The Semiconductor Model Hierarchy in Optimal Dopant Profiling

Concetta R. Drago* René Pinnau†

September 26, 2006

Abstract. We consider optimal design problems for semiconductor devices which are simulated using the energy transport model. We develop a descent algorithm based on the adjoint calculus and present numerical results for a ballistic diode. Further, we compare the optimal doping profile with results computed on basis of the drift diffusion model. Finally, we exploit the model hierarchy and test the space mapping approach, especially the aggressive space mapping algorithm, for the design problem. This yields a significant reduction of numerical costs and programming effort.

Keywords: Optimal semiconductor design, adjoints, energy transport, drift diffusion, numerics, descent algorithm, aggressive space mapping.

AMS(MOS) subject classification: 35J50, 49J20, 49K20, 90C48.

1 Introduction

In the last four decades, the mathematical modeling and simulation of charge transport in semiconductors has become a thriving research area in applied mathematics. Semiconductor device modeling started in the early fifties when the Van Roosbroeck drift diffusion equations were formulated, which became the most popular model for the simulation of semiconductors [16, 12].

Due to the ongoing miniaturization of semiconductor devices, one comes closer to the limit of validity of the drift-diffusion equations and, in order to improve the physical description of the device, various others models have been derived. Meanwhile, there is a whole hierarchy of semiconductor models available ranging from microscopic models, like the Boltzmann–Poisson or the Wigner–Poisson model, to macroscopic models, like the energy

*Università di Catania, Dipartimento di Matematica e Informatica, Catania, Italy, email: *drago@dmi.unict.it*

†Technische Universität Kaiserslautern, D–67663 Kaiserslautern, Germany, email: *pinnau@mathematik.uni-kl.de*

transport, the hydrodynamic and the drift diffusion model (see [12] and the references therein). Here, we focus on the energy transport system. This model, unlike the drift diffusion one, that is based on the assumption of isothermal motion, takes into account also the thermal effects related to the electron flow through the semiconductor crystal.

Naturally, the ever increasing computing power and the development of fast simulations tools, has contributed to an increasing interest in the optimal design of semiconductor devices, which can nowadays be based on mathematical optimization. A major objective in the optimal design is to improve the current flow over some contacts by a slight modification of the device doping profile, which enters as a source term in the mathematical model for semiconductor devices and plays here the role of the design variable. The earliest approaches in the engineering literature were based on black box optimization tools or non-linear least-squares methods [13, 18]. They gave reasonable results, but the principal disadvantage of such an approach is the large computational cost due to the large number of the state equation solves.

Recently, the problem of optimal semiconductor doping profiling was embedded into the framework of optimal control of systems governed by partial differential equations [9, 10, 4, 3]. Using the adjoint calculus it was possible to derive fast algorithms for the solution of the overall design problem. First, only the drift diffusion system was used to model the semiconductor device. Meanwhile, it has been also extended to the energy transport model, which allows to handle variable carrier temperatures [6]. In this paper, we present the first numerical results for optimal semiconductor design based on the energy transport model. Further, we compare the new optimal designs with the ones obtained based on the drift diffusion model.

Finally, we suggest to exploit this classical model hierarchy to speed up the convergence of the optimization algorithms using the idea of space mapping, which was first introduced by *Bandler* [1]. The main idea is to combine the advantage of simple (*coarse*, less accurate, easy to evaluate) models, and more complex (*fine*, more accurate, expensive) models, to accelerate the numerical computation of a fine model optimum. Here, the coarse model will be given by the drift diffusion equations, while the energy transport model acts as the fine model. The advantage of this approach is that it allows for the optimization of the energy transport model without implementing the adjoint code for this fine model. Especially, from the industrial point of view this is a major advantage since it allows to combine easily adjoint optimization techniques with commercial device simulators.

The paper is organized as follows. In the remainder of this section, we present the energy transport model. The design problem is formulated in Section 3, where we also state the first–order optimality system. Numerical results for the optimization of a one–dimensional n^+–n–n^+ ballistic diode are given in Section 4 and compared with the corresponding ones obtained by the drift diffusion model. Finally, we present the space mapping approach to optimal semiconductor design in Section 5, where encouraging numerical results are presented. Concluding remarks are given in Section 6.

1.1 The Energy Transport Model

The dimensionless, stationary energy transport model for charge carriers in a semiconductor enclosed in a bounded domain $\Omega \subset \mathbb{R}^d$, $d = 1, 2, 3$, is given, in the unipolar case, by the following equations for the electron density n and the temperature T, coupled to the Poisson equation for the electrostatic potential V [12]:

$$\mathrm{div} J_1 = 0 \tag{1.1a}$$

$$\mathrm{div} J_2 = J_1 \cdot \nabla V + W(\mu, T) \tag{1.1b}$$

$$\lambda^2 \Delta V = n(\mu, T) - C(x) \tag{1.1c}$$

where J_1 is the carrier flux density, J_2 the energy flux density, W the energy production, μ the chemical potential, $C(x)$ the doping concentration and λ the Debye length. The energy relaxation term $W(\mu, T)$ satisfies $W(\mu, T)(T - T_L) \leq 0$, where T_L is the lattice temperature.

For sufficiently small deviation from the thermal equilibrium state, the general form of the constitutive equations are given by

$$J_1 = -L_{11} \left(\nabla \left(\frac{\mu}{T} \right) - \frac{\nabla V}{T} \right) - L_{12} \nabla \left(-\frac{1}{T} \right), \tag{1.2a}$$

$$J_2 = -L_{21} \left(\nabla \left(\frac{\mu}{T} \right) - \frac{\nabla V}{T} \right) - L_{22} \nabla \left(-\frac{1}{T} \right). \tag{1.2b}$$

The coefficients L_{ij} depend on μ and T. Moreover, the diffusion matrix $L = (L_{ij})$ is symmetric, as a consequence of the Onsanger reciprocity relations, and positive definite, due to the second law of thermodynamics [12]. Assuming the parabolic band approximation one has for the electron density $n(\mu, T) = T^{3/2} e^{\mu/T}$ and the previous constitutive equations can be written in terms of n, T and V as

$$\begin{aligned} J_1 &= -\left(\nabla n - \frac{n}{T} \nabla V \right) \\ J_2 &= -\frac{3}{2} \left(\nabla (nT) - n \nabla V \right) \end{aligned}$$

and the energy relaxation term is simply

$$W(\mu, T) = -\frac{3}{2} \frac{n(\mu, T)(T - T_L)}{\tau_w},$$

where $\tau_w = \tau_0 \mu_0 U_t / L^2$ is the scaled energy relaxation time.

Remark 1.1. The energy transport model with the above relations is called the *Chen model* [5]. The diffusion matrix in term of n and T reads

$$L = n \begin{pmatrix} 1 & \frac{3}{2} T \\ \frac{3}{2} T & \frac{15}{4} T^2 \end{pmatrix}.$$

In order to obtain symmetric equations let us introduce the entropy variables (cf. [12] and the references therein):

$$u_1 = \mu/T \quad \text{and} \quad u_2 = -1/T.$$

Hence, the constitutive equations can be written as

$$J_i = -L_{i1}(u,T)\,(\nabla u_1 + u_2 \nabla V) - L_{i2}(u,T)\nabla u_2, \tag{1.3}$$

where $u = (u_1, u_2)$.

