CWI Quarterly Volume 10 (3&4) 1997, pp. 301 – 315

Stabilization of Mixed Finite Elements for Convection-Diffusion Problems

Riccardo Sacco
Fausto Saleri
Dipartimento di Matematica "F. Brioschi",
Politecnico di Milano
Via Bonardi 9, 20133 Milano, Italy

A new family of stabilized mixed finite volume methods (St-MFV) is proposed for the approximation of the Dirichlet problem for the convection-diffusion operator $Lu = -\operatorname{div}(\varepsilon \underline{\nabla} u - \underline{\beta} u) + \sigma u$ using the lowest-order Raviart-Thomas (RT) finite element space. The stabilization procedure is based on the use of a quadrature formula to diagonalize the stress matrix and on the addition of an artificial viscosity across each edge of the triangulation. A discrete maximum principle and a stability estimate in a discrete energy norm are proved to hold for the new formulation. A special member of the family that generalizes to the two-dimensional case the Scharfetter-Gummel scheme (SG-MFV scheme) is then examined. For this latter method, a $\mathcal{O}(h)$ convergence theorem in the standard mixed finite element norm is established under the assumption $\mathrm{curl}\underline{\beta} = 0$. The nodal superconvergence of the SG-MFV scheme is also demonstrated by the fact that it passes the Constant-Current Patch Test. Such a property provides a sound indication of good behaviour of the method in presence of an advection-dominated flow. The numerical results include an experimental error analysis, the study of some benchmark test problems in convection-dominated flows and the simulation of two semiconductor devices at high field regimes, namely, a one-dimensional p-n diode and a realistic state-of-the-art nMOS transistor with channel length $L_{ch} = 1\mu\mathrm{m}$.

1. Introduction

In this paper we deal with the dual mixed formulation of a convection-diffusion model problem

$$\begin{cases} \text{Find}\underline{J} \in H(\text{div};\Omega),\ u \in L^2(\Omega) \text{ such that:} & \\ (\underline{J},\underline{\tau}) + (u\underline{\beta},\underline{\tau}) + \varepsilon(\operatorname{div}\underline{\tau}, u) = 0, & \forall \underline{\tau} \in H(\text{div};\Omega), \\ -(\operatorname{div}\underline{J}, v) + (\sigma u, v) = (f, v), & \forall v \in L^2(\Omega). \end{cases} \tag{1}$$

where Ω is a polygonal open set in $\mathbb{R}^2$, $(\cdot,\cdot)$ denotes the inner product in $(L^2(\Omega))^n$ $(n = 1, 2)$, ε is a positive given parameter, $f \in L^2(\Omega)$, $\underline{\beta} \in (W^{1,\infty}(\Omega))^2$ and σ is a nonnegative given function in $W^{1,\infty}(\Omega)$ such that $\sigma + \operatorname{div}\underline{\beta}/2 \geq \mu_0 > 0$ almost everywhere in Ω. This latter assumption ensures existence and uniqueness of $u \in H^2(\Omega) \cap H^1_0(\Omega)$ solution in the distributional sense of the differential problem $Lu = f$ (see, e.g., [14], pag.165).

The lowest-order RT mixed finite element space is considered in the numerical approximation (5) [16]. This provides the interelement continuity of the normal traces of the discrete vector $\underline{J}_h$ and yields optimal order convergence rates for both the discrete scalar u_h and $\underline{J}_h$ [6]. Two main difficulties arise, however, when solving problem (5). First, the method is a centered scheme that becomes unstable when $\|\underline{\beta}\|_{0,\infty} h/\varepsilon$ is large. The second drawback is connected with the algebraic structure of (5)

$$\begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{pmatrix} \begin{pmatrix} \Phi_h \\ U_h \end{pmatrix} = \begin{pmatrix} 0 \\ F_h \end{pmatrix}, \tag{2}$$

where Φ_h and U_h are respectively the vector of the unknown fluxes and of the unknown values of u_h on each element of the triangulation. Eliminating Φ_h leads to the following scheme

$$(-\mathbf{C}\mathbf{A}^{-1}\mathbf{B} + \mathbf{D})U_h = F_h. \tag{3}$$

The matrix $\mathbf{M} = -\mathbf{C}\mathbf{A}^{-1}\mathbf{B} + \mathbf{D}$ is full and, in general, neither symmetric nor positive definite; moreover, even in the case $\underline{\beta} = \underline{0}$, $\mathbf{M}$ may not be an M-matrix for any value of σ [15], [3].

To cure these difficulties, we develop in Section 3 a stabilization procedure for the discrete version of (1) by adding to the first equation an artificial diffusion term that can be written in terms of the jumps of u_h across each edge of the triangulation and of the convective flux [21]. Our stabilization is suggested by the quadrature formula proposed in [1] for diagonalizing the stress matrix and leads to a family of cell-centered finite volume methods for which we establish in Section 3.1 a discrete maximum principle and a stability estimate in a discrete energy norm. A special choice of the artificial diffusion term is then considered in Section 3.2 that recovers the exponentially-fitted Scharfetter-Gummel scheme [22]. For this latter method we establish a $\mathcal{O}(h)$ convergence theorem in the standard mixed finite element norm under the assumption or an irrotational convective field [9]. Moreover, an indication for good behaviour in case of an advection-dominated flow is also provided by the fact that the method reproduces the exact solution of the differential problem in the case of constant coefficients and $\sigma = f = 0$ (see also [25], [21] and [19]).

Some numerical results concerning the performances of the SG-MFV scheme are reported in Section 4. They include an experimental convergence analysis and the study of two benchmark problems in convection-dominated flows and of two realistic semiconductor devices.

2. Notations and dual mixed approximation

In view of the numerical approximation, let us introduce a regular family $\{\mathcal{T}_h\}_h$ of decompositions of $\overline{\Omega}$ into triangles T of diameter $h_T \leq h$. We assume for the sake of simplicity each triangulation $\mathcal{T}_h$, for a fixed $h > 0$, to be strictly acute, although weakly acute or Delaunay-type meshes can be easily handled as explained in [20]. We denote by $\mathcal{E}_h$ the set of the edges of $\mathcal{T}_h$, distinguishing between internal edges ($\mathcal{E}_{h,int}$) and boundary edges ($\mathcal{E}_{h,\partial\Omega}$). Also, `Nel` and `Ned` are respectively the total number of triangles and edges in $\mathcal{T}_h$ and $\mathcal{E}_h$.

