MATHEMATICS OF COMPUTATION
VOLUME 51, NUMBER 184
OCTOBER 1988, PAGES 431-449

Inverse-Average-Type Finite Element Discretizations of Selfadjoint Second-Order Elliptic Problems

By Peter A. Markowich and Miloš A. Zlámal

Abstract. This paper is concerned with the analysis of a class of "special purpose" piecewise linear finite element discretizations of selfadjoint second-order elliptic boundary value problems. The discretization differs from standard finite element methods by inverse-average-type approximations (along element sides) of the coefficient function $a(x)$ in the operator $-\operatorname{div}(a(x)\operatorname{grad} u)$. The derivation of the discretization is based on approximating the flux density $J = a \operatorname{grad} u$ by constants on each element. In many cases the flux density is well behaved (moderately varying) even if $a(x)$ and $u(x)$ are fast varying.

Discretization methods of this type have been used successfully in semiconductor device simulation for many years; however, except in the one-dimensional case, the mathematical understanding of these methods was rather limited.

We analyze the stiffness matrix and prove that—under a rather mild restriction on the mesh—it is a diagonally dominant Stieltjes matrix. Most importantly, we derive an estimate which asserts that the piecewise linear interpolant of the solution u is approximated to order 1 by the finite element solution in the H^1-norm. The estimate depends only on the mesh width and on derivatives of the flux density and of a possibly occurring inhomogeneity.

1. Introduction. We present and analyze a "special purpose" finite element discretization of two-dimensional scalar second-order elliptic boundary value problems of the form

$$\operatorname{div}(a(x)\operatorname{grad} u) = f(x), \qquad x = (x_1, x_2) \in \Omega \subseteq \mathbf{R}^2, \tag{1.1}$$

$$u|_{\partial\Omega_D} = u_D|_{\partial\Omega_D}, \qquad \frac{\partial u}{\partial \nu}\Big|_{\partial\Omega_N} = 0. \tag{1.2}$$

The scalar function a is bounded away from zero and positive. f denotes an inhomogeneity. Ω is a bounded domain in $\mathbf{R}^2$ whose boundary splits into a Dirichlet part $\partial\Omega_D$ and a Neumann part $\partial\Omega_N$. ν denotes the exterior unit normal vector of $\partial\Omega$. A homogeneous Neumann boundary condition is prescribed on $\partial\Omega_N$ and a possibly inhomogeneous Dirichlet condition on $\partial\Omega_D$.

In many practical applications the coefficient function a and derivatives of the solution u (or even u itself) vary extremely rapidly in certain subdomains of Ω while the flux density J, which is defined by

$$J := a \operatorname{grad} u, \tag{1.3}$$

is well behaved, i.e., moderately varying in Ω (note that $J \equiv \text{const}$ holds in the one-dimensional homogeneous case $f \equiv 0$!). For problems of this type it normally cannot be guaranteed that the variations of the quantities a and u are resolved

Received May 27, 1986; revised October 27, 1987.
1980 *Mathematics Subject Classification* (1985 *Revision*). Primary 65N30, 35J25.

accurately by a computationally feasible mesh on which a discretization of (1.1) is performed. In this situation the solution u is not approximated well by functions which are piecewise polynomials on the given mesh. Thus, standard finite element or finite difference discretizations neither yield practically relevant error estimates nor give useful numerical results.

These difficulties were encountered in the area of semiconductor device modeling in the sixties, since the electron and hole continuity equations of the Van Roosbroeck model (see Van Roosbroeck [14]) are (after an appropriate transformation) of the type described above (see Markowich [7], Selberherr [12]). Scharfetter and Gummel [11] published a discretization of the one-dimensional model equations, which is based on approximating the flux density J by a constant on each mesh interval. This method differs from standard finite difference schemes by an inverse-average-type approximation of the coefficient function a.

Since then, various multidimensional finite difference, finite element and box-scheme analogues have been employed extensively and successfully in device simulations, even on rather coarse grids (see Selberherr [12] for a collection of references). However, except in the one-dimensional case, the mathematical understanding of these methods was extremely limited. Only recently, Mock [8], [9] published an analysis of Scharfetter-Gummel type box-schemes, which shows that ony the variations of the flux density J and the inhomogeneity f have to be resolved accurately by the mesh in order to obtain reasonable discretization errors.

Zlámal [18] derived finite element discretizations of the Van Roosbroeck semiconductor device model in which he generalizes the approach of Scharfetter and Gummel to two and three space dimensions. The flux density J is approximated by a constant vector on each finite element (triangles or quadrilaterals in two dimensions and tetrahedra or hexahedra in three dimensions with piecewise linear or, resp., piecewise bilinear basis and test functions using the isoparametric technique). Analogously to the one-dimensional case, the coefficient function a is approximated by inverse averages over edges of the finite elements.

We remark that Babuška and Osborn [2], [3] derive finite element methods for elliptic problems with strongly varying coefficients which have a similar flavor (their methods also involve inverse averages of a coefficient function). They use different trial and test spaces to set up the method and their convergence proofs are based on variational approximation. In this paper we generalize Zlámal's approach to two-dimensional problems of the form (1.1), (1.2) and derive an inverse-average-approximation-type finite element discretization using piecewise linear test and basis functions on a triangular mesh. We prove (under a rather mild assumption on the mesh) that the corresponding stiffness matrix has the same qualitative properties as the stiffness matrix of Galerkin's method, i.e., it is a diagonally dominant Stieltjes matrix.

The main thrust of the analysis is towards the convergence performance of the discretization. We prove (again, under a mild geometric assumption on the mesh) convergence of order 1 in the H^1-norm and, most importantly, that the H^1-norm of the difference of the finite element solution and the piecewise linear interpolant

of the exact solution is

$$O\left(h\left\|\left|\frac{\partial J}{\partial x}\right|\right\|_{L^\infty(\Omega)} + h^2\|f\|_{H^2(\Omega)}\right),$$

where h denotes the maximal length of the sides of the triangles of the partition. Thus, under the assumptions made, the error is independent of derivatives of "fast" quantities, it only depends on derivatives of the flux density and of the inhomogeneity f on the given mesh (the error term involving f stems from numerical integration !). An analogous error estimate is obtained for the flux density. These results explain mathematically why inverse-average-type finite element discretizations of (1.1), (1.2) are appropriate in the sense that simulations on rather coarse meshes give reasonably accurate results if only the flux density and the inhomogeneity are well behaved.

We remark that the methods employed for the proofs carry over to three-dimensional problems of the form (1.1), (1.2), only the required calculations are more involved.

The paper is organized as follows. In Section 2 we state assumptions and introduce notations, in Section 3 we derive the finite element scheme, Section 4 is concerned with the properties of the stiffness matrix and Section 5 with the convergence analysis. The application to the semiconductor device equations is discussed in Section 6.

2. Preliminaries. We shall employ the following assumptions on the boundary value problem (1.1), (1.2):

$$0 < \underline{a} \le a(x) \le \bar{a} < \infty, \qquad x \in \Omega;\ a \in C(\overline{\Omega}). \tag{2.1}$$

$$f \in H^2(\Omega); \quad u_D \in H^1(\Omega), \quad u_D|_{\partial\Omega_D} \in C(\partial\Omega_D). \tag{2.2}$$

$\Omega \subseteq \mathbf{R}^2$ is a bounded domain with a polygonal boundary $\partial\Omega = \partial\Omega_D \cup \partial\Omega_N$, $\partial\Omega_D \cap \partial\Omega_N = \varnothing$. $\partial\Omega_D$ is closed and $\partial\Omega_N$ is open with respect to $\partial\Omega$. $\partial\Omega_D$ is the union of a finite number of line segments and the arclength of $\partial\Omega_D$ is positive. (2.3)

The problem (1.1), (1.2) has a (unique) solution $u \in H^1(\Omega) \cap C(\overline{\Omega})$ for which $J = a \operatorname{grad} u \in C^{0,1}(\overline{\Omega})$. (2.4)

In the case of a mixed Neumann-Dirichlet boundary value problem ($\partial\Omega_N \neq \varnothing$) (2.4) represents a condition on the angles at which $\partial\Omega_D$ and $\partial\Omega_N$ meet (see, e.g., Grisvard [5], Kawohl [6]).

