Journal of Computational Physics **156,** 241–256 (1999)
Article ID jcph.1999.6358, available online at http://www.idealibrary.com on IDEAL®

An Optimization Approach to a Finite Dimensional Parameter Estimation Problem in Semiconductor Device Design

W. R. Lee,* S. Wang,† and K. L. Teo‡

*School of Mathematics and Statistics, Curtin University of Technology, GPO Box U 1987, Perth 6845, Australia; †Department of Mathematics and Statistics, The University of Western Australia, Nedlands, Western Australia 6907, Australia; ‡Department of Applied Mathematics, The Hong Kong Polytechnic University, Kowloon, Hong Kong, China

Received August 12, 1998; revised August 12, 1999

In this paper the parameter selection in semiconductor device design is posed as an optimization problem: given an ideal voltage-current (V-I) characteristic, find one or more physical and geometrical parameters such that the V-I characteristic of the device matches the ideal one optimally with respect to a prescribed performance criterion. The voltage-current characteristic of a semiconductor device is governed by a set of nonlinear partial differential equations (PDE), and thus a black-box approach is taken for the numerical solution of the PDEs. Various existing numerical methods are proposed for the solution of the nonlinear optimization problem. The Jacobian of the cost function is ill-conditioned and a scaling technique is thus proposed to stabilize the resulting linear system. Numerical experiments, performed to show the usefulness of this approach, demonstrate that the approach always gives optimal or near-optimal solutions to the test problems in both two and three dimensions.

1. INTRODUCTION

The electrical behaviour of a semiconductor device is governed by the following system of (scaled) nonlinear second-order elliptic equations [14, 15]

$$\nabla^2\psi - n + p = -N, \tag{1.1}$$

$$\nabla \cdot \boldsymbol{J}_n - R(\psi, n, p) = 0, \tag{1.2}$$

$$\nabla \cdot \boldsymbol{J}_p + R(\psi, n, p) = 0, \tag{1.3}$$

in $\Omega \subset \mathbb{R}^m\,(m = 2, 3)$ with appropriate boundary conditions, where the current densities $\boldsymbol{J}_n$

0021-9991/99 $30.00

and $\boldsymbol{J}_p$ are defined respectively by

$$\boldsymbol{J}_n = \nabla n - n\nabla\psi,$$
$$\boldsymbol{J}_p = -(\nabla p + p\nabla\psi).$$

Here ψ is the electrostatic potential, n is the electron concentration, p is the hole concentration, N denotes the doping function, and R denotes the recombination/generation rate which is assumed to be monotone with respect to n and p. One example is the (scaled) Shockley–Read–Hall type recombination (cf. [14, Chap. 4]) given by

$$R(n, p) = \frac{np - 1}{\tau_n(p + 1) + \tau_p(n + 1)},$$

where τ_n and τ_p are positive constants. We assume that the boundary $\partial\Omega$ of the device region Ω is polygonal or polyhedral, and let $\partial\Omega_D \subset \partial\Omega$ denote the union of all ohmic contacts (i.e., terminals of a device) and $\partial\Omega_N$ the part of boundary such that $\overline{\partial\Omega}_D \cup \overline{\partial\Omega}_N = \partial\Omega$. On $\partial\Omega_D$ the values of ψ, n, and p are given as functions of the applied bias V and on $\partial\Omega_N$ the normal derivatives of ψ, n, and p vanish because it is normally insulated. The current flowing in or out of a terminal $c \in \partial\Omega_D$ is given by

$$I = \int_c (\boldsymbol{J}_n + \boldsymbol{J}_p) \cdot \boldsymbol{n}\, ds, \tag{1.4}$$

where $\boldsymbol{n}$ denotes the outward unit normal vector of $\partial\Omega_D$. A typical 2D p-n diode with two ohmic contacts is depicted in Fig. 1 where the interior curve, called the p-n junction, represents the interface of the p and n regions. We denote this curve by C. For simplicity, we assume, in the rest of this paper, that the doping function N is a step function of the form

$$N = \begin{cases} a & \text{in } p\text{-region} \\ -b & \text{in } n\text{-region,} \end{cases} \tag{1.5}$$

where a and b are constants in the range from 10^{10} to 10^{20}. It is expected that all the techniques presented here are also applicable to the case of continuous doping profiles,

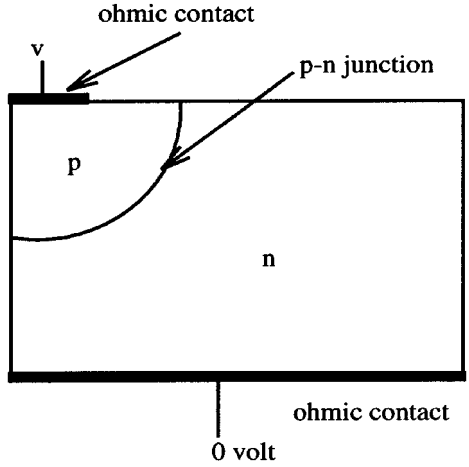

FIG. 1. A typical two-dimensional diode.

though the latter case is not discussed in the present paper. Note that N is also a function of the p-n junction C. Given N, the dimension of a device, and the applied bias V, we can solve (1.1)–(1.3) numerically using an appropriate numerical method and then evaluate the terminal current $I(V)$ flowing in or out of the device using (1.4) (cf. [10, 11]). The problem can be solved for various biases V to obtain the V-I characteristic for a given device.

In the conventional design cycle of a device, the above process will be repeated using different doping functions N and different geometries until the V-I characteristic matches the required one within a given error range. This approach is time-consuming, because after each iteration in the design cycle, the new values of the parameters to be used in the next iteration are chosen empirically from previous experience which may be far away from the optimal choice. In this paper the problem is posed as the following optimization problem: given an ideal V-I characteristic curve $I_g(v)$, find some physical and/or geometric parameters such that the V-I characteristic of the device matches the ideal one optimally with respect to a specified performance criterion. This problem is formulated as a nonlinear optimization problem, and the cost function of the problem consists of two competing quadratic terms with penalty parameters. By a judicious choice of these parameters, one can balance the competing costs. Various existing efficient numerical methods are proposed for the numerical solution of this nonlinear optimization problem. Due to the large variations in the magnitude of parameters and the solutions to the semiconductor device equations, the Jacobian of the optimization problem is ill-conditioned. To overcome this difficulty, a scaling technique is proposed to balance the entries of the Jacobian so that the problem is numerically more stable. This approach is applied to some test problems and all the numerical results demonstrate the usefulness of the approach. To our best knowledge, little work on this approach to semiconductor device design can be found in the literature (cf. [6] and the references therein), though, in practice, it will dramatically reduce the time required in the design cycle of a semiconductor device.

