\text{cm}^3$. The doping profile was specified analytically as

$$C = (N + P_s)\, e^{-al^2} - P_s, \tag{64}$$

where

$$\begin{aligned} l &= x, \qquad z \leqslant 0.5\ \mu\text{m} \\ l &= \sqrt{x^2 + (z - 0.5)^2}, \qquad z > 0.5\ \mu\text{m} \end{aligned} \tag{65}$$

and

$$N_o P_o = N_i^2. \tag{66}$$

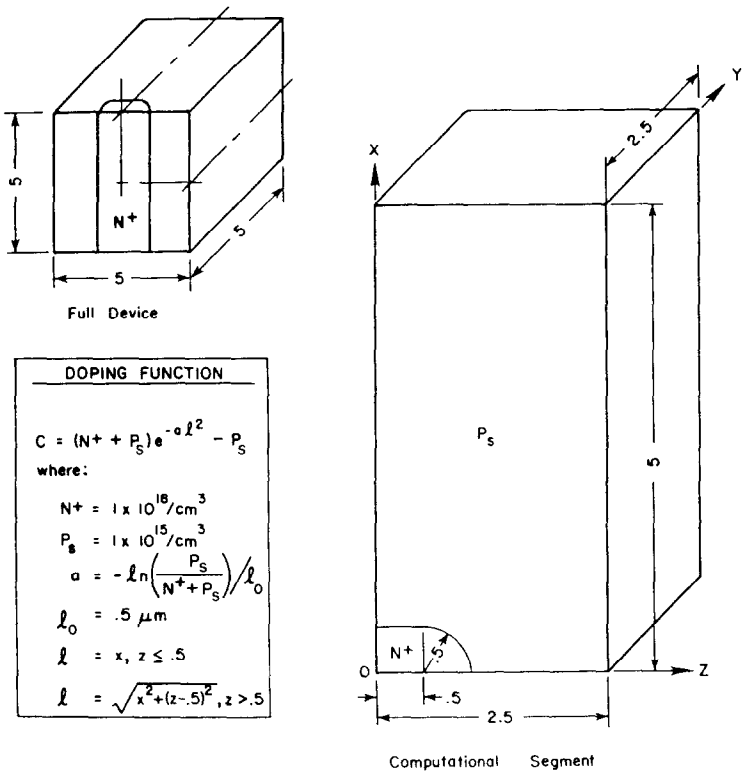

FIG. 16. Schematic representation of the three-dimensional device and computational segment. (all dimensions in μm unless noted.)

In Eq. (66), N_i is the intrinsic concentration of electrons and holes, and is taken as $1.0 \times 10^{11}/\text{cm}^3$ for the present simulation. Equations (64) and (66) are solved simultaneously to obtain N_0 and P_0. The contacts, which cover the N^+ region on the surface, and the opposite P surface are assumed to be ohmic, thus the carrier densities are assumed equal to the doping densities on the contacts. A potential of 0 volts is applied to the P contract and $+3$ volts is applied to the N^+ contact. All other surfaces are assumed to be insulating and the corresponding boundary conditions (discussed previously) are applied.

Due to the symmetry of the device structure, only a quadrant of the device need be considered in the computation, as is also shown in Fig. 16. The path of the ionizing particle is assumed to lie along the x axis of the computational segment. A mesh of 28 x points by 25 y points by 25 z points was used. The mesh point distributions in the x–z, x–y, and y–z planes are shown in Figs. 17a–c, respectively. The mesh points are distributed to provide adequate resolution of the track of the ionizing particle, as well as the depletion region surrounding the junction. The reference quantities and dimensionless constants used in this simulation are given in Table III.