We cover $\overline{\Omega}$ by a triangular mesh $\Delta = \{T_1, \ldots, T_N\}$ which is such that all sides which have a nonempty intersection with $\partial\Omega_D$ have an empty intersection with $\partial\Omega_N$. We denote by h_T the maximal length of the three sides of the triangle T and define

$$h := \max_{T \in \Delta} h_T. \tag{2.5}$$

We denote by M the number of vertices $P \in \Omega \cup \partial\Omega_N$ and by K the number of vertices $P \in \overline{\Omega}$. X_Δ shall denote the space of all functions ϕ_Δ, which are continuous in $\overline{\Omega}$ and linear on each of the finite elements T_i and $X_{\Delta,0}$ the subspace of X_Δ

consisting of all continuos piecewise linear functions which vanish on $\partial\Omega_D$. By $\phi_\Delta^{(i)} \in X_\Delta$ we denote the basis function, which assumes the value 1 at the vertex P_i and 0 at all other vertices. Obviously, $\dim X_\Delta = K$ and $\dim X_{\Delta,0} = M$. We denote by v_I the piecewise linear interpolant of the function $v \in C(\overline{\Omega})$, i.e.,

$$v_I = \sum_{P_i \in \overline{\Omega}} v(P_i)\phi_\Delta^{(i)}. \tag{2.6}$$

Most of the calculations performed in the sequel are done "elementwise". They are expedited by transforming the corresponding finite element T (in the (x_1, x_2)-plane) into the reference triangle $\hat{T}$ with the vertices $(0,0)$, $(1,0)$, $(0,1)$ (in the (ξ_1, ξ_2)-plane). A possible transformation reads:

$$x(\xi) = P_1 M_1(\xi) + P_2 M_2(\xi) + P_3 M_3(\xi), \tag{2.7}$$

where P_1, P_2, P_3 are the coordinate vectors of the vertices of T and M_1, M_2, M_3 are the shape functions

$$M_1(\xi) = 1 - \xi_1 - \xi_2, \quad M_2(\xi) = \xi_1, \quad M_3(\xi) = \xi_2. \tag{2.8}$$

By (2.7), P_1 is mapped into $(0,0)$, P_2 into $(1,0)$ and P_3 into $(0,1)$.

We denote the Jacobian $\frac{\partial x}{\partial \xi}$ by I_T (and its transpose by I_T'). x_{G_T} denotes the center of gravity of the triangle T. We write $x \cdot y$ for the scalar product of the two n-dimensional vectors $x, y \in \mathbf{R}^n$ and $|x|$ for the Euclidean norm of x. The corresponding matrix norm is denoted by the same symbol. The standard notation for Sobolev spaces and associated norms is employed (see, e.g., Adams [1]). $H^m(\Omega)$ stands for the space of real (or vector-valued) functions, whose weak derivates of order up to m are in $L^2(\Omega)$, and

$$\|v\|_{H^m(\Omega)} := \left(\sum_{0 \le \alpha_1 + \alpha_2 \le m} \left\| \left| \frac{\partial^{\alpha_1+\alpha_2} v}{\partial x_1^{\alpha_1} \partial x_2^{\alpha_2}} \right| \right\|_{L^2(\Omega)}^2 \right)^{1/2}, \tag{2.9}$$

$$|v|_{H^m(\Omega)} := \left(\sum_{\alpha_1 + \alpha_2 = m} \left\| \left| \frac{\partial^{\alpha_1+\alpha_2} v}{\partial x_1^{\alpha_1} \partial x_2^{\alpha_2}} \right| \right\|_{L^2(\Omega)}^2 \right)^{1/2}. \tag{2.10}$$

$H_0^1(\Omega \cup \partial\Omega_N)$ denotes the subspace of $H^1(\Omega)$ consisting of functions which vanish (in the weak sense) on $\partial\Omega_D$. Note that $|\cdot|_{H^1(\Omega)}$ is a norm on $H_0^1(\Omega \cup \partial\Omega_N)$ because of (2.3). The weak formulation of the boundary value problem (1.1), (1.2) is given by:

$$b(u, \phi) = -(f, \phi)_{2,\Omega} \quad \forall \phi \in H_0^1(\Omega \cup \partial\Omega_N), \tag{2.11 (a)}$$

$$u|_{\partial\Omega_D} = u_D|_{\partial\Omega_D}, \tag{2.11 (b)}$$

where $(\cdot,\cdot)_{2,\Omega}$ denotes the $L^2(\Omega)$-scalar product and b the bilinear form on $H^1(\Omega)$:

$$b(u, \phi) := \int_\Omega a(x) \operatorname{grad} u \cdot \operatorname{grad} \phi \, dx; \qquad u, \phi \in H^1(\Omega). \tag{2.12}$$

3. Derivation of the Finite Element Discretization. We shall derive the "fitted" finite element discretization of (1.1), (1.2) by generalizing the approach of Zlámal [17], [18]. The basic idea is to approximate the flux density J by a constant

vector on each finite element T. Therefore, let $T \in \Delta$ be a triangle with vertices P_1, P_2, P_3. We write

$$
\begin{aligned}
J(x) &= J(x_{G_T}) + \varepsilon_T(x), \qquad x \in T; \\
\varepsilon_T(x) &= \int_0^1 \frac{\partial J}{\partial x}(x_{G_T} + t(x - x_{G_T}))\, dt (x - x_{G_T}),
\end{aligned} \tag{3.1}
$$

or equivalently, by using the definition (1.3) of J:

$$
a(x) \operatorname{grad} u(x) = J(x_{G_T}) + \varepsilon_T(x), \qquad x \in T. \tag{3.2}
$$

Here $\frac{\partial J}{\partial x}$ denotes the Jacobian of J. We transform the equation (3.2) to the reference triangle $\hat{T}$ by employing (2.7) and obtain

$$
a(x(\xi))(I_T')^{-1} \operatorname{grad}_\xi u(x(\xi)) = J(x_{G_T}) + \varepsilon_T(x(\xi)), \qquad \xi \in \hat{T}. \tag{3.3}
$$

Premultiplication of (3.3) by $\frac{1}{a} I_T'$ gives

$$
\operatorname{grad}_\xi u(x(\xi)) = \frac{1}{a(x(\xi))} I_T' J(x_{G_T}) + \frac{1}{a(x(\xi))} I_T' \varepsilon_T(x(\xi)). \tag{3.4}
$$

We set

$$
J := \begin{pmatrix} J_1 \\ J_2 \end{pmatrix}, \quad \varepsilon_T := \begin{pmatrix} \varepsilon_{1,T} \\ \varepsilon_{2,T} \end{pmatrix}, \quad I_T = \begin{pmatrix} b_T & c_T \\ d_T & e_T \end{pmatrix} \tag{3.5}
$$

and integrate the ξ_1-component of (3.4) with respect to ξ_1 over the interval $[0, 1]$ and the ξ_2-component with respect to ξ_2 over $[0, 1]$:

$$
\begin{pmatrix} u(P_2) - u(P_1) \\ u(P_3) - u(P_1) \end{pmatrix} = \Lambda_T(a) I_T' J(x_{G_T}) + \delta_T, \tag{3.6}
$$

where $\Lambda_T(a)$ is the diagonal matrix

$$
\Lambda_T(a) := \operatorname{diag}\left(\int_0^1 \frac{d\xi_1}{a(x(\xi_1, 0))}, \int_0^1 \frac{d\xi_2}{a(x(0, \xi_2))} \right) \tag{3.7}
$$

and

$$
\delta_T := \begin{pmatrix} b_T \displaystyle\int_0^1 \frac{\varepsilon_{1,T}(x(\xi_1, 0))}{a(x(\xi_1, 0))}\, d\xi_1 + d_T \int_0^1 \frac{\varepsilon_{2,T}(x(\xi_1, 0))}{a(x(\xi_1, 0))}\, d\xi_1 \\ c_T \displaystyle\int_0^1 \frac{\varepsilon_{1,T}(x(0, \xi_2))}{a(x(0, \xi_2))}\, d\xi_2 + e_T \int_0^1 \frac{\varepsilon_{2,T}(x(0, \xi_2))}{a(x(0, \xi_2))}\, d\xi_2 \end{pmatrix}. \tag{3.8}
$$

We solve (3.6) for $J(x_{G_T})$ and obtain

$$
J(x_{G_T}) = (I_T')^{-1} \Lambda_T^{-1}(a) I_T' \operatorname{grad} u_I - (I_T')^{-1} \Lambda_T^{-1}(a) \delta_T. \tag{3.9 (a)}
$$

The crucial step in deriving the discretization scheme lies in neglecting that term on the right-hand side of (3.9) (a) which involves δ_T. Obviously, (3.1) implies that neglecting δ_T corresponds to approximating the flux density J by a constant vector on T. Intuitively, this is justified by the assumption that J is a "well-behaved" function (i.e., slowly varying in Ω).

We approximate $J(x)$ on T by

$$
J_T[u_\Delta] := (I_T')^{-1} \Lambda_T^{-1}(a) I_T' \operatorname{grad} u_\Delta, \tag{3.9 (b)}
$$

where u_Δ denotes a piecewise linear approximation to u with nodal values u_1, u_2, u_3 on P_1, P_2, and P_3, respectively. This process of deriving an approximate element flux density is repeated by employing different reference element transformations.