2. THE FORMULATION OF THE PROBLEM

We consider the formulation of the semiconductor device parameter design. For simplicity we assume hereafter that the geometry of a device is rectangular or brick, and thus the device region is $\bar{\Omega} = [0, L_x] \times [0, L_y]$ in two dimensions or $\bar{\Omega} = [0, L_x] \times [0, L_y] \times [0, L_z]$ in three dimensions. We also assume that the p-n junction of the device consists of line segments or facets parallel to one of the axes. In this case, the p-n junction is uniquely determined by its intercept on each of the axes in the interval $[0, L_x]$, $[0, L_y]$, or $[0, L_z]$. Let c_x denote the ratio of the the intercept on the x-axis and L_x, c_y the ratio of the intercept on the y-axis and L_y, and c_z the ratio of the intercept on the z-axis and L_z. Then, the p-n junction is uniquely determined by c_x, c_y, and c_z. Using this notation, we formulate the above parameter selection problem in semiconductor design as the following optimization problem:

Problem 1. Given an ideal V-I characteristic function $I_g(v)$ on $[0, V_{\max}]$, find the doping parameters a, b, and the geometrical parameters L_x, L_y, L_z, c_x, c_y, and c_z such that

$$F(a, b, c_x, c_y, c_z, L_x, L_y, L_z) = \alpha \int_0^{V_{\max}} (I(v) - I_g(v))^2 \, dv + \beta\left(L_x^2 + L_y^2 + L_z^2\right)$$

is minimized subject to the bound constraints

$$10^{10} \le a \le 10^{20}, \tag{2.1}$$

$$10^{10} \le b \le 10^{20}, \tag{2.2}$$

$$c_x^{\min} \le c_x \le c_x^{\max}, \tag{2.3}$$

$$c_y^{\min} \le c_y \le c_y^{\max}, \tag{2.4}$$

$$c_z^{\min} \le c_z \le c_z^{\max}, \tag{2.5}$$

$$L_x^{\min} \le L_x \le L_x^{\max}, \tag{2.6}$$

$$L_y^{\min} \le L_y \le L_y^{\max}, \tag{2.7}$$

$$L_z^{\min} \le L_z \le L_z^{\max}, \tag{2.8}$$

where L_x, L_y, and L_z are the length, width, and height of the device, c_x, c_y, and c_z are the ratios defined before, α, β are two positive constants, and I is the terminal current.

This is a continuous least squares problem and the cost function F contains two competing performance criteria. This is because $L_x = L_y = L_z = 0$ is the obviously optimal solution for the second term of the cost. Thus, we need to choose α and β properly to balance the two terms in F. In practice, Problem 1 can be approximated by taking a set of appropriate sampling points v_i, $i = 1, 2, \ldots, m$, in $[0, V_{\max}]$, leading to

Problem 2. Given an ideal V-I characteristic function $I_g(v)$ on $[0, V_{\max}]$, find the doping parameters a, b, and the geometrical parameters c_x, c_u, c_z, L_x, L_y, and L_z such that

$$E(\theta) = (\boldsymbol{I} - \boldsymbol{I}_g)^T A(\boldsymbol{I} - \boldsymbol{I}_g) + \beta\left(L_x^2 + L_y^2 + L_z^2\right) \tag{2.9}$$

is minimized subject to (2.1)–(2.8), where

$$\begin{aligned} \boldsymbol{I} &= (I(v_1, \theta), I(v_2, \theta), \ldots, I(v_m, \theta)), \\ \boldsymbol{I}_g &= (I_g(v_1), I_g(v_2), \ldots, I_g(v_m)), \\ \theta &= (a, b, c_x, c_y, c_z, L_x, L_y, L_z) \end{aligned}$$

and $A = \operatorname{diag}(\alpha_i)$ is a diagonal matrix.

Here $\alpha_i > 0$ ($i = 1, 2, \ldots, m$) and $\beta > 0$ are weights to be chosen later.

Problem 2 is a nonlinear optimization problem with only bound constraints. The nonlinear differential equations (1.1)–(1.3) do not appear explicitly in the formulation, but the current $\boldsymbol{I}$ depends on these equations through the expression (1.4). The dependence of the cost function on the parameter θ and the applied bias v is complicated, and thus the solvability of Problem 2 is theoretically difficult. (Even the solvability of the nonlinear PDE system (1.1)–(1.3) is still an open problem unless under some restrictive assumptions (cf., for example, [9]).) However, from our computational experience, Problem 2 is computable, though local minima may exist, as will be seen in Section 4.

3. THE NUMERICAL METHODS

We now consider the numerical solution of Problem 2.

Starting from an initial guess θ_0, Problem 2 can be solved iteratively. At each step an increment $\delta\theta_i$ is calculated such that

$$E(\theta_i + \delta\theta_i)$$

is minimized with respect to $\delta\theta_i$, where θ_i and $\delta\theta_i$ are the ith approximation and ith increment of θ, respectively. The iterative procedure continues until the relative error $\|\boldsymbol{I} - \boldsymbol{I}_g\|_2 / \|\boldsymbol{I}_g\|_2$ and the change in (L_x, L_y, L_z) in the Euclidean norm $\|L\|_2$ between two consecutive iterations are smaller than a given tolerance.

The Gauss–Newton Method

To calculate the increment $\delta\theta_i$ at each step, the Gauss–Newton Method is used. Let

$$\begin{aligned}
\hat{\boldsymbol{I}} &= (I(v_1, \theta), I(v_2, \theta), \ldots, I(v_m, \theta), L_x, L_y, L_z)^T, \\
\hat{\boldsymbol{I}}_g &= (I_g(v_1), I_g(v_2), \ldots, I_g(v_m), 0, 0, 0)^T, \\
B &= \operatorname{diag}(\alpha_1, \ldots, \alpha_m, \beta, \beta, \beta).
\end{aligned}$$

Then

$$E(\theta) = (\hat{\boldsymbol{I}} - \hat{\boldsymbol{I}}_g)^T B (\hat{\boldsymbol{I}} - \hat{\boldsymbol{I}}_g).$$

