J. H. Hohl

Variational Principles for Semiconductor Device Modeling with Finite Elements

Abstract: Variational principles related to three areas of semiconductor device modeling by the finite element method are presented. Some subtle points which are crucial to the successful application of the method are explored. It is suggested that the validity of the selected variational formulations must be carefully ensured, and that the physics disciplines provide the best guidance for the right selections.

Introduction

The finite element method has been widely and successfully used for quite some time for the analysis of heat transfer, mechanical deformation and similar phenomena; but to date, little has been published on its application in the modeling of semiconductor device behavior, although it would seem predestined for this purpose. References [1–3] are characteristic of the work in this area that has been published to date. This state of affairs may be due, in part, to the natural inertia which the introduction of new techniques into any field must overcome. A closer look, however, reveals some more basic obstacles.

The extensive theories of semiconductor technology are almost exclusively cast in differential formulations, which lend themselves most readily to numerical calculations by means of finite difference schemes. By contrast, the finite element method is preferably based on variational descriptions of the phenomena to be modeled, or on mathematical approximation methods involving the minimization of errors, such as Rayleigh-Ritz or Galerkin schemes. In most applications, it is a numerical technique for extremizing a functional. To this end the region over which the functional is defined, e.g., a large rectangle, is subdivided into subregions of simple geometric shape, e.g., small rectangles. These are called finite elements. Within each finite element the argument function of the functional is approximated by a simple function, e.g., a low order polynomial. The coefficients of this interpolation function are chosen such that the approximation, and perhaps some of its derivatives, are continuous across the boundaries of the finite elements. In the case of the rectangle, one may use a product of linear interpolation functions such that the approximation along each edge is a linear interpolation between the values at the adjacent vertices. This leads to continuity across the boundaries of the elements for the approximation, but not for its derivatives. Next, the contribution of each element to the functional is determined formally by integrating over the finite element, using the approximation. For a linear or linearized problem, the result is quadratic in the values of the argument function at particular points of the finite element, e.g., the vertices of the rectangular elements. Let these points be named nodes. As the contributions of all elements are summed, the value of the functional is obtained formally, as a function of all node values of the approximated argument function. The functional can now be extremized formally by setting its partial derivatives with respect to each node value equal to zero. This leads to a system of coupled equations. In essence, the finite element method is a collection of structured, algorithmic techniques for performing the steps just described.

Variational formulations are rare in semiconductor device theory. Short of developing the formulations from principles of physics, investigators are forced either to develop Rayleigh-Ritz or Galerkin schemes or to adapt formulations from related fields. The latter approach, although sometimes favored or recommended in the literature [4], is often treacherous.

This contribution reviews some of the variational principles applicable to semiconductor device modeling, together with others that seem to be useful, but are not. A variational formulation of transient diffusion processes that does not produce valid results and a similar, not very widely known formulation that can solve such problems

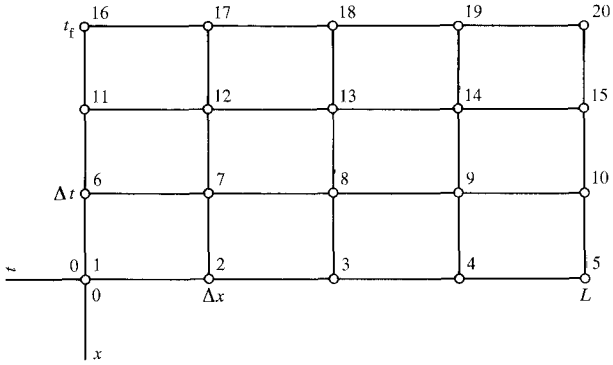

Figure 1 Space-time lattice for finite element solution of one-dimensional diffusion problem.

are outlined in the second section, which also demonstrates the subtlety of the differences in the two finite element discretizations. The third section is devoted to the development of a variational principle describing the state of thermostatic equilibrium in a region of arbitrarily doped semiconductor material. This example highlights the fact that a seemingly straightforward argument can easily lead to an erroneous formulation. The last section presents a variational description of current flow conditions in a semiconductor region and points out the stringent limitations of that formulation.

Transient diffusion problems

With few exceptions, the impurity profiles forming planar semiconductor devices are fabricated by diffusion processes. Predeposition diffusions are typically controlled by boundary conditions of a constant impurity concentration over parts of the boundary surface and a vanishing impurity flux over the remainder of the boundary surface. Drive-in diffusions typically evolve from an initial profile and boundary conditions of vanishing flux. Neither process involves explicit sources.

Mathematically the impurity profiles are described by solutions of the diffusion equation

$$D\nabla^2 C(\mathbf{r}, t) - \partial C(\mathbf{r}, t)/\partial t = 0, \tag{1}$$

with initial and boundary conditions

$$C(\mathbf{r}, 0) = C_0(\mathbf{r}), \tag{2}$$

$$C(\mathbf{r}_{s1}, t) = C_s, \qquad \partial C(\mathbf{r}_{s2}, t)/\partial n = 0, \tag{3}$$

where D and C respectively denote diffusivity and impurity concentration; $\mathbf{r}_{s1}$ and $\mathbf{r}_{s2}$ define mutually exclusive regions on the boundary surface; and where $\partial C/\partial n$ denotes the directed derivative normal to and evaluated at the surface.