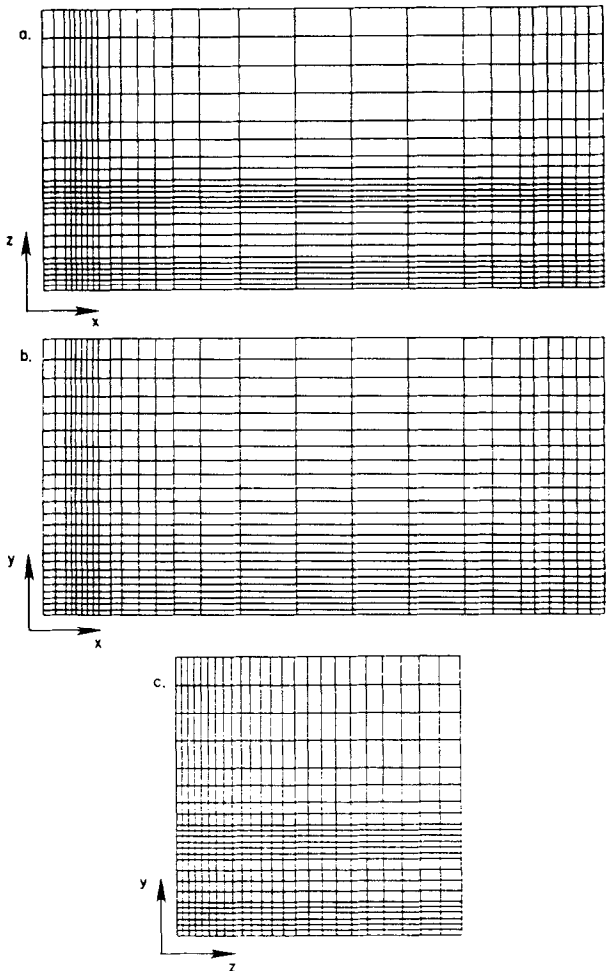

FIG. 17. Grid structure in (a) x–z plane, (b) x–y plane, and (c) y–z plane.

TABLE III

Quantity	Value
N_r	$1.0 \times 10^{18}/\text{cm}^3$
ψ_r	3.0V
X_r	5.0×10^{-4} cm
V_r	1.0×10^{7} cm/sec
μ_r	1300.0 cm^2/V-sec
D_r	34.4 cm^2/sec
t_r	50.0 psec
Cn	0.798
Re	145.0
Sn	1.36×10^{4}

Since the device structure prior to being struck by incident radiation has no variation in the y coordinate direction, the equilibrium distribution of the carriers and the electrostatic potential is not a function of that coordinate. The equilibrium solution, which serves as the initial condition for the three-dimensional transient simulation, was therefore computed two-dimensionally for only a single x–z plane. The resulting undisturbed equilibrium solution at any plane of constant y is shown as contour plots of electron and hole density, and of potential in Fig. 18. The

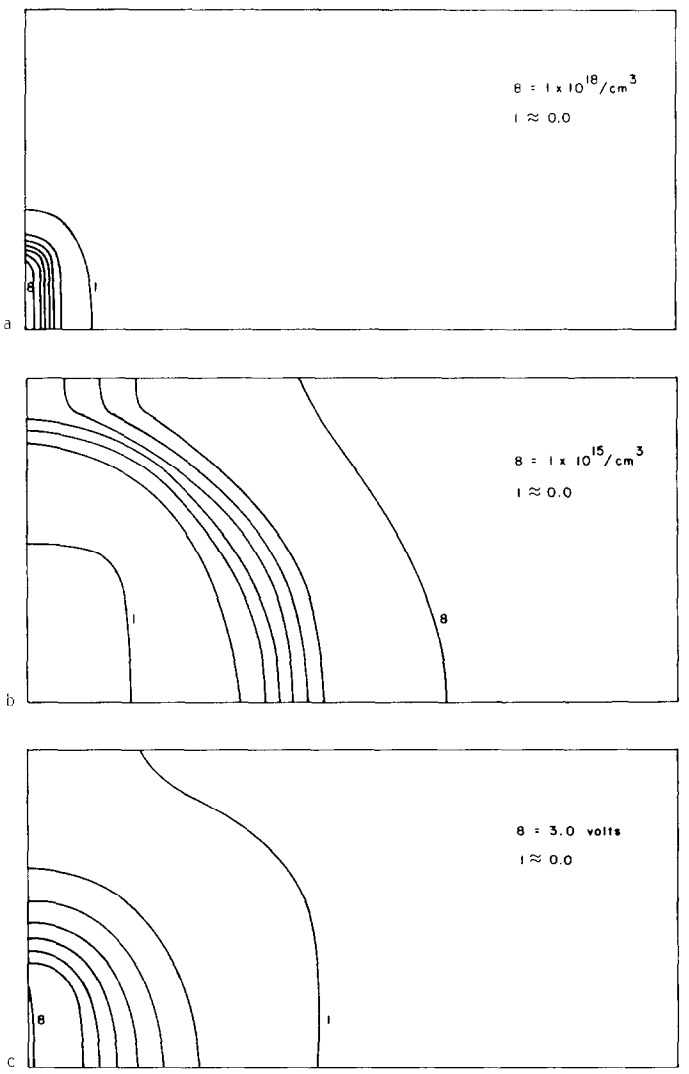

FIG. 18. Equilibrium solution contour of (a) electron density, (b) hole density, (c) potential. Contours are equally spaced.

potential is shown in Fig. 19.