On the other hand, let us observe that from the first equation of system (1.1), one gets

$$J_1 \cdot \nabla V = \mathrm{div}(J_1 V) - V \mathrm{div} J_1 = \mathrm{div}(J_1 V).$$

Then, the second equation of system (1.1) becomes

$$\mathrm{div}(J_2 - J_1 V) = W(n,T).$$

Now we define the dual entropy variables, or electro–chemical potentials, as

$$w_1 = u_1 + u_2 V \quad \text{and} \quad w_2 = u_2.$$

By elementary calculations, one gets the equivalence of equations (1.1), (1.3) with the following system

$$\mathrm{div} I_1 = 0, \tag{1.4a}$$

$$\mathrm{div} I_2 = Q(w,V), \tag{1.4b}$$

$$\lambda^2 \Delta V = N(w,V) - C(x), \tag{1.4c}$$

where

$$I_1 = -\sum_{k=1}^{2} D_{1k}(w,V)\nabla w_k, \quad I_2 = -\sum_{k=1}^{2} D_{2k}(w,V)\nabla w_k. \tag{1.4d}$$

Here, we have $w = (w_1, w_2)$ and

$$D_{11} = L_{11}, \quad D_{12} = D_{21} = L_{12} - V L_{11}, \quad D_{22} = L_{22} - 2V L_{21} + V^2 L_{11},$$

as well as

$$Q(w,V) = W(\mu,T) \quad \text{and} \quad N(w,V) = (-1/w_2)^{3/2} \mathrm{exp}(w_1 - w_2 V).$$

To get a well posed problem, system (1.4) has to be supplemented with appropriate boundary conditions. We assume that the boundary $\partial\Omega$ of the domain Ω splits into two disjoint parts Γ_D and Γ_N, where Γ_D models the Ohmic contacts of the device and Γ_N represents the insulating parts of the boundary. Let ν denote the unit outward normal vector along the boundary, we consider the following mixed boundary conditions

$$w_1 = w_{1D}, \;\; w_2 = w_{2D}, \;\; V = V_D \qquad \text{on } \Gamma_D, \tag{1.5a}$$

$$I_i \cdot \nu = \nabla V \cdot \nu = 0 \qquad i = 1,2 \qquad \text{on } \Gamma_N, \tag{1.5b}$$

where w_{1D}, w_{2D} and V_D are the $H^1(\Omega)$–extensions of fixed functions defined on Γ_D.

2 The Design Problem

Now we put our optimal semiconductor design problem into a mathematical framework. The design objective is to adjust the current I_1 at some given Ohmic contact $\Gamma_O \subset \Gamma_D$ via a small change of the reference doping profile $\overline{C}$. At the contact Γ_O we prescribe the desired current I_g and allow deviations, in some suitable norm, of the doping profile from $\overline{C}$ in order to achieve this current flow. In other words, we intend to minimize the cost functional

$$F(w,V,C) = \frac{1}{2}\left|\int_{\Gamma_O} I_1(w,V)\cdot d\nu - I_g\right|^2 + \frac{\gamma}{2}\int_\Omega |\nabla(C-\bar{C})|^2\, dx \tag{2.1}$$

under the constraint that (w,V) is a solution of the energy transport model (1.4) supplemented with (1.5). Note that the parameter $\gamma > 0$ allows to balance the effective cost. Altogether, this yields a constrained optimization problem in the framework of the mathematical theory for the control of systems governed by partial differential equations [11]. In order to get a solution to this problem we introduce the state $y \stackrel{\text{def}}{=} (w_1, w_2, V)$ and the space of states

$$Y = y_D + Y_0,$$

where $y_D \stackrel{\text{def}}{=} (w_D, V_D)$ denotes the boundary data introduced above, and $Y_0 = [H_0^1(\Omega\cup\Gamma_N)]^2 \times (H_0^1(\Omega\cup\Gamma_N)\cap L^\infty(\Omega))$ is equipped with the norm $\|y\|_{X_0} \stackrel{\text{def}}{=} \|w\|_{[H^1(\Omega)]^2} + \|V\|_{H^1(\Omega)} + \|V\|_{L^\infty(\Omega)}$. An admissible set of controls is given by

$$\mathcal{C} = \{C \in H^1(\Omega)\cap L^\infty(\Omega) : C = \bar{C} \text{ on } \Gamma_D\} \subset H^1(\Omega). \tag{2.2}$$

We rewrite the state equations (2.2) shortly as $e(y,C) = 0$. Further, we introduce $Z \stackrel{\text{def}}{=} [H^1(\Omega)]^3$. Then the nonlinear mapping $e : Y\times\mathcal{C} \to Z^*$ defined via

$$e(y,C) \stackrel{\text{def}}{=} \begin{pmatrix} \operatorname{div}\left(\sum_{k=1}^2 D_{1k}(w,V)\nabla w_k\right) \\ \operatorname{div}\left(\sum_{k=1}^2 D_{2k}(w,V)\nabla w_k\right) + Q(w,V) \\ \lambda^2\Delta V - N(w,V) + C(x) \end{pmatrix}, \tag{2.3}$$

is wellposed.
In [6] the existence of a minimizer $(y^*, C^*) \in Y\times\mathcal{C}$ to the minimization problem

$$\min_{Y\times\mathcal{C}} F(y,C) \qquad \text{s.t.} \qquad e(y,C) = 0 \tag{2.4}$$

is proved, i.e. we have

Theorem 2.1. *The constrained minimization problem* (2.4) *admits at least one solution* $(w^*, V^*, C^*) \in Y\times\mathcal{C}$.

3 The First–order Optimality System

In this section we briefly discuss the first–order optimality system which yields the basis for all optimization methods seeking at least a stationary point. Since we want to tackle a constrained optimization problem, we write the first–oder optimality system using the Lagrangian $\mathcal{L} : Y \times \mathcal{C} \times Z \to \mathbb{R}$ associated to problem (2.4) defined by

$$\mathcal{L}(y, C, \xi) \stackrel{\text{def}}{=} F(y, C) + \langle e(y, C), \xi \rangle_{Z^*, Z} ,$$

where ξ denotes the adjoint variable.

Remark 3.1. For the existence of a Lagrange multiplier associated to an optimal solution (y^*, C^*) of (2.4) it is sufficient that the operator $e'(y^*, C^*)$ is surjective. Note the equivalence

$$e'(y, C)[(v, \tilde{C})] = g \quad \text{in } Z^* \quad \Leftrightarrow \quad e_y(y, C)[v] = g - e_C(y, C)[\tilde{C}] \quad \text{in } Z^*.$$

For the energy transport model this does in general not hold, but one can ensure the bounded invertibility of $e'(y^*, C^*)$ near to the thermal equilibrium state [12]. This idea was used in [6] to prove the unique existence of adjoint states. Hence, at least for small current densities and small deviations of the temperature from the lattice temperature, there exists a unique Lagrange multiplier ξ^* such that together with an optimal solution (y^*, C^*) it fulfills the first–order optimality system

$$\mathcal{L}'(y^*, C^*, \xi^*) = 0. \tag{3.1}$$

In fact one can rewrite this equations in a more concise form [10]:

$$\begin{aligned} e(y^*, C^*) &= 0 \quad \text{in } Z^*, \\ e_y^*(y^*, C^*)\xi^* + F_y(y^*, C^*) &= 0 \quad \text{in } Y^*, \\ e_C(y^*, C^*)\xi^* + F_C(y^*, C^*) &= 0 \quad \text{in } \mathcal{C}^*. \end{aligned}$$