Next, we denote by $\underline{x} = (x_1, x_2)^T \in \mathbb{R}^2$, by $\mathbb{P}_k$, $k \geq 0$, the space of polynomials of degree $\leq k$ in the variables x_1, x_2 and by $\mathbb{D}_k = (\mathbb{P}_{k-1})^2 \oplus \underline{x}\ \mathbb{P}_{k-1}$, $k \geq 1$. For any triangulation $\mathcal{T}_h$, $h > 0$, we introduce the lowest-order RT mixed finite element space [16]

$$\begin{aligned} \mathbb{Q}_h &= \left\{\underline{\tau}_h \in H(\mathrm{div};\Omega) \mid \underline{\tau}_h \,|_T \in \mathbb{D}_1 \ \forall T \in \mathcal{T}_h\right\}, \\ V_h &= \left\{v_h \in L^2(\Omega) \mid v_h \,|_T \in \mathbb{P}_0 \ \forall T \in \mathcal{T}_h\right\}, \end{aligned} \tag{4}$$

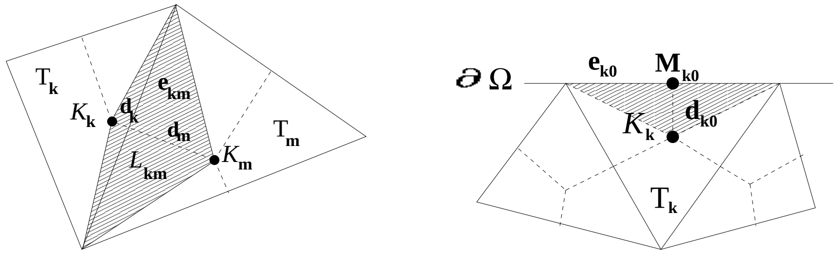
and assume from now on that $\underline{\beta}|_T \in \mathbb{D}_1$ for every $T \in \mathcal{T}_h$.

The discrete version of problem (1) reads

$$\begin{cases} \text{Find } \underline{J}_h \in \mathbb{Q}_h,\ u_h \in V_h \text{ such that:} \\ (\underline{J}_h, \underline{\tau}_h) + (u_h\underline{\beta}, \underline{\tau}_h) + \varepsilon(\,\mathrm{div}\,\underline{\tau}_h, u_h) = 0, \quad \forall \underline{\tau}_h \in \mathbb{Q}_h, \\ -(\,\mathrm{div}\,\underline{J}_h, v_h) + (\sigma u_h, v_h) = (f, v_h), \qquad \forall v_h \in V_h. \end{cases} \tag{5}$$

For every triangle $T_k \in \mathcal{T}_h$, let η_{T_k} be the associated index set; we denote by $\underline{e}_{km} = \partial T_k \cap \partial T_m \in \mathcal{E}_{h,int}$ the common edge between T_k and every $T_m \in \eta_{T_k}$ (internal edge) and by $\underline{e}_{k0}$ the common edge between T_k and $\partial\Omega$ (boundary edge). We then define along each edge $\underline{e}_{km}$ two distinct unit normal vectors, $\underline{n}_{km}$ and $\tilde{\underline{n}}_{km}$, such that $\underline{n}_{km}$ is always *outward* oriented (i.e. from T_k to T_m) and $\tilde{\underline{n}}_{km}$ is fixed for the sake of convenience in order to ensure that $\beta_{km} \equiv \int_{\underline{e}_{km}} \underline{\beta} \cdot \tilde{\underline{n}}_{km}\, ds \ \geq 0$ for every $\underline{e}_{km} \in \mathcal{E}_h$. Let $\mathcal{S}_{km} \equiv \mathrm{sign}(\underline{n}_{km} \cdot \tilde{\underline{n}}_{km})$; clearly, the properties $\underline{n}_{mk} = -\underline{n}_{km}$, $\tilde{\underline{n}}_{km} = -\tilde{\underline{n}}_{mk}$ and $\beta_{mk} = \beta_{km}$ hold.

We denote respectively by d_k, d_m the distances between the circumcenters $\mathcal{K}_k$, $\mathcal{K}_m$ of T_k, T_m and the edge $\underline{e}_{km}$; we let $d_{km} = d_k + d_m$. Next, we associate to every internal edge $\underline{e}_{km} \in \mathcal{E}_{h,int}$ a *lumping region* $\mathcal{L}_{km}$ defined as the parallelogram joining $\mathcal{K}_k$, $\mathcal{K}_m$ and the vertices of $\underline{e}_{km}$. An obvious modification of this definition must be done when $\underline{e}_{km} \in \mathcal{E}_{h,\partial\Omega}$ (see Figure 1).



(a) Internal edge

(b) Boundary edge

FIGURE 1. Lumping regions

We denote by $\mathcal{L}_h$ the set of the lumping regions defined over the primal triangulation $\mathcal{T}_h$, distinguishing between $\mathcal{L}_{h,int}$ (lumping regions associated to an internal edge) and $\mathcal{L}_{h,\partial\Omega}$ (lumping regions associated to a boundary edge). Accordingly with the dual tessellation $\mathcal{L}_h$, we let

$$W_h = \left\{ w_h \in L^2(\Omega) \mid w_h|_{\mathcal{L}_{km}} \in \mathbb{P}_0 \ \forall \mathcal{L}_{km} \in \mathcal{L}_h \right\}, \tag{6}$$

and we introduce the *lumping operator* $L_h : \ V_h \to W_h$ such that

$$L_h \varphi_h|_{\mathcal{L}_{km}} = \frac{\varphi_k + \varphi_m}{2}, \qquad \forall \varphi_h \in V_h, \ \forall \mathcal{L}_{km} \in \mathcal{L}_{h,int}, \tag{7}$$

where $\varphi_k \equiv \varphi_h(\mathcal{K}_k)$ for every $T_k \in \mathcal{T}_h$. In the case of a boundary lumping region, the lumping operator L_h is defined as

$$L_h \varphi_h|_{\mathcal{L}_{k0}} = \frac{\varphi_k + \varphi_{k0}}{2}, \qquad \varphi_h \in V_h, \ \forall \mathcal{L}_{k0} \in \mathcal{L}_{h,\partial\Omega}, \tag{8}$$

where $\mathcal{M}_{k0}$ is the midpoint of the boundary edge $\underline{e}_{k0}$ and $\varphi_{k0} \equiv \varphi_h(\mathcal{M}_{k0}) = 0$ (see Figure 1).