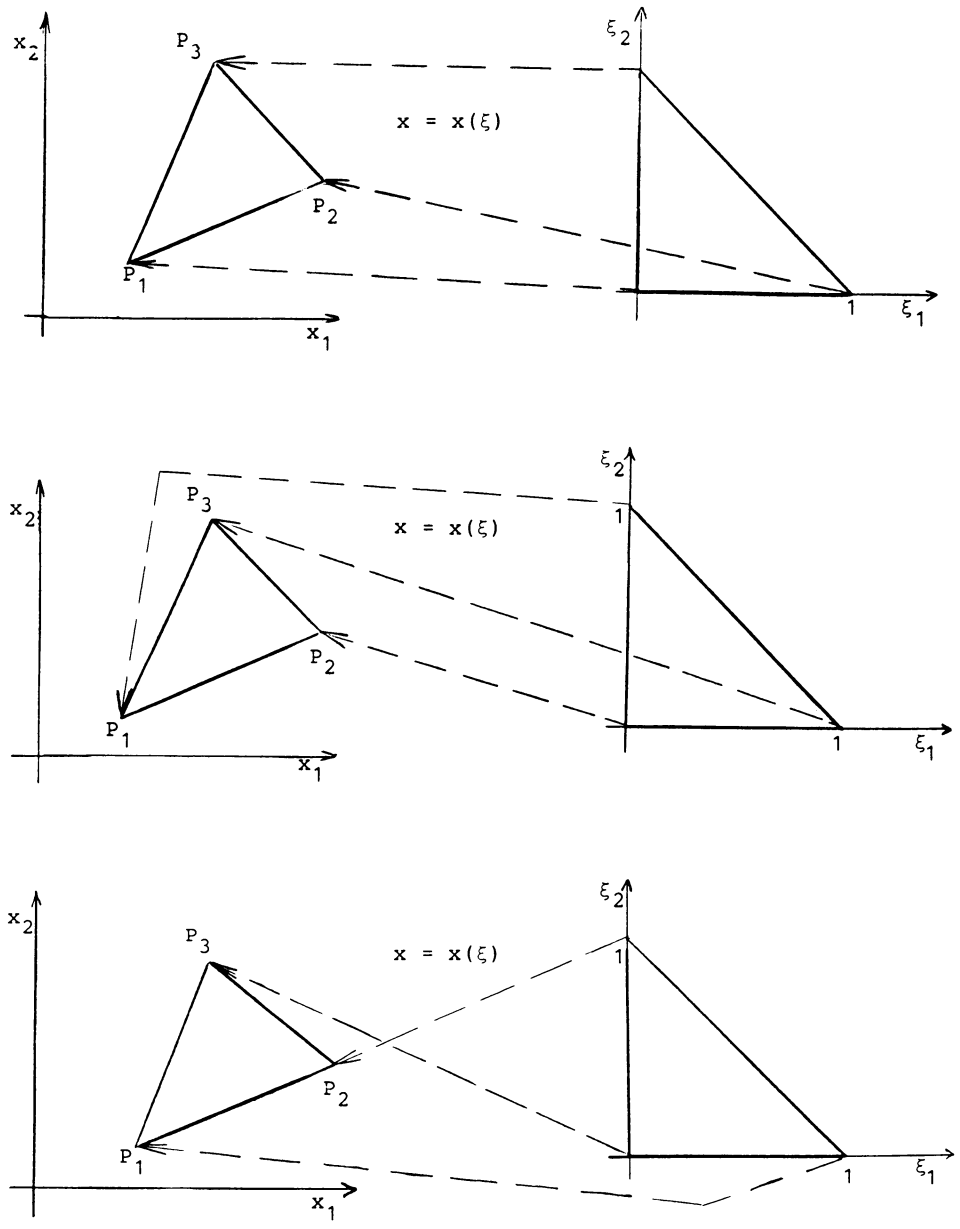

FIGURE 1

Element transformations

In addition to (2.7) we set up the transformation $x = x(\xi)$ such that P_2 is mapped into $(0,0)$ (and P_1, P_3 into $(0,1)$ and $(1,0)$, respectively). Also we map P_3 into $(0,0)$ (and P_1, P_2 into $(1,0)$ and $(0,1)$, respectively) (see Figure 1). Altogether, three different element transformations with associated Jacobians $I_{T,1}$, $I_{T,2}$, and $I_{T,3}$, respectively determine three approximate element flux densities (obtained from (3.9)) by

$$J_{T,i}[u_\Delta] := (I'_{T,i})^{-1}\Lambda_{T,i}^{-1}(a)I'_{T,i}\,\mathrm{grad}\,u_\Delta, \qquad i = 1,2,3. \tag{3.10}$$

The diagonal matrices $\Lambda_{T,i}(a)$ are given by

(3.11) (a) $$\Lambda_{T,1}^{-1}(a) = \operatorname{diag}(A_1, A_3),$$

(3.11) (b) $$\Lambda_{T,2}^{-1}(a) = \operatorname{diag}(A_2, A_1),$$

(3.11) (c) $$\Lambda_{T,3}^{-1}(a) = \operatorname{diag}(A_3, A_2),$$

(3.11) (d) $$A_i := \left(\frac{1}{l_i}\int_{e_i} a^{-1}\, dl\right)^{-1}.$$

Here, e_1, e_2, e_3 denote the sides of T and l_1, l_2, l_3 their lengths (see Figure 2). dl stands for the arclength differential. Thus, A_i is the inverse average of the function a over the side e_i.

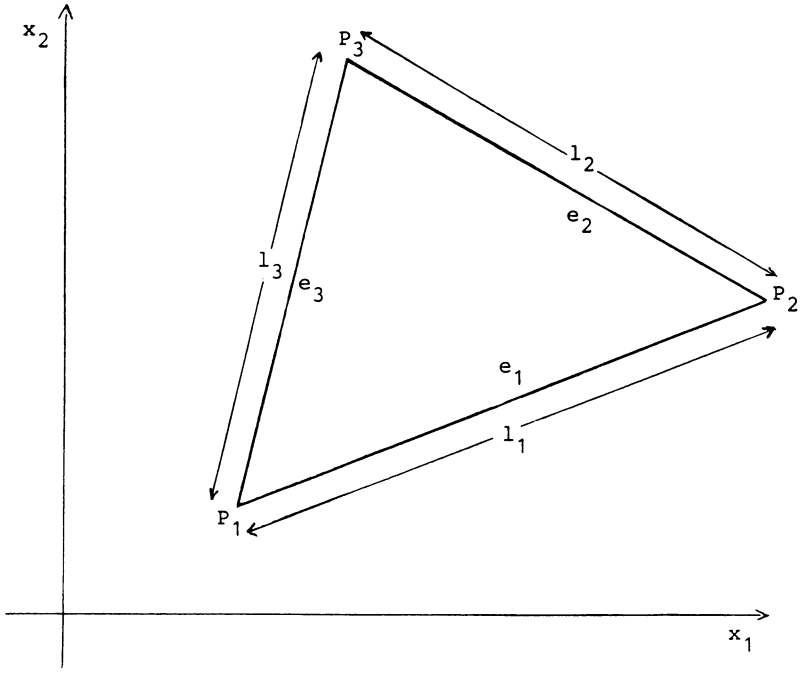

FIGURE 2
Element notation

Note that the vertex which is mapped into the origin of the (ξ_1, ξ_2)-plane completely determines the approximate element flux density. Exchanging the images of the other two vertices corresponds to an exchange of the columns of I_T and $\Lambda_T(a)$, which is annihilated by the similarity transformation $(I_T')^{-1}\Lambda_T^{-1}(a)I_T'$.

The derived approximate flux densities differ from standard discretizations by the special way of approximating the coefficient function $a(x)$. The matrices $(I_{T,i}')^{-1}\Lambda_{T,i}^{-1}(a)I_{T,i}'$, which multiply $\operatorname{grad} u_\Delta$ in (3.10), can be regarded as inverse-average-type approximations to $\operatorname{diag}(a(x), a(x))$ on T.