I.e., a critical point of the Lagrangian has to satisfy the state system (1.4) with boundary data given in (1.5), as well as the adjoint system

$$\begin{aligned} &\operatorname{div}(D_{11}(x, y)\nabla\xi^{w_1}) + \operatorname{div}(D_{21}(x, y)\nabla\xi^{w_2}) \\ &- \sum_{k=1}^{2} \left(\frac{\partial D_{1k}}{\partial w_1}\nabla w_k\right) \cdot \nabla\xi^{w_1} - \sum_{k=1}^{2} \left(\frac{\partial D_{2k}}{\partial w_1}\nabla w_k\right) \cdot \nabla\xi^{w_2} + \frac{\partial Q}{\partial w_1}\xi^{w_2} = \frac{\partial N}{\partial w_1}\xi^V, \end{aligned} \tag{3.2a}$$

$$\begin{aligned} &\operatorname{div}(D_{12}(x, y)\nabla\xi^{w_1}) + \operatorname{div}(D_{22}(x, y)\nabla\xi^{w_2}) \\ &- \sum_{k=1}^{2} \left(\frac{\partial D_{1k}}{\partial w_2}\nabla w_k\right) \cdot \nabla\xi^{w_1} - \sum_{k=1}^{2} \left(\frac{\partial D_{2k}}{\partial w_2}\nabla w_k\right) \cdot \nabla\xi^{w_2} + \frac{\partial Q}{\partial w_2}\xi^{w_2} = \frac{\partial N}{\partial w_2}\xi^V, \end{aligned} \tag{3.2b}$$

$$-\lambda^2\Delta\xi^V + \frac{\partial N}{\partial V}\xi^V = -\sum_{k=1}^{2} \left(\frac{\partial D_{1k}}{\partial V}\nabla w_k\right) \cdot \nabla\xi^{w_1} - \sum_{k=1}^{2} \left(\frac{\partial D_{2k}}{\partial V}\nabla w_k\right) \cdot \nabla\xi^{w_2} + \frac{\partial Q}{\partial V}\xi^{w_2} \tag{3.2c}$$

supplemented with appropriate boundary data. Further, we have the optimality condition

$$\gamma\Delta\left(C-\bar{C}\right)=\xi^V \quad \text{in } \Omega, \tag{3.3a}$$
$$C=\bar{C} \quad \text{on } \Gamma_D, \quad \nabla C\cdot\nu=\nabla\bar{C}\cdot\nu \quad \text{on } \Gamma_N. \tag{3.3b}$$

Remark 3.2. For our choice of the cost functional (2.1) we get the following boundary conditions

$$\nabla\xi^{w_1}\cdot\nu=\nabla\xi^{w_2}\cdot\nu=0 \qquad \text{on } \Gamma_N. \tag{3.4a}$$

$$\xi^{w_2}=0 \quad \text{on } \Gamma_D \quad \text{and} \quad \xi^{w_1}=\begin{cases} 0 & \text{on } \Gamma_D\backslash\Gamma_O, \\ \left(\int_{\Gamma_O} I_1 d\nu - I_g\right) & \text{on } \Gamma_O \end{cases} \tag{3.4b}$$

These might change for a different choice of F.

The first two equations of system (3.2) can be written in the simplified form

$$\operatorname{div}\left(-\sum_{k=1}^{2} D_{ki}(w,V)\nabla\xi^{w_k}\right)+\sum_{k=1}^{2}\mathbf{b}_{ki}\cdot\nabla\xi^{w_k}-\mathbf{c}_i\cdot\xi^w=-s_i\xi^V, \tag{3.5}$$

where $i=1,2$ and

$$\mathbf{b}_{ki}=\sum_{j=1}^{2}\frac{\partial D_{kj}}{\partial w_i}\nabla w_j, \quad \mathbf{c}_i=\left(0,\frac{\partial Q}{\partial w_i}\right), \quad s_i=\frac{\partial N}{\partial w_i}, \quad \xi^w=(\xi^{w_1},\xi^{w_2}).$$

Remark 3.3. The matrix (D_{ki}) is symmetric positive definite and there exists a $\delta=\delta(V)>0$ such that

$$\sum_{i,k=1}^{2} D_{ki}\xi_k\xi_i \geq \delta(V)|\xi|^2 \quad \text{for all } \xi\in\mathbb{R}^2.$$

Moreover, taking into account the $L^\infty(\Omega)$–bound on V, there exists some $\delta_0>0$ such that $\delta(V)\geq\delta_0$ (see [12]).

If we define

$$\mathbf{h}=\left(\sum_{k=1}^{2}\left(\frac{\partial D_{1k}}{\partial V}\nabla w_k\right),\sum_{k=1}^{2}\left(\frac{\partial D_{2k}}{\partial V}\nabla w_k\right)\right) \quad \text{and} \quad \mathbf{g}=(0,\frac{\partial Q}{\partial V}),$$

equation (3.2c) can be written as

$$-\lambda^2\Delta\xi^V+\frac{\partial N}{\partial V}\xi^V=-\mathbf{h}\cdot\nabla\xi^w+\mathbf{g}\cdot\xi^w, \tag{3.6}$$

where $\nabla\xi^w=(\nabla\xi^{w_1},\nabla\xi^{w_2})$.

4 Numerical Method and Results

An adequate and easy to implement numerical method for the solution of (2.4) is the following gradient algorithm.

Algorithm 1.

1. *Choose* $C_0 \in \mathcal{C}$.
2. *For* $k = 1, 2, \ldots$ *compute* $C_k = C_{k-1} - \alpha_k \hat{F}'(C_{k-1})$.

Here, $\hat{F}(C) \overset{\text{def}}{=} F(y(C), C)$ denotes the reduced cost functional, which can be introduced near to the thermal equilibrium state, and $\hat{F}'(C)$ is the Riesz representative of its first variation. The evaluation of

$$\hat{F}'(C) = F_C(y(C), C) + e_C^* \xi$$

requires the solution of the nonlinear state system (1.4) for y, as well as a solution of the linear adjoint system (3.2) for ξ, and finally a linear solve of a Poisson problem to get the correct Riesz representative.

Remark 4.1. The choice of the step–length parameters α_k is critical to ensure the convergence of this descent algorithm. The overall numerical performance of this algorithm relies on an appropriate choice of the step–size rule for α_k, since these methods require in general consecutive evaluations of the cost functional requiring additional solves of the nonlinear state system [15].

4.1 Numerical Optimal Designs

Now we are showing some numerical results for the optimal design of an one–dimensional n^+–n–n^+ ballistic silicon diode, which is a simple model for the channel of a MOS transistor [16]. The semiconductor domain is given by the interval $\Omega = (0, L)$, with $L > 0$. In the n^+-regions a maximal doping concentration of $C_m = 5 \cdot 10^{17}$ cm^{-3} is prescribed. In the n–channel the minimal doping density is $2 \cdot 10^{15}$ cm^{-3}. The length of the n^+–regions is 0.1μm, whereas the length of the channel region equals 0.4μm. The numerical values of the physical parameters are given in Table 1.

PLEASE provide the values of λ and τ_w

We solve the constrained optimization problem (2.4) using algorithm Algorithm 1. For the parameter γ we chose 10^{-3} and use the constant step–size $\alpha = 10^{-2}$. The iteration stops when the difference of two consecutive iterates is below some specified threshold.

