A Second-Order Stabilized Control Volume Finite Element Method for Self-Heating Effects Simulation of Semiconductor Devices based on Triangular Elements

Da-Miao Yu, Xiao-Min Pan*, and Xin-Qing Sheng

Center for Electromagnetic Simulation
Beijing Institute of Technology, Beijing, 100081, China
*xmpan@bit.edu.cn

Abstract — A second-order control volume finite element method combined with the multiscale flux approximation (CVFEM-MS) based on triangular elements is proposed to numerically investigate the self-heating effects of semiconductor devices. The multiscale fluxes are combined with a selected set of second-order vector basis functions to stabilize the discretization of carrier continuity equations with respect to triangular elements. Numerical results reveal that the proposed method is robust and accurate, even on the mesh of low-quality, where the detrimental impacts caused by the severe self-heating on the terminal currents can be obviously observed for a bipolar transistor model.

Index Terms — CVFEM-MS, self-heating effects, semiconductor devices, triangular element.

I. INTRODUCTION

Nowadays, self-heating effects simulation plays a vital role in the design of modern high-performance semiconductor devices, because of the miniaturization and integration of electron devices, as well as the newly developed device architectures such as silicon-on-insulator (SOI) transistors, FinFETs and resistive random access memory (RRAM) [1-3].

Physically, the carrier transport in semiconductor devices can be described by the drift-diffusion equation [1-6]. To deal with the strong advection effects, many discretization schemes based on Scharfetter-Gummel (S-G) upwinding were employed to enhance the stability and to ease the requirement of a mesh of high quality, such as the finite volume S-G (FVSG) method and the control volume finite element with S-G upwinding method (CVFEM-SG) [2-8]. In these methods, an analytically derived one-dimensional solution is utilized to generate an exponential type of upwinding. Nevertheless, due to the additional diffusion originating from the classical S-G upwinding, the above methods can only deliver first-order accuracy, which limits their ability to resolve the interior and boundary layers. To overcome this limitation, many schemes were proposed, such as the streamline-upwind Petrov-Galerkin method (SUPG), the spectral element time-domain (SETD) method, and the local discontinuous Galerkin (LDG) method [9-13]. But these methods have their own limitations. For example, the choice of stability parameters in the SUPG is problem dependent; the SETD is based on quadrilateral/hexahedron grids; and the auxiliary variables related to fluxes are introduced to generate a larger system in the LDG.

Recently, a parameter-free second-order control volume finite element method combined with multiscale flux approximation (CVFEM-MS) was proposed [14]. The quadrilateral elements were employed and the currents were interpolated by second-order quadrilateral edge elements with fluxes of second-order accuracy. It has been proved that the additional diffusion can be effectively suppressed. Generally, triangular elements are more versatile than quadrilateral ones in modeling complex structures. It is thus important to study CVFEM-MS on triangular meshing. Therefore, in this work, we implement the CVFEM-MS with respect to triangular elements to solve drift-diffusion equations. Different from the case of quadrilateral elements, the number of vector basis functions are more than what are required to expand the currents within an element. A special form of second-order curl-conforming vector basis functions is developed to remove the redundancy without affecting the stabilization.

The rest of the paper is organized as follows. The employed thermodynamic model is presented in Section II, and the discretization scheme is discussed in detail in Section III. The simulation results are presented and discussed in Section IV and the conclusions are finally drawn in Section V.

II. PHYSICAL MODEL

The thermodynamic drift-diffusion model is utilized in this paper to simulate the self-heating effects. The employed equations can be written as [5,6].
\[-\nabla \cdot \varepsilon \nabla \psi = q\left(p - n + N_u^+ - N_p^+\right), \tag{1}\]
\[q \frac{\partial \hat{\psi}}{\partial t} - \nabla \cdot \mathbf{J}_n = -q\left(U_n - G_n\right), \tag{2}\]
\[q \frac{\partial \rho}{\partial t} + \nabla \cdot \mathbf{J}_p = -q\left(U_p - G_p\right), \tag{3}\]
\[\rho C \frac{\partial T}{\partial t} - \nabla \cdot \kappa \nabla T = Q, \tag{4}\]
where \(q\) is the elementary charge, \(\varepsilon\) is permittivity, \(\psi\) is electric potential, \(n\), \(p\) and \(T\) are electric current, hole current, and lattice temperature respectively; \(N_u^+\) and \(N_p^+\) are doping concentrations; \(\mathbf{J}_n, p\) denote electron/hole current densities; \(G_n, p\) and \(U_n, p\) are generation and recombination rates corresponding to electrons/holes; \(\rho\), \(C\), and \(\kappa\) are, respectively, mass density, specific heat capacity and thermal conductivity. \(Q\) is the generated heat power, which is defined as [5,12]:

\[Q = (\mathbf{J}_n + \mathbf{J}_p) \cdot \mathbf{E} + (U - G)(E_x + 3k_b T), \tag{5}\]

where \(E_x\) is the energy gap, \(k_b\) is the Boltzmann constant, \(U = (U_n + U_p) / 2\) and \(G = (G_n + G_p) / 2\).