The three-dimensional simulation was initialized from the equilibrium solution through introduction of the excess carriers along the track of the ionizing particle. The generation of the carriers was assumed to take place over a period of approximately 10 psec through introduction of an exponential source term in the governing equations at each grid point along the track. The track was 4.5 μm in length and the effective diameter was 1130 Å. The total excess charge implanted in the device was approximately 0.0072 pcoul. A time step of 0.05 psec was used for the first 50 steps of the solution. The time step was allowed to increase to 0.25 psec over the next 50 time steps, to 0.5 psec from time step 101 to 150, and to 1.25 psec from step 151 to the end of the simulation. A total of 350 time steps were taken to a final time of approximately 290 psec.

Before discussing the results of the transient computation it is appropriate to describe the basic aspects of the device response. Prior to introduction of the excess electron-hole pairs the equilibrium potential and space charge distributions within the device satisfy Poisson's equation,

$$\nabla^2 \psi_e = -\frac{e}{\varepsilon} \rho_e, \tag{67}$$

where the subscript, e, refers to the equilibrium state. Since no current is flowing and since the gradients of the carrier densities normal to the contacts are

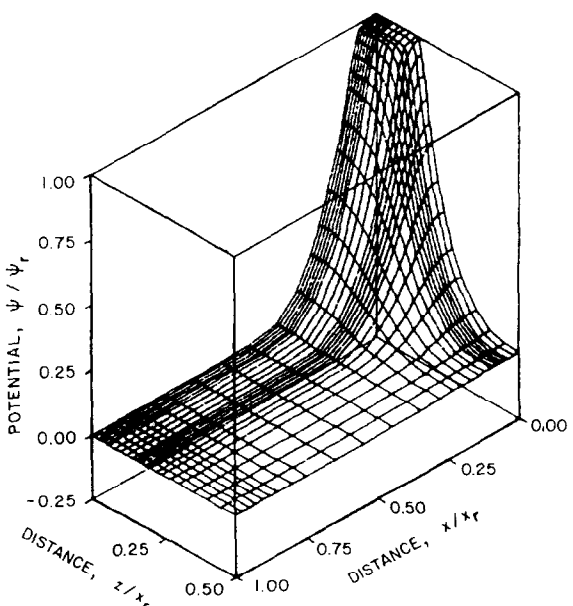

FIG. 19. Surface plot of equilibrium solution potential.

(analytically) zero, the normal components of the electric field at the contacts are also zero. Furthermore, on all other surfaces of the device zero normal field boundary conditions are enforced. Applying Green's theorem to Eq. (64)

$$\int_V \nabla \cdot (-\mathbf{E}_e)\, dV = -\int_V \frac{e}{\varepsilon} \rho_e\, dV = \oint_S -\mathbf{E}_e \cdot d\mathbf{S}, \tag{68}$$

which states that the volume integral of the equilibrium space charge is equal to the surface integral of the normal component of the electric field, ($\mathbf{E} = -\nabla\psi$). Since $\mathbf{E} \cdot \mathbf{S}$ is zero everywhere, the integral of the equilibrium space charge is also zero. Now, the introduction of the excess carriers does not initially upset this balance since an equal number of electrons and holes are generated. However, during the initial phase of the transient, the excess carriers will respond to the equilibrium field and will also migrate by diffusion. A portion of the electrons which are deposited in the depletion region respond to the equilibrium field and are rapidly drawn to the N^+ region. Holes are correspondingly forced into the substrate. This separation of carriers gives rise to a local upset in the depletion region charge balance. The equilibrium junction field begins to collapse, and electric fields at the contacts begin to appear. Since no particle current has yet begun to flow out of the contacts, Green's theorem again shows that the area averaged fields at the N^+ and P contacts are equal! If the mobilities and contact concentrations of both carriers were equal, then the flux of electrons out the N^+ contact would equal the flux of holes out of the P contact and the volume integral of the space charge would remain unaltered. However, the electron concentration at the N^+ contact is significantly higher than that of the holes at the P contact, and the electron mobility is also greater than that of the holes. As a result, electrons are initially collected more rapidly than holes. This alters the space charge integral, and further field distortion results. Since the electron currents at the N^+ contact are greater than the hole currents at the substrate contact, conservation of total current requires that the additional field distortions manifest themselves as displacement currents at the substrate contact. This charge imbalance and the field at the substrate contact continue to grow until the field at the P contact is high enough to yield hole currents which balance the electron current at the N^+ contact. Once this is established, a quasi-equilibrium situation is established during which the majority of excess carriers are collected. As the last few electrons are collected, the remaining holes continue to flow out of the P contact balanced by displacement currents resulting from the return of the field to its initial equilibrium state, as the original space charge distribution is restored. The displacement currents at the N^+ contact are negligible since the high doping concentration prevents significant field distortions.