The state system was discretized by a variant of the well–known exponentially fitted Scharfetter–Gummel scheme [17, 12]. The computations were performed on a uniform

Parameter	Physical meaning	Numerical value
q	elementary charge	$1.6 \cdot 10^{-19}$As
ε_s	permittivity constant	10^{-12}AsV^{-1}cm^{-1}
μ_0	(low field) mobility constant	$1.4 \cdot 10^3$cm^{2}V^{-1}s^{-1}
U_T	thermal voltage at $T_0 = 300$K	0.026V
τ_0	energy relaxation time	$0.4 \cdot 10^{-12}$s
L	length of the device	0.6μm

Table 1: Physical Parameters

grid of 301 points to be sure to have no grid effects, but the same results can be obtained with a coarser grid of 100 point. For the biasing voltage at the working point we chose $\bar{U} = 0.5, 1,$ or $3V$, respectively, and tried to gain an amplification of the current $\bar{I}$ by 50%, i.e. we set $I_g = \bar{I} \cdot 1.5$.

In Figure 4.1 we present the optimal doping profile for the working point $\bar{U} = 0.5$V as well as the reference doping $\bar{C}$. Note, that already a very slight change in doping profile, i.e. an increase of background charges in the channel, yields the desired amplified current as can be seen from the evolution of the observation. The overall performance of the algorithm is very promising, since already 35 gradient steps are sufficient to reach the optimum. Further, we depict in Figure 4.1 the densities, velocities and temperatures before and after the optimization. They almost coincide due to the very small change in the doping density. Nevertheless, we reach our objective as can be also seen from the given current–voltage characteristics (IVC).

Analogous results for the biasing voltages $\bar{U} = 1$ and $3V$ can be seen in Figure 4.2 and Figure 4.3, respectively. It is noteworthy, that in all three cases the optimized doping density yields a reduction of the electron temperature in the channel, which is responsible for the increased current.

4.2 Comparisons with Optimal Designs based on Drift Diffusion

As the adjoint based optimal dopant profiling using the standard drift diffusion model is meanwhile well understood [9, 4, 10, 3], naturally the interesting question arises, how the optimized doping profiles derived from the two different models compare. The standard drift diffusion model with constant mobilities and without generation–recombination rates can be derived from the energy transport model (1.1) in the special case of a constant electron temperature $T \equiv T_L$. Then the reduced system reads

$$\begin{aligned} &\mathrm{div} J_1 = 0, \quad J_1 = -\left(\nabla n - n\nabla V\right), \\ &\lambda^2 \Delta V = n - C(x). \end{aligned}$$

Again, we use an adjoint based descent algorithm to compute the optimal doping profile (for details we refer to [10]). For the biasing voltage at the working point we choose

here $\bar{U} = 1V$ and again we tried to gain an amplification of the current by 50%. The corresponding results can be found in Figure 4.4, where we depict the optimal doping profiles, the electron densities before and after the optimization, as well as the electron velocities and the current–voltage characteristics. Note that the optimal doping profile for the energy transport model is much more symmetric than the one for the drift diffusion model. Comparing the electron velocities, we see a strong change for the drift diffusion model while the one for the energy transport model is almost unchanged. We conclude that in the optimization of the latter case the increased current is gained due to a lower electron temperature, while in the drift diffusion case it is increased due to a higher electron velocity. Note, that than the well–known velocity overshoot of the drift diffusion model will be even more pronounced.

Naturally, the question arises if we can use the optimal design for the (simple) drift diffusion model also in the energy transport model, and vice versa. Here, we can give a positive answer as can be seen from Figure 4.5 and Figure 4.6. In both cases, we get a current–voltage characteristics which is in the spirit of our optimization objective. This is remarkable, since we are considering a drift diffusion model with constant mobility, which is only a crude model for a ballistic diode. But this result is very encouraging also from the application point of view, since it shows us that the design question considered is even robust under changes of the model equations.

5 Space Mapping Optimization

As we have seen in the preceding sections, optimal design of semiconductor devices is in general a difficult task, due to the high computational cost required for each forward solve of the state system. Next we want to combine the two different models, energy transport and drift diffusion, to speed up the convergence of the optimization routine. In many applications semiconductor design companies rely on a given device simulator, where the implementation of adjoints is not straight–forward. Hence, there is a strong need for optimization routines which combine the reliabilty of the commercial software package with the adjoint approach. The main idea is to combine the advantage of simple (*coarse*, less accurate, easy to evaluate) models, for which one can implement the adjoints more easily, and more complex (*fine*, more accurate, expensive) models, to accelerate the numerical computation of a fine–model optimum.

This leads to the so–called space mapping approach which was introduced by *Bandler* [1] in the field of microwave filter design. Although the space mapping technique has been mainly applied in electromagnetics, the underlying principles are quite general and suitable to be used also in other areas [7, 2].

Here, we apply the space mapping technique – to our knowledge – for the first time in the field of optimal semiconductor design. We present some results obtained by using the aggressive space mapping algorithm for the minimization problem (2.4). Especially, we use

the drift diffusion equations as the coarse model and the energy transport model as the fine one.

Let $\bar{C}$ denote the reference doping profile and I^* the target current. Further, let $(n, V) = c(\zeta)$ denote the solution of the (coarse) drift diffusion model, where $\zeta \in \mathcal{C}$ denotes the coarse model doping profile design variable and let

$$\phi(\zeta) = \phi(\Upsilon(c(\zeta)), I^*, \zeta, \bar{C}) = \frac{1}{2}\left|\int_{\Gamma_0} \Upsilon d\nu - I^*\right|^2 + \frac{\gamma}{2}\int_{\Omega} |\nabla(\zeta - \bar{C})|^2 dx$$

denote the coarse model cost functional. Moreover, let $(w_1, w_2, V) = f(C)$ denote the solution of the (fine) energy transport model, where $C \in \mathcal{C}$ denotes the fine model doping profile design variable and let

$$\Phi(C) = \Phi(J(f(C)), I^*, C, \bar{C}) = \frac{1}{2}\left|\int_{\Gamma_0} J d\nu - I^*\right|^2 + \frac{\gamma}{2}\int_{\Omega} |\nabla(C - \bar{C})|^2 dx$$

denote the fine model cost functional. Here, Υ and J denote the electron current densities given by the drift diffusion and the energy transport model, respectively.

In the foolowing we want to approximate the solution of the fine model by an appropriate solution of the coarse model for which we define the misalignment function

$$r(\zeta, C) = |\phi(\zeta) - \Phi(C)| .$$

For a given $C \in \mathcal{C}$ we look for $\zeta \in \mathcal{C}$ such that $r(\zeta, C)$ is minimal, i.e. we define the space mapping function

$$\begin{aligned} p : \mathcal{C} &\to \mathcal{C} \\ C &\mapsto p(C) = \operatorname{argmin}_{\zeta \in \mathcal{C}} r(\zeta, C). \end{aligned}$$

Remark 5.1. Clearly, p will not exist as a function, if the minimizer is not unique. For optimal semiconductor design, this cannot be expected in general (for a detailed discussion of this issue we refer to [10, 3]). But we can ensure the uniqueness at least locally, when we restrict the set of admissible controls.