In the above model, the electric currents \(\mathbf{J}_{ni, p}\) are given as in [5],

\[\mathbf{J}_n = qn \mu_n \mathbf{E} + qD_n \nabla n + qnD_{ni} \nabla T, \tag{6}\]
\[\mathbf{J}_p = qp \mu_p \mathbf{E} - qD_p \nabla p - qpD_{pi} \nabla T, \tag{7}\]

where \(\mu_{ni, p}\), \(D_{ni, p}\) and \(D_{pi}\) are the mobilities, diffusion coefficients and thermal diffusion coefficients of electron/holes respectively, and \(\mathbf{E}\) is the electric field. According to the Einstein relation, the diffusion coefficients are connected with the mobilities \(D_{ni, p} = k_b T \mu_{ni, p} / q\) and the coefficients for thermal diffusion are given by \(D_{pi} = D_{ni} / 2T\).

### III. OUR DISCRETIZATION SCHEME

The main difficulty in the numerical simulation of the above model is the accurate and stable discretization of equations (2) and (3). The performance of the widely used FVSG method depends on the mesh quality [3]. To ease the requirement on the mesh quality, CVFEM-SG was proposed [8], but it can only deliver first-order accuracy due to the additional artificial diffusion originating from the classical S-G upwinding. Recently, a second-order accurate scheme, denoted as CVFEM-MS, was developed on the quadrilateral meshing [14]. Considering that triangular elements are more suitable in modeling the complex structures, CVFEM-MS on the triangular meshing is developed to discretize equations (2) and (3) in this paper. Besides, the Galerkin finite element method is utilized to discretize equations (1) and (4), which has been detailed in many textbooks. Different from what was discussed in [14] for quadrilateral elements, this work defines the second-order accurate multiscale edge currents in terms of triangular elements to avoid unphysical oscillations in carrier continuity equations by the so-called multiscale flux approximation.

### A. Second-order basis functions with respect to triangular elements

This subsection discusses the choice of the second-order basis functions with respect to triangular elements to construct the MS currents. To this end, a general drift-diffusion equation is considered, which can be written as:

\[\frac{\partial \hat{\psi}}{\partial t} + \nabla \cdot \mathbf{J} = -R, \tag{8}\]

\[\mathbf{J} = D \nabla n - F n,\]

where \(n\) is carrier density, \(R\) is net recombination rate, \(D\) and \(F\) are diffusion coefficient and drift coefficient respectively, and \(\mathbf{J}\) is the scaled current density which is scaled by the elementary charge \(q\).

Based on the conformal triangular meshing, the configuration of a given interpolation point \(x_i\), its control volume \(C_i\) and the intersection points \(\mathbf{m}_i\) on the control volume boundary edges can be defined, as sketched in Fig. 1 (a). The integration operation over (8) in terms of the control volume \(C_i\) for point \(x_i\), yield:

\[\frac{\partial \hat{\psi}}{\partial t} \int_{C_i} n d\Omega + \int_{\partial C_i} \hat{n} \cdot \mathbf{J} d\Omega = -\int_{C_i} R d\Omega, \tag{9}\]

where \(\partial C_i\) and \(\hat{n}\) are the boundary and the associated outward normal vector respectively.

![Fig. 1. Notations used in the discretization.](image)

In the second-order CVFEM, the carrier densities \(n\) are interpolated by second-order Lagrangian nodal basis functions, while the current densities \(\mathbf{J}\) are written as:

\[\mathbf{J} = \sum_{\text{subedges}} \mathbf{N}_f \mathbf{J}_f, \tag{10}\]
where \( J_y \) are the tangential current components on the subedge midpoints and \( \mathbf{N}_y \) are the second-order curl-conforming vector basis functions corresponding to the subedges \( \mathbf{e}_y \). As illustrated in Fig. 1 (b), the subedges \( \mathbf{e}_y \) with respect to triangular patches are defined by triangle vertices or midpoints of element segments. If the commonly used set of vector basis functions [15] are employed, it is not trivial to construct the suitably stabilized currents corresponding to the face basis functions because the face interpolation point is located at the barycenter of triangle. This is the reason why we give up the well-defined basis functions in [15].