Simultaneous with the collapse of the junction field and establishment of a field at the P contact the electrons and holes begin to migrate radially outward from the

axis of the track, initially by diffusion. Due to the difference in the diffusion coefficients, the electrons spread more rapidly, which gives rise to a secondary charge imbalance all along the axis of the track, particularly in the substrate. This generates a radially directed field which augments the radial outward flow of the holes by drift while drawing the electrons back towards the center of the track. As a result, the electrons and holes tend to spread radially at a uniform rate which is greater than that of the diffusion of holes, but less than that of the diffusion of electrons.

The description of the charge collection transient given above is illustrated graphically in the following figures. Fig. 20 shows the current pulse and charge collected at the N^+ contact. This current pulse is predominenty due to the flux of electrons, which were initially deposited along the track, out of the device. The initial peak in the current is associated with the rapid distortion of the junction field

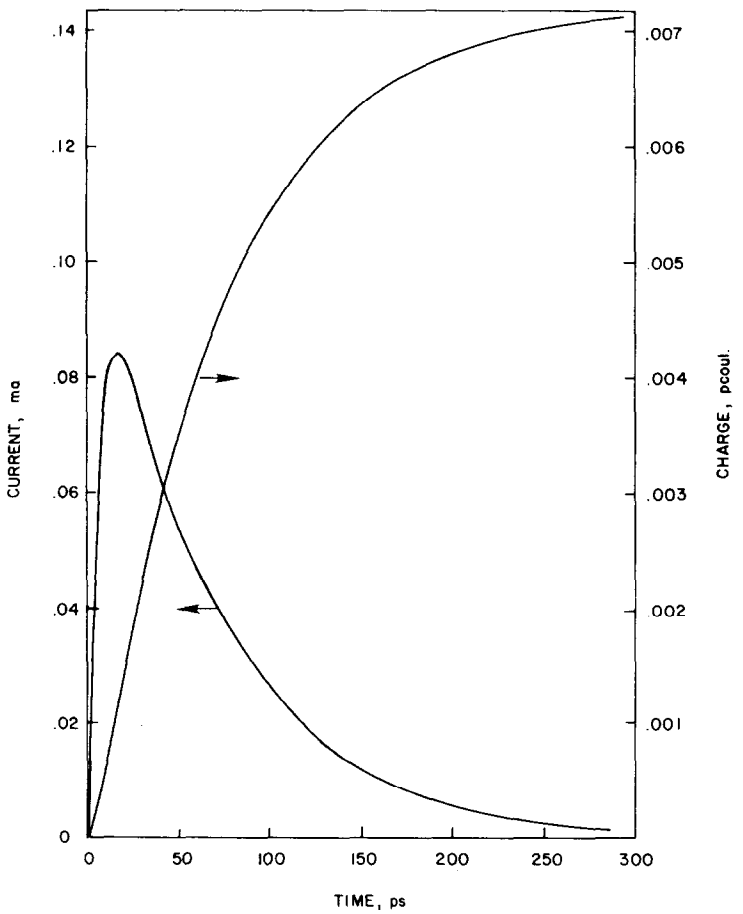

FIG. 20. Current and charge vs time at N^+ contact.

and establishment of the field at the P contact. The particle and displacement current pulses at the P contact are shown in Fig. 21. During the initial phase of the transient the current is primarily the displacement current, which is associated with the rapid establishment of the field at the P contact. From both Figs. 20 and 21 it is apparent that after 290 psec the majority of the excess carriers (approximately 99%) have been collected. This is further borne out by examination of the "displacement charge" at the P contact, shown in Fig. 22. The displacement charge (the time integral of the displacement current) represents the departure from the equilibrium field distribution at the contact

$$\begin{aligned} Q_D &= \int_0^t \int_A -\varepsilon \frac{\partial \nabla \psi}{\partial t} \cdot d\mathbf{A}\, dt \\ &= -\int_A \varepsilon(\nabla\psi|_t - \nabla\psi|_0) \cdot d\mathbf{A}. \end{aligned} \tag{69}$$

Since after 290 psec the displacement charge is nearly zero, it is apparent that the field at the P contact has almost returned to its initial state.