Since we want to evaluate p only a few times, we assume $\Phi(C^*) \approx \phi(\zeta^*)$, such that

$$p(C^*) = \operatorname{argmin}_{\zeta \in \mathcal{C}} r(\zeta, C^*) \approx \zeta^*.$$

Hence, we first determine ζ^* and then solve for $p(C^*) = \zeta^*$. But in general it holds $p(C^*) \neq \zeta^*$, such that we solve instead for

$$C^* = \operatorname{argmin}_{C \in \mathcal{C}} \|p(C) - \zeta^*\| .$$

This is done iteratively and the space mapping p is updated using a Broyden–rank–1 update yielding the so–called ASM (aggressive space mapping) algorithm (for details we refer to [7]):

Algorithm 2.

1. *Evaluate* $C_0 = \zeta^* = argmin_{\zeta \in \mathcal{C}}\, \phi(\Upsilon, I^*, \zeta, \bar{C})$ *and let* B_0 *be the identity matrix.*
2. *while* $\|p(C_k) - \zeta^*\|_{L^2} / \|\zeta^*\|_{L^2} > tolerance$
   (a) *Evaluate the electron density current* J^* *by the fine model* $f(C_k)$
   (b) *Determine* $\zeta_k = p(C_k) = argmin_{\zeta \in \mathcal{C}}\, \phi(\Upsilon, J^*, \zeta, \bar{C})$
   (c) *Solve* $B_k h_k = -(p(C_k) - \zeta^*)$ *for* h_k
   (d) *Set* $C_{k+1} = C_k + h_k$
   (e) *Update* $B_{k+1} = B_k + \frac{(p(C_{k+1}) - z^*) h_k^T}{h_k^T h_k}$
   (f) *Set* $k \to k + 1$.

Remark 5.2. On each iteration level we need one evaluation of the fine model and one solve of the optimal control problem for the coarse model. I.e. it is sufficient to implement an adjoint code for the coarse model.

We tested the performance of the ASM algorithm for the same ballistic diode as in the previous section. Especially, we tried to achieve an amplification of the current by 50 % for different values of the applied voltage $\bar{U} = 0.26, 0.52, 1, 1.5V$. Note that for increasing biasing voltages the coarse and fine model yield different responses. The convergence history for the four test cases can be found in Figure 5.1. The results underline clearly the feasibility of this approach. Compared with the direct optimization approach presented in Section 4 we reduced the evaluation of the fine energy transport model by a factor 2 to 4. To be fair in this comparison we also have to consider the number of inner iterations used for the solution of the coarse optimization problem. These can be found Table 2 and we observe that the overall number of gradient steps is comparable; but note that the computation of the coarse model gradient is much cheaper than the fine model gradient.

Voltage	Step 1	Step 2	Step 3	Step 4	Step 5	Step 6
0.26	3	3	3	3	3	
0.52	3	3	3	3	3	
1	3	3	4	4	4	4
1.5	4	3	3	3	2	2

Table 2: Inner iterations in the ASM algorithm

6 Conclusion and future works

Optimal semiconductor design is a growing research field for applied mathematicians which poses several analytical and numerical challenges. This is due to the fact that one has

to choose the appropriate models for the specific semiconductor device and the correct numerical schemes for their solution. So far, the optimal design was mainly based on the drift diffusion model. Here, we presented for the first time numerical results for the energy transport model, which takes into account temperature and mobility effects. Further, we did a thorough comparison with results based on the drift diffusion model.

From the engineering point of view it is most interesting to exploit the available hierarchy of semiconductor models to speed up the optimization and to lower the programming costs. A possible approach is given by the so–called space mapping optimization, which we used here – to our knowledge – for the first time in optimal semiconductor design. The presented results are very promising for industrial applications.

Future work will focus on the numerical study of more sophisticated devices, especially two dimensional and bipolar ones, and a more detailed analytical investigation of the space mapping approach in the semiconductor design context.

Acknowledgments. This work was supported by European Network HYKE, funded by EC under contract HPRN-CT-2002-00282.

References

[1] J.W. Bandler, R.M. Biernacki, S.H. Chen, P.A. Grobelny and R.H. Hemmers. Space mapping technique for electromagnetic optimization. *IEEE Trans. Microwave Theory Tech.* 43:2874–2882, 1995.

[2] J.W. Bandler, S. Koziel and K. Madsen. Space mapping for engineering optimization. *SIAG/OPT Views-and-News*, 17(1):19–26, 2006.

[3] M. Burger, M. Hinze, R. Pinnau. Optimization models for semiconductor dopant profiling. *Submitted*, 2006.

[4] M. Burger, R. Pinnau. Fast optimal design for semiconductor devices. *SIAM J. Appl. Math.*, 64(1):108–126, 2003.

[5] D. Chen, E. Kan, U. Ravaioli, C. Shu, R. Dutton. An improved energy transport model including nonparabolicity and non-Maxwellian distribution effects. *IEEE Electr. Dev. Letters*, 13:26–28, 1992.

[6] C. R. Drago, A. M. Anile. An optimal Control approach for an Energy Transport Model in Semiconductor Design. *Proceedings SCEE 2004, Mathematics in Industry*, Springer Verlag (under press), 2004.

[7] D. Echeverría, P.W. Hemker. Space mapping and defect correction. *Comput. Methods Appl. Math.*, 5(2):107–136, 2005.

[8] H.K. Gummel. A self-consistent iterative scheme for one-dimensional steady state transistor calculations. *IEEE Trans. Elec. Dev.*, ED–11:455–465, 1964.

[9] M. Hinze, R. Pinnau. An optimal control approach to semiconductor design. Math. Mod. Meth. Appl. Sc., 12(1):89–107, 2002.

[10] M. Hinze, R. Pinnau. Mathematical Tools in Optimal Semiconductor Design. *to appear in TTSP*, 2006.

[11] K. Ito and K. Kunisch. Augmented Lagrangian-SQP-methods for nonlinear optimal control problems of tracking type. *SIAM J. Control and Optimization* 34:874–891, 1996.

[12] A. Jüngel. Quasi-hydrodynamic Semiconductor Equations. PNDEA, Birkhäuser, 2001.

[13] W. R. Lee, S. Wang, K. L. Teo. An optimization approach to a finite dimensional parameter estimation problem in semiconductor device design. *Journal of Computational Physics*, 156:241–256, 1999.

[14] J.L. Lions. Optimal Control of Systems Governed by Partial Differential Equations. Springer-Verlag, New York, 1971.

[15] D. G. Luenberger. *Linear and Nonlinear Programming.* Addison-Wesley, Reading, second edition, 1989.

[16] P.A. Markowich, C.A. Ringhofer, C. Schmeiser. *SemiconductorEquations*, Springer, Wien, New York, 1990.

[17] D.L. Scharfetter and H.K. Gummel. Large signal analysis of a silicon read diode oscillator. *IEEE Trans. Electr. Dev.*, 15:64–77, 1969.

[18] M. Stockinger, R. Strasser, R. Plasun, A. Wild, S. Selberherr. A qualitative study on optimized MOSFET doping profiles. *In Proocedings SISPAD 98 Conf.*, pp. 77–80, 1998.

[19] G. M. Troianello. Elliptic Differential Equations and Obstacle Problems. Plenum Press, New York, first edition, 1987.