Let $\hat{\boldsymbol{I}}^i = \hat{\boldsymbol{I}}(\theta_i)$. Then Taylor's formula for vector valued functions gives

$$\hat{\boldsymbol{I}} = \hat{\boldsymbol{I}}^i + J_i \delta\theta_i + \frac{1}{2}\left\{\delta\theta_i^T G_1 \delta\theta_i, \ldots, \delta\theta_i^T G_{m+3} \delta\theta_i\right\}^T, \tag{3.1}$$

where

$$J_i = \begin{pmatrix}
\frac{\partial \boldsymbol{I}(v_1)}{\partial a_i} & \frac{\partial \boldsymbol{I}(v_1)}{\partial b_i} & \frac{\partial \boldsymbol{I}(v_1)}{\partial c_{x_i}} & \frac{\partial \boldsymbol{I}(v_1)}{\partial c_{y_i}} & \frac{\partial \boldsymbol{I}(v_1)}{\partial c_{z_i}} & \frac{\partial \boldsymbol{I}(v_1)}{\partial L_{x_i}} & \frac{\partial \boldsymbol{I}(v_1)}{\partial L_{y_i}} & \frac{\partial \boldsymbol{I}(v_1)}{\partial L_{z_i}} \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
\frac{\partial \boldsymbol{I}(v_m)}{\partial a_i} & \frac{\partial \boldsymbol{I}(v_m)}{\partial b_i} & \frac{\partial \boldsymbol{I}(v_m)}{\partial c_{x_i}} & \frac{\partial \boldsymbol{I}(v_m)}{\partial c_{y_i}} & \frac{\partial \boldsymbol{I}(v_m)}{\partial c_{z_i}} & \frac{\partial \boldsymbol{I}(v_m)}{\partial L_{x_i}} & \frac{\partial \boldsymbol{I}(v_m)}{\partial L_{y_i}} & \frac{\partial \boldsymbol{I}(v_m)}{\partial L_{z_i}} \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1
\end{pmatrix},$$

and the matrix G_j denotes the Hessian of $\hat{I}_j^i$ evaluated at $\theta_i + r\delta\theta_i$ with $0 \le r \le 1$. Omitting the second order term in (3.1), we have

$$\hat{\boldsymbol{I}} = \hat{\boldsymbol{I}}^i + J_i \delta\theta_i$$

when $\delta\theta_i$ is small. Using this, $E(\theta_i + \delta\theta_i)$ can be approximated by

$$\begin{aligned}
E(\theta_i + \delta\theta_i) &= \left(\hat{\boldsymbol{I}}^i + J_i\delta\theta_i - \hat{\boldsymbol{I}}_g\right)^T B\left(\hat{\boldsymbol{I}}^i + J_i\delta\theta_i - \hat{\boldsymbol{I}}_g\right) \\
&= \left(\hat{\boldsymbol{I}}^i - \hat{\boldsymbol{I}}_g\right)^T B\left(\hat{\boldsymbol{I}}^i - \hat{\boldsymbol{I}}_g\right) + \left(\hat{\boldsymbol{I}}^i - \hat{\boldsymbol{I}}_g\right)^T B J_i \delta\theta_i \\
&\quad + (J_i\delta\theta_i)^T B(\hat{\boldsymbol{I}}^i - \hat{\boldsymbol{I}}_g) + (J_i\delta\theta_i)^T B(J_i\delta\theta_i).
\end{aligned}$$

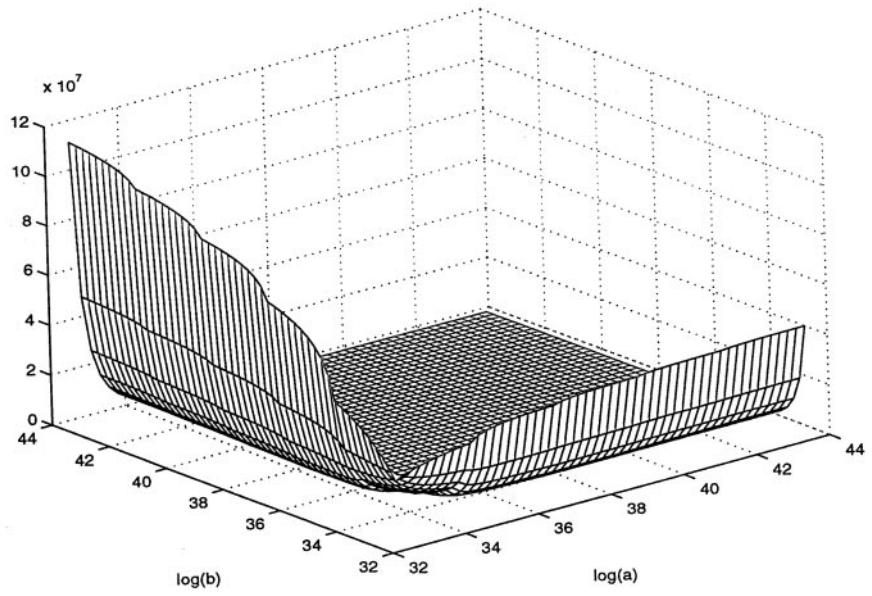

FIG. 2. The cost function against doping concentrations a and b.

This is a quadratic form in $\delta\theta_i$, and the minimum point $\delta\theta_i^*$ of this quadratic function satisfies

$$\nabla E(\theta_i + \delta\theta_i^*) = 0,$$

which leads to

$$\left(J_i^T B J_i\right)\delta\theta_i^* = -J_i^T B(\hat{\boldsymbol{I}} - \hat{\boldsymbol{I}}_g). \tag{3.2}$$

The solution to (3.2) defines the ith search direction called the Gauss–Newton direction.

The Levenberg–Marquardt Method

From computational studies we see that the partial derivatives of the cost function E with respect to the parameters a and b in the doping function are large when a and b are close to their lower bounds and small when a and b are close to their upper bound (see Fig. 2). Thus, initial guesses for a and b are always chosen to be their lower bound 10^{10}, and at the first few iterations, we also need to restrict the step size to avoid oscillations. Mathematically, this can be formulated as

$$\min E(\theta_i + \delta\theta_i)$$

$$\text{subject to } (\delta\theta_i)^T(\delta\theta_i) \le \Delta,$$

where Δ is a positive constant. The above inequality constraint can be added to the cost function to form

$$\min E(\theta_i + \delta\theta_i) + \lambda\left\{\Delta - (\delta\theta_i)^T(\delta\theta_i)\right\},$$

where $\lambda > 0$ is a penalty parameter called the Marquardt parameter. The optimal point $\delta\theta_i^*$

of this problem is given by

$$\left(J_i^T B J_i + \lambda I\right)\delta\theta_i^* = -J_i^T B(\hat{\boldsymbol{I}} - \hat{\boldsymbol{I}}_g),$$

where I denotes the unit matrix. This method is called the Levenberg–Marquardt Method (cf. [7, 8]) which is a combination of the Gauss–Newton Method and the Steepest Descent Method (cf. [13]). The Marquardt parameter λ can be chosen properly to avoid unbounded oscillation in the reestimation procedure. The same technique can also be applied to the decision variables a and b in the original cost function (2.9). This yields a cost function

$$\hat{E}(\theta) = (\boldsymbol{I} - \boldsymbol{I}_g)^T A(\boldsymbol{I} - \boldsymbol{I}_g) + \beta\left(L_x^2 + L_y^2 + L_z^2\right) + \gamma(a^2 + b^2)$$

corresponding to (2.9).