In order to apply the finite element method to the calculation of such impurity profiles, it is desired to describe the diffusion process in terms of a variational principle. In searching the literature, one is easily led toward a canonical Hamiltonian formulation [5] and a technique for finite element discretization [6]. The canonical functional is in the form

$$\Omega = \int\int\int_V \int_{t_0}^{t_f} [D\nabla\bar{C} \cdot \nabla C + \tfrac{1}{2}(\bar{C}\partial C/\partial t - C\partial\bar{C}/\partial t) + S\bar{C} + S\bar{C}]dtd\mathbf{r}, \tag{4}$$

where V denotes the volume under consideration.

Extremizing Eq. (4) with respect to $\bar{C}$ reproduces Eq. (1), with an added source term $S(\mathbf{r}, t)$, and extremizing with respect to C leads to

$$D\nabla^2\bar{C} + \partial\bar{C}/\partial t = \bar{S}(\mathbf{r}, t), \tag{5}$$

which is called the adjoint equation to (1), with adjoint variables $\bar{C}$ and $\bar{S}$. A finite element algorithm can be derived from Eq. (4), extremized with respect to $\bar{C}$, by standard techniques.

Nothing is lost for the thoughts to be developed if attention is focused on a one-dimensional and linear example. To be specific, consider a plate of semiconductor of thickness L, with the impurity concentrations fixed at both surfaces, and with the diffusion process starting at $t = 0$. Using four length elements and three time steps leads to rectangular elements and to the space-time lattice illustrated in Fig. 1.

The functional, Eq. (4), now reduces to a double integral with limits 0, L and 0, t_f, respectively. Extremizing its finite element approximation, with respect to those node parameters which are not fixed by initial or boundary conditions, generates a system of linear equations for those parameters. The equations for the parameters at the nodes 7, 8, 9, 12, 13, 14, 17, 18, and 19 of Fig. 1 are derived in Appendix A. All equations are coupled.

On the other hand, the physics of the diffusion process under consideration suggests an iterative solution: The parameters at the nodes 7, 8, and 9 must be independent of the future conditions at the nodes 11 to 20. This contradiction with the above system of equations arouses suspicion.

The validity of the equations can be tested with a spatially linear, time-invariant impurity profile. Since the linear interpolation functions used exactly represent such a profile, the equations must analytically reproduce the profile. Consequently, inserting such a solution into the equations, e.g., using the values 1, 2, 3, 4, and 5, respectively, at the nodes of the first, second, third, fourth, and fifth columns in Fig. 1, must convert the equations into identities. One finds, however, that the equations for the nodes 17, 18, and 19 fail this test; hence, the finite element formulation does not properly model the diffusion process.

The reason for this failure is buried in the canonical formulation of the variational principle. The functional (4) could, in principle, be used to calculate the sources which would have to be applied to the surfaces of the plate in order to evolve from a given initial impurity profile to a specified final profile at the time t_f. Indeed, the finite element approximation will yield a system of equations which solves such problems. But, since the physical diffusion processes are causal, they strongly restrict the variety of final profiles that can be reached from a given initial profile under the influence of physical sources. The canonical formulation, (4), removes these restrictions by virtue of the inclusion of the adjoint variables, which satisfy the noncausal, nonphysical equation, (5), thereby rendering the formulation useless for solving real diffusion problems.

It is possible to apply the finite element method to the spatial domain only. This results in a system of ordinary differential equations in time which, in turn, can be solved by either a finite element or a finite difference scheme [7]. Such an approach, however, is more cumbersome than a straightforward finite difference solution.

A variational formulation that is explicitly causal has been published by Gurtin [8]. The functional for the diffusion problems considered here is of the form

$$\Omega = \int \int \int_V (C*C + D*\nabla C*\nabla C - 2C_0*C)d\mathbf{r}. \tag{6}$$

It explicitly contains the initial condition, C_0. The time integration appears in the form of convolutions and inevitably leads to an iterative solution, time-step by time-step, because the convolution is a function of past values only. The finite element discretization of the functional, (6), is also presented in Appendix A. The resulting algorithm analytically reproduces a linear, time-invariant profile.

In conclusion, this example demonstrates that caution is necessary when finite element techniques are introduced into new fields. In particular, the existence of a system of equations for the unknown node parameters does not ensure the convergence of the approximation to the true solution.

In passing, it might also be pointed out that, for the same numbers of spatial elements and time steps, the finite element method is less accurate than the finite difference method for calculating impurity profiles over the ranges of concentrations of interest for predepositions [9].

Equilibrium carrier distributions in a doped semiconductor

The equilibrium condition to be considered is thermostatic equilibrium, defined as the absence of all carrier and heat fluxes through the semiconductor region under investigation. Such conditions are important in understanding the behavior of metal-insulator-semiconductor capacitors and in the investigation of subthreshold characteristics of insulated gate field effect transistors, to mention two examples.