It is also possible to determine the average field at the P contact at any time during the transient from the displacement charge. Thus, it is observed that the

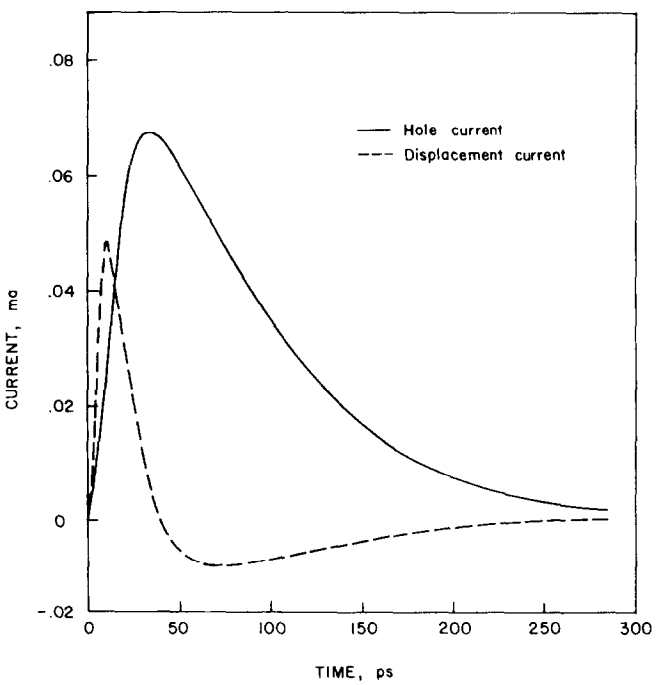

FIG. 21. Hole and displacement current vs time at P contact.

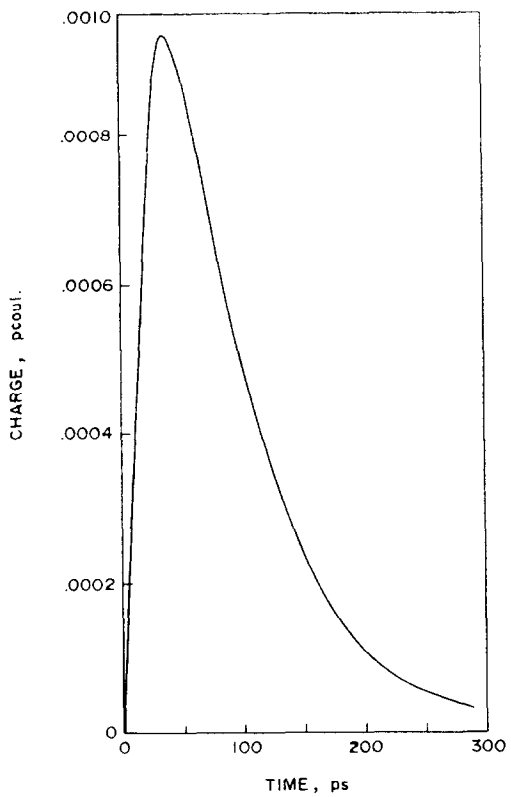

FIG. 22. Displacement charge vs time at P contact.

maximum average field at the contact is approximately 4.2×10^3 volts/cm at 35 psec into the transient. This is well below the value of electric field at which the hole velocity saturates and gives a hole velocity of the order of 1.65×10^6 cm/sec. With the density of holes at the contact equal to $1 \times 10^{15}/\text{cm}^3$, and the contact area of $2.5 \times 10^{-7}\,\text{cm}^2$ the drift component of current is 0.066 mA. From Fig. 21 the total current at the P contact at this time is 0.067 mA, thus the holes are collected predominantly by drift at the contact.

The details of the distortion of the electric field are shown in Figs. 23 and 24. Figure 23 shows the potential distribution at several instants during the transient in the x–z plane, which contains the track of the ionizing particle. Figure 23a shows the potential distribution after 2.5 psec. Here, the small radial fields along the axis of the track are evident as is the initiation of the axial field between the end of the track and the P contact. The junction field has also begun its collapse, although this is not clearly evident at this time. Figure 23b shows the same result after 14 psec. The collapse of the junction field and the high field at P contact are shown clearly here. After 30 psec the potential distribution in this plane appears as shown in