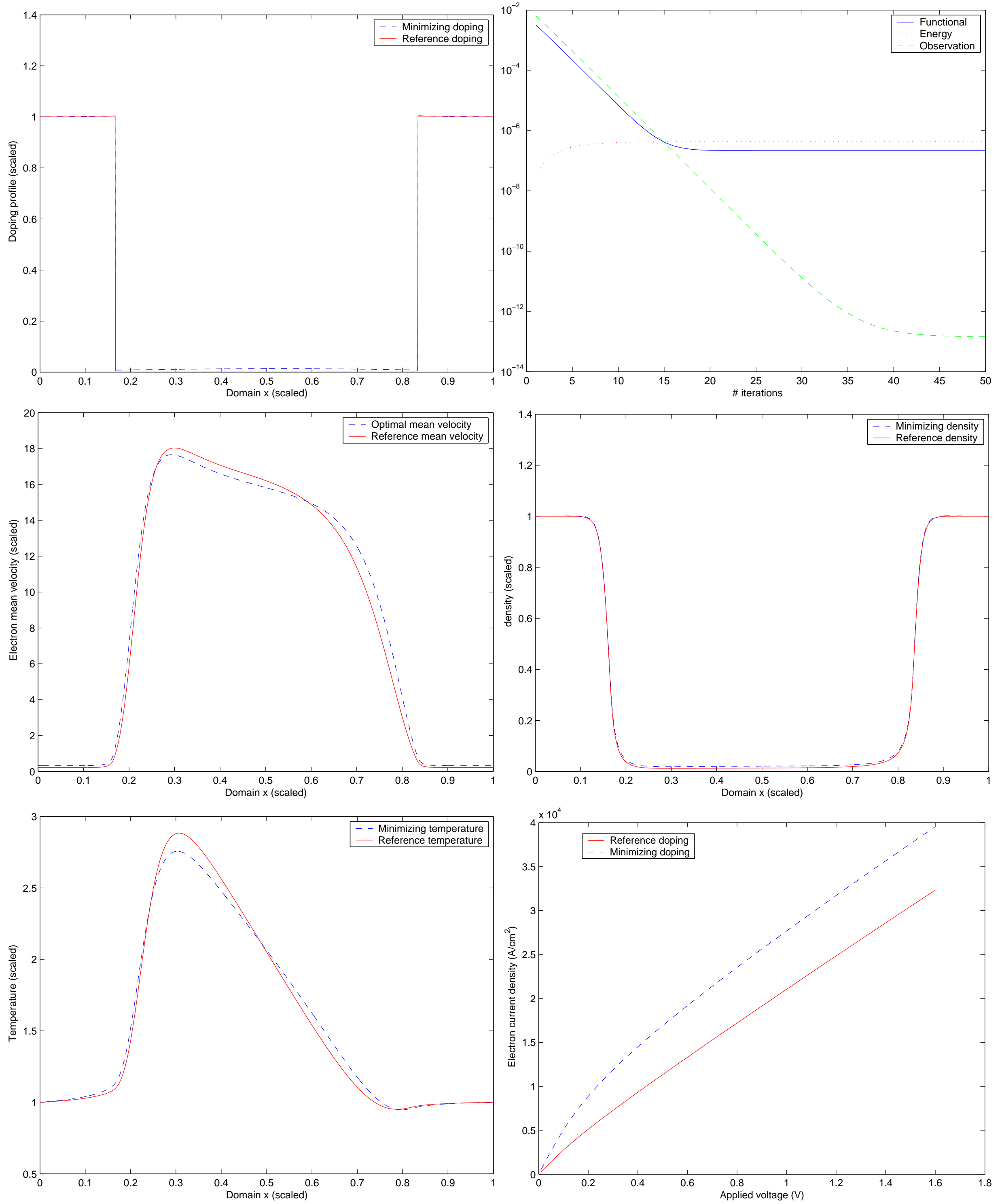

Figure 4.1: Optimized doping profile, evolution of the cost functional, electron mean velocity, electron density and temperature for a biasing voltage of 0.5 V, and the corresponding IVCs

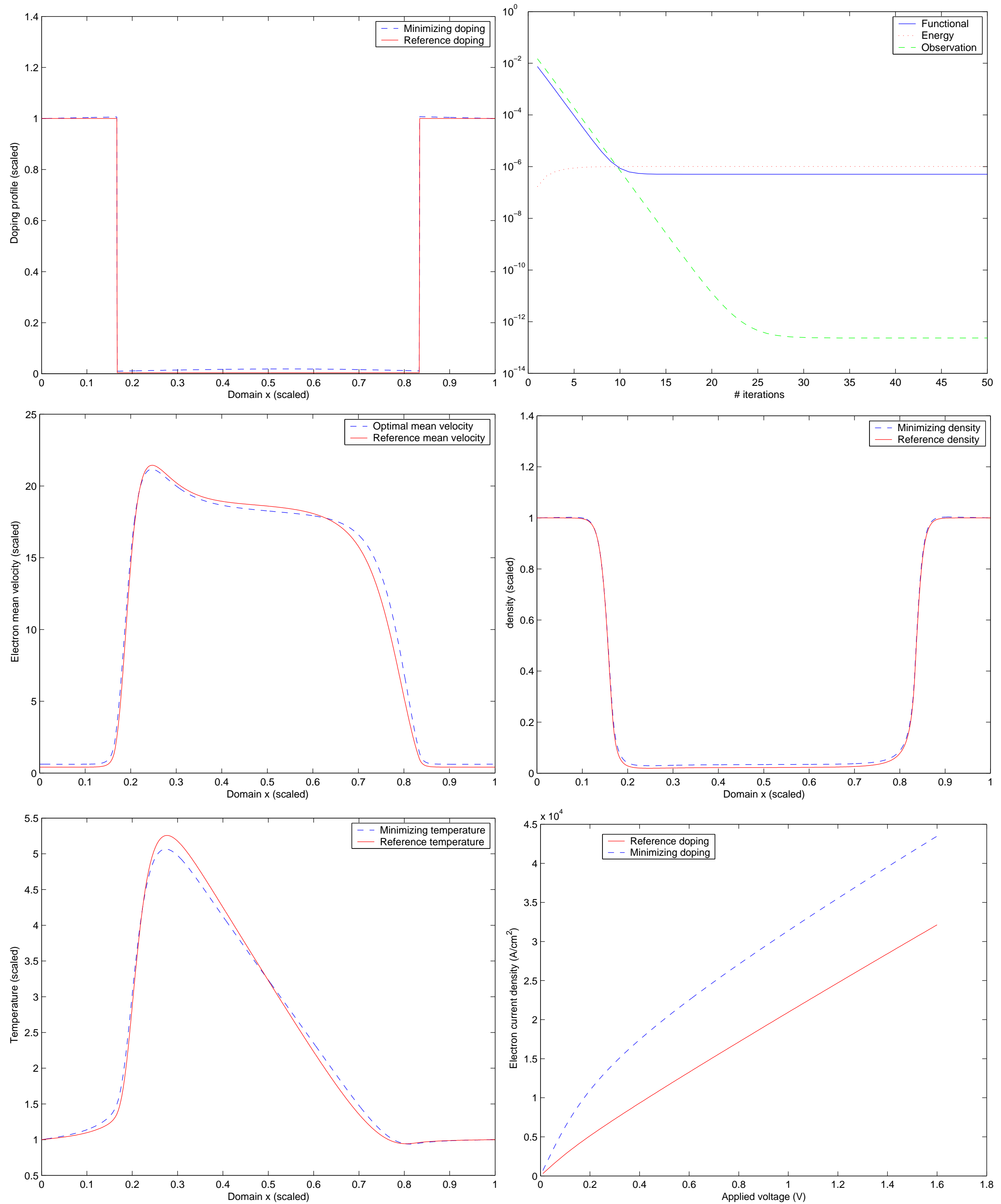

Figure 4.2: Optimized doping profile, evolution of the cost functional, electron mean velocity, electron density and temperature for a biasing voltage of 1 V, and the corresponding IVCs