In the interest of simplicity, nondegenerate carrier concentrations will be considered, so that the Boltzmann approximation is applicable to the carrier statistics. Then the electron density $n(\mathbf{r})$, the hole density $p(\mathbf{r})$, the net concentration of positive impurity ions, $N(\mathbf{r}) = N_D^+(\mathbf{r}) - N_A^-(\mathbf{r})$, and the electrostatic potential, $\phi(\mathbf{r})$, are related by the familiar equations

$$n = n_i \exp [e\phi/(kT)], \tag{7}$$

$$p = n_i \exp [-e\phi/(kT)], \tag{8}$$

$$\rho = p - n + N, \tag{9}$$

$$\nabla \cdot (\epsilon\nabla\phi) = -\rho = e(n - p - N), \tag{10}$$

where n_i is the intrinsic carrier concentration, a temperature dependent parameter of the material; e is the charge of the proton; and where N is fixed by the impurity profiles in the material. Again, the mathematical description is in differential form, while the application of the finite element method requires a variational formulation.

Since the variational principle corresponding to Poisson's equation, Eq. (10), is known as

$$W = \int \int \int_V \{\rho(\mathbf{r})\phi(\mathbf{r}) - \tfrac{1}{2}\epsilon(\mathbf{r})[\nabla\phi(\mathbf{r})]^2\}d\mathbf{r}, \tag{11}$$

it is very tempting to insert Eqs. (7) and (8) into (9) and then (9) into (11). This maneuver results in a functional of the form

$$W = \int \int \int_V \{\rho(\phi)\phi(\mathbf{r}) - \tfrac{1}{2}\epsilon(\mathbf{r})[\nabla\phi(\mathbf{r})]^2\}d\mathbf{r}. \tag{12}$$

Extremizing with respect to ϕ in turn generates an Euler-Lagrange equation of the form

$$\phi\partial\rho/\partial\phi + \rho(\phi) + \nabla \cdot (\epsilon\nabla\phi) = 0, \tag{13}$$

which deviates from Poisson's equation, (10), by the term $\phi(\partial\rho/\partial\phi)$, unless either ϕ vanishes or ρ is not a function of ϕ. Neither of these two conditions is applicable to the case of an arbitrarily doped semiconductor.

Conceptually, Eq. (12) attempts to describe a system of fixed charges (the impurity ions) and mobile charges (the carriers) in a conducting region. However, it is clearly not possible to maintain a nonzero charge distribution inside a conductor by means of an electrostatic field and to maintain zero current densities at the same time. If a nonzero charge distribution—with the automatically resultant electrostatic potential distribution—is to be maintained, forces other than electrostatic forces must also be present. However, the potentials associated with those other

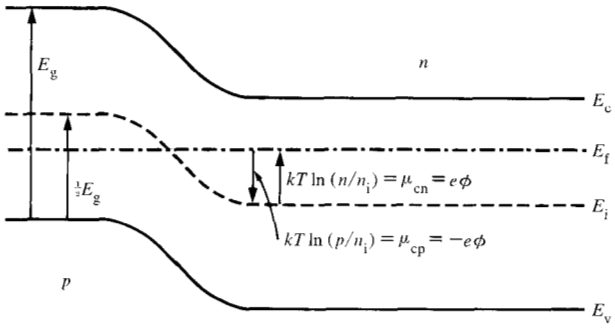

Figure 2 Energy band picture of a p-n junction in equilibrium.

forces are missing in Eqs. (12) and (13). These are the chemical potentials μ_{cp} and μ_{cn}, and their introduction highlights the fact that the problem under consideration is thermodynamic.

The definitions of the electrostatic and chemical potential energies are illustrated in Fig. 2, which shows the familiar energy band picture of a p-n junction in equilibrium. The formulation becomes simplest when the band center of the intrinsic semiconductor is chosen as the zero point of the energy scales.

If one considers that an electrically insulated semiconductor region can exchange heat, but cannot exchange carriers with its environment, one can conclude that its state of thermostatic equilibrium is that of minimum Helmholtz free energy. (The specimen and its environment will also be at a uniform temperature.) The integral (12) represents the electrostatic contribution to the free energy. The derivation of the chemical free energies by methods of statistical mechanics [10] is shown in Appendix B. The results are

$$f_{cn} = n(\mu_{cn} - kT), \quad \text{and} \tag{14}$$

$$f_{cp} = p(\mu_{cp} - kT). \tag{15}$$

Adding these contributions to the integrand of Eq. (12), considering the definitions of Fig. 2 and Eqs. (7) and (8), leads to the free energy density

$$\begin{aligned} f = {} & e\phi\{n_i \exp[\mu_{cp}/(kT)] - n_i \exp[\mu_{cn}/(kT)] + N\} \\ & - \tfrac{1}{2}\epsilon(\nabla\phi)^2 + n_i(\mu_{cp} - kT) \exp[\mu_{cp}/(kT)] \\ & + n_i(\mu_{cn} - kT) \exp[\mu_{cn}/(kT)]. \end{aligned} \tag{16}$$

Extremizing the resulting functional with respect to μ_{cn} and μ_{cp} yields the equations

$$n_i \exp[\mu_{cp}/(kT)][e\phi/(kT) + \mu_{cp}/(kT) + 1 - 1] = 0, \quad \text{and} \tag{17}$$

$$n_i \exp[\mu_{cn}/(kT)][-e\phi/(kT) + \mu_{cn}/(kT) + 1 - 1] = 0, \tag{18}$$

from which one confirms at once the equilibrium conditions

$$e\phi = \mu_{cn} = -\mu_{cp}. \tag{19}$$

Extremizing with respect to ϕ results in

$$e\{n_i \exp[\mu_{cp}/(kT)] - n_i \exp[\mu_{cn}/(kT)] + N\} + \nabla \cdot (\epsilon\nabla\phi) = 0, \tag{20}$$

and, after insertion of the conditions, (19),

$$e\{n_i \exp[-e\phi/(kT)] - n_i \exp[e\phi/(kT)] + N\} + \nabla \cdot (\epsilon\nabla\phi) = 0, \tag{21}$$

which is, indeed, Poisson's equation for doped semiconductor material.