Let

$$\begin{aligned}
\hat{\hat{\boldsymbol{I}}} &= (I(v_1, \theta), I(v_2, \theta), \ldots, I(v_m, \theta), L_x, L_y, L_z, a, b)^T, \\
\hat{\boldsymbol{I}}_g &= (I_g(v_1), I_g(v_2), \ldots, I_g(v_m), 0, 0, 0, 0, 0)^T, \\
\hat{B} &= \operatorname{diag}(\alpha_1, \ldots, \alpha_m, \beta, \beta, \beta, \gamma, \gamma).
\end{aligned}$$

Then $\hat{E}(\theta)$ can be rewritten as

$$\hat{E}(\theta) = (\hat{\hat{\boldsymbol{I}}} - \hat{\hat{\boldsymbol{I}}}_g)^T \hat{B}(\hat{\hat{\boldsymbol{I}}} - \hat{\hat{\boldsymbol{I}}}_g).$$

The corresponding Levenberg–Marquardt correction $\delta\theta^*$ satisfies

$$\left(\hat{J}_i^T \hat{B} \hat{J}_i + \lambda I\right) \cdot \delta\theta_i^* = -\hat{J}_i^T \hat{B}(\hat{\hat{\boldsymbol{I}}} - \hat{\hat{\boldsymbol{I}}}_g), \tag{3.3}$$

where

$$\hat{J}_i = \begin{pmatrix}
\frac{\partial \boldsymbol{I}(v_1)}{\partial a_i} & \frac{\partial \boldsymbol{I}(v_1)}{\partial b_i} & \frac{\partial \boldsymbol{I}(v_1)}{\partial c_{x_i}} & \frac{\partial \boldsymbol{I}(v_1)}{\partial c_{y_i}} & \frac{\partial \boldsymbol{I}(v_1)}{\partial c_{z_i}} & \frac{\partial \boldsymbol{I}(v_1)}{\partial L_{x_i}} & \frac{\partial \boldsymbol{I}(v_1)}{\partial L_{y_i}} & \frac{\partial \boldsymbol{I}(v_1)}{\partial L_{z_i}} \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
\frac{\partial \boldsymbol{I}(v_m)}{\partial a_i} & \frac{\partial \boldsymbol{I}(v_m)}{\partial b_i} & \frac{\partial \boldsymbol{I}(v_m)}{\partial c_{x_i}} & \frac{\partial \boldsymbol{I}(v_m)}{\partial c_{y_i}} & \frac{\partial \boldsymbol{I}(v_m)}{\partial c_{z_i}} & \frac{\partial \boldsymbol{I}(v_m)}{\partial L_{x_i}} & \frac{\partial \boldsymbol{I}(v_m)}{\partial L_{y_i}} & \frac{\partial \boldsymbol{I}(v_m)}{\partial L_{z_i}} \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 & 0
\end{pmatrix}.$$

Using the correction $\delta\theta_i^*$, we update $\hat{\hat{\boldsymbol{I}}}$ by

$$\hat{\hat{\boldsymbol{I}}} = \hat{\hat{\boldsymbol{I}}}^i + \hat{J}_i \delta\theta_i^*. \tag{3.4}$$

The values of the elements in $\hat{J}_i$ may differ from each other by several orders of magnitude. This may cause some stability problems in real computation. To avoid this, we scale $\hat{J}_i$ by a

diagonal matrix M with positive diagonal entries so that $\hat{J}_i$ has more balanced entries. The scaled equation corresponding to (3.3) is

$$\left(\hat{\hat{J}}_i^T \hat{B} \hat{\hat{J}}_i + \lambda I\right)\delta\hat{\theta}_i^* = -\hat{\hat{J}}_i^T \hat{B}(\hat{\hat{\boldsymbol{I}}} - \hat{\hat{\boldsymbol{I}}}_g),$$

where

$$\hat{\hat{J}}_i = \hat{J}_i M \qquad \text{and} \qquad \delta\hat{\theta}_i^* = M^{-1}\delta\theta_i^*. \tag{3.5}$$

The updating formula corresponding to (3.4) then becomes

$$\hat{\hat{\boldsymbol{I}}} = \hat{\hat{\boldsymbol{I}}}^i + \hat{\hat{J}}_i \delta\hat{\theta}_i^*.$$

We comment that all the partial derivatives in $\hat{J}_i$ are approximated by forward finite differences. More specically,

$$\frac{\partial \boldsymbol{I}(v_i, \theta)}{\partial \theta_j} \simeq \frac{\boldsymbol{I}(v_i, \theta + h \cdot \boldsymbol{e}_j) - \boldsymbol{I}(v_i, \theta)}{h}, \tag{3.6}$$

where $\boldsymbol{e}_j$ is a unit vector and h is a small positive increment.

We also comment that the above method is based on the assumption that the independent variable, $\boldsymbol{v} = \{v_1, v_2, \ldots, v_m\}$, does not contain any observation errors. This is because it is normally not from an experimental observation. In the case that $\boldsymbol{v}$ does contain observation errors, the Orthogonal Distance Regression Method may be used, instead of the above Gauss–Newton or Levenberg–Marquardt Method. For details of the Orthogonal Distance Regression Method, we refer to [1].

Gummel's Method

The solvability of the nonlinear system (1.1)–(1.3) in general is a long-standing open problem, but in the case that the $R = 0$ and the applied bias is close to zero, it can be shown that the system is unique solvable (cf., for example, [9]). In practice, this nonlinear system can be solved iteratively by Gummel's method [5] defined as

1. Given an initial value (ψ^0, n^0, p^0) let $k = 0$.
2. Solve the following system sequentially for $(\psi^{k+1}, n^{k+1}, p^{k+1})$

$$\begin{aligned}
\nabla^2\psi^{k+1} &= n^k - p^k - N,\\
\nabla\cdot(\nabla n^{k+1} - n^{k+1}\nabla\psi^{k+1}) - R(\psi^{k+1}, n^k, p^k) &= 0,\\
\nabla\cdot(\nabla p^{k+1} + p^{k+1}\nabla\psi^{k+1}) - R(\psi^{k+1}, n^{k+1}, p^k) &= 0,
\end{aligned}$$

with appropriate boundary conditions.

3. Test for convergence. If failed, increase k and repeat step 2.

In Step 2 of the above algorithm we deal with equations of the form

$$-\nabla\cdot(\nabla u - \boldsymbol{c}u) + Gu = F \qquad \text{in } \Omega \tag{3.7}$$

$$u\big|_{\partial\Omega_D} = 0, (\nabla u - \boldsymbol{c}u)\cdot\boldsymbol{n}|_{\partial\Omega_N} = 0. \tag{3.8}$$

This problem can be solved effectively by the exponentially fitted finite volume method discussed in [10, 11, 12]. We now give a brief account of this method in two dimensions.