Let $P_l \in \overline{\Omega}$ be an arbitrary node. For all those triangles which have P_l as vertex we choose the element transformation (2.7) such that P_l is mapped into $(0,0)$. We define the following functional on X_Δ:

(3.12) $$b_l(u_\Delta) = \sum_{T \in P_l} \int_T J_{T,l}[u_\Delta] \cdot \operatorname{grad} \phi_\Delta^{(1)}\, dx,$$

where $J_{T,l}[u_\Delta]$ is given by (3.10) and $\phi_\Delta^{(1)}$ is the basis function which corresponds to P_l. By linearity we set up the bilinear form

$$b_\Delta(u_\Delta, \phi_\Delta) := \sum_{P_i \in \overline{\Omega}} \phi_i b_i(u_\Delta); \qquad u_\Delta, \phi_\Delta \in X_\Delta, \tag{3.13}$$

where ϕ_i are the nodal values of ϕ_Δ, i.e.,

$$\phi_\Delta = \sum_{P_i \in \overline{\Omega}} \phi_i \phi_\Delta^{(i)}. \tag{3.14}$$

We shall regard b_Δ as finite element approximation of the form b defined in (2.12). The approximation of the $L^2(\Omega)$-scalar product can be done in one of the usual ways. We choose the following numerical quadrature rule,

$$\int_T f(x)\,dx \approx Q_T(f) := \frac{1}{3}\,\mathrm{area}(T)(f_1 + f_2 + f_3), \tag{3.15}$$

which is exact if f is linear on T. We define the discrete scalar product

$$(f, g)_{2,\Omega,\Delta} := \sum_{T \in \Delta} Q_T(fg). \tag{3.16}$$

The finite element approximation of (2.11) then reads:

$$b_\Delta(u_\Delta, \phi_\Delta) = -(f, \phi_\Delta)_{2,\Omega,\Delta} \quad \forall \phi_\Delta \in X_{\Delta,0}, \tag{3.17 (a)}$$

$$u_\Delta|_{\partial\Omega_D} = (u_D)_I|_{\partial\Omega_D}. \tag{3.17 (b)}$$

We remark that the discretization (3.17) is a generalization of the scheme presented by Zlámal [17], [18], who dealt with the special case of the function a being the exponential of a piecewise linear function ψ_Δ. This application occurs in the modeling of semiconductor devices and will be discussed in Section 7. In fact, the analysis presented in this paper was motivated by the fundamental semiconductor device equations (see Markowich [7]). Note that the form b_Δ reduces to the Galerkin approxmation $b_\Delta(u_\Delta, \phi_\Delta) = \int_\Omega \operatorname{grad} u_\Delta \cdot \operatorname{grad} \phi_\Delta\,dx$ of $-\Delta u$ for $a \equiv 1$.

4. Analysis of the Bilinear Form b_Δ. First, we rewrite b_Δ as a sum over all triangles in Δ (instead of as a sum over the nodes). By observing that three integrals over each triangle occur in (3.13), we obtain by using (3.12):

$$\begin{aligned} b_\Delta(u_\Delta, \phi_\Delta) = \sum_{T \in \Delta} \Big(& \phi_1 \int_T J_{T,1}[u_\Delta] \cdot \operatorname{grad} \phi_\Delta^{(1)}\,dx \\ & + \phi_2 \int_T J_{T,2}[u_\Delta] \cdot \operatorname{grad} \phi_\Delta^{(2)}\,dx \\ & + \phi_3 \int_T J_{T,3}[u_\Delta] \cdot \operatorname{grad} \phi_\Delta^{(3)}\,dx \Big). \end{aligned} \tag{4.1}$$

As in the previous section, we use "local notation", i.e., the vertices and sides of T are denoted as in Figure 2; ϕ_1, ϕ_2, ϕ_3 are the corresponding nodal values and $\phi_\Delta^{(1)}, \phi_\Delta^{(2)}, \phi_\Delta^{(3)}$ the corresponding basis functions.

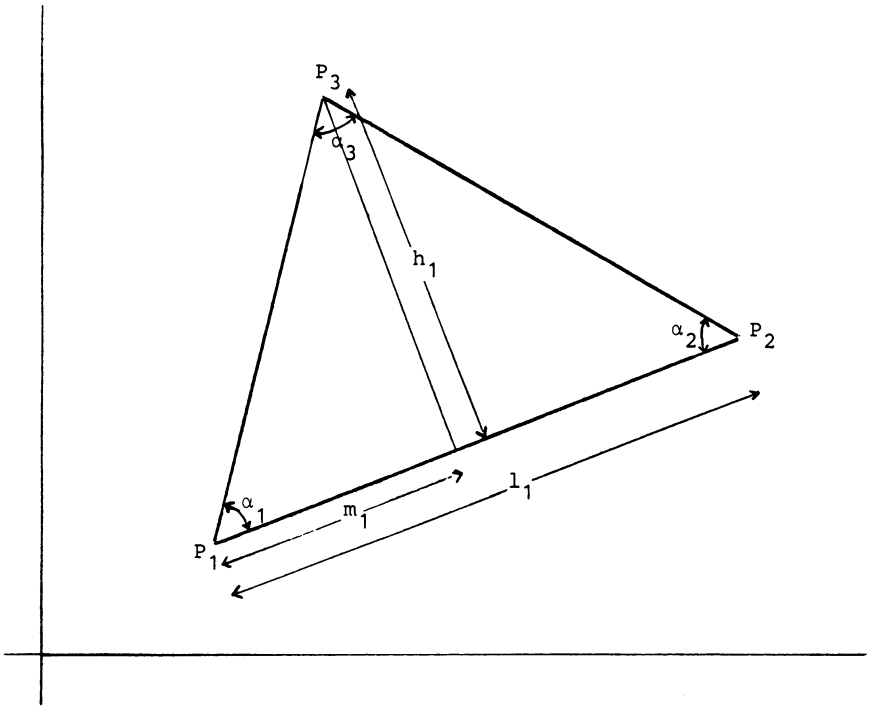

FIGURE 3
Triangle parameters

It is easy to show (by using (3.10)) that the integrals $\int_T J_{T,i}[u_\Delta] \cdot \operatorname{grad} \phi_\Delta^{(i)}\, dx$, $i = 1, 2, 3$, are invariant under translations and rotations of T. Thus, they can be expressed in terms of any set of three parameters, which determines the triangle T, e.g., in terms of the length l_1 of e_1, the height h_1 and the segment length m_1 (see Figure 3). Then the integrals occurring in (4.1) are obtained by a simple but tedious calculation:

$$\int_T J_{T,1}[u_\Delta] \cdot \operatorname{grad} \phi_\Delta^{(1)}\, dx$$

(4.2) (a)

$$= \frac{\operatorname{area}(T)}{(l_1 h_1)^2}(l_1(l_1 - m_1)A_3(u_1 - u_3) + (h_1^2 - l_1 m_1 + m_1^2)A_1(u_1 - u_2)),$$

$$\int_T J_{T,2}[u_\Delta] \cdot \operatorname{grad} \phi_\Delta^{(2)}\, dx$$

(4.2) (b)

$$= \frac{\operatorname{area}(T)}{(l_1 h_1)^2}(l_1 m_1 A_2(u_2 - u_3) + (h_1^2 - l_1 m_1 + m_1^2)A_1(u_2 - u_1)),$$

$$\int_T J_{T,3}[u_\Delta] \cdot \operatorname{grad} \phi_\Delta^{(3)}\, dx$$

(4.2) (c)

$$= \frac{\operatorname{area}(T)}{(l_1 h_1)^2}(l_1 m_1 A_2(u_3 - u_2) + l_1(l_1 - m_1)A_3(u_3 - u_1)).$$

We insert (4.2) into (4.1) and obtain by an easy manipulation

$$
\begin{aligned}
b_\Delta(u_\Delta,\phi_\Delta) = \sum_{T\in\Delta} \frac{\text{area}(T)}{(l_1h_1)^2}\big(&A_1(h_1^2 - l_1m_1 + m_1^2)(u_1-u_2)(\phi_1-\phi_2)\\
&+ A_2 l_1 m_1 (u_2-u_3)(\phi_2-\phi_3)\\
&+ A_3 l_1 (l_1-m_1)(u_1-u_3)(\phi_1-\phi_3)\big).
\end{aligned}
\tag{4.3}
$$

By inspection of (4.3) we conclude:

LEMMA 4.1. *b_Δ is symmetric on X_Δ.*

An even more convenient form of b_Δ is derived by employing trigonometric arguments to express those terms in (4.3) which depend on the geometry of T only in terms of the interior angles of T. We obtain

$$
\begin{aligned}
b_\Delta(u_\Delta,\phi_\Delta) = \frac{1}{2}\sum_{T\in\Delta} \big(&A_1\cot(\alpha_3)(u_1-u_2)(\phi_1-\phi_2)\\
&+ A_2\cot(\alpha_1)(u_2-u_3)(\phi_2-\phi_3)\\
&+ A_3\cot(\alpha_2)(u_1-u_3)(\phi_1-\phi_3)\big),
\end{aligned}
\tag{4.4}
$$

where α_i denote the interior angles of the triangle T (see Figure 3). We call a triangular mesh Δ of acute type, if all interior angles α of all triangles $T \in \Delta$ satisfy $0 < \alpha \le \pi/2$. We prove

LEMMA 4.2. *Let Δ be of acute type and assume that (2.1) holds. Then b_Δ is bounded on X_Δ and coercive on $X_{\Delta,0}$. In particular,*

$$
|b_\Delta(u_\Delta,\Phi_\Delta)| \le \bar{a}|u_\Delta|_{H^1(\Omega)}|\phi_\Delta|_{H^1(\Omega)} \quad \forall u_\Delta,\phi_\Delta \in X_\Delta, \tag{4.5}
$$

$$
b_\Delta(\phi_\Delta,\phi_\Delta) \ge \underline{a}|\phi_\Delta|_{H^1(\Omega)} \quad \forall \phi_\Delta \in X_\Delta. \tag{4.6}
$$