Finally, one can insert the result, (19), into the expression for the energy density, (16), and realize a simplification of the latter equation to

$$\begin{aligned} f = {} & e\phi N - \tfrac{1}{2}\epsilon(\nabla\phi)^2 \\ & - n_i kT\{\exp[-e\phi/(kT)] + \exp[e\phi/(kT)]\} \\ = {} & e\phi N - \tfrac{1}{2}\epsilon(\nabla\phi)^2 - nkT - pkT. \end{aligned} \tag{22}$$

Equation (22) shows that all the free energy is related to the electrostatic potential and to the fixed ions, while the free carriers possess no free energy; their kinetic energies subtract from the total energy.

Carrier transport in semiconductors

The essence of modeling the electrical behavior of semiconductor devices from their physical structure lies in the determination of the carrier fluxes due to externally applied potentials. The following considerations will be restricted to stationary operating conditions. In addition, the temperature will be assumed constant throughout, an assumption usually made in semiconductor device modeling, although the theory would allow the inclusion of thermal effects.

The relationships between the forces acting on the carriers and the resulting electric current densities are given by the transport equations

$$\mathbf{j}_p = -e(D_p\nabla p + pm_p\nabla\phi), \quad \text{and} \tag{23}$$

$$\mathbf{j}_n = e(D_n\nabla n - nm_n\nabla\phi), \tag{24}$$

D_p and D_n representing the carrier diffusivities, and m_p and m_n denoting the carrier mobilities. Under the assumption of nondegenerate carrier concentrations, these quantities are related by the Einstein relations

$$D_p/m_p = D_n/m_n = kT/e. \tag{25}$$

Using these relations to express the diffusivities in terms of the mobilities and then factoring out the mobilities and carrier densities changes the form of Eqs. (23) and (24) to

$$\mathbf{j}_p = -pm_p\nabla[(kT)\ln(p/n_i) + e\phi], \quad \text{and} \tag{26}$$

$$\mathbf{j}_n = nm_n\nabla[(kT)\ln(n/n_i) - e\phi]. \tag{27}$$

The terms in the brackets are recognized as the chemical and electrostatic potential energies, which combine to the electrochemical potentials

$$\mu_p - (kT)\ln(p/n_i) + e\phi = \mu_{cp} + e\phi = E_{fp}, \text{ and} \tag{28}$$

$$\mu_n = (kT)\ln(n/n_i) - e\phi = \mu_{cn} - e\phi = E_{fn}. \tag{29}$$

In semiconductor device engineering these quantities are better known as Imrefs or quasi Fermi levels.

Equations (26) and (27) are in the form of Ohm's Law. Associating, for instance, the electron charge with the first two factors on the right of Eq. (26) leads to the conductivity $\sigma = pm_p/e$, which is a transport coefficient and is a function of the chemical potential through the hole density p. The remaining quantity is the electromotive force, $-(1/e)\cdot\nabla[(kT)\ln(p/n_i) + e\phi]$, which is the negative of the gradient of the electrochemical potential.

The carrier fluxes satisfy the conservation equations

$$\nabla\cdot\mathbf{j}_p = e(g_p - \partial p/\partial t), \quad \text{and} \tag{30}$$

$$\nabla\cdot\mathbf{j}_n = -e(g_n - \partial n/\partial t), \tag{31}$$

g_p and g_n denoting hole and electron generation rates. Stationary conditions are defined by the time invariance of the distributions of the potentials, carrier densities and ionization levels of the impurities; hence

$$\partial p/\partial t = \partial n/\partial t = 0. \tag{32}$$

The stationary ionization levels prevent transitions into and out of impurity levels from contributing to the carrier generation rates, which are thus governed by pair generation or annihilation only. This leads to

$$g_p = g_n. \tag{33}$$

The theory of irreversible thermodynamics postulates the minimization of the entropy generation rate as the variational principle that governs carrier transport. This theory, however, is linearized and only applies as long as the conductivities and carrier generation rates can be approximated by the first term in a Taylor expansion about their values at thermostatic equilibrium. For the carrier concentrations (contained in the conductivities) this leads to the approximations

$$p = p_0\exp[(\mu_{cp} - \mu_{cp0})/(kT)] \approx p_0, \quad \text{and} \tag{34}$$

$$n = n_0\exp[(\mu_{cn} - \mu_{cn0})/(kT)] \approx n_0, \tag{35}$$

the index 0 denoting quantities at thermostatic equilibrium.

For weak nonequilibrium states, the generation rates are described well by the expression

$$g_p = g_n = ak(n_i^2 - np)/n_i^2 = ak\{1 - \exp[(\mu_{cp} + \mu_{cn})/(kT)]\}, \tag{36}$$

where the factor a is a material constant; and where the identity $n_0p_0 = n_i^2$ has been used. The first-order approximation amounts to

$$g_p = g_n \approx -(a/T)(\mu_{cp} + \mu_{cn}). \tag{37}$$

The entropy generation rate per unit volume is composed of products of generalized flux densities and associated driving forces, all divided by the absolute temperature [11]. The flux densities, in turn, are products of the transport coefficients with the corresponding driving forces, so that the density of the entropy generation rate becomes a quadratic function of the driving forces. The transport coefficient in Eq. (37) is a/T, and the driving force is $-(\mu_{cp} + \mu_{cn})$.