The Exponentially Fitted Finite Volume Scheme

To discuss the exponential fitted finite volume scheme (cf. [10, 11, 12]), we first define some meshes on Ω. Let T be any partition of $\bar{\Omega}$ by a set of triangles. Let $X = \{x_i\}_1^N$ be the set of all vertices of T and $E = \{e_i\}_1^M$ the set of edges of T. Without loss of generality we assume that the nodes in X and the edges in E are numbered such that $X' = \{x_i\}_1^{N'}$ and $E' = \{e_i\}_1^{M'}$ are respectively the set of nodes in X not on $\partial\Omega_D$ and the set of edges in E not on $\partial\Omega_D$.

DEFINITION 3.1. T is a Delaunay mesh if, for every $t \in T$, the circumcircle of the element contains no other vertices in X (cf. [2]).

We assume henceforth that T is a Delaunay triangulation.

DEFINITION 3.2. The Dirichlet tessellation D, corresponding to the triangulation T is defined by $D = \{d_i\}_1^N$ where the tile

$$d_i = \{x \in \Omega : |x - x_i| < |x - x_j|, x_j \in X, j \neq i\}$$

for all $x_i \in X$ (cf. [3]).

We remark that for each $x_i \in X$, the boundary ∂d_i of the tile d_i is the polygon having as its vertices the circumcentres of all triangles with common vertex x_i. Each segment of ∂d_i is perpendicular to one of the edges sharing the vertex x_i (see Fig. 3). The Dirichlet tessellation D is a polygonal mesh dual to the Delaunay mesh T. For each $i = 1, 2, \ldots, N'$, integrating (3.7)–(3.8) over d_i and applying Green's formula to the first term we have

$$-\int_{\partial d_i} (\nabla u - \boldsymbol{c}u) \cdot \boldsymbol{n}\, ds + \int_{d_i} Gu\, d\Omega = \int_{d_i} F\, d\Omega.$$

For $i = 1, 2, \ldots, N'$, let u_i be the approximate value of $u(x)$ at x_i. Using the one-point

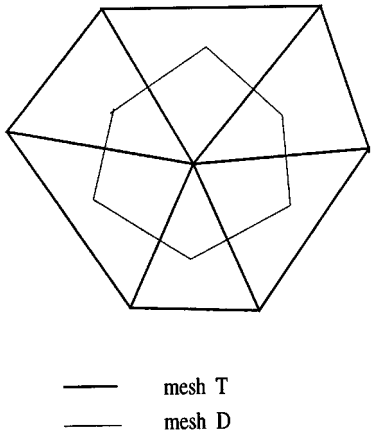

FIG. 3. Part of a Delaunay mesh T and dual Dirichlet tessellation D.

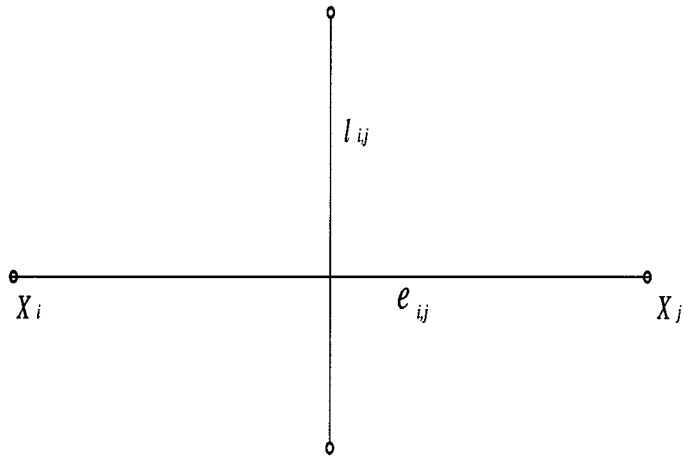

FIG. 4. Notation for edges and nodes.

quadrature rule we have from the above

$$-\int_{\partial d_i} (\nabla u - \boldsymbol{c}u)\cdot \boldsymbol{n}\, ds + G_i u_i |d_i| = F_i |d_i|, \tag{3.9}$$

where $G_i = G(x_i)$ and $F_i = F(x_i)$. We now consider the approximation of the first term in (3.9). Let $I_i = \{j : e_{i,j} \in E\}$ denote the index set of neighbouring nodes of x_i, where $e_{i,j}$ denotes the edge joining x_i and x_j, as shown in Fig. 4. Since ∂d_i is polygonal and each of its sides is perpendicular to one of the edges joining x_i, we have

$$\int_{\partial d_i} (\nabla u - \boldsymbol{c}u)\cdot \boldsymbol{n}\, ds = \sum_{j\in I_i} \left(\int_{l_{i,j}} (\nabla u - \boldsymbol{c}u)\cdot \boldsymbol{e}_{i,j}\, ds \right), \tag{3.10}$$

where $l_{i,j}$ denotes the segment of ∂d_i perpendicular to the edge $e_{i,j}$ and is oriented counterclockwise with respect to x_i (see Fig. 4) and $\boldsymbol{e}_{i,j}$ denotes the unit vector from x_i to x_j. For any $j \in I_i$ we now consider the two-point boundary value problem

$$\nabla(\nabla u \cdot \boldsymbol{e}_{i,j} - c_{i,j} u)\cdot \boldsymbol{e}_{i,j} = 0 \qquad \text{on } e_{i,j} \tag{3.11}$$

$$u(x_i) = u_i, \qquad u(x_j) = u_j, \tag{3.12}$$

where $c_{i,j}$ is a constant approximation to $\boldsymbol{c}\cdot \boldsymbol{e}_{i,j}$ on $e_{i,j}$. Solving this equation analytically we obtain

$$f_{i,j} \equiv \nabla u \cdot \boldsymbol{e}_{i,j} - c_{i,j} u = \frac{1}{|e_{i,j}|}(B(c_{i,j}|e_{i,j}|)u_j - B(-c_{i,j}|e_{i,j}|)u_i), \tag{3.13}$$

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

$$B(x) = \begin{cases} \frac{x}{e^x - 1} & \text{if } x \neq 0, \\ 1 & \text{if } x = 0. \end{cases} \tag{3.14}$$

Obviously $f_{i,j}$ defines a constant approximation to the integrand on the right side of (3.11). Furthermore the solution of (3.10) also defines a piecewise exponential approximation to the solution of (3.7) on $e_{i,j}$. Substituting (3.13) into (3.10) and the result into (3.9) we obtain

$$\sum_{j\in I_i} \frac{|l_{i,j}|}{|e_{i,j}\|d_{i,j}|}(B(-c_{i,j}|e_{i,j}|)u_i - B(c_{i,j}|e_{i,j}|u_j) + G_i u_j = F_i, \tag{3.15}$$

for all $i = 1, 2, \ldots, N'$. In matrix form, we have

$$(E + D)\boldsymbol{U} = \boldsymbol{F},$$

where E and D denote the matrices corresponding respectively to the first and second terms of (3.15), $\boldsymbol{U} = (u_1, u_2, \ldots, u_{N'})^T$ and $\boldsymbol{F} = (f_1, f_2, \ldots, f_{N'})^T$. The matrix $E + D$ is unsymmetric unless $c_{i,j} = 0$ for all $i = 1, 2, \ldots, N'$ and all $j \in I_i$. However, it is easy to verify that E is an M-matrix (cf. [11]).