Similar to what was done in [16], the midpoints of the subedges are chosen as the interpolation points, thus the basis functions have the form:

\[
\begin{align*}
N_{y1} &= W_{y1} (2L_y - 0.5)l_{i12}, \\
N_{y2} &= W_{y2} (2L_y - 0.5)l_{j12}, \\
N_{y3} &= W_{y3} (2L_y - 0.5)l_{k12}, \\
N_{y4} &= W_{y4} (2L_y - 0.5)l_{i20}, \\
N_{y5} &= W_{y5} (2L_y - 0.5)l_{j20}, \\
N_{y6} &= W_{y6} (2L_y - 0.5)l_{k20}, \\
N_{y7} &= 4W_{y7}l_{i32}, \\
N_{y8} &= 4W_{y8}l_{j32}, \\
N_{y9} &= 4W_{y9}l_{k32}, \\
N_{y10} &= 4W_{y10}l_{i240}, \\
N_{y11} &= 4W_{y11}l_{j240}, \\
N_{y12} &= 4W_{y12}l_{k240}, \\
N_{y13} &= 4W_{y13}l_{i240}, \\
N_{y14} &= 4W_{y14}l_{j240}, \\
N_{y15} &= 4W_{y15}l_{k240}, \\
N_{y16} &= 4W_{y16}l_{i240}, \\
N_{y17} &= 4W_{y17}l_{j240}, \\
N_{y18} &= 4W_{y18}l_{k240}.
\end{align*}
\]

As revealed in [15], the number of curl-conforming vector basis functions to reach the second-order accuracy for a triangle element is 8. In contrast, (11) defines 9 basis functions. It is obvious that one of them is redundant. The reason lies in that the three basis functions corresponding to \( \mathbf{e}_{y1} \), \( \mathbf{e}_{y2} \), and \( \mathbf{e}_{y3} \) are not independent. A common way to remove the redundancy is to discard one of them, as in [15]. However, this method cannot offer accurate result for our applications according to our numerical experiments in Section IV. To solve the problem, we establish nine equations by integrating (10) over the subedges:

\[
\sum_{y} J_{y} \int_{c_{y}} \mathbf{N}_{y} \cdot \mathbf{e}_{y} \, ds = \int_{c_{y}} \mathbf{J} \cdot \mathbf{e}_{y} \, ds. 
\]  

The right-hand side of (12) can be approximated analytically through the currents on the subedges, whose detailed discussion will be given in the next subsection. In short, the interpolation quantities \( J_y \) can be obtained by solving (12). Since the matrix system of (12) is singular due to the redundancy of basis functions, we employ the Moore-Penrose pseudoinverse to solve it.

### B. CVFEM-MS with respect to triangular elements

As shown in Fig. 1 (c), there are two types of segments in the triangle. Accordingly, two types of tangential current expressions should be defined. Similar to what was done in [14], for segments connecting three points \( s_{ijk} \), the multiscale current can be written as:

\[
\begin{align*}
J_{y} &= \frac{D_{y}}{l_{y}} \left( n_{ij} - n_{i} \right) - F_{y} \frac{n_{j} + n_{i}}{2} + L_{y} \Psi(-q_{ij}) \left( n_{i} - 2n_{j} + n_{i} \right), \\
J_{y} &= \frac{D_{y}}{l_{y}} \left( n_{ij} - n_{i} \right) - F_{y} \frac{n_{j} + n_{i}}{2} + L_{y} \Psi(q_{ij}) \left( n_{i} - 2n_{j} + n_{i} \right). 
\end{align*}
\]

where the indices \( ijk \in \{132, 240, 051\} \), \( D_{y} \) and \( F_{y} \) are average diffusion coefficients and average drift coefficients over the segment \( s_{ijk} \), \( l_{y} \) is the segment length, \( \Psi(x) = (\coth x + 1)(\coth x - 1)/x \), and

\[
q_{ij} = F_{y}l_{y} / 4D_{y}.
\]

For the segments connecting only two points as shown in Fig. 1 (c), the tangential currents have no counterparts in [14] due to the employment of the triangle elements. Here, the classical S-G approximation is used:

\[
\begin{align*}
J_{y} &= \frac{D_{y}}{l_{y}} \left( n_{ij} - n_{i} \right) - F_{y} \frac{n_{j} + n_{i}}{2} + L_{y} \left( p_{ij} \coth p_{ij} - 1 \right) \left( n_{i} - n_{j} \right), \\
J_{y} &= F_{y} l_{y} / 2D_{y}.
\end{align*}
\]

Similar to that in [14], the third term in the right-hand side of (13) makes the numerical diffusion weaker compared with the case of the traditional first-order S-G current. Our proposed method is thus expected to resolve the interior and boundary layer better than the CVFEM-SG scheme for triangular elements, which will be illustrated later by numerical results in Fig. 3. However, the first-order S-G current (14) is used in (12), resulting in a relative larger numerical diffusion compared with the original CVFEM-MS for quadrilateral elements. Further study should be carried out to remedy it.