We conclude introducing the L^2-projection operator $\mathcal{P}_h : \ L^2(\Omega) \to V_h$ such that, for any function $\varphi \in L^2(\Omega)$

$$(\varphi - \mathcal{P}_h \varphi, w_h) = 0, \qquad \forall w_h \in V_h. \tag{9}$$

We shall denote in the following

$$\sigma_k \equiv (\mathcal{P}_h \sigma)|_{T_k} = \frac{\int_{T_k} \sigma d\underline{x}}{|T_k|}, \qquad f_k \equiv (\mathcal{P}_h f)|_{T_k} = \frac{\int_{T_k} f d\underline{x}}{|T_k|},$$

where $|T_k| = \text{meas}(T_k)$.

3. Finite volume stabilization of the Raviart-Thomas method

Let us consider the second equation in (5) and denote by δ_{mr} the Kronecker symbol. Taking $v_h = \chi(T_k)$ (the characteristic function of triangle T_k), applying the divergence theorem and reminding that the RT basis function $\underline{\tau}_{km}$ satisfies $\int_{\underline{e}_{kr}} \underline{\tau}_{km} \cdot \underline{\tilde{n}}_{kr} ds = \delta_{mr}$, we get the discrete conservation law

$$- \sum_{m \in \eta_{T_k}} \mathcal{S}_{km} \mathcal{J}_{km} + \sigma_k u_k |T_k| = f_k |T_k|, \qquad \forall T_k \in \mathcal{T}_h, \tag{10}$$

where $\mathcal{J}_{km}$ is the degree of freedom for $\underline{J}_h$ associated with the edge $\underline{e}_{km} = \partial T_k \cap \partial T_m$ and such that $\mathcal{J}_{km} = \mathcal{J}_{mk}$. To end up with a finite volume scheme we must write out $\mathcal{J}_{km}$ as a function of u_k and of the value u_m assumed by u_h over each element T_m adjacent to T. To this aim we employ the quadrature formula proposed in [1] to diagonalize the local stress matrix $a^{T_k}(\underline{\tau}_{km}, \underline{\tau}_{kr})$

$$\begin{aligned} & a^{T_k}(\underline{J}_h, \underline{\tau}_{kr}) = \sum_{m \in \eta_{T_k}} \mathcal{J}_{km} \int_{T_k} \underline{\tau}_{km} \cdot \underline{\tau}_{kr} \, d\underline{x} \simeq \\ & \sum_{m \in \eta_{T_k}} \mathcal{J}_{km} \delta_{mr} \frac{d_{km}}{|\underline{e}_{km}|} \equiv a_h^{T_k}(\underline{J}_h, \underline{\tau}_{kr}), \qquad r \in \eta_{T_k}, \ \forall T_k \in \mathcal{T}_h. \end{aligned} \tag{11}$$

It can be shown [1] that the introduced quadrature error is $\mathcal{O}(h_T)$. To the aim of stabilizing (5) or, equivalently, introducing flux upwinding, let us now consider a general *artificial diffusion* function $\rho^h : \mathcal{L}_h \rightarrow \mathbb{R}$, with $\rho^h \in W_h$, such that, for every lumping region $\mathcal{L}_{km} \in \mathcal{L}_h$, $\rho^h_{km} = \rho^h_{mk}$, $\rho^h|_{\mathcal{L}_{km}} \geq 0$ and $\lim_{h \to 0} \rho^h|_{\mathcal{L}_{km}} = 0$. Next, we introduce the stabilized dual mixed discretization

$$\begin{cases} \text{Find } \underline{J}_h^* \in \mathbb{Q}_h, \ u_h^* \in V_h \text{ such that:} \\ (\alpha \underline{J}_h^*, \underline{\tau}_h) + (u_h^* \underline{\beta}, \underline{\tau}_h) + \varepsilon(\operatorname{div} \underline{\tau}_h, u_h^*) \\ \\ + \varepsilon \displaystyle\sum_{T_k \in \mathcal{T}_h} \int_{\partial T_k} \rho^h \gamma_0(u_h) \underline{\tau}_h \cdot \underline{n}_{\partial T_k} \, ds = 0, \forall \underline{\tau}_h \in \mathbb{Q}_h, \\ -(\operatorname{div} \underline{J}_h^*, v_h) + (\sigma u_h^*, v_h) = (f, v_h), \forall v_h \in V_h, \end{cases} \tag{12}$$

where $\gamma_0(u_h)$ is the continuous extension of u_h on the boundary ∂T_k of each element $T_k \in \mathcal{T}_h$ and $\underline{n}_{\partial T_k}$ is the outward normal unit vector along ∂T_k. We let $\gamma_0(u_h) = 0$ on every boundary edge $\underline{e}_{km} \in \mathcal{E}_{h, \partial\Omega}$. The stabilizing term introduced in the dual mixed approximation can be written as

$$\begin{aligned} & \varepsilon \sum_{T_k \in \mathcal{T}_h} \int_{\partial T_k} \rho^h \gamma_0(u_h) \underline{\tau}_h \cdot \underline{n}_k \, ds = \varepsilon \sum_{\underline{e}_{km} \in \mathcal{E}_h} \\ & \mathcal{S}_{km} \left(\gamma_0(u_k) - \gamma_0(u_m) \right) \rho^h_{km} = \varepsilon \sum_{\underline{e}_{km} \in \mathcal{E}_h} \mathcal{S}_{km} [u_h]_{km} \rho^h_{km}, \end{aligned} \tag{13}$$

where $[u_h]_{km} \equiv \gamma_0(u_k) - \gamma_0(u_m)$ denotes the jump of u_h across every interelement edge $\underline{e}_{km} \in \mathcal{E}_h$. Our goal is to choose ρ^h_{km} in such a way that problem

(12) becomes stable irrespectively of the strength of the *local Péclet number* $\mathbb{P}\mathrm{e}_{km} = \hat{\beta}_{km} d_{km}/2\varepsilon$, where $\hat{\beta}_{km} \equiv \beta_{km}/|\underline{e}_{km}| \geq 0$. Continuing to denote by $(\underline{J}_h, u_h)$ the unknowns $(\underline{J}_h^*, u_h^*)$ of the modified problem (12) and using the quadrature formula (11) and the lumping operator L_h, we obtain

$$\int_{T_k \cup T_m} u_h \underline{\beta} \cdot \underline{\tau}_{km}\, d\underline{x} \simeq \int_{T_k \cup T_m} \chi_{km}\, L_h u_h\, \underline{\beta} \cdot \underline{\tau}_{km}\, d\underline{x} \simeq \beta_{km} \left(\frac{u_k + u_m}{2} \right) \frac{d_{km}}{|\underline{e}_{km}|}. \quad (14)$$