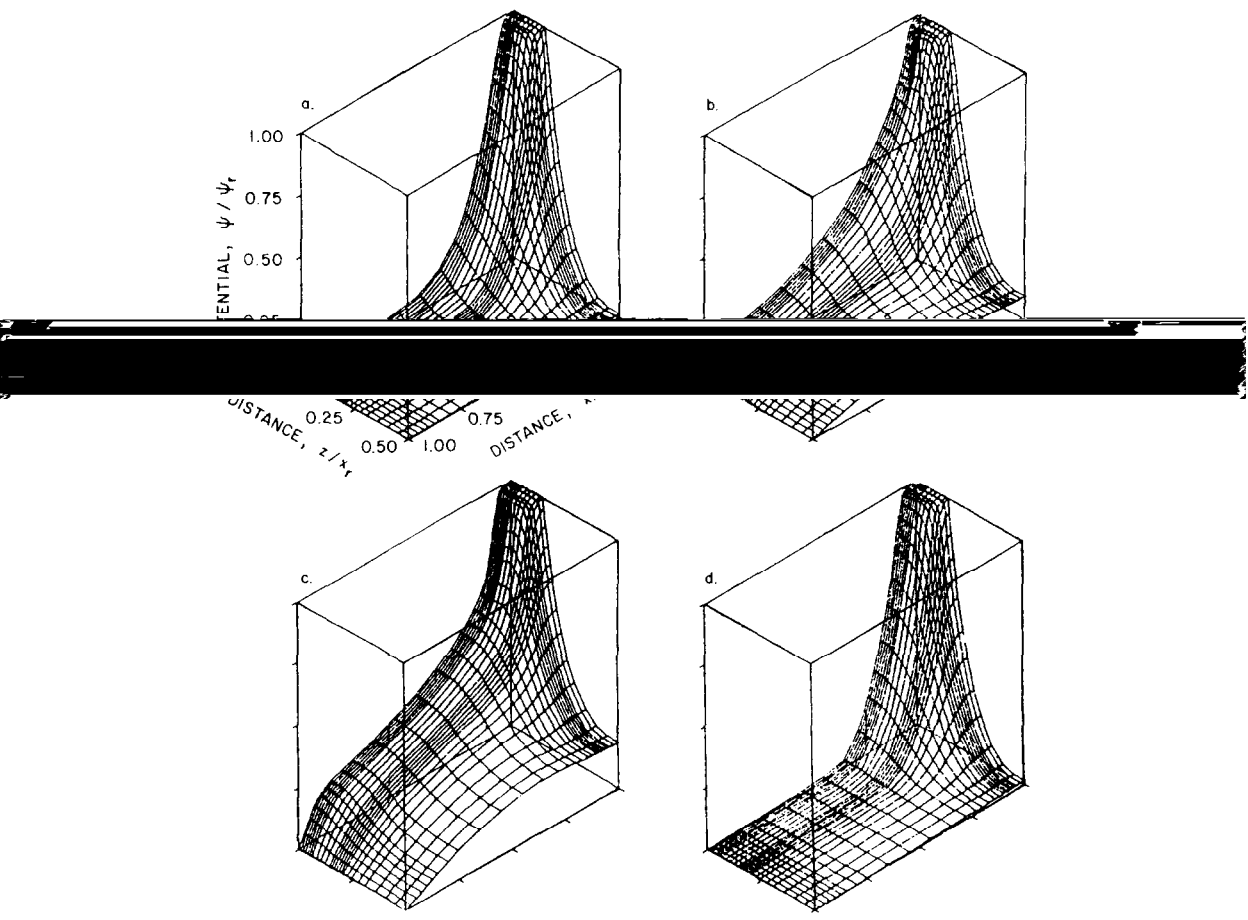

FIG. 23. Potential surface in the x–z plane of the track at (a) 2.5 psec, (b) 14 psec, (c) 38 psec, and (d) 290 psec.

Fig. 23c. The potential surface in the plane of the track retains this distorted shape, more or less, until field restoration occurs, as shown in Fig. 23d, after 290 psec.

In Fig. 24 contour plots of the potential in the y–z plane 1.4 μm below the N^+ contact are shown at 2.5, 14.0, 38.0, and 100 psec. This y–z plane is normal to the axis of the track of the ionizing particle, which passes through this plane at the upper-left-hand corner of the plots. The contour values are equally spaced and the maximum and minimum values are noted in each frame. The contours clearly show that initially the spreading of the disturbance is axisymmetric in nature. However, as time progresses, and the disturbance propagates outward further from the axis of the track, the structure of the device exerts a significant influence on the potential distribution and the axisymmetric structure is lost.