Proof. Since Δ is assumed to be of acute type, we have $\cot\alpha_i > 0$. Thus, the Cauchy-Schwarz inequality (applied to (4.4)) gives

$$
\begin{aligned}
|b_\Delta(u_\Delta,\phi_\Delta)| \le \frac{1}{2}\sum_{T\in\Delta} &\max(A_1,A_2,A_3)\\
&\times [\cot(\alpha_3)(u_1-u_2)^2 + \cot(\alpha_1)(u_2-u_3)^2\\
&\qquad + \cot(\alpha_2)(u_1-u_3)^2]^{1/2}\\
&\times [\cot(\alpha_3)(\phi_1-\phi_2)^2 + \cot(\alpha_1)(\phi_2-\phi_3)^2\\
&\qquad + \cot(\alpha_2)(\phi_1-\phi_3)^2]^{1/2}.
\end{aligned}
\tag{4.7}
$$

A simple computation gives

$$
\begin{aligned}
|\operatorname{grad} u_\Delta| = \frac{1}{\sqrt{2}}\,&\frac{1}{(\text{area}(T))^{1/2}}\\
&\times (\cot(\alpha_3)(u_1-u_2)^2 + \cot(\alpha_1)(u_2-u_3)^2\\
&\qquad + \cot(\alpha_2)(u_1-u_3)^2)^{1/2} \quad \text{on } T
\end{aligned}
\tag{4.8}
$$

and, since (2.1), (3.11) (d) imply

$$
\underline{a} \le A_i \le \bar{a}, \qquad i = 1,2,3, \tag{4.9}
$$

we derive

$$|b_\Delta(u_\Delta, \phi_\Delta)| \leq \bar{a} \sum_{T \in \Delta} \int_T |\operatorname{grad} u_\Delta| \, |\operatorname{grad} \phi_\Delta| \, dx. \tag{4.10}$$

Also, we obtain from (4.4), (4.8) that

$$b_\Delta(\phi_\Delta, \phi_\Delta) \geq \sum_{T \in \Delta} \min(A_1, A_2, A_3) \int_T |\operatorname{grad} \phi_\Delta|^2 \, dx \geq \underline{a} \int_\Omega |\operatorname{grad} \phi_\Delta|^2 \, dx,$$

from which (4.6) follows. □

The existence and uniqueness of a finite element solution u_Δ follows from Lemma 4.2.

THEOREM 4.1. *Let the assumption* (2.1), (2.2), (2.3) *hold and assume that the mesh* Δ *is of acute type. Then the finite element discretization* (3.17) *has a unique solution* $u_\Delta \in X_\Delta$.

Proof. The estimate (4.6) implies the invertibility of the stiffness matrix. □

The element stiffness matrices S_T can be easily calculated from the expression (4.4). We have

$$b_\Delta(u_\Delta, \phi_\Delta) = \sum_{T \in \Delta} (u_1, u_2, u_3) S_T \begin{pmatrix} \phi_1 \\ \phi_2 \\ \phi_3 \end{pmatrix}, \tag{4.11) (a}$$

where the element stiffness marix S_T is given by

(4.11) (b)

$$S_T := \frac{1}{2} \left(\begin{array}{c|c|c} A_1 \cot(\alpha_3) + A_3 \cot(\alpha_2) & -A_1 \cot(\alpha_3) & -A_3 \cot(\alpha_2) \\ \hline -A_1 \cot(\alpha_3) & A_1 \cot(\alpha_3) + A_2 \cot(\alpha_1) & -A_2 \cot(\alpha_1) \\ \hline -A_3 \cot(\alpha_2) & -A_2 \cot(\alpha_1) & A_2 \cot(\alpha_1) + A_3 \cot(\alpha_2) \end{array} \right).$$

The following maximum-minimum principle for b_Δ is proven by proceeding as in Zlámal [18].

LEMMA 4.3. *Let* (2.1), (2.3) *hold and assume that* Δ *is of acute type. If* $w_\Delta \in X_\Delta$ *assumes a maximum (minimum) at the node* $P_i \in \overline{\Omega}$, *then*

$$b_\Delta(w_\Delta, \phi_\Delta^{(i)}) \geq 0 \quad (\leq 0) \tag{4.12}$$

holds.

Proof. From (4.4) we derive, setting $w_i = w_\Delta(P_i)$,

$$b_\Delta(w_\Delta, \phi_\Delta^{(i)}) = \frac{1}{2} \sum_{T \in P_i} (A_1 \cot(\alpha_3)(w_1 - w_2) + A_3 \cot(\alpha_2)(w_1 - w_3)). \tag{4.13}$$

Since $\cot(\alpha_3) \geq 0, \cot(\alpha_2) \geq 0$, we conclude $b_\Delta(w_\Delta, \phi_\Delta^{(i)}) \geq 0$ if w_1 is a (local) maximum of w_Δ and $b_\Delta(w_\Delta, \phi^{(i)}) \leq 0$ if w_1 is a local minimum of w_Δ. □

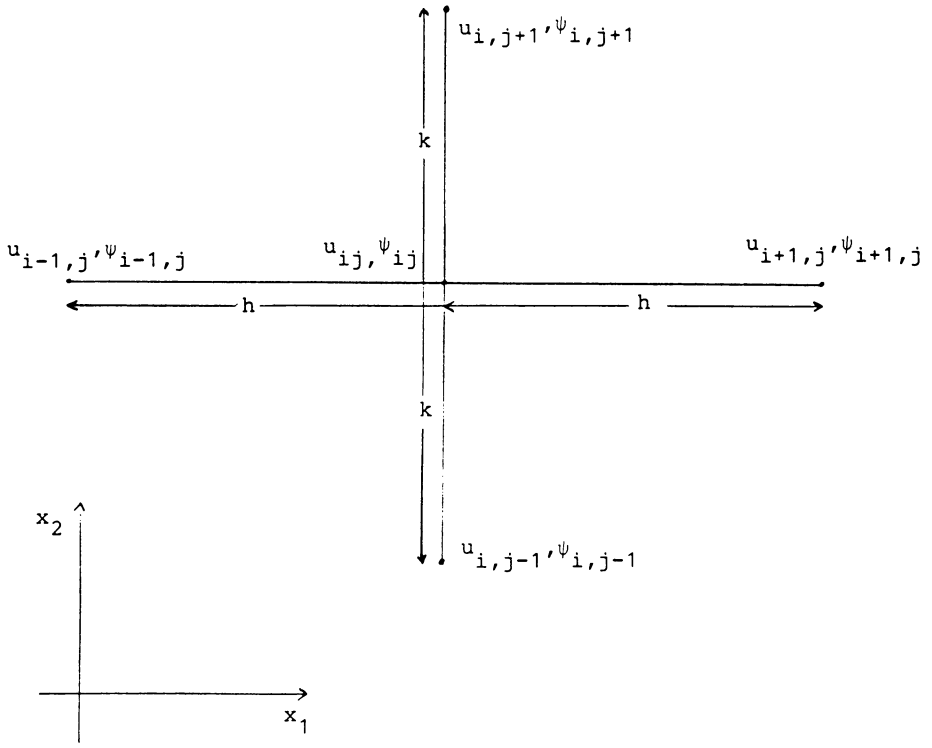

FIGURE 4
A Finite Difference Star

The maximum-minimum principle implies important properties of the stiffness matrix S which is associated with b_Δ (S is an $M \times M$-matrix with entries $S_{ij} := b_\Delta(\phi_\Delta^{(i)}, \phi_\Delta^{(j)})$).

THEOREM 4.2. *If Δ is of acute type and if* (2.1), (2.3) *hold, then S is a diagonally dominant Stieltjes matrix.*

Proof. The symmetry of b_Δ implies that S is symmetric and the coercivity estimate implies positive definiteness. Let $i \neq j$. Since the basis function $\phi_\Delta^{(i)}$ assumes a minimum (namely 0) at P_j, we conclude from the minimum principle $S_{ij} = b_\Delta(\phi_\Delta^{(i)}, \phi_\Delta^{(j)}) \leq 0$. Thus, all off-diagonal entries of S are nonpositive (the diagonal entires S_{ii} are, obviously, positive) and, consequently, S is a Stieltjes matrix (see Varga [15]). The proof of the diagonal dominance of S is standard (see, e.g., Zlámal [18]). □

The results of this section assert that the most important qualitative properties of the "piecewise linear" Galerkin approximation of the operator $-\operatorname{div}(a(x)\operatorname{grad} u)$ carry over to the bilinear form b_Δ if the mesh is of acute type.

5. Convergence. The well-known convergence result for Galerkin's method (see, e.g., Strang and Fix [13]) asserts that the error of the Galerkin approximation of u is minimal in the energy norm. Thus, an interpretation of (3.17) as a perturbation of Galerkin's approximation cannot produce error estimates in the energy norm, which are independent of the variation of u between the nodes. In fact,

estimates obtained in this way depend on the H^2-norms of u on the triangles of the partition (just as the Galerkin approximation error estimate) and on $W^{1,\infty}$-norms of the coefficient function a on the triangles.