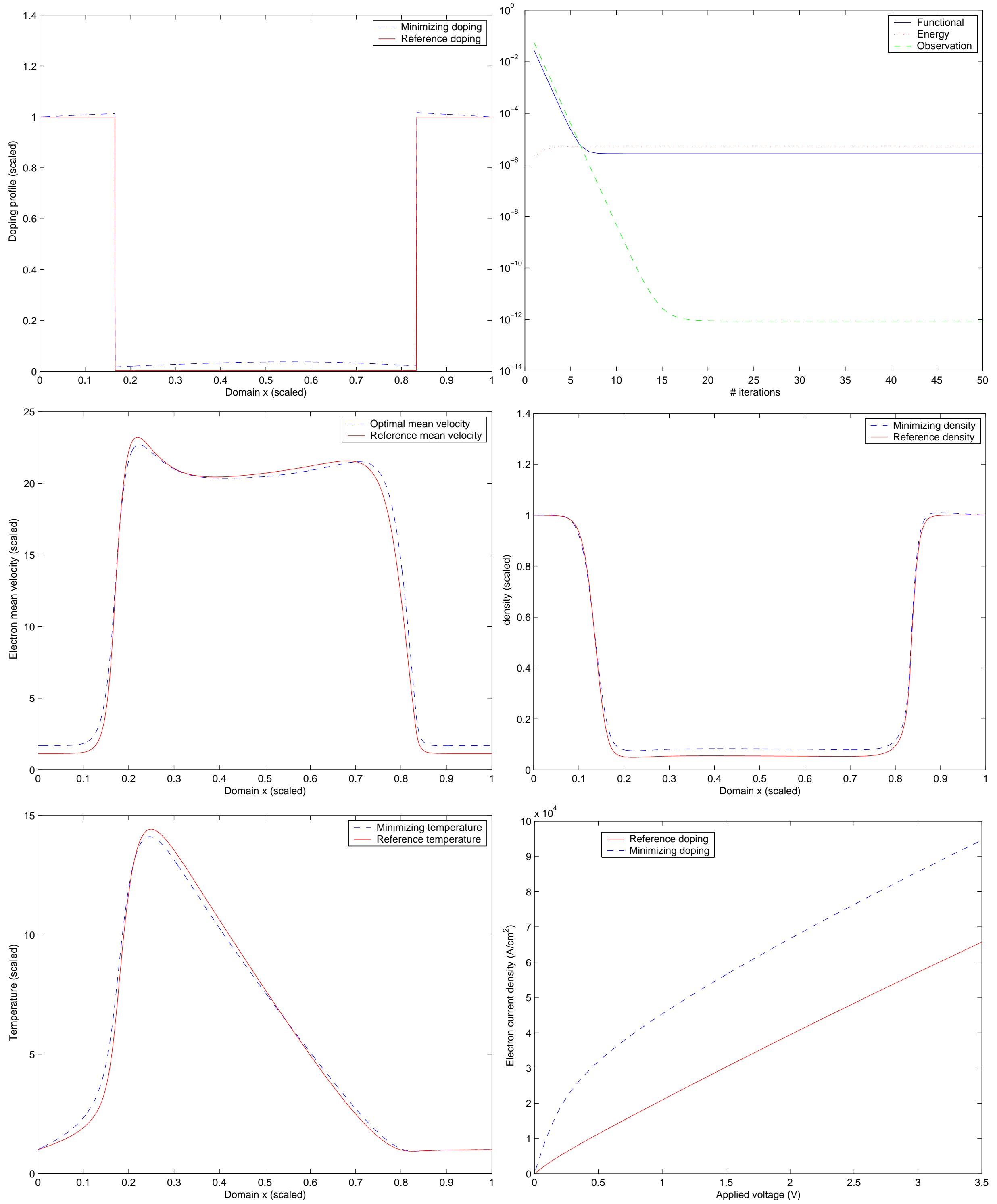

Figure 4.3: Optimized doping profile, evolution of the cost functional, electron mean velocity, electron density and temperature for a biasing voltage of 3 V, and the corresponding IVCs