In view of Eqs. (26), (27), (34), (35), and (37), the entropy generation rate assumes the form

$$g_s = (1/T)\int\int\int_V \{(p_0m_p/e)[\nabla(\mu_{cp} + e\phi)]^2 + (n_0m_n/e) \times [\nabla(\mu_{cn} - e\phi)]^2 + (a/T)(\mu_{cp} + \mu_{cn})^2\}d\mathbf{r}. \tag{38}$$

Minimizing with respect to ϕ, μ_{cp} and μ_{cn} results in the equations

$$0 = \nabla\cdot[2e(p_0m_p/e)\nabla(\mu_{cp} + e\phi) - 2e(n_0m_n/e)\nabla(\mu_{cn} - e\phi)] = 2\nabla\cdot(\mathbf{j}_p + \mathbf{j}_n) = 2\nabla\cdot\mathbf{j}, \tag{39}$$

$$0 = 2(a/T)(\mu_{cp} + \mu_{cn}) - \nabla\cdot[2(p_0m_p/e)\nabla(\mu_{cp} + e\phi)] = 2(eg_p - \nabla\cdot\mathbf{j}_p), \tag{40}$$

$$0 = 2(a/T)(\mu_{cp} + \mu_{cn}) - \nabla\cdot[2(n_0m_n/e)\nabla(\mu_{cn} - e\phi)] = 2(eg_n - \nabla\cdot\mathbf{j}_n). \tag{41}$$

These are the continuity equations for the total current, for the hole current and for the electron current. It is worth noting that the conservation laws are more basic than the transport equations, (23) and (24).

The restrictions introduced with Eqs. (34), (35), and (37) decisively limit the range of validity of the variational principle. It only applies to operating conditions, where the chemical potentials deviate by substantially less than kT from thermostatic equilibrium conditions. This is tantamount to operating voltages of the devices of less than 25 mV, which excludes almost all cases of practical interest.

Irreversible thermodynamics lacks the basis for the formal extension of the variational principle to strong nonequilibrium, although the concept of minimizing the entropy generation rate itself would not be limited to weak

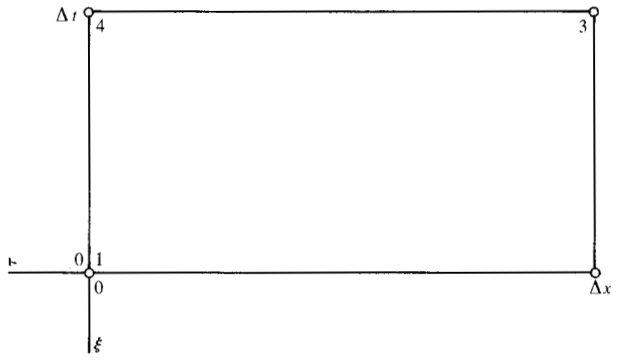

Figure A1 Finite element with local variables and parameters.

nonequilibrium. Also, diodes, transistors and other semiconductor elements respond stably and consistently over wide ranges of stationary operating conditions; this demonstrates restoring mechanisms governed by a variational principle. Its formulation, however, will at least require a formally self-consistent derivation of the carrier mobilities and diffusivities and the generation rate coefficients from, e.g., a collision model. Until such results become available, the efficient application of the finite element method to carrier transport problems in semiconductor devices will remain in doubt.

In present applications the finite element method is not used to directly solve carrier transport problems in the way suggested in this section. Instead, Poisson's equation and the carrier continuity equations are solved in sequence, in an iterative loop, until self-consistent carrier and flux distributions are achieved. In this scheme the finite element method is used to solve the individual partial differential equations by using a method of residuals, e.g., Galerkin's method.

Summary

Three variational principles related to semiconductor device modeling by the finite element method have been discussed in some detail in order to highlight some subtle points which are crucial to the successful application of the method. The derivation and evaluation of finite element algorithms for the simulation of transient diffusion processes demonstrate that the mere existence of a set of equations for calculating unknown node parameters does not ensure the correctness of the algorithm. Rather, a proof of the convergence of the approximation to the true solution must be verified, unless the variational principle is known to be the correct one beyond any doubt. The development of a variational formulation of the thermostatic equilibrium condition in a semiconductor region points to the danger in extending a known variational principle by the mere substitution of some of the variables. The last example shows the limitations inherent in today's variational description of transport processes in semiconductors.

The success of finite element calculations in semiconductor device modeling depends strongly on the care with which the validity of underlying variational principles is ensured. The best guidance for this task is provided by the theories of physics, although contemporary formulations do not appear favorable to the rapid and broadbased introduction of finite element calculations into semiconductor device modeling schemes. Growing interest in the finite element method will surely stimulate further progress in semiconductor theories.

Acknowledgments

Discussions with D. J. Hamilton, H. A. Kamel and R. A. Young of the University of Arizona, with N. D. Lubart of IBM Austin, Texas and with A. V. Ferris-Prabhu of IBM Burlington, Vermont contributed to the development and clarification of the ideas presented in this contribution.