We comment that the above method can be extended to three dimensions easily. A detailed discussion of this can be found, for example, in [4]. We also comment that using the numerical solution from the above discretization scheme, we can evaluate the currents flowing in or out of a device, based on (1.4). For detailed discussions of this, we refer to [10, 11, 4].

4. NUMERICAL EXPERIMENTS

The numerical methods described in the previous section are applied to some two- and three-dimensional test problems. All computations were performed in Fortran 77 double precision on a Unix Workstation. In what follows, we use *objective characteristic* to denote the discrete I-V characteristic generated by directly solving Eqs. (1.1)–(1.3) using a given parameter set. This given parameter set is referred to as *ideal solution*, and the solution to Problem 2 using the objective characteristic is referred to as *optimal solution*. In all the examples below, the bounds in the constraints (2.6)–(2.8) are chosen to be

$$L_x^{\min} = L_y^{\min} = L_z^{\min} = 2\,\mu\text{m}, \qquad L_x^{\max} = L_y^{\max} = L_z^{\max} = 15\,\mu\text{m}.$$

Also, the increment h in the finite difference approximation (3.6) is chosen to be 10^5 for the doping parameters a and b and 10^{-9} μm for L_x, L_y, and L_z. The units for the doping concentration parameters a and b and the geometric parameters L_x, L_y, and L_z are respectively $1/\text{cm}^2$ (or $1/\text{cm}^3$ in 3D) and μm. However, these are omitted below for brevity.

EXAMPLE 1. A Two-Dimensional Diode. A two-dimensional rectangular diode with width L_x and height L_y, depicted in Fig. 5, is chosen to be our first test problem. The p-region of the device is rectangular with width $\frac{3}{10}L_x$ and height $\frac{3}{10}L_y$, as shown in Fig. 5. (Correspondingly, $c_x = c_y = \frac{7}{10}$ in (2.3) and (2.4).) The parameters $\alpha_i (i = 1, \ldots, m)$, β, and γ are chosen to be 10^{26}, 10^{10}, and 10^{-18}, respectively, and the scaling matrix M in (3.5) is chosen to be diag $(10^5, 10^5, 10^{-9}, 10^{-9})$. This example is studied numerically in the following two different situations.

Case 1. Decision variables a, b, and $L_x = L_y$. In this case, we assume that the device is a square and the doping function is piecewise constant. Thus, we are to determine three

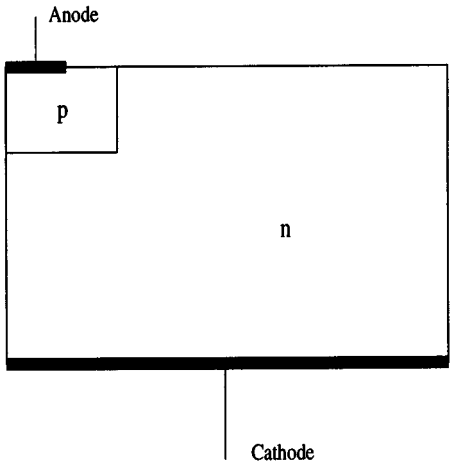

FIG. 5. A two-dimensional rectangular diode.

parameters: a, b in the doping function, and the dimension of the device $L_x = L_y$. Two different discrete V-I characteristics $I_{g,1}$ and $I_{g,2}$, listed in Tables I and II, respectively, are used as the objective characteristics. These objective characteristics are generated by solving Eqs. (1.1)–(1.3) using the parameter sets

$$\left(a = 10^{14}, b = 10^{14}, L_x = 10, L_y = 10\right)$$

and

$$\left(a = 10^{16}, b = 10^{16}, L_x = 10, L_y = 10\right)$$

respectively for the applied forward biases listed in Tables I and II.

To solve the optimization problems, we choose the following stopping criteria:

1. the relative error $\|\boldsymbol{I} - \boldsymbol{I}_g\|_2 / \|\boldsymbol{I}_g\|_2 \leq 10^{-4}$ (10^{-11} for case 1, using $I_{g,1}$ and 10^{-6} for Case 2, using $I_{g,1}$), and
2. the difference in $(L_x^2 + L_y^2)^{1/2}$ between two consecutive iterates is smaller than 10^{-3}.

The solution procedure stops when both of the above criteria are satisfied. The results obtained for various initial values are listed in Tables III and IV. From the tables we see that the relative error is always smaller than 10^{-4}. From the tables we also see that the optimal solutions are very accurate for $I_{g,1}$, and are reasonably close to the ideal solution for $I_{g,2}$.

Case 2. Decision variables a, b, L_x, and L_y. This case differs from Case 1 in the way that we do not assume that $L_x = L_y$. Thus, there are four independent parameters to be determined. The problem is solved using the objective characteristics listed in Tables I and II, the same criterion as in Case 1 is used in this case, and the results are listed in Tables V and VI.

TABLE I

The V-I Characteristic $I_{g,1}$

V	0.2143	0.2857	0.3571	0.4286	0.5
I	$6.4903 \cdot 10^{-11}$	$8.7707 \cdot 10^{-10}$	$1.1387 \cdot 10^{-8}$	$1.0729 \cdot 10^{-7}$	$4.8370 \cdot 10^{-7}$

TABLE II
The *V-I* Characteristic $I_{g,2}$

V	0.2143	0.2857	0.3571	0.4286	0.5
I	$4.1961 \cdot 10^{-13}$	$6.6184 \cdot 10^{-12}$	$1.0421 \cdot 10^{-10}$	$1.6397 \cdot 10^{-9}$	$2.5754 \cdot 10^{-8}$

TABLE III
The Results for Example 1, Case 1, Using *V-I* Characteristic $I_{g,1}$