Finally, the matrix form of (9) can be written as:

\[
[M][n] + [K][n] = [b],
\]

where \([M]\) is the mass matrix, \([K]\) is the stiffness matrix, and \([b]\) is the right-hand side vector. Employing the Backward Euler (BE) method, the time evolution equation is written as:

\[
([M] + \Delta t[K])[n]_{i, n+1} = [M][n]_{i, n} + \Delta t[b]_{i, n+1},
\]

where \( \Delta t \) is the time step.

In this work, Gummel iteration is employed to solve the coupled nonlinear equations. The equations for electrical properties are firstly solved self-consistently to calculate the carrier and current distributions. Subsequently, the heat generation rate is evaluated according to equation (5). The temperature-dependent parameters for electric potentials and carrier densities
are then updated through the obtained temperature distribution. In the above process, the proposed method is employed to handle the carrier continuity equations, while the Poisson equation for electric potentials and heat equation for temperatures are discretized by second-order classical nodal FEM.

IV. RESULTS AND DISCUSSIONS

The validity and the numerical stability of the developed algorithm are validated here. Numerical experiments on the electrothermal characteristics of the bipolar transistor are also presented to show the performance of the developed method on the low-quality mesh. In the computations, each single equation is solved by the multifrontal sparse direct solver MUMPS [17]. Since all computations can be finished in short time, we will not detail the CPU time in what follows.

Fig. 2. The basis functions remained after removing one basis function, and the profile of $n$ of equation (8) obtained by using these basis functions to expand the currents.

![Figure 2](image1)

(a) $N_{33}$ is removed
(b) The numerical results

Fig. 3. The $n$ of equation (8) obtained from different schemes. In all the figures, the overshoots and undershoots are highlighted in gray and black respectively. Note that the SUPG-FEM results are obtained by COMSOL.

![Figure 3](image2)

(a) 2nd-order SUPG-FEM (b) Quadrilateral CVFE-MS
(c) Triangular CVFE-SG (d) The proposed method

We begin with a homogeneous test problem to examine qualitatively the performance of the proposed method for the carrier continuity equations by setting $R = 0 \, \text{cm}^{-3} \cdot \text{s}^{-1}$ where coefficients and boundary conditions of (8) are pre-specified. In particular, less-diffusive effect is considered by setting $D = 1 \times 10^{-5} \, \text{cm}^{2} \cdot \text{s}^{-1}$. The advective velocity and the boundary conditions are given by:

$$\mathbf{F} = (-\sin \pi / 6, \cos \pi / 6) \, \text{cm} \cdot \text{s}^{-1},$$

and

$$n = \begin{cases} 0, & \Gamma_L \cup (\Gamma_g \cap \{x \leq 0.5 \mu m\}) \\ 1 \times 10^{19} \, \text{cm}^{-3}, & \Gamma_T \cup \Gamma_R \cup (\Gamma_g \cap \{x \geq 0.5 \mu m\}) \end{cases},$$

where the problem domain $\Omega = [0,1 \mu m] \times [0,1 \mu m]$; $\Gamma_T$, $\Gamma_R$, $\Gamma_L$, and $\Gamma_g$ denote respectively the top, bottom, left and right boundary of $\Omega$. As stated in [14], the exact solution of this problem is bounded between 0 and 1, and has an internal layer corresponding to the discontinuity at the boundary. The performance of our proposed method to deal with the singular matrix system of (12) is compared with the naive manner by removing one of the three interior basis functions.

In naive computation, $N_{33}$, as depicted in Fig. 2 (a), is removed and the associated numerical result is given in Fig. 2 (b). Apparently, tremendous unphysical oscillations appear. The similar numerical instability can be also observed when other two interior basis functions are deleted respectively. Clearly, the singularity of the matrix system of (12) cannot be overcome through deleting the redundant basis functions directly.