Appendix A: Finite element discretization for diffusion problem

Using local variables and node numbers of a rectangular element, as shown in Fig. A1, and defining the linear interpolation functions as

$$p = (\Delta x - \xi)/\Delta x, \qquad q = \xi/\Delta x, \quad \text{and} \tag{A1}$$

$$r = (\Delta t - \tau)/\Delta t, \qquad s = \tau/\Delta t, \tag{A2}$$

one obtains the local approximation of the concentration

$$C(\xi, \tau) = (C_1 r + C_4 s)p + (C_2 r + C_3 s)q, \tag{A3}$$

and similarly $\bar{C}(\xi, \tau)$. The integrand, (4), of the second section for the element is composed of products of the node parameters and interpolation functions. The integrals operate only on the interpolation functions, leading to

$$\int_{\Delta x} p^2 d\xi = \int_{\Delta x} q^2 d\xi = 2 \int_{\Delta x} pq d\xi = \Delta x/3, \text{ and} \tag{A4}$$

$$\int_{\Delta t} d\tau = 2 \int_{\Delta t} r d\tau = 2 \int_{\Delta t} s d\tau = 3 \int_{\Delta t} r^2 d\tau$$

$$= 3 \int_{\Delta t} s^2 d\tau = 6 \int_{\Delta t} rs d\tau = \Delta t. \tag{A5}$$

After straightforward algebraic manipulation, one eventually obtains the partial derivatives with respect to the node parameters of the elemental contribution Ω_e to the functional:

$$\frac{\partial \Omega_e}{\partial \bar{C}_1} = \frac{\Delta t}{6\Delta x}\left[2C_1 - 2C_2 - C_3 + C_4 + \tfrac{1}{2}\frac{\Delta x^2}{D\Delta t}(C_3 + 2C_4)\right], \tag{A6}$$

$$\frac{\partial \Omega_e}{\partial \bar{C}_2} = \frac{\Delta t}{6\Delta x}\left[-2C_1 + 2C_2 + C_3 - C_4 + \tfrac{1}{2}\frac{\Delta x^2}{D\Delta t}(2C_3 + C_4)\right], \tag{A7}$$

$$\frac{\partial \Omega_e}{\partial \bar{C}_3} = \frac{\Delta t}{6\Delta x}\left[-C_1 + C_2 + 2C_3 - 2C_4 - \tfrac{1}{2}\frac{\Delta x^2}{D\Delta t}(C_1 + 2C_2)\right], \tag{A8}$$

$$\frac{\partial \Omega_e}{\partial \bar{C}_4} = \frac{\Delta t}{6\Delta x}\left[C_1 - C_2 - 2C_3 + 2C_4 - \tfrac{1}{2}\frac{\Delta x^2}{D\Delta t}(2C_1 + C_2)\right]. \tag{A9}$$

Letting $\frac{1}{2}\Delta x^2/(D\Delta t) = d$ and translating to the global system of Fig. 1, one obtains, e.g., by extremizing with respect to the node parameters C_7 and C_{17},

$$\begin{aligned} 0 = {} & -C_1 + 2C_2 - C_3 - 4C_6 + 8C_7 - 4C_8 - C_{11} \\ & + 2C_{12} - C_{13} - d(C_1 + 4C_2 + C_3 - C_{11} \\ & - 4C_{12} - C_{13}), \quad \text{and} \end{aligned} \tag{A10}$$

$$\begin{aligned} 0 = {} & -C_{11} + 2C_{12} - C_{13} - 2C_{16} + 4C_{17} \\ & - 2C_{18} - d(C_{11} + 4C_{12} + C_{13}). \end{aligned} \tag{A11}$$

The equations for the unknown parameters C_7, C_8, C_9, C_{12}, C_{13}, C_{14}, C_{17}, C_{18}, and C_{19} are now given by

$$\begin{bmatrix} 8 & -4 & 0 & (2+4d) & (d-1) & 0 & 0 & 0 \\ -4 & 8 & -4 & (d-1) & (2+4d) & (d-1) & 0 & 0 \\ 0 & -4 & 8 & 0 & 0 & (2+4d) & 0 & 0 \\ (2-4d) & (-1-d) & 0 & 8 & -4 & 0 & (2+4d) & 0 \\ (-1-d) & (2-4d) & (-1-d) & -4 & 8 & -4 & (d-1) & (d-1) \\ 0 & (-1-d) & (2-4d) & 0 & 0 & 8 & 0 & (2+4d) \\ 0 & 0 & 0 & (2-4d) & (-1-d) & 0 & 4 & 0 \\ 0 & 0 & 0 & (-1-d) & (2-4d) & (-1-d) & -2 & -2 \\ 0 & 0 & 0 & 0 & (-1-d) & (2-4d) & 0 & 4 \end{bmatrix} \begin{bmatrix} C_7 \\ C_8 \\ C_9 \\ C_{12} \\ C_{13} \\ C_{14} \\ C_{17} \\ C_{18} \\ C_{19} \end{bmatrix}$$

$$= \begin{bmatrix} (1+d) & (4d-2) & (1+d) & 0 & 0 & 4 & 0 & (1-d) & 0 & 0 & 0 \\ 0 & (1+d) & (4d-2) & (1+d) & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & (1+d) & (4d-2) & (1+d) & 0 & 4 & 0 & (1-d) & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & (1+d) & 0 & 4 & 0 & (1-d) & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & (1+d) & 0 & 4 & 0 & (1-d) \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & (1+d) & 0 & 4 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & (1+d) & 0 & 4 \end{bmatrix} \begin{bmatrix} C_1 \\ C_2 \\ C_3 \\ C_4 \\ C_5 \\ C_6 \\ C_{10} \\ C_{11} \\ C_{15} \\ C_{16} \\ C_{20} \end{bmatrix} \tag{A12}$$