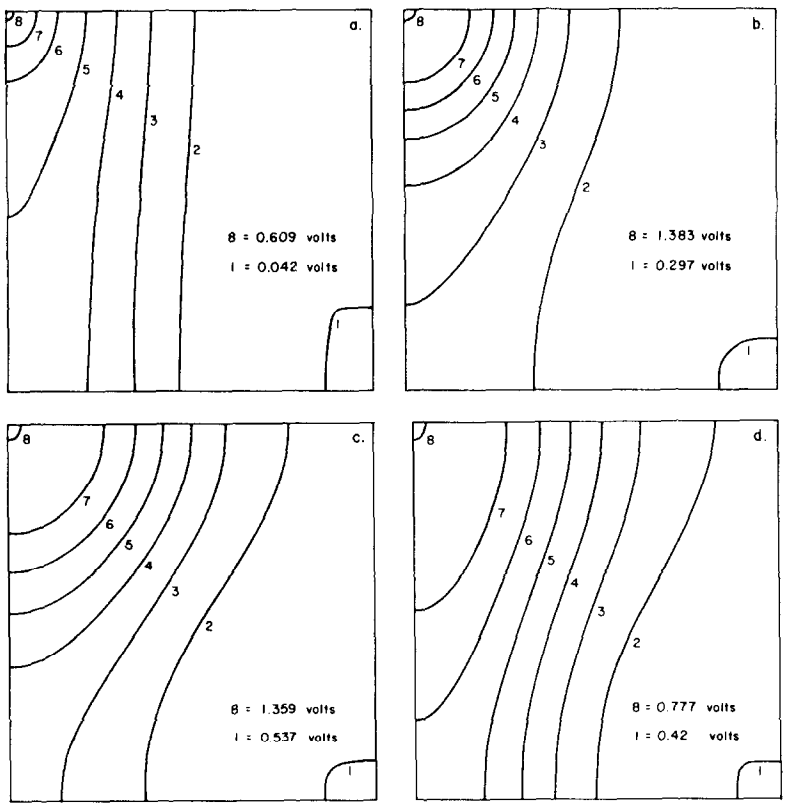

FIG. 24. Comparison of potential contours in a plane normal to the axis of the track, 1.4 μm, below the N^+ contact, at (a) 2.5 psec, (b) 14.0 psec, (c) 38.0 psec, (d) 100.0 psec. Contours are equally spaced.

The distribution of the electron concentration in the plane of the track is shown in Fig. 25 at various times following the strike. The lateral spreading of the electrons is clearly evident, as is the shortening of the track at latter times indicating the collection of the excess electrons. At 290 psec following the strike, the concentration of the remaining excess electrons is too small to appear on the plots for the chosen contour values, further indicating that the majority of the excess charge has been collected, and the electron distribution has almost returned to the initial state. The distribution of the excess holes closely follows that of the electrons.

This simulation demonstrates the capability of the present algorithm, particularly with regard to the important consideration of current conservation. The current flux and charge collected at the N^+ and P contacts agreed to within five significant figures throughout the computation. With regard to the field distortions during the transient, it should be noted that the present authors have predicted qualitatively similar results for two-dimensional simulations [26], as have Hsieh, Murley, and O'Brien [27].

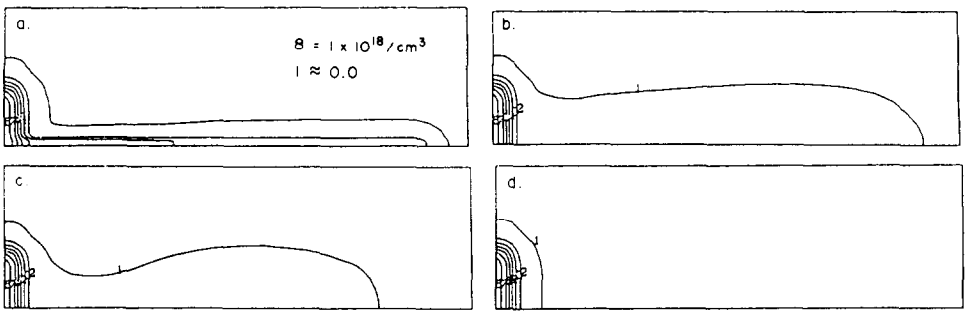

FIG. 25. Comparison of electron density contours in the plane of the track at (a) 2.5 psec, (b) 38 psec, (c) 100 psec, and (d) 290 psec. Contour values are equally spaced.

The computation time required for the present simulation was 3.4 CPU hours on a CRAY-1 computer. For the 17,500 grid points and 350 time steps used in the simulation this is approximately 0.002 CPU sec per grid point per time step. Although the simulation was performed on a CRAY-1, the present version of the code used was not developed for vector machines, thus when vectorized a substantial reduction in run time is anticipated. Additionally, the present code was based on a general LBI algorithm which is coded to handle arbitrary systems of coupled PDEs and a substantial penalty in CPU time is incurred for this generality.