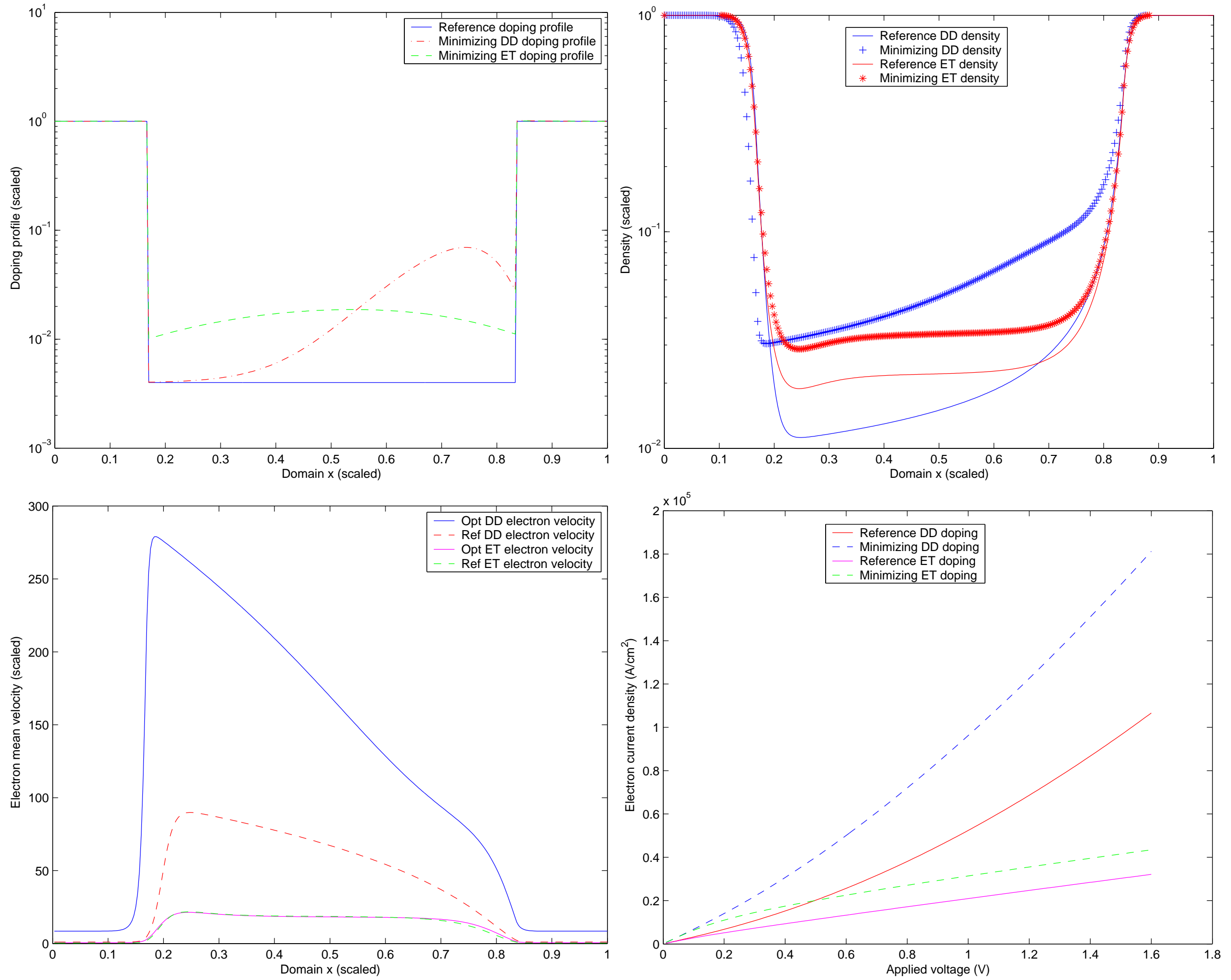

Figure 4.4: Comparisons for the ET and the DD model for a biasing voltage of 1 V

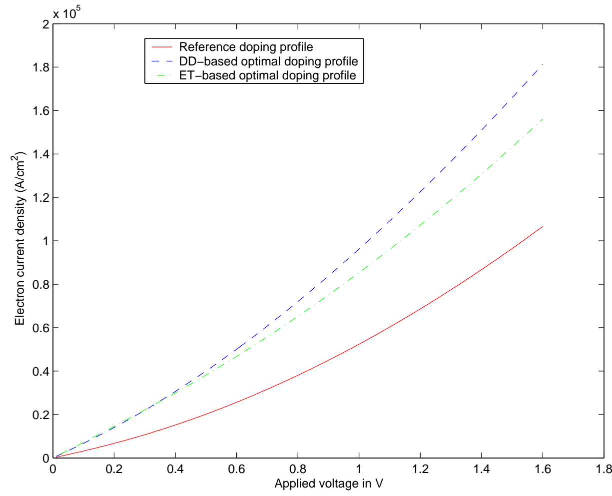

Figure 4.5: CV characteristics computed via DD model, with the ET-based optimal doping profile and the DD-based optimal doping profile for a biasing voltage of 1 V and a gain of 50%

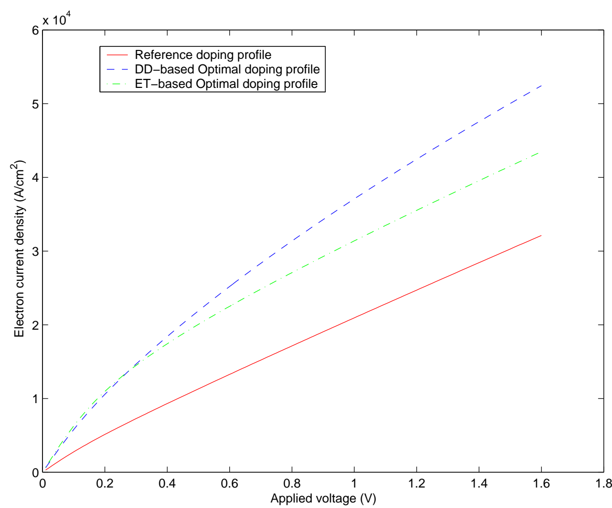

Figure 4.6: CV characteristics computed via ET model, with the ET-based optimal doping profile and the DD-based optimal doping profile for a biasing voltage of 1 V and a gain of 50%

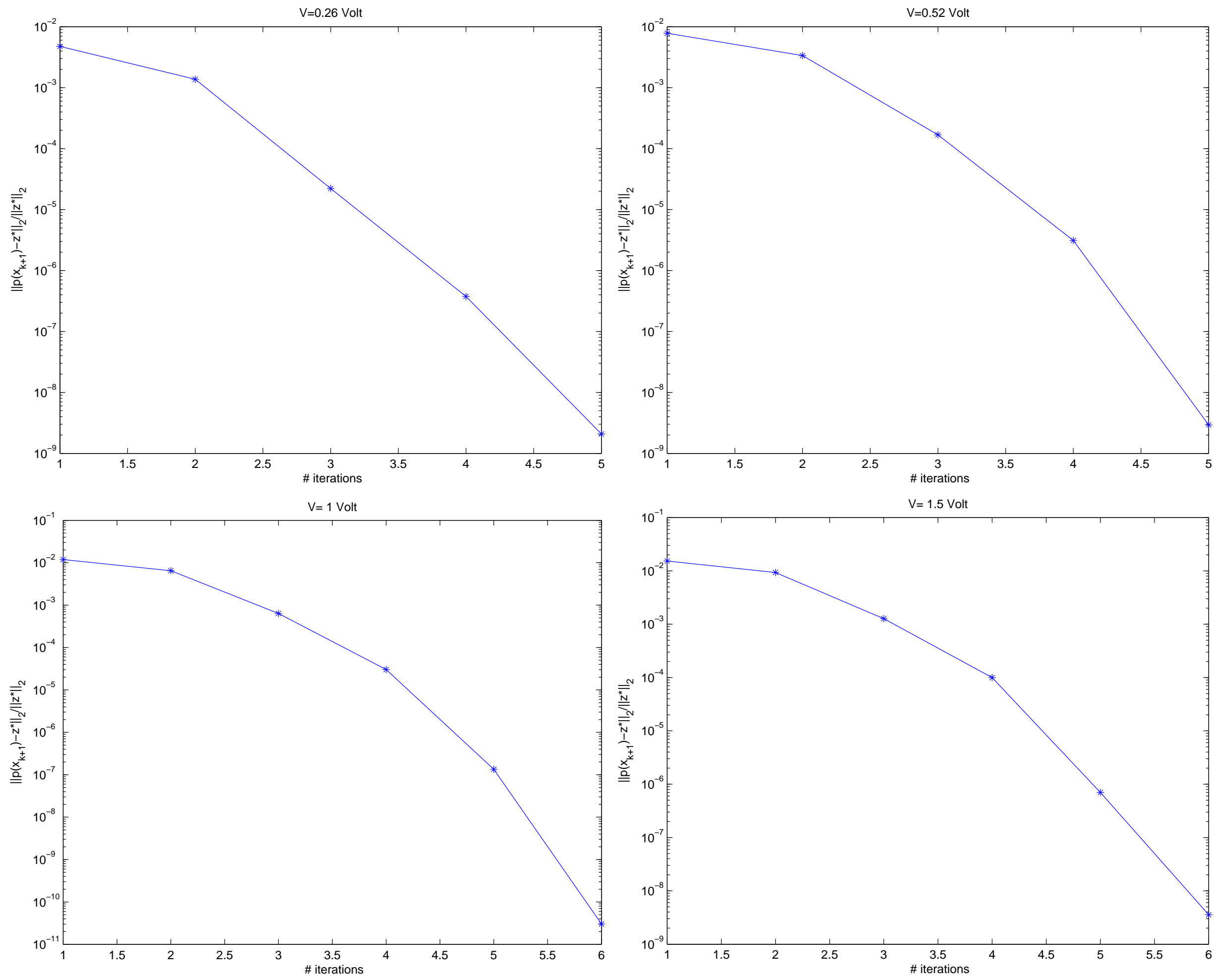

Figure 5.1: Convergence history of the space mapping optimization for a applied voltage of 0.26, 0.52, 1, 1.5 Volt and a gain in the current of 50 %