The validity test outlined in the second section should change these equations into identities, independent of the value of d, because the choice of the factors Δx and Δt should be free. The node parameter values suggested in the second section change the first six equations into identities, with d indeterminate, while the last three equations lead to the inconsistencies $d = -6$, $d = 1/11$ and $d = -5/12$. This demonstrates the failure of the system of equations, (A12), to model the diffusion process in question.

The integrand (6) in the second section decomposes into products of the node parameters and interpolation functions p and q, and convolutions of the interpolation functions r and s. By noting that

$$r(\Delta t - \tau) = s(\tau), \qquad s(\Delta t - \tau) = r(\tau), \tag{A13}$$

the convolutions are recognized as time integrations over appropriate products of these functions; e.g.,

$$\begin{aligned} r*s &= \int_0^{\Delta t} [r(\Delta t - \tau)s(\tau)d\tau \\ &= \int_0^{\Delta t} s^2(\tau)d\tau; \qquad 1*r = \int_0^{\Delta t} r d\tau. \end{aligned} \tag{A14}$$

The convolutions appearing in the integrand reduce to the integrals

$$\int_{\Delta t} d\tau = 2\int_{\Delta t} r d\tau = 2\int_{\Delta t} s d\tau = 3\int_{\Delta t} r^2 d\tau = 3\int_{\Delta t} s^2 d\tau$$
$$= 4\int_{\Delta t} r^3 d\tau = 4\int_{\Delta t} s^3 d\tau = 6\int_{\Delta t} rs d\tau$$
$$= 12\int_{\Delta t} r^2 s d\tau = 12\int_{\Delta t} rs^2 d\tau = \Delta t, \quad \text{(A15)}$$

while the integrals over Δx lead to the same quantities as before.

The partial derivatives of the elemental contributions to the functional Ω_e, with respect to the node parameters, become

$$\frac{\partial \Omega_e}{\partial C_1} = \frac{\Delta t \Delta x}{18}[-10C_1 - 5C_2 - C_3 - 2C_4] + \frac{D\Delta t^2}{12\Delta x}[5C_1 - 5C_2 - 3C_3 + 3C_4], \quad \text{(A16)}$$

$$\frac{\partial \Omega_e}{\partial C_2} = \frac{\Delta t \Delta x}{18}[-5C_1 - 10C_2 - 2C_3 - C_4] - \frac{D\Delta t^2}{12\Delta x}[5C_1 - 5C_2 - 3C_3 + 3C_4], \quad \text{(A17)}$$

$$\frac{\partial \Omega_e}{\partial C_3} = \frac{\Delta t \Delta x}{18}[-C_1 - 2C_2 + 2C_3 + C_4] - \frac{D\Delta t^2}{12\Delta x}[3C_1 - 3C_2 - C_3 + C_4], \quad \text{(A18)}$$

and

$$\frac{\partial \Omega_e}{\partial C_4} = \frac{\Delta t \Delta x}{18}[-2C_1 - C_2 + C_3 + 2C_4] + \frac{D\Delta t^2}{12\Delta x}[3C_1 - 3C_2 - C_3 + C_4]. \quad \text{(A19)}$$

Since the convolution integral is a function of past values of the variables only, the node parameters C_7 to C_9 only depend on the conditions at the nodes 1 to 6 and 10, and the equations are derived only from the bottom row of elements in Fig. 1. Letting $3D\Delta t/(2\Delta x^2) = d$ and translating to the global system (Fig. 1), one obtains, e.g., by extremizing with respect to the node parameter C_7,

$$C_1 + 4C_2 + C_3 - C_6 - 4C_7 - C_8 - d(-3C_1 + 6C_2 - 3C_3 - C_6 + 2C_7 - C_8) = 0. \quad \text{(A20)}$$

This leads to the equations for the unknown parameters C_7, C_8, and C_9,

$$\begin{bmatrix} (4+2d) & (1-d) & 0 \\ (1-d) & (4+2d) & (1-d) \\ 1 & (1-d) & (4+2d) \end{bmatrix} \begin{bmatrix} C_7 \\ C_8 \\ C_9 \end{bmatrix} = \begin{bmatrix} (1+3d) & (4+6d) & (1+3d) & 0 & 0 & (d-1) & 0 \\ 0 & (1+3d) & (4+6d) & (1+3d) & 0 & 0 & 0 \\ 0 & 0 & (1+3d) & (4+6d) & (1+3d) & 0 & (d-1) \end{bmatrix} \begin{bmatrix} C_1 \\ C_2 \\ C_3 \\ C_4 \\ C_5 \\ C_6 \\ C_{10} \end{bmatrix} \quad \text{(A21)}$$

A spatially linear and time-invariant solution changes this system of equations into an exact identity, with the factor d indeterminate. This demonstrates convergence of the algorithm toward the solutions of the diffusion equation.