Therefore, we have to proceed differently in order to obtain an estimate which only depends on the variation of the flux density J and of the right-hand side f (and on the diameters of the triangles). The basic idea is to estimate the difference $u_I - u_\Delta$ of the piecewise linear interpolant u_I of the solution u of (1.1), (1.2) and the finite element solution u_Δ of (3.17) and to interpret b_Δ directly as perturbation of b. We have

$$b_\Delta(u_I - u_\Delta, \phi_\Delta) = b_\Delta(u_I, \phi_\Delta) + (f, \phi_\Delta)_{2,\Omega,\Delta} \quad \forall \phi_\Delta \in X_{\Delta,0}. \tag{5.1}$$

In particular, we obtain, with $\phi_\Delta = u_I - u_\Delta$ ($\in X_{\Delta,0}$!),

$$b_\Delta(u_I - u_\Delta, u_I - u_\Delta) = b_\Delta(u_I, u_I - u_\Delta) + (f, u_I - u_\Delta)_{2,\Omega,\Delta} \tag{5.2}$$

and the estimate (4.6) gives

$$\underline{a}|u_I - u_\Delta|^2_{H^1(\Omega)} \leq |b_\Delta(u_I, u_I - u_\Delta) + (f, u_I - u_\Delta)_{2,\Omega,\Delta}|. \tag{5.3}$$

The following consistency type lemma provides the basic ingredient for an estimate of the right-hand side of (5.3).

LEMMA 5.1. *Let* (2.1)–(2.4) *hold and assume that the mesh* Δ *is of acute type. Then*

$$\begin{aligned} &|b_\Delta(u_I, \phi_\Delta) - b(u, \phi_\Delta)| \\ &\quad \leq 3(\operatorname{area}(\Omega))^{1/2} \max_{T \in \Delta} \left(h_T \left\| \left| \frac{\partial J}{\partial x} \right| \right\|_{L^\infty(T)} \right) |\phi_\Delta|_{H^1(\Omega)} \end{aligned} \tag{5.4}$$

holds for every $\phi_\Delta \in X_\Delta$.

Proof. By employing the three reference element transformations of Figure 1 we obtain from (3.9) (a), (b):

$$J_{T,i}[u_I] = J(x_{G_T}) + \sigma_{T,i}, \qquad i = 1, 2, 3, \tag{5.5}$$

where $\sigma_{T,i}$ is given by

$$\sigma_{T,i} = (I_{T,i})^{-1} \Lambda_{T,i}^{-1}(a) \delta_{T,i}, \qquad i = 1, 2, 3. \tag{5.6}$$

$\delta_{T,i}$ is of the form (3.8). By inserting (5.5) into (4.1) we obtain

$$b_\Delta(u_I, \phi_\Delta) = \sum_{T \in \Delta} \int_T J(x_{G_T}) \cdot \operatorname{grad} \phi_\Delta \, dx + \sum_{T \in \Delta} \omega_T[\phi_\Delta] \tag{5.7}$$

with

$$\begin{aligned} \omega_T[\phi_\Delta] = \phi_1 \int_T \sigma_{T,1} \cdot \operatorname{grad} \phi_\Delta^{(1)} \, dx + \phi_2 \int_T \sigma_{T,2} \cdot \operatorname{grad} \phi_\Delta^{(2)} \, dx \\ + \phi_3 \int_T \sigma_{T,3} \cdot \operatorname{grad} \phi_\Delta^{(3)} \, dx. \end{aligned} \tag{5.8}$$

We substitute $J(x) + (J(x_{G_T}) - J(x))$ for $J(x_{G_T})$ into the first sum on the right-hand side of (5.7) and obtain

$$\begin{aligned} b_\Delta(u_I, \phi_\Delta) = \int_\Omega J(x) \cdot \operatorname{grad} \phi_\Delta \, dx + \sum_{T \in \Delta} \int_T (J(x_{G_T}) - J(x)) \cdot \operatorname{grad} \phi_\Delta \, dx \\ + \sum_{T \in \Delta} \omega_T[\phi_\Delta]. \end{aligned} \tag{5.9}$$

By (2.12) we have

$$\int_\Omega J(x)\cdot \operatorname{grad}\phi_\Delta\,dx = b(u,\phi_\Delta), \qquad \phi_\Delta \in X_\Delta. \tag{5.10}$$

The second term on the right-hand side of (5.9) can be estimated by

$$\begin{aligned}&\left|\sum_{T\in\Delta}\int_T (J(x_{G_T})-J(x))\cdot\operatorname{grad}\phi_\Delta\,dx\right| \\ &\quad\le (\operatorname{area}(\Omega))^{1/2}\max_{T\in\Delta}\left(h_T\left\|\left|\frac{\partial J}{\partial x}\right|\right\|_{L^\infty(T)}\right)|\phi_\Delta|_{H^1(\Omega)}.\end{aligned} \tag{5.11}$$

A somewhat involved but essentially trivial computation gives

$$\begin{aligned}\omega_T[\phi_\Delta] &= \tfrac{1}{2}(A_1\cot(\alpha_3)((\bar x_1-\bar x_2)E_{1,1}+(\bar y_1-\bar y_2)E_{1,2})(\phi_1-\phi_2)\\ &\quad + A_2\cot(\alpha_1)((\bar x_2-\bar x_3)E_{2,1}+(\bar y_2-\bar y_3)E_{2,2})(\phi_2-\phi_3)\\ &\quad + A_3\cot(\alpha_2)((\bar x_1-\bar x_3)E_{3,1}+(\bar y_1-\bar y_3)E_{3,2})(\phi_1-\phi_3)),\end{aligned} \tag{5.12}$$

where $P_i=(\bar x_i,\bar y_i)'$, $i=1,2,3$ are the vertices of T (see Figure 2) and

$$E_{i,j}=\frac{1}{l_i}\int_{e_i}\frac{\varepsilon_{j,T}}{a}\,dl; \qquad i=1,2,3,\ j=1,2, \tag{5.13}$$

with $\varepsilon_{j,T}$ defined in (3.1), (3.5). We obtain from (4.8) and (5.12):

$$\begin{aligned}|\omega_T[\phi_\Delta]| &\le \frac{(\operatorname{area}(T))^{1/2}}{\sqrt{2}}\\ &\quad\times(\cot(\alpha_3)A_1^2((\bar x_1-\bar x_2)E_{1,1}+(\bar y_1-\bar y_2)E_{1,2})^2\\ &\qquad+\cot(\alpha_1)A_2^2((\bar x_2-\bar x_3)E_{2,1}+(\bar y_2-\bar y_3)E_{2,2})^2\\ &\qquad+\cot(\alpha_2)A_3^2((\bar x_1-\bar x_3)E_{3,1}+(\bar y_1-\bar y_3)E_{3,2})^2)^{1/2}\\ &\quad\times|\operatorname{grad}\phi_\Delta| \quad\text{on } T\end{aligned} \tag{5.14}$$

and, consequently,

$$\begin{aligned}|\omega_T[\phi_\Delta]| &\le \frac{(\operatorname{area}(T))^{1/2}}{\sqrt{2}}\max_{i=1,2,3}\left(A_i\sqrt{|E_{i,1}|^2+|E_{i,2}|^2}\right)\\ &\quad\times(\cot(\alpha_3)l_1^2+\cot(\alpha_1)l_2^2+\cot(\alpha_2)l_3^2)^{1/2}|\operatorname{grad}\phi_\Delta| \quad\text{on } T.\end{aligned} \tag{5.15}$$

Since

$$(\cot(\alpha_3)l_1^2+\cot(\alpha_1)l_2^2+\cot(\alpha_2)l_3^2)^{1/2}=2(\operatorname{area}(T))^{1/2} \tag{5.16}$$

holds, we derive

$$|\omega_T[\phi_\Delta]|\le 2\operatorname{area}(T)\max_{\substack{i=1,2,3\\ j=1,2}}(A_i|E_{i,j}|)|\operatorname{grad}\phi_\Delta| \quad\text{on } T. \tag{5.17}$$

From (5.13), (3.11) (d), (3.1) and (3.5) we obtain

$$A_i|E_{i,j}|\le\sup_{e_i}|\varepsilon_{j,T}|\le h_T\left\|\left|\frac{\partial J}{\partial x}\right|\right\|_{L^\infty(T)} \tag{5.18}$$

and thus

$$\left|\sum_{T\in\Delta}\omega_T[\phi_\Delta]\right|\le 2(\operatorname{area}(\Omega))^{1/2}\max_{T\in\Delta}\left(h_T\left\|\left|\frac{\partial J}{\partial x}\right|\right\|_{L^\infty(T)}\right)|\phi_\Delta|_{H^1(\Omega)}. \tag{5.19}$$

The estimate (5.4) follows from (5.8), (5.9), (5.10), (5.11) and (5.19). □

Estimates for the error in approximating the $L^2(\Omega)$-scalar product $(\cdot,\cdot)_{2,\Omega}$ by the discrete scalar product $(\cdot,\cdot)_{2,\Omega,\Delta}$ are completely standard. The following lemma is a special case of Theorem 2 in Nedoma [10].