Initial values				Optimal solution					
a	b	L_x	L_y	a	b	L_x	L_y	No. iter.	Rel. err.
1.0e10	1.0e10	15.0	15.0	1.0e14	1.0e14	10.0	10.0	23	5.89e-13
1.0e11	1.0e11	15.0	15.0	1.0e14	1.0e14	10.0	10.0	22	1.74e-13
1.0e12	1.0e12	15.0	15.0	1.0e14	1.0e14	10.0	10.0	24	8.82e-14
1.0e13	1.0e13	15.0	15.0	1.0e14	1.0e14	10.0	10.0	21	7.92e-13
1.0e12	1.0e12	8.0	8.0	1.0e14	1.0e14	10.0	10.0	19	8.69e-12

TABLE IV
The Results for Example 1, Case 1, Using *V-I* Characteristic $I_{g,2}$

Initial values				Optimal solution					
a	b	L_x	L_y	a	b	L_x	L_y	No. iter.	Rel. err.
1.0e12	1.0e12	15.0	15.0	9.75e15	1.03e16	9.39	9.39	57	5.50e-5
1.0e13	1.0e13	15.0	15.0	1.08e16	3.12e16	6.53	6.53	40	2.90e-5
1.0e11	1.0e11	15.0	15.0	5.27e15	7.12e16	11.8	11.8	65	1.51e-5
1.0e10	1.0e10	15.0	15.0	8.07e15	1.29e16	10.3	10.3	100	2.35e-5

TABLE V
The Results for Example 1, Case 2, Using *V-I* Characteristic $I_{g,1}$

Initial values				Optimal solution					
a	b	L_x	L_y	a	b	L_x	L_y	No. iter.	Rel. err.
1.0e12	1.0e12	15.0	15.0	1.0e14	1.0e14	10.0	10.0	22	1.59e-8
1.0e10	1.0e10	15.0	15.0	1.0e14	1.0e14	10.0	10.0	19	8.18e-7
1.0e10	1.0e10	15.0	5.0	1.0e14	1.0e14	10.0	10.0	40	2.55e-8
1.0e12	1.0e12	15.0	5.0	1.0e14	1.0e14	10.0	10.0	84	3.96e-8

TABLE VI
The Results for Example 1, Case 2, Using *V-I* Characteristic $I_{g,2}$

Initial values				Optimal solution					
a	b	L_x	L_y	a	b	L_x	L_y	No. iter.	Rel.err.
1.0e11	1.0e11	15.0	15.0	1.32e16	1.44e16	12.3	6.61	52	3.35e-5
1.0e12	1.0e12	15.0	15.0	1.42e16	1.38e16	12.2	6.40	53	5.53e-5
1.0e13	1.0e13	15.0	15.0	5.95e15	1.67e16	11.3	13.8	43	5.73e-5
1.0e12	1.0e12	5.0	5.0	1.77e16	2.25e16	12.5	3.79	40	9.61e-5
1.0e14	1.0e14	15.0	5.0	1.55e16	1.35e16	12.3	6.11	48	9.15e-5
1.0e10	1.0e10	15.0	5.0	1.85e16	1.07e16	13.7	7.41	46	6.93e-5

TABLE VII
The *V-I* Characteristic $I_{g,3}$

V	0.2143	0.2857	0.3571	0.4286	0.5
I	$3.2050 \cdot 10^{-11}$	$4.1269 \cdot 10^{-10}$	$5.2201 \cdot 10^{-9}$	$4.5746 \cdot 10^{-8}$	$1.8943 \cdot 10^{-7}$

EXAMPLE 2. A Three-Dimensional Diode. The second test problem is chosen to be a three-dimensional rectangular prismatic diode with width L_x, depth L_y, and height L_z. The p-region is also chosen to be a rectangular prism with width $\frac{3}{10}L_x$, depth $\frac{3}{10}L_y$, and height $\frac{3}{10}L_z$. (Correspondingly, the constraints (2.3)–(2.5) are replaced by $c_x = c_y = c_z = 7/10$.) For simplicity, we assume that $L_x = L_y$. Thus the decision variables are a, b in the doping function, and the dimensions of the device $L_x = L_y$ and L_z. The parameters α_i $(i = 1, \ldots, m)$, β, and γ are chosen to be 10^{26}, 10^{8}, and 10^{-18}, respectively, and the scaling matrix M in (3.5) is chosen to be diag$(10^5, 10^5, 10^{-9}, 10^{-9}, 10^{-9})$. Two different V-I Characteristics $I_{g,3}$ and $I_{g,4}$ listed in Tables VII and VIII, respectively, are used as the objective characteristics. These objective characteristics are generated by solving (1.1)–(1.3) using the parameter sets

$$\big(a = 10^{14}, b = 10^{14}, L_x = 10, L_y = 10, L_z = 10\big)$$

and

$$\big(a = 10^{16}, b = 10^{16}, L_x = 10, L_y = 10, L_z = 10\big),$$

respectively, for various applied forward biases. The solution procedure stops if the following items are satisfied:

1. relative error $\|\boldsymbol{I} - \boldsymbol{I}_g\|_2 / \|\boldsymbol{I}_g\|_2 \le 1.2 \times 10^{-4}$ (10^{-6} or the case using $I_{g,3}$) and
2. the difference in $(L_x^2 + L_y^2 + L_z^2)^{1/2}$ between two consecutive iterates is smaller than 10^{-3}.

Various initial values are used, and the results are listed in Tables IX and X, respectively. From the tables we see that the relative errors are always smaller than 1.2×10^{-4}. It is also seen that the optimal solutions are very accurate for $I_{g,3}$, and are reasonably close to the ideal solution for $I_{g,4}$.

We remark that the optimal solutions for $I_{g,2}$ and $I_{g,4}$ are harder to obtain than for $I_{g,1}$ and $I_{g,3}$, since the partial derivatives of the cost function E with respect to the doping parameter a and b are very small when the doping parameters are close to their upper bounds (cf. Fig. 2).