Collecting (11), (14) and (13), we end up with the stabilized Galerkin mixed finite volume system $\mathbf{M}_h^{Gstab} U_h = F_h$, where, for $k = 1, \dots, \mathtt{Nel}$

$$\begin{cases} (\mathbf{M}_h^{Gstab})_{kk} = \displaystyle\sum_{m \in \eta_{T_k}} \left(\frac{\varepsilon(1+\rho_{km}^h)}{d_{km}} + \frac{\mathcal{S}_{km}}{2} \hat{\beta}_{km} \right) |\underline{e}_{km}| + \sigma_k |T_k|, \\ (\mathbf{M}_h^{Gstab})_{km} = \left(-\dfrac{\varepsilon(1+\rho_{km}^h)}{d_{km}} + \dfrac{\mathcal{S}_{km}}{2} \hat{\beta}_{km} \right) |\underline{e}_{km}|, \quad m \neq k,\, m \in \eta_{T_k}. \end{cases} \quad (15)$$

3.1. Stability analysis for the St-MFV scheme

Let us determine sufficient conditions for $\mathbf{M}_h^{Gstab}$ defined in (15) to be a nonsingular M-matrix. Such a property ensures the stability of the St-MFV scheme and provides a discrete maximum principle for u_h (see, e.g., [17], Section 1.2), which is quite relevant in real-life applications where u has the physical meaning of a density or a concentration. We start observing that

$$\begin{cases} (\mathbf{M}_h^{Gstab})_{kk} = \displaystyle\sum_{m \in \eta_{T_k}} \varepsilon(1+\rho_{km}^h) \frac{|\underline{e}_{km}|}{d_{km}} + \int_{T_k} \left(\sigma + \frac{1}{2} \operatorname{div} \underline{\beta} \right) d\underline{x} > 0, \\ k = 1, \dots, \mathtt{Nel} \\ \displaystyle\sum_{k=1,\mathtt{Nel}} (\mathbf{M}_h^{Gstab})_{km} = \sigma_{T_m} |T_m| \geq 0, \qquad m = 1, \dots, \mathtt{Nel} \\ (\textit{column sum}). \end{cases} \quad (16)$$

Therefore, requiring the off-diagonal entries of $\mathbf{M}_h^{Gstab}$ to be nonpositive, enforces $\mathbf{M}_h^{Gstab}$ to be an M-matrix. This is stated in the following

PROPOSITION 3.1. *Let the edge artificial viscosity ρ^h be chosen in such a way that*

$$\rho_{km}^h = \rho_{mk}^h \geq \frac{\hat{\beta}_{km} d_{km}}{2\varepsilon} - 1 = \mathbb{P}\mathrm{e}_{km} - 1, \quad \forall \mathcal{L}_{km} \in \mathcal{L}_h. \quad (17)$$

Then, the stiffness matrix $\mathbf{M}_h^{Gstab}$ of the stabilized dual mixed finite volume scheme turns out to be an irriducible diagonally dominant M-matrix with respect to its colums [26].

A first immediate consequence of Proposition 3.1 is that the linear system $\mathbf{M}_h^{Gstab} U_h = F_h$ is uniquely solvable; moreover, provided $F_h \geq 0$, the nodal values U_k are nonnegative for $k = 1, \dots, \mathtt{Nel}$ irrespectively of the strength of the local Péclet number (discrete maximum principle). The next step is to examine the coerciveness of the discrete bilinear form associated with the St-MFV method. To this aim let us write out the discrete equations as

$$\lambda_h^{Gstab}(u_h, v_h) = (f, v_h), \qquad \forall v_h \in V_h \tag{18}$$

where the discrete bilinear form $\lambda_h^{Gstab}(u_h, v_h) : V_h \times V_h \to \mathbb{R}$ is defined as

$$\begin{cases} \lambda_h^{Gstab}(u_h, v_h) = -\displaystyle\int_\Omega \operatorname{div} \underline{J}_h^{Gstab}(u_h) v_h \, d\underline{x} + \int_\Omega \sigma u_h v_h \, d\underline{x} \\ \underline{J}_h^{Gstab}(u_h) = \displaystyle\sum_{\underline{e}_{km} \in \mathcal{E}_h} \mathcal{J}_{km}^{Gstab}(u_h) \underline{\tau}_{km}(\underline{x}) \\ \mathcal{J}_{km}^{Gstab} = \left\{ \varepsilon (1 + \rho_{km}^h) \mathcal{S}_{km} \dfrac{u_m - u_k}{d_{km}} - \dfrac{u_k + u_m}{2} \hat{\beta}_{km} \right\} |\underline{e}_{km}|, \\ \forall \underline{e}_{km} \in \mathcal{E}_h. \end{cases} \tag{19}$$

Since $\operatorname{div} \mathbb{Q}_h = V_h$, the discrete problem (18) supplied with the definition (19) can be interpreted as a nonconforming approximation of problem (1) on the space of the piecewise constant functions, provided the fluxes have been eliminated at the edge level by means of the static condensation procedure described in the previous Section. For any function $u_h = \sum_{k=1}^{\mathtt{Nel}} U_k \chi(T_k)$ in V_h, let us introduce the *discrete energy norm* on $\mathrm{H}_0^1(\Omega)$

$$|u_h|_{1,h}^2 \equiv \sum_{\mathcal{L}_{km} \in \mathcal{L}_h} \left(\frac{u_m - u_k}{d_{km}} \right)^2 |\mathcal{L}_{km}| = \frac{1}{2} \sum_{\mathcal{L}_{km} \in \mathcal{L}_h} \left(\frac{u_m - u_k}{d_{km}} \right)^2 |\underline{e}_{km}| d_{km}, \tag{20}$$

and the "stabilized" discrete energy norm

$$\begin{aligned} |u_h|_{1,h,stab}^2 &\equiv |u_h|_{1,h}^2 + \sum_{\mathcal{L}_{km} \in \mathcal{L}_h} \left(\frac{u_m - u_k}{d_{km}} \right)^2 \rho_{km}^h |\mathcal{L}_{km}| = \\ \frac{1}{2} &\sum_{\mathcal{L}_{km} \in \mathcal{L}_h} \left(1 + \rho_{km}^h \right) \left(\frac{u_m - u_k}{d_{km}} \right)^2 |\underline{e}_{km}| d_{km}. \end{aligned} \tag{21}$$