LEMMA 5.2. *Let $f \in H^2(\Omega)$ and assume that there is $\alpha_0 > 0$ such that $\alpha \geq \alpha_0$ holds for every interior angle α of every triangle $T \in \Delta$ ("minimum angle condition"). Then the estimate*

$$|(f,\phi_\Delta)_{2,\Omega} - (f,\phi_\Delta)_{2,\Omega,\Delta}| \leq Ch^2\|f\|_{H^2(\Omega)}\|\phi_\Delta\|_{H^1(\Omega)} \tag{5.20}$$

holds for every $\phi_\Delta \in X_\Delta$. The constant C depends only on Ω and on α_0.

We have

$$\begin{aligned} &|b_\Delta(u_I, u_I - u_\Delta) + (f, u_I - u_\Delta)_{2,\Omega,\Delta}| \\ &\quad \leq |b_\Delta(u_I, u_I - u_\Delta) - b(u, u_I - u_\Delta)| \\ &\qquad + |b(u, u_I - u_\Delta) - (f, u_I - u_\Delta)_{2,\Omega,\Delta}|. \end{aligned} \tag{5.21}$$

Lemmas 5.1, 5.2, (2.11) (with $\phi_\Delta = u_I - u_\Delta \in X_{\Delta,0}$) and (5.3) imply the main result of this paper.

THEOREM 5.1. *Assume that (2.1)–(2.4) hold and that the mesh Δ is of acute type and satisfies the minimum angle condition (of Lemma 5.2). Then the (uniquely defined) finite element approximation u_Δ of u satisfies the error estimate*

$$\|u_I - u_\Delta\|_{H^1(\Omega)} \leq \frac{K}{\underline{a}} \left(h \left\| \left| \frac{\partial J}{\partial x} \right| \right\|_{L^\infty(\Omega)} + h^2 \|f\|_{H^2(\Omega)} \right). \tag{5.22}$$

The constant K depends only on the minimum angle α_0, on Ω and on $\partial\Omega_D$.

The error estimate (5.22) is of order h and independent of the derivatives of the exact solution u. It only depends on first derivatives of the flux density J and—owing to the error in approximating $(\cdot,\cdot)_{2,\Omega}$ by $(\cdot,\cdot)_{2,\Omega,\Delta}$—on derivatives of the right-hand side f.

Obviously, only nodal value differences $u(P_i) - u_\Delta(P_i)$ enter the left-hand side of (5.22). We cannot expect a convergence result of the type (5.22) to hold for $\|u - u_\Delta\|_{H^1(\Omega)}$ since the inverse average finite element method only depends on the values of the coefficient function a along element edges. An estimate for $\|u - u_\Delta\|_{H^1(\Omega)}$ certainly depends on derivatives of the solution u and of the coefficient function a. The strength of the method (3.17) is the good nodal approximation property reflected by the estimate (5.22). In the one-dimensional homogeneous case ($f \equiv 0$) the flux $J(x) = a(x)u'(x)$ is a constant. Thus the one-dimensional analogue of the inverse-average-finite element scheme (3.17) gives exact nodal values, i.e., $u_\Delta(P_i) = u(P_i)$ (this follows immediately by setting $\frac{\partial J}{\partial x} = 0$ in (5.22)). The behavior of u *between* nodes may be arbitrarily bad.

For the (multi-dimensional) semiconductor device equations (see, e.g., Van Roosbroeck [14]) there is a lot of computational evidence indicating the fast variation of $a, u, \nabla u$ and the slow variation of J (see Selberherr [12]). Also, these results are confirmed by a singular perturbation analysis (see Markowich [7]). It turns out that a and ∇u exhibit internal and boundary layers of fast variation while the flux density J varies significantly more slowly. Rigorous regularity-type estimates of this type do not—to the authors' knowledge—exist in the literature as yet.

6. Approximation of the Flux. In many applications an accurate approximation of the flux is extremely important (e.g. for the semiconductor device problem, see Markowich [7], where $J = a \operatorname{grad} u$ represents a current density). Since our finite element method is based on approximating J by constant vectors on each finite element, the corresponding convergence results are rather easily obtained.

First we remark that the discrete flux density is not uniquely defined. On every finite element T each of the three quantities $J_{T,i} := J_{T,i}[u_\Delta]$, $i = 1,2,3$ defined in (3.10) can be regarded as an approximation to J. We prove

LEMMA 6.1. *Let the assumptions of Theorem* 5.1 *hold. Then, given an arbitrary numbering of the vertices in each triangle T, the estimate*

$$\left(\sum_{T\in\Delta}\int_T |J(x) - J_{T,i}|^2\,dx\right)^{1/2} \le \text{const}\,\frac{\bar a}{\underline a}\left(h\left\|\left|\frac{\partial J}{\partial x}\right|\right\|_{L^\infty(\Omega)} + h^2\|f\|_{H^2(\Omega)}\right) \tag{6.1}$$

holds for $i = 1,2,3$.

Proof. We estimate

$$\begin{aligned}|J(x) - J_{T,i}| \le{}& |J(x) - J(x_{G_T})| + |J_{T,i}[u_I] - J(x_{G_T})| \\ &+ |J_{T,i}[u_I] - J_{T,i}[u_\Delta]|.\end{aligned}$$

Obviously,

$$|J(x) - J(x_{G_T})| \le h_T\left\|\left|\frac{\partial J}{\partial x}\right|\right\|_{L^\infty(T)}, \qquad x \in T,$$

and, by (5.5), (5.6), (3.8),

$$|J_{T,i}[u_I] - J(x_{G_T})| = |\sigma_{T,i}| \le \text{const}\,h_T\left\|\left|\frac{\partial J}{\partial x}\right|\right\|_{L^\infty(T)}.$$

Also, (3.10) gives

$$|J_{T,i}[u_I] - J_{T,i}[u_\Delta]| = |(I'_{T,i})^{-1}\Lambda_{T,i}^{-1}(a)I'_{T,i}\operatorname{grad}(u_I - u_\Delta)|.$$

The "minimum angle condition" implies $|(I'_{T,i})^{-1}|\,|I'_{T,i}| \le \text{const}$ and (4.9) gives

$$|J_{T,i}[u_I] - J_{T,i}[u_\Delta]| = \text{const}\,\bar a|\operatorname{grad}(u_I - u_\Delta)|.$$

The estimate (6.1) follows from (5.22). □

The discrete flux density J_Δ defined by

$$J_\Delta(x) = J_{T,i}, \qquad x \in T,\ i = 1,2 \text{ or } 3, \tag{6.2}$$

converges to J of order 1 in the L^2-norm:

$$\|J - J_\Delta\|_{L^2(\Omega)} = O\left(h\left\|\left|\frac{\partial J}{\partial x}\right|\right\|_{L^\infty(\Omega)} + h^2\|f\|_{H^2(\Omega)}\right). \tag{6.3}$$

The error bound only depends on the variation of J and on the variation of the right-hand side f.

In semiconductor device simulation the computation of outflow currents is often the final goal. Usually, the Dirichlet boundary $\partial\Omega_D$ is the union of finitely many

(closed and disconnected) contact segments. Let $C \in \partial\Omega_D$ be one of these contacts. Then

$$J_C := \int_C J \cdot \nu \, dl \tag{6.4}$$

is the outflow current at C. We choose a function $\phi_C \in H^1(\Omega) \cap C(\overline{\Omega})$ which satisfies

$$\phi_C(x) = \begin{cases} 1, & x \in C, \\ 0, & x \in \partial\Omega_D - C. \end{cases} \tag{6.5}$$

Multiplying (2.1) by ϕ_C and integrating by parts gives

$$(f, \phi_C)_{2,\Omega} = -b(u, \phi_C) + \int_{\partial\Omega} \phi_C a \operatorname{grad} u \cdot \nu \, dl.$$

We derive from (2.2), (6.4), (6.5)

$$J_C = b(u, \phi_C) + (f, \phi_C)_{2,\Omega}. \tag{6.6}$$

An approximation of the outflow current is obtained by substituting the bilinear form b by its finite element discretization b_Δ, u by the finite element solution u_Δ, ϕ_C by its piecewise linear interpolant $(\phi_C)_I$, and $(\cdot,\cdot)_{2,\Omega}$ by $(\cdot,\cdot)_{2,\Omega,\Delta}$:

$$J_{C,\Delta} := b_\Delta(u_\Delta, (\phi_\Delta)_I) + (f, (\phi_C)_I)_{2,\Omega,\Delta}. \tag{6.7}$$

$J_{C,\Delta}$ is computed via post-processing by an additional evaluation of the bilinear form b_Δ and of the discrete scalar product $(\cdot,\cdot)_{2,\Omega,\Delta}$.