![Figure 4](image3)

(a) Schematic of the device (b) Net doping profile

Fig. 4. The sketch of the bipolar transistor and the net doping profile in the right side of the device.

Figure 3 compare the results obtained from different schemes for the homogeneous test problem. The unstructured grid is generated by SUPG-FEM in COMSOL with the maximum edge length 0.01. The comparison indicates that the proposed method resolves the internal layer accurately. However, compared with the CVFEM-MS with quadrilateral elements, our scheme produces a wider internal layer due to the larger numerical diffusion. The reason for it has been discussed in Section III.B. One positive effect of the larger
diffusion is that the unphysical oscillations in our method are confined at the boundary discontinuity, as shown in Fig. 3. In contrast, such oscillations in SUPG-FEM and quadrilateral CVFEM-MS methods propagate along the crosswind direction.

As shown in Fig. 5 (a), the obtained temperature distributions by the developed simulator agree well with that generated by the COMSOL. The results show that the base currents depend heavily on self-heating effects. More specifically, the inhomogeneous temperature distribution leads to lower terminal currents, especially when the base voltage increases. According to the temperature profiles at $V_{\text{base}} = 5\, \text{V}$, a hot spot with 1300 K appears in the area between the base and the collector.

In the following, a more complicated model, e.g., a bipolar transistor is employed to demonstrate the performance of the proposed method where the self-heating effects play an important role. The sketch of the example is given in Fig. 4 (a) and the doping profile is shown in Fig. 4 (b). The doping-dependent and temperature-dependent carrier mobilities are taken from [18], which can be written as:

$$\mu_e(T) = \frac{88T_n^{-0.57}}{1 + \left(C/1.26 \times 10^{17} T_n^{2.4}\right)^{0.88T_n^{-0.46}}},$$

$$\mu_p(T) = \frac{54.3T_n^{-0.57}}{1 + \left(C/2.35 \times 10^{17} T_n^{2.4}\right)^{0.88T_n^{-0.46}}},$$

where $T_n = T/300$, $C$ is the doping concentration. The Auger process plays a vital role in this simulation.

\[ R_{\text{Auger}} = U - G = c_n n_p n_i^2 + c_p n_p n_i^2, \]

where $n_i$ is the intrinsic carrier concentration, $c_n = 2.8 \times 10^{-31} \text{cm}^6 \cdot \text{s}^{-1}$ and $c_p = 9.9 \times 10^{-32} \text{cm}^6 \cdot \text{s}^{-1}$ are the Auger recombination coefficients. The thermal conductivity is $\kappa(T) = 354/(T - 68) \text{W/cm} \cdot \text{K}$. In the computations, the emitter is grounded and the collector voltage $V_c$ is set to 0.5 V. All the terminals are treated as ideal heat sink and the environment temperature is fixed at 300 K throughout the simulations.

To study the performance of the proposed method with bad mesh quality, we generate a set of low-quality mesh with many randomly distributed obtuse triangles for the bipolar transistor example. The mesh, as shown in Fig. 6 (a), is generated by the Delaunay tessellation method provided in COMSOL. According to the maximum-angle-rule [19], an element is an obtuse triangle if its quality is less than 0.75. In the Newton solver of COMSOL, a sufficiently small step size $\Delta V_{\text{base}} = 0.1\, \text{V}$ is chosen and the solutions of previous steps are reused to improve its convergence. As highlighted in Fig. 6 (c), the Newton iteration failed to converge at $V_{\text{base}} = 1.8\, \text{V}$ due to the low quality of the mesh, whereas our method converges successfully. An abrupt variation of electron concentration caused by large electric potential gradients can be observed clearly from our results, as shown in Fig. 6 (d). In contrast, the obtuse triangles in the low-quality mesh makes the finite volume method employed in COMSOL solver fail in resolving such an internal layer, as depicted in Fig. 6 (c), and the resultant unphysical negative electron concentration densities lead to the divergence of Newton
iteration. In brief, our proposed method can resolve the internal layers successfully and deliver accurate results even when the mesh becomes low in quality.

V. CONCLUSION

A control volume finite element method equipped with the multiscale flux (CVFEM-MS) in terms of triangular elements is developed to simulate the self-heating effects on semiconductor devices by the thermodynamic drift-diffusion model. The qualitative numerical studies suggest that the proposed CVFEM-MS with respect to triangular elements is acceptably robust and accurate. The stability of our method is superior to the widely used finite volume method, especially on the low-quality mesh.

ACKNOWLEDGMENT

This work was supported in part by the NSFC under Grant 61771053, Grant U-1730102, and Grant 61421001.

REFERENCES