Appendix B: Derivation of the free energy of carriers in a semiconductor

In deriving thermodynamic potentials of ensembles of particles, one must be careful to use the correct statistics and the correct boundary conditions. Equations (7) and (8),

$$p = n_i \exp[\mu_{cp}/(kT)] \quad \text{and} \quad \text{(B1)}$$

$$n = n_i \exp[\mu_{cn}/(kT)], \quad \text{(B2)}$$

imply the Boltzmann approximation to the carrier distributions, which means that the carriers are considered as a quasi-classical gas, maintaining indistinguishability but disregarding the Pauli exclusion principle. The boundary conditions of a volume element of the semiconductor allow for the free exchange of energy and of carriers, which means that statistics of the grand canonical ensemble must be applied.

The Helmholtz free energy of such an ensemble of electrons is given by [12]

$$f_{cn} = (\bar{n}kT)\ln(z) - kT\ln[Q_g(z, v, T)], \quad \text{(B3)}$$

where Q_g is the grand partition function, v is the volume of the element, and the bar denotes the expected value of the randomly fluctuating density n. The quantity z is the fugacity, given by

$$z = \exp[\mu_{cn}/(kT)] = \bar{n}/n_i. \quad \text{(B4)}$$

Since $\bar{n}$ can be derived from Q_g by

$$\bar{n} = z\partial\{\ln[Q_g(z, v, T)]\}/\partial z, \quad \text{(B5)}$$

knowledge of Q_g solves the problem.

The grand partition function Q_g, the canonical partition function Q_c, and the single particle partition function Q are related by

$$Q_g(z, v, T) = \sum_{n=0}^{\infty} z^n Q_{cn}(v, T) \quad \text{and} \tag{B6}$$

$$Q_{cn}(v, T) = Q^n(v, T)/(n!), \tag{B7}$$

and, for completeness, Q is given by

$$Q = v\left(\sqrt{2\pi mkT}\,\Big/ h\right)^3 = v/\lambda_{Th}^3, \tag{B8}$$

λ_{Th} being the de Broglie wavelength of a particle with thermal velocity.

The grand partition function is obtained as

$$Q_g(z, v, T) = \sum_{0}^{\infty} z^n Q^n/(n!) = \exp(zQ), \tag{B9}$$

which, in view of Eq. (B5), leads to

$$\bar{n} = zQ = [v/\lambda_{Th}^3] \exp[\mu_{cn}/(kT)]. \tag{B10}$$

Comparison with Eq. (B4) further yields

$$Q = n_i. \tag{B11}$$

Finally, inserting Eqs. (B9) and (B4) into (B3) yields

$$f_{cn} = \bar{n}\mu_{cn} - \bar{n}kT, \tag{B12}$$

and by implication

$$f_{cp} = \bar{p}\mu_{cp} - \bar{p}kT. \tag{B13}$$

Recognizing that carrier densities are specified as expected values, we can omit the bar to obtain Eqs. (14) and (15).

References

1. E. M. Buturla and P. E. Cottrell, "Two Dimensional Finite Element Analysis of Semiconductor Steady State Transport Equations," *International Conference on Computer Methods in Nonlinear Mechanics*, Austin, TX, 1974.
2. G. Hachtel, R. O'Brien, and M. Mack, "Semiconductor Device Analysis via Finite Elements," *Eighth Asilomar Conference on Circuits and Systems*, Pacific, Grove, CA, 1974.
3. J. J. Barnes, R. J. Lomax, and G. I. Haddad, "Finite Element Simulation of GaAs MESFET's with Lateral Doping Profiles and Submicron Gates," *IEEE Trans. Electron Devices* **ED-23,** 455 (1976).
4. O. C. Zienkiewicz, *The Finite Element Method in Engineering Science*, McGraw-Hill Book Co., Inc., London, 1971.
5. P. M. Morse and H. Feshbach, *Methods of Theoretical Physics*, Part I, McGraw-Hill Book Co., Inc., New York, 1953.
6. J. T. Oden, *Finite Elements of Nonlinear Continua*, McGraw-Hill Book Co., Inc., New York, 1972.
7. G. Strang and G. J. Fix, *An Analysis of the Finite Element Method*, Prentice-Hall, Inc., Englewood Cliffs, NJ, 1973.
8. M. E. Gurtin, "Variational Principles for Linear Initial Value Problems," *Quart. Appl. Math.* **72,** 252 (1964).
9. J. H. Hohl and D. J. Hamilton, "Errors in Calculations of Predeposition Diffusion Profiles by Iterative Numerical Methods," *J. Electrochem. Soc.* **124,** 1912 (1977).
10. J. H. Hohl, "Techniques for Modeling Multidimensional Effects in Integrated Devices," Ph.D. Thesis, University of Arizona, Tucson, AZ, 1976.
11. A. C. Smith, J. F. Janak, and R. B. Adler, *Electronic Conduction in Solids*, McGraw-Hill Book Co., Inc., New York, 1967.
12. K. Huang, *Statistical Mechanics*, John Wiley & Sons, Inc., New York, 1967.

Received April 14, 1977; revised August 8, 1977

The author is with the IBM General Technology Division, Essex Junction, Vermont 05452.