ACKNOWLEDGMENTS

The authors wish to thank their colleagues, W. R. Briley and B. C. Weinberg for many fruitful discussions which contributed to the development of this procedure. This work was supported in part by the National Science Foundation under Grant ECS-811439 and by the Defense Nuclear Agency under Contract DNA001-83-C-0410.

REFERENCES

1. P. E. COTTRELL AND E. M. BUTURLA, in *Proceedings, NASECOD I Conference, Dublin, Ireland, 1979*, edited by B. T. Brown and J. J. H. Miller (Boole Press, Dublin, 1979), p. 31.
2. W. R. BRILEY AND H. MCDONALD, *J. Comput. Phys.* **34**, 54 (1980).
3. W. R. BRILEY AND MCDONALD, *J. Comput. Phys.* **24**, 372 (1977).
4. I. LINDEMUTH AND J. KILLEEN, *J. Comput. Phys.* **13**, 181 (1973).
5. H. L. GRUBIN, J. P. KRESKOVSKY, AND B. C. WEINBERG, *IEEE Trans. Nuclear Sci.* **NS-31**, 1161 (1984).
6. J. P. KRESKOVSKY AND H. L. GRUBIN, *IEEE Trans. Nuclear Sci.* **NS-32**,4140 (1985).
7. J. P. KRESKOVSKY AND H. L. GRUBIN, *Solid State Electron.* **29** (1986).
8. M. S. MOCK, *Analysis of Mathematical Models of Semiconductor Devices* (Boole Press, Dublin, 1983).
9. A. DEMARI, *Solid State Electron.* **11**, 1021 (1968).
10. J. P. KRESKOVSKY, W. R. BRILEY, AND H. MCDONALD, *AIAA J.* **22**, 374 (1984).

11. S. J. SHAMROTH AND H. J. GIBELING, NASA Contractor Report 3183 (1979).
12. H. MCDONALD AND W. R. BRILEY, in *Numerical and Physical Aspects of Aerodynamic Flows*, edited by T. Cebeci (Springer-Verlag, New York, 1982), p. 99.
13. J. DOUGLAS AND J. E. GUNN, *Numer. Math.* **6**, 428 (1964).
14. D. W. PEACEMAN AND H. H. RACHFORD, *J. Soc. Indus. Appl. Math.* **3**, 28 (1955).
15. M. S. MOCK, in *Proceddings, NASECODE I Conference, Dublin, Ireland, 1979*, edited by B. T. Brown and J. J. H. Miller (Boole Press, Dublin, 1979), p. 120.
16. R. E. BANK, D. J. ROSE, AND W. FICHTNER, *IEEE Trans. Electron. Devices* **30**, 1031 (1983).
17. D. L. SCHARFETTER AND H. K. GUMMEL, *IEEE Trans. Electron. Devices* **16**, 64 (1969).
18. P. J. ROACHE, *Computational Fluid Dynamics* (Hermosa, New Mexico, 1972).
19. D. N. DE, G. ALLEN, AND R. V. SOUTHWELL, *Quart. J. Mech. Appl. Math.* **8**, 129 (1955).
20. T. M. EL-MISTIKAWY AND M. J. WERLE, *AIAA J.* **16**, 749 (1978).
21. J. P. KRESKOVSKY, *IEEE Trans. Electron. Devices*, to appear (1987).
22. E. ISAACSON AND H. B. KELLER, *Analysis of Numerical Methods* (Wiley, New York, 1966).
23. D. M. CHAUGHEY AND R. F. THOMAS, *Proc. IEEE*, 2192 (1967).
24. M. KURATA, *Numerical Analysis for Semiconductor Devices* (Lexington Books, Lexington, Mass., 1982).
25. J. P. KRESKOVSKY, W. R. BRILEY, AND H. MCDONALD, in *Computers in Flow Predictions and Fluid Dynamics Experiments*, Edited by K. N. Ghia, T. J. Mueller, and B. R. Patel (ASME, New York, 1981), p. 129.
26. H. L. GRUBIN, J. P. KRESKOVSKY, AND B. C. WEINBERG, Scientific Research Associates Report, R84-940002-F (1984), unpublished.
27. C. M. HSIEH, P. C. MURLEY, AND R. R. O'BRIEN, *IEEE Electron. Devices Lett.* **2**, 103 (1981).