1992

R.R.P. van Nooyen

A finite volume discretisation scheme with a-posteriori error estimates for the symmetrised continuity equation

Department of Numerical Mathematics Report NM-R9203 March

CWI is the research institute of the Stichting Mathematisch Centrum, which was founded on February 11, 1946, as a non-profit institution aiming at the promotion of mathematics, computer science, and their applications. It is sponsored by the Dutch Government through the Netherlands organization for scientific research (NWO).

A Finite Volume Discretisation Scheme with a-posteriori Error Estimates for the Symmetrised Continuity Equation

R. R. P. van Nooyen

CWI
P. O. box 4079, 1009 AB Amsterdam,
The Netherlands

The purpose of this paper is the derivation of an a-posteriori error estimate for the Scharfetter-Gummel discretisation of the continuity equations in the semiconductor problem. We use the deferred correction method to derive an a-posteriori error estimate. We also prove stability and consistency of the discretisation scheme.

1980 AMS Subject Classification: 65N30
1990 CR category: G.1.8 Finite Element Methods.
Key Words and Phrases: Finite Volumes, Mixed Finite Elements, Continuity equations, Semi-conductors, Quadrature Rules, A Priori Error Estimates, A Posteriori Error Estimates.

Note: This report will be submitted for publication elsewhere.

1 Introduction.

Before we can discuss the approach used to obtain an a-posteriori error estimate, we must give our interpretation of the well-known extension of the one-dimensional Scharfetter-Gummel scheme[1] to two dimensions. We take as our starting point the continuity equation for electrons in the stationary case,

$$-a(\mathbf{grad}\, u + u\, \mathbf{grad}\, \psi) = \boldsymbol{\sigma} ,$$

$$\operatorname{div} \boldsymbol{\sigma} = f .$$

We sketch the derivation of the discretisation on a rectangular grid. Consider two adjacent cells. We assume that a and the component of $\mathbf{grad}\,\psi$ along the line segment $\hat{\Gamma}$ connecting the cell centres are constant. Furthermore we assume that the component of $\boldsymbol{\sigma}$ parallel to $\hat{\Gamma}$ is constant on $\hat{\Gamma}$ and on the common cell edge. Under these assumptions we can give an expression for $\boldsymbol{\sigma}$ in terms of u. Furthermore we can calculate the integral over the common cell edge of the component of $\boldsymbol{\sigma}$ orthogonal to the common cell edge. This gives us a finite volume scheme for the above equations. Note that along the line segment $\hat{\Gamma}$ we get an exponential fitting scheme as described by Il'in[2]. The resulting discretisation scheme is equivalent to one of the schemes discussed in the articles by Bank et al. [3,4].

For the error analysis we choose