EXAMPLE 3. A Two-Dimensional Diode with a Variable p-n Junction. The third test problem is chosen to be a two-dimensional rectangular diode with a variable p-n junction. The configuration of the device is the same as that in Fig. 5, but the p-region has the width

TABLE VIII
The *V-I* Characteristic $I_{g,4}$

V	0.2143	0.2857	0.3571	0.4286	0.5
I	$2.1158 \cdot 10^{-13}$	$3.3410 \cdot 10^{-12}$	$5.2657 \cdot 10^{-11}$	$8.2842 \cdot 10^{-10}$	$1.2987 \cdot 10^{-8}$

TABLE IX
The Results for Example 2, Using Objective *V-I* Characteristic $I_{g,3}$

Initial values				Optimal solution					
a	b	$L_x = L_y$	L_z	a	b	$L_x = L_y$	L_z	No. iter.	Rel. err.
1.0e10	1.0e10	15.0	15.0	9.9e13	9.7e13	10.2	12.0	19	4.3e-7
1.0e12	1.0e12	15.0	15.0	1.0e14	1.0e14	10.0	10.0	24	6.6e-8
1.0e11	1.0e11	15.0	12.0	9.9e13	9.7e13	10.2	12.0	23	3.6e-7
1.0e12	1.0e12	15.0	10.0	1.0e14	1.0e14	10.0	10.0	19	2.2e-7
1.0e11	1.0e13	14.0	14.0	9.9e13	9.7e13	10.2	12.0	27	6.0e-7

$(1 - c_x)L_x$ and the height $(1 - c_y)L_y$ where c_x and c_y are the ratios defined in Section 2. For simplicity, we assume that $L_x = L_y$ and $c_x = c_y$. Thus, the decision variables are a, b in the doping function, the dimension parameter $L_x = L_y$, and the parameter for the p-n junction $c_x = c_y$. The penalty parameters α_i $(i = 1, \ldots, m)$, β, and γ are chosen to be 10^{26}, 10^{10}, and 10^{-18}, respectively, and the scaling matrix M in (3.5) is chosen to be diag$(10^5, 10^5, 10^{-9}, 4.0 \times 10^{-9})$. The bounds in (2.3) and (2.4) are chossen to be

$$c_x^{\min} = c_y^{\min} = 0.6 \qquad \text{and} \qquad c_x^{\max} = c_y^{\max} = 0.75.$$

The increment h in (3.6) for c_x is chosen to be 4×10^{-3}.

To solve the problem, we choose the V-I characteristic $I_{g,1}$, listed in Table I as the objective characteristic, which is generated by solving (1.1)–(1.3) using the parameter set

$$\left(a = 10^{14}, b = 10^{14}, L_x = 10, L_y = 10, c_x = 0.7, c_y = 0.7\right).$$

The solution procedure stops when both of the conditions

1. the relative error $\|\boldsymbol{I} - \boldsymbol{I}_g\|_2 / \|\boldsymbol{I}_g\|_2 \le 10^{-5}$, and
2. the difference in $(L_x^2 + L_y^2)^{1/2}$ between two consecutive iterates is smaller than 10^{-3}

are satisfied. Two initial guesses are used, and the results are listed in Table XI. From the table we see that the relative errors are always smaller than 6.0×10^{-6}.

5. CONCLUSION

In this paper we posed the parameter selection in semiconductor device design as an optimization problem with competing costs. Various existing efficient methods for

TABLE X
The Results for Example 2, Using Objective *V-I* Characteristic $I_{g,4}$

Initial values				Optimal solution					
a	b	$L_x = L_y$	L_z	a	b	$L_x = L_y$	L_z	No. iter.	Rel. err.
1.0e10	1.0e10	15.0	15.0	1.1e16	9.9e15	11.5	15.0	31	1.0e-4
1.0e10	1.0e10	15.0	10.0	1.1e16	7.7e15	10.0	9.81	30	7.9e-5
1.0e11	1.0e11	15.0	9.0	1.6e16	6.0e15	10.7	10.6	30	6.7e-5
1.0e11	1.0e11	14.0	14.0	1.2e16	6.3e15	10.2	14.9	30	3.6e-5
1.0e13	1.0e11	14.0	14.0	1.2e16	8.8e15	11.4	14.8	30	8.9e-5

TABLE XI
The Resuts for Example 3

Initial values			Optimal solution					
$a=b$	$L_x=L_y$	$c_x=c_y$	a	b	$L_x=L_y$	$c_x=c_y$	No. iter	Rel. err.
1.0e10	15.0	0.72	9.98e13	1.00e14	10.002	0.7008	25	5.6e-6
1.0e10	15.0	0.68	1.07e14	9.61e13	10.129	0.6773	22	5.5e-6

nonlinear optimization, nonlinear partial differential equations were discussed for the numerical solution of this problem. A scaling technique was also proposed to avoid numerical instability in computation. Numerical experiments for various model devices were performed and the numerical results showed the effectiveness of the optimization approach to the semiconductor device design.

ACKNOWLEDGMENT

The authors thank the anonymous referees for several observations and suggestions on the improvement of this paper.

REFERENCES

1. P. T. Boggs, R. H. Byrd, and R. B. Schnabel, A stable and efficient algorithm for nonlinear orthogonal distance regression, *SIAM J. Sci. Stat. Comput.* **8**(6), 1057(1987).
2. B. Delaunay, Sur la sphere vide, *Izv. Akad. Nauk SSSR Mat. Nat. Sci.* **6**, 793 (1934).
3. G. L. Dirichlet, Über die Reduction der positive quadratischen Formen mit drei unbestimmten ganzen Zahlen, *J. Reine Angew. Math.* **40**, 209 (1850).
4. C. J. Fitzsimons, J. J. H. Miller, S. Wang, and C. H. Wu, Hexahedral discretisations of the stationary semiconductor device equations, *Comp. Methods Appl. Mech. Eng.* **84**, 43 (1990).
5. H. K. Gummel, A self-consistent iterative scheme for one-dimensional steady state transistor calculation, *IEEE Elec. Dev.* **11**, 455 (1964).
6. N. Khalil, *ULSI Characterization with Technology Computer-Aided Design*, Ph.D. Dissertation, Technische Universität Wien, 1995.
7. K. Levenberg, A method for the solution of certain nonlinear problems in least squares, *Quart. Appl. Math.* **2**, 164 (1994).
8. D. W. Marquardt, An algorithm for least squares estimation of non-linear parameters, *SIAM J. Appl. Math.* **11**, 431 (1963).
9. P. A. Markowich, *The Stationary Semiconductor Device Equations* (Springer-Verlag, Wien/New York, 1986).
10. J. J. H. Miller and S. Wang, A triangular mixed finite element method for the stationary semiconductor device equations, *RAIRO Modél. Math. Anal. Numér*. **25**, (4), 441 (1991).
11. J. J. H. Miller and S. Wang, An analysis of the Scharfetter–Gummel box method for the stationary semiconductor device equations, *RAIRO Modél. Math. Anal. Numér*. **28**(2), 123 (1994).
12. J. J. H. Miller and S. Wang, An exponentially fitted finite volume method for the numerical solution of 2d incompressible flow problems, *J. Comput. Phys.* **115**, 56 (1994).
13. W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in FORTRAN: The Art of Scientific Computing*, 2nd ed. (Cambridge Univ. Press, New York, 1992).
14. S. Selberherr, *Analysis and Simulation of Semiconductor Devices* (Springer-Verlag, Wien/New York, 1984).
15. W. V. Van Roosbroeck, Theory of flow of electrons and holes in germanium and other semiconductors, *Bell Syst. Tech. J.* **29**, 560 (1950).