The proof for the convergence of $J_{C,\Delta}$ to J_C as $h \to 0$ is similar to the proof of Lemma 6.1. We obtain

$$|J_C - J_{C,\Delta}| = O\left(h \left\| \left| \frac{\partial J}{\partial x} \right| \right\|_{L^\infty(\Omega)} + h^2 \|f\|_{H^2(\Omega)} \right). \tag{6.8}$$

We remark that the approximation of the projection of the flux density onto the sides of the triangles is uniquely defined, although the discrete flux density is not. Let e_i be an arbitrary side of length l_i connecting the vertices P_{i_1} and P_{i_2}, and let σ_i be the unit vector parallel to e_i pointing from P_{i_1} to P_{i_2}. Then the "intrinsic" approximation of $J \cdot \sigma_i|_{e_i}$ is

$$J \cdot \sigma_i|_{e_i} \approx A_i \frac{u_\Delta(P_{i_2}) - u_\Delta(P_{i_1})}{l_i}, \tag{6.9}$$

where A_i is given by (3.11) (d). The right-hand side of (6.9) is equal to $J_{T,i_1}[u_\Delta] \cdot \sigma_i$ and $J_{T,i_2}[u_\Delta] \cdot \sigma_i$ for both triangles T which have the edge e_i.

In stress analysis it is common to approximate the stress at an interior node P_i by the arithmetic mean of approximate stresses (obtained from the piecewise linear finite element solution) over all triangles which contain P_i (see Zienkiewicz [16, p. 104]). In fact, superconvergence of this approximation holds for Galerkin's method on uniform meshes. Therefore, we suggest to approximate the current density $J(P_i)$, $P_i \in \Omega$, by the arithmetic mean $J_\Delta(P_i)$ of the approximate current densities $J_{T,i}[u_\Delta]$ over all triangles T with $P_i \in T$:

$$J(P_i) \approx J_\Delta(P_i) := \frac{1}{N_i} \sum_{T \in P_i} J_{T,i}[u_\Delta] \tag{6.10}$$

(N_i denotes the number of all triangles T which contain P_i).

7. An Application. A typical example for the application of inverse-average-type finite element schemes is provided by the electron and hole continuity equations of the Van Roosbroeck semiconductor device model (see Van Roosbroeck [14], Selberherr [12], Markowich [7]). In a simplified setup we obtain for the electron continuity equation

$$a = e^{\psi}, \tag{7.1}$$

where ψ denotes the electrostatic potential and $n = e^{\psi}u$ the electron concentration. $J = e^{\psi}\nabla u$ is the electron current density. Typically, the potential ψ, which satisfies a singularly perturbed Poisson's equation, exhibits thin space charge layers, within which it varies rapidly. Outside these layers, ψ varies moderately. The electron current density J exhibits layer behavior in some cases, too; however, its variation is usually much weaker than that of ψ.

Usually, approximations ψ_i of the nodal values $\psi(P_i)$ of ψ are obtained as solutions of a finite element or finite difference discretization of Poisson's equation. We set up the piecewise linear interpolant ψ_Δ of ψ_i satisfying

$$\psi_\Delta(P_i) := \psi_i \tag{7.2}$$

and define

$$a_\Delta(x) := e^{\psi_\Delta(x)}. \tag{7.3}$$

Let T be the triangle of Figure 2. Then the inverse averages of a_Δ along the edges are given by

$$A_i = \left(\frac{1}{l_i}\int_{e_i} a_\Delta^{-1}\,dl\right)^{-1} = \left(\frac{1}{l_i}\int_{e_i} e^{-\psi_\Delta}\,dl\right)^{-1}. \tag{7.4}$$

A simple calculation gives

$$A_1 = e^{\psi_1}B(\psi_1 - \psi_2), \quad A_2 = e^{\psi_2}B(\psi_2 - \psi_3), \quad A_3 = e^{\psi_3}B(\psi_3 - \psi_1), \tag{7.5}$$

where $B(z)$ is the Bernoulli function defined by

$$B(z) = \frac{z}{e^z - 1}. \tag{7.6}$$

If the triangular mesh Δ is generated by a rectangular grid with mesh sizes h in the x_1-direction and k in the x_2-direction, then the finite element discretization reduces to the finite difference scheme

$$\begin{aligned}
&\frac{1}{h}\left(e^{\psi_{i+1,j}}B(\psi_{i+1,j} - \psi_{i,j})\frac{u_{i+1,j} - u_{i,j}}{h}\right.\\
&\qquad\left. - e^{\psi_{i,j}}B(\psi_{i,j} - \psi_{i-1,j})\frac{u_{i,j} - u_{i-1,j}}{h}\right)\\
&+\frac{1}{k}\left(e^{\psi_{i,j+1}}B(\psi_{i,j+1} - \psi_{i,j})\frac{u_{i,j+1} - u_{i,j}}{k}\right.\\
&\qquad\left. - e^{\psi_{i,j}}B(\psi_{i,j} - \psi_{i,j-1})\frac{u_{i,j} - u_{i,j-1}}{k}\right) = f_{ij}
\end{aligned} \tag{7.7}$$

(see Figure 4 for notation). The scheme (7.7) is widely employed in semiconductor device simulation (see, e.g., Selberherr [12], Markowich [7]). Originally, its one-dimensional analogue was obtained by Scharfetter and Gummel [11], who developed the idea of approximating the current density by a constant on each mesh interval.

The finite element schemes presented in this paper represent a straightforward generalization of Scharfetter-Gummel type difference schemes. The "nice" properties of the coefficient matrix carry over to the stiffness matrix if the mesh is of acute type (see Theorem 4.2), and the convergence performance only depends on the resolution of the current density and of the inhomogeneity by the mesh.

Institut für Angewandte und Numerische Mathematik
Technische Universität Wien
Wiedner Hauptstrasse 6-10
A-1040 Wien, Austria
E-mail: Ell5N.5@AWITUW.1 EARN

Technical University
Obrancu Miru 21
60200 Brno, Czechoslovakia

1. R. A. ADAMS, *Sobolev Spaces*, Academic Press, New York, 1975.

2. I. BABUŠKA & J. E. OSBORN, "Generalized finite element methods: their performance and their relation to mixed methods," *SIAM J. Numer. Anal.*, v. 20, 1983, pp. 510–536.

3. I. BABUŠKA & J. E. OSBORN, *Finite Element Methods for the Solution of Problems with Rough Data*, Lecture Notes in Math., Vol. 1121, Springer-Verlag, Berlin and New York, 1985, pp. 1–18.

4. P. G. CIARLET, *The Finite Element Method for Elliptic Problems*. North-Holland, Amsterdam, 1978.

5. P. GRISVARD, "Behavior of the solutions of an elliptic boundary value problem in a polygonal or polyhedral domain," in *Numerical Solution of Partial Differential Equations* III (B. Hubbard, ed.), Academic Press, New York, 1976, pp. 207–274.

6. B. KAWOHL, *Über nichtlineare gemischte Randwertprobleme für elliptische Differentialgleichungen zweiter Ordnung auf Gebieten mit Ecken*, Dissertation, TH Darmstadt, BRD, 1978.

7. P. A. MARKOWICH, *The Stationary Semiconductor Device Equations*, Springer-Verlag, Wien and New York, 1986.

8. M. S. MOCK, "Analysis of a discretization algorithm for stationary continuity equations in semiconductor device models. I," *Compel*, v. 2, 1983, pp. 117–139.

9. M. S. MOCK, "Analysis of a discretization algorithm for stationary continuity equations in semiconductor models. II," *Compel*, v. 3, 1984, pp. 137–149.

10. J. NEDOMA, "The finite element solution of parabolic equations," Apl. Mat., v. 23, 1978, pp. 408–438.

11. D. L. SCHARFETTER & H. K. GUMMEL, "Large signal analysis of a silicon read diode oscillator," *IEEE Trans. Electron Devices*, v. ED-16, 1969, pp. 64–77.

12. S. SELBERHERR, *Analysis and Simulation of Semiconductor Devices*, Springer-Verlag, Wien and New York, 1984.

13. G. STRANG & G. J. FIX, *An Analysis of the Finite Element Method*, Prentice-Hall, Englewood Cliffs, N. J., 1973.

14. W. V. VAN ROOSBROECK, "Theory of flow of electrons and holes in germanium and other semiconductors," *Bell Syst. Techn. J.*, v. 29, 1950, pp. 560–607.

15. R. S. VARGA, *Matrix Iterative Analysis*, Prentice-Hall, Englewood Cliffs, N. J., 1962.

16. O. C. ZIENKIEWICZ, *The Finite Element Method*, McGraw-Hill, London, 1977.

17. M. A. ZLÁMAL, "Finite element solution of the fundamental equations of semiconductor devices. II," submitted for publication, 1985.

18. M. A. ZLÁMAL, "Finite element solution of the fundamental equations of semiconductor devices. I," *Math. Comp.*, v. 46, 1986, pp. 27–43.