The discrete energy norms on $\mathrm{H}^1(\Omega)$ corresponding to (20) and (21) can be defined as $\|u_h\|_{1,h}^2 = |u_h|_{1,h}^2 + \|u_h\|_0^2$ and $\|u_h\|_{1,h,stab}^2 = |u_h|_{1,h,stab}^2 + \|u_h\|_0^2$. The bilinear form $\lambda_h^{Gstab}(\cdot, \cdot)$ can be proved to be coercive over $V_h \times V_h$ [21], as is stated in the following

THEOREM 3.1. *There exists a positive constant α^*, independent of h, such that, for any given regular and strictly acute triangulation $\mathcal{T}_h$, $h > 0$*

$$\lambda_h^{Gstab}(u_h, u_h) \geq \alpha^* \|u_h\|_{1,h,stab}^2, \qquad \forall u_h \in V_h. \tag{22}$$

Theorem 3.1 provides a theoretical explanation of the good stability property of the St-MFV scheme with respect to its corresponding nonstabilized counterpart

$$\left\{\begin{array}{l} \lambda_h^{Gal}(u_h, v_h) = -\int\limits_{\Omega} \mathrm{div}\, \underline{J}_h^{Gal}(u_h) v_h \, d\underline{x} + \int\limits_{\Omega} \sigma u_h v_h \, d\underline{x} \\ \underline{J}_h^{Gal}(u_h) = \sum\limits_{\underline{e}_{km} \in \mathcal{E}_h} \mathcal{J}_{km}^{Gal}(u_h) \underline{\tau}_{km}(\underline{x}) \\ \mathcal{J}_{km}^{G} = \left\{ \varepsilon \mathcal{S}_{km} \dfrac{u_m - u_k}{d_{km}} - \dfrac{u_k + u_m}{2} \hat{\beta}_{km} \right\} |\underline{e}_{km}|, \qquad \forall \underline{e}_{km} \in \mathcal{E}_h. \end{array}\right. \tag{23}$$

Indeed, it can be easily checked that $\lambda_h^{Gal}(u_h, u_h) \geq \alpha^* \|u_h\|_{1,h}^2$, stating that, for fixed h and ε, $\|u_h\|_{1,h,stab}^2 \gg \|u_h\|_{1,h}^2$, while, as $h \to 0$, the two estimates become equivalent since $\rho_{km}^h \to 0$ for every lumping region $\mathcal{L}_{km} \in \mathcal{L}_h$ (see [21]).

3.2. The SG-MFV scheme

We provide in this section a choice of ρ^h that fulfils the stability requirement (17) and extends to the two-dimensional case the upwinding scheme proposed in [22]. Precisely, we recover the Scharfetter-Gummel exponentially fitted scheme taking for every $\mathcal{L}_{km} \in \mathcal{L}_h$

$$(\rho_{km}^h)^{SG} = \mathbb{P}\mathrm{e}_{km} - 1 + \mathcal{B}(2\mathbb{P}\mathrm{e}_{km}), \qquad \forall \mathcal{L}_{km} \in \mathcal{L}_h, \tag{24}$$

where for any $z \in \mathbb{R}$, $\mathcal{B}(z) = z/(e^z - 1)$ denotes the Bernoulli function. Plugging (24) into (15) we obtain the following expressions of the stiffness matrix $\mathbf{M}_h^{SG}$ acting on U_h

$$\left\{\begin{array}{l} (\mathbf{M}_h^{SG})_{kk} = \sum\limits_{m \in \eta_{T_k}} \mathcal{B}\left(-\mathcal{S}_{km} \dfrac{\hat{\beta}_{km} d_{km}}{\varepsilon} \right) \dfrac{|\underline{e}_{km}|}{d_{km}} + \sigma_k |T_k|, \\ k = 1, \dots, \mathtt{Nel} \\ (\mathbf{M}_h^{SG})_{km} = -\mathcal{B}\left(\mathcal{S}_{km} \dfrac{\hat{\beta}_{km} d_{km}}{\varepsilon} \right) \dfrac{|\underline{e}_{km}|}{d_{km}}, \\ m \neq k, \, m \in \eta_{T_k}. \end{array}\right. \tag{25}$$

As far as the amount of the introduced artificial viscosity is concerned, it is worth noting that for high Péclet numbers the SG-MFV scheme degenerates into the classical upwinding method of Engquist-Osher (EO) [7]. On the other hand, as $\mathbb{P}\mathrm{e}_{km} \to 0$, the artificial diffusion introduced by the SG method is $\mathcal{O}(h^2)$, while the corresponding term in the EO scheme is $\mathcal{O}(h)$ (see also [24]).

A convergence estimate for the SG-MFV method can be obtained assuming that curl$\underline{\beta} = 0$, as happens in the drift-diffusion model for semiconductor device simulation (see [11], [12]). In such a case $\underline{\beta} = \nabla\psi$, being ψ the electric potential, so that the Slotboom change of variable $u = \rho e^{\psi/\varepsilon}$ allows us to write the convection-diffusion problem in symmetrized form as

$$\begin{cases} L\rho = -\operatorname{div}(\varepsilon e^{\psi/\varepsilon} \underline{\nabla} \rho) + \sigma \rho e^{\psi/\varepsilon} = f, \qquad \underline{x} \in \Omega, \\ \rho = 0 \qquad \underline{x} \in \partial\Omega. \end{cases} \tag{26}$$

Standard mixed finite element error analysis (see [10], Thm. 11.2, pag. 575, [3]) can be applied to (26) to end up with the following result [9]:

THEOREM 3.2. *Let $(\rho, \underline{J})$ be the solution of problem (26) and $(\rho_h, \underline{J}_h)$ be the solution of the corresponding discrete problem. Then there exists a constant C, independent of h, such that:*

$$\|\rho - \rho_h\|_0 + \|\underline{J} - \underline{J}_h\|_0 \leq Ch\|\rho\|_2 \tag{27}$$

To conclude, it is worth noting that the SG-MFV scheme gives the *exact solution* at the nodes when $\underline{\beta}$ is constant and $\sigma = f = 0$ (suitable Dirichlet and Neumann boundary conditions must of course be supplied in such a case). This is an instance of the so-called Constant-Current Patch-Test (see [19]) and provides a sound indication for a good behaviour of the numerical method in presence of steep layers arising in advection-dominated flows, as previously remarked in [25], [21].

4. NUMERICAL RESULTS

In this section we demonstrate the performance of the SG-MFV scheme on five benchmark problems in convection-dominated flows. In the first three examples $\Omega = (0,1)^2$, while x, y denote the space coordinates. For graphical purposes the computed (piecewise constant) solution u_h has been reinterpolated at the nodes of the triangulation by piecewise linear continuous splines. An average value has been computed at the barycenter of each mesh triangle to represent the approximate flux $\underline{J}_h$.

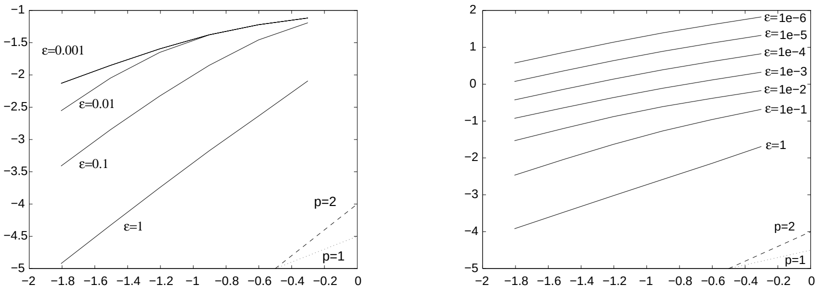
4.1. A Dirichlet problem with analytical solution

We study on the unit square the Dirichlet problem (1) considered in [13], where $\underline{\beta} = (1,1)^T$, $\sigma = 2$ and

$$f(x,y) = x(1 - e^{(x-1)/\varepsilon}) \left[1 + e^{(y-1)/\varepsilon} + y(1 - e^{(y-1)/\varepsilon})\right]$$
$$+ y(1 - e^{(y-1)/\varepsilon}) \left[1 + e^{(x-1)/\varepsilon} + x(1 - e^{(x-1)/\varepsilon})\right].$$

The exact solution is $u_{ex}(x,y) = xy(1 - e^{(x-1)/\varepsilon})(1 - e^{(y-1)/\varepsilon})$; for small values of ε it exhibits sharp layers along the outflow boundary $x = 1$ and $y = 1$.
An experimental analysis has been performed on a uniform tensor-product triangulation of mesh size $h_k = 2^{-k}$, for $k = 1, \ldots, 6$, to measure the absolute error in the maximum nodal norm

$$\|u_{ex} - u_h\|_{\infty,h} = \max_{\underline{x}_k \in X_h} |u_{ex}(\underline{x}_k) - u_h(\underline{x}_k)| \tag{28}$$

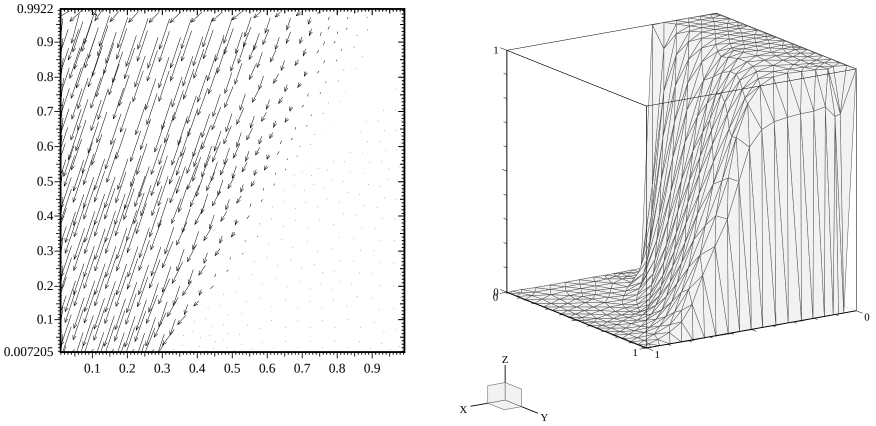


(a) Maximum norm

(b) Discrete energy norm

FIGURE 2. Error curves

and in the stabilized discrete $\mathrm{H}^1_0(\Omega)$ norm (21). We let ε assume the decreasing values $\varepsilon_j = 10^{-j}$, with $j = 0, \dots, 6$ and show the convergence results in figure 2 where the values of $\log_{10}(\|u_{ex} - u_h\|)$ versus $\log_{10}(h)$ are plotted for each value of ε_j. To provide an immediate reading of the plots, two straight lines with slopes $p = 1$ and $p = 2$ are added in the right corners of the figures, denoting respectively linear and quadratic convergence.



(a) $\underline{J}_h$

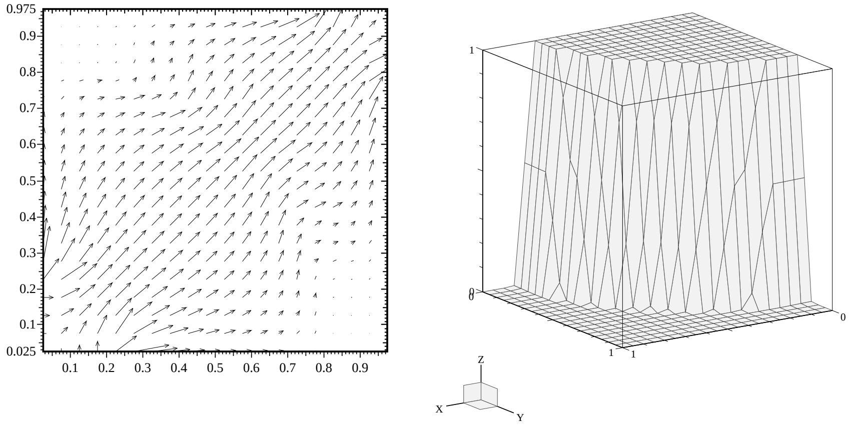
(b) u_h

FIGURE 3. Discontinuity transport problem.

The method clearly exhibits an asymptotic $\mathcal{O}(h)$ convergence for small values of ε. Notice how the log-curves of the error measured in the maximum norm cannot be distinguished for $\varepsilon \leq 10^{-3}$, since the regime of the flow is highly convection-dominated and, as a consequence, the SG-MFV scheme degenerates into the Engquist-Osher standard upwinding method. On the other hand, second-order convergence is obtained when the local Péclet number gets smaller, in agreement with the fact that the extra-viscosity added by the SG-MFV method is in such a case of $\mathcal{O}(h^2)$.

4.2. Transport of discontinuous data

We deal with the transport of discontinuity test case considered in [5], where SUPG and bubble stabilization (BS) methods have been employed in the computations. We take $\varepsilon = 10^{-6}$, $\underline{\beta} = (1,3)^T$, $f = \sigma = 0$ and use an unstructured grid where the size h_T of each triangle is $\simeq 1/20$. The boundary data are $u = 1$ on $\{(x,y) : x = 0,\, 0 \leq y \leq 1\} \cup \{(x,y) : 0 \leq x \leq 1/3,\, y = 0\}$ and $u = 0$ elsewhere. The solution $(\underline{J}_h, u_h)$ is shown in figure 3. There is no presence of spurious oscillations in the graph of u_h, although some crosswind dissipation is visible along the internal layer, if compared with the SUPG solution exhibited in [5]. This latter, however, as well as the BS solution, is affected by some "wiggles" instabilities, due to the nonmonotonicity of the schemes.



(a) $\underline{J}_h$ (b) u_h

FIGURE 4. Curved $p-n$ junction.

4.3. Simulation of a model curved p-n diode

This test case has been taken from [4] where the classical dual mixed method with Lagrange multipliers (MML) and exponential fitting is employed in the computations. The problem models the flow of free charges (electron and holes) in a semiconductor curved *p-n* junction under high reverse bias. (See [23] and

[12] for a complete discussion of the physical and mathematical aspects of the problem). The convective field $\underline{\beta} = \underline{\nabla}\psi_0$, where $\psi_0(\underline{x})$ is a piecewise linear continuous function equal respectively to 0 and 0.2 for $r \leq 0.8$ and $r \leq 0.9$, being $r = \sqrt{x^2 + y^2}$. We take $\varepsilon = 10^{-2}$, $f = \sigma = 0$ and use a structured grid of isoscele right-angled triangles of side $h = 1/20$. The boundary data are $u = 1$ on $\{(x, y) : 0 \leq x \leq 0.25,\, y = 0\} \cup \{(x, y) : x = 0,\, 0 \leq y \leq 0.25\}$, $u = 0$ on $\{(x, y) : x = 1,\, 0.75 \leq y \leq 1\} \cup \{(x, y) : y = 1,\, 0.75 \leq x \leq 1\}$ and $\underline{J} \cdot \underline{n} = 0$ elsewhere. The solution $(\underline{J}_h, u_h)$ is shown in figure 4. An abrupt jump is attained by u_h around $r = 0.8$, while the flow field lines are mainly directed along the diagonal $x_1 = x_2$, due to the simmetry of the problem, becoming almost negligible around the corners of the unit square. The comparison with the MML is quite favorable, as far as u_h is concerned, while no plot of $\underline{J}_h$ is reported in [4]. As for the computational cost, we remark that the SG-MFV scheme is much cheaper than the MML since the number of elements `Nel` is typically much less than the number of edges `Ned`. Moreover, being a cell-centered method, the sparsity pattern of the stabilized mixed finite volume scheme exhibits *at most* four nonzero entries for each matrix row.

4.4. Simulation of two realistic semiconductor devices

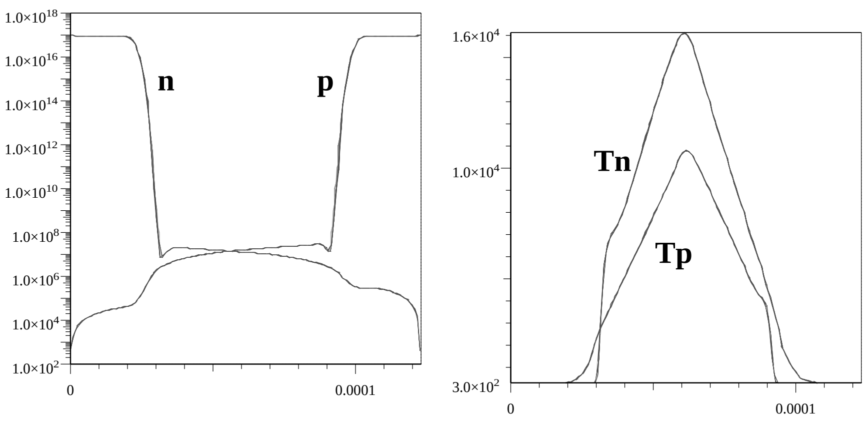
In this section we present two numerical examples obtained by applying the SG-MFV method for the discretization of to the numerical solution of the Energy-Balance transport equations for semiconductors [11], [2], [8].

The first device is a one-sided *p-n* diode with a 1D geometry analyzed in [18]. The diode has been simulated under reverse-bias conditions, with the applied voltage varying from $0V$ to $50V$. Figure 5 shows the carrier concentrations n (electrons) and p (holes) and the corresponding temperatures T_n and T_p at an applied bias of $28V$. The comparison with the results of [18] is quite satisfactory. In particular, we point out the strong carrier heating in the middle of the device due to the presence of a high electric field.

The next example refers to a realistic 1μm channel-length nMOS transistor, with a bulk (B) doping equal to $-3 \cdot 10^{15}\mathrm{cm}^{-3}$, whereas in the source (S) and drain (D) regions the doping amounts to $2 \cdot 10^{20}\mathrm{cm}^{-3}$. A channel implantation of $-3 \cdot 10^{16}\mathrm{cm}^{-3}$ is placed under the gate (G) oxide. The device has been simulated with $5V$ applied between the gate and the source and with a drain-source voltage varying from $0V$ to $5V$. In figures 6 and 7 the electron concentration, electric field and carrier temperatures are shown, relatively to a drain-source voltage of $4.8V$. The carrier heating around the drain region, where a high peak of electric field exists, is clearly visible. This, in turn, is the reason for the large flooding of electrons near the drain end of the channel, due to their increased thermal diffusivity.

References

1. J. Baranger, J. F. Maitre, F. Oudin (1993). Application de la théorie des éléments finis mixtes à l'étude d'une classe de schémas aux volumes



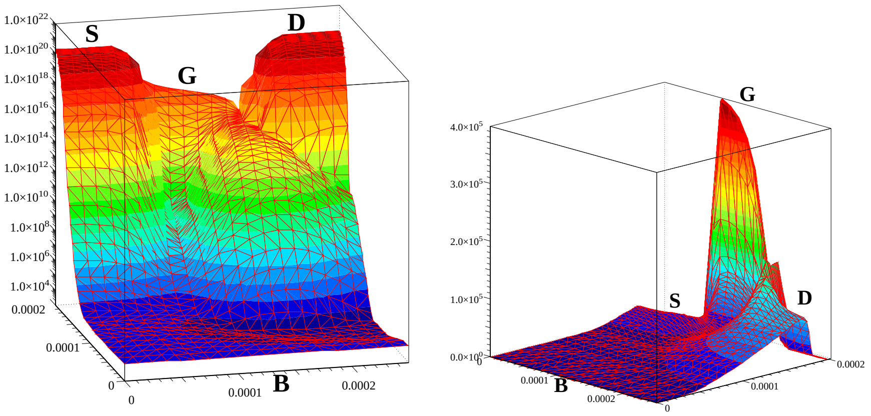
(a) Carrier concentrations

(b) Carrier temperatures

FIGURE 5. Energy-balance simulation of a p-n diode

différences finis pour les problèmes elliptiques, *C. R. Acad. Sci. Paris*, **316**, série I, 509–512.

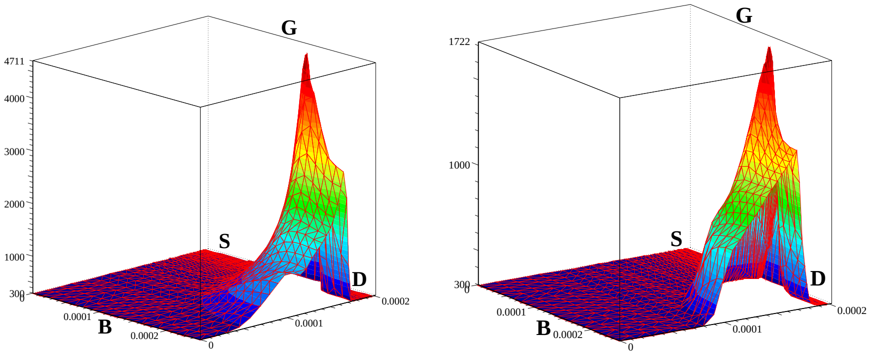
2. K. BLØTAKJÆR (1970). Transport equations in two-valley semiconductors, *IEEE Trans. on El. Dev.* ED-**17**(1), 38–47.
3. F. BREZZI, M. FORTIN (1991). *Mixed and Hybrid Finite Element Methods*, Springer-Verlag, New York.
4. F. BREZZI, L. D. MARINI, P. PIETRA (1989). Two-dimensional exponential fitting and applications to drift-diffusion models. *SIAM J. Numer. Anal.* **26**, 1342–1355.
5. F. BREZZI, A. RUSSO (1994). Choosing bubbles for advection-diffusion problems. *Math. Models Meths. Appl. Sci.* **4**, 571–587.
6. J. DOUGLAS, JR., J. E. ROBERTS (1985). Global estimates for mixed methods for second order elliptic equations. *Math. Comp.* **44**, 39–52.
7. B. ENGQUIST, S. OSHER (1981). One-sided difference approximations for nonlinear conservation laws. *Math. Comp.* **36**, 321–351.
8. F. BOSISIO, E. GATTI, R. SACCO, F. SALERI (1997). Exponentially fitted mixed finite volumes for energy balance models in semiconductor device simulation. *ENUMATH97, Second European Conference on Numerical Mathematics and Advanced Applications*, Heidelberg, 250–251.
9. F. BOSISIO, S. MICHELETTI, R. SACCO, F. SALERI (1998). *Work in preparation.*
10. P. G. CIARLET, J. L. LIONS (Editors) (1991). *Handbook of Numerical Analysis, Volume II, Finite Element Methods (Part 1)*, North-Holland, Amsterdam.



(a) Electron concentrations

(b) Electric field

FIGURE 6. Energy-balance simulation of a MOS transistor



(a) Electron temperature

(b) Hole temperature

FIGURE 7. Energy-balance simulation of a MOS transistor

11. J. W. JEROME (1996). *Analysis of charge Transport. A Mathematical Study of Semiconductor Devices*, Springer-Verlag, Wien, New York.

12. P. MARKOWICH, C. A. RINGHOFER, C. SCHMEISER (1990). *Semiconductor Equations*, Springer-Verlag, Wien-New York.

13. J. J. H. MILLER, S. WANG (1994). A new non-conforming Petrov-Galerkin finite-element method with triangular elements for a singularly perturbed advection-diffusion problem. *IMA J. Num. An.* **14**, 257–276.

14. A. Quarteroni, A. Valli (1994). *Numerical Approximation of Partial Differential Equations*, Springer-Verlag, Berlin.
15. S. J. Polak (1990). Mixed FEM for $\triangle u = \alpha u$. *Mathematical Modelling and Simulation of Electrical Circuits and Semiconductor Devices* (R. E. Bank, R. Bulirsch, K. Merten, eds.) **90**, Birkhäuser Verlag, Basel, 247–255.
16. P. A. Raviart, J. M. Thomas (1977). A mixed finite element method for second order elliptic problems. *Mathematical Aspects of the Finite Element Method* (I. Galligani, E. Magenes, eds.), Lectures Notes in Math. **606**, Springer-Verlag, New York, 292–315.
17. H.-G.Roos, M. Stynes, L. Tobiska (1996). *Numerical methods for singularly perturbed differential equations. Convection-diffusion and flow problems*, Springer-Verlag, Berlin.
18. W. Quade, M. Rudan, E. Schöll (1991). Hydrodynamic simulation of impact ionization effects in P-N junctions. *IEEE Trans. on CAD*, **10**, 1287–1294.
19. R. Sacco, E. Gatti, L. Gotusso (1995). The Patch-Test as a validation of a new finite element for the solution of convection-diffusion equations, *Comp. Meth. Appl. Mech. Engrg.* **124**, 113–124.
20. R. Sacco, F. Saleri (1997). Mixed finite volume methods for semiconductor device simulation. *Numer. Meth. Part. Diff. Eq.* **13**, 215–236.
21. R. Sacco, F. Saleri (1997). Stabilized mixed finite volume methods for convection-diffusion problems. *East West Jour. Num. Math.* **5**, No.4.
22. D.L. Scharfetter, H. K. Gummel (1969). Large-signal analysis of a silicon Read diode oscillator. *IEEE Trans. on Electr. Dev.* **ED-16**, 64–77.
23. S. M. Sze (1981). *Physics of Semiconductor Devices*, 2nd Edition, Wiley-Interscience, New York.
24. L. Tobiska (1995). A note on the artificial viscosity of numerical schemes. *Comp. Fluid Dyn.* **5**, 281–290.
25. R. R. P. van Nooyen (1995). A Petrov-Galerkin mixed finite element method with exponential fitting. *Numer. Meth. Part. Diff. Eq.* **11**, 501–524.
26. R. S. Varga (1962). *Matrix iterative analysis*, Englewood Cliffs, NJ, Prentice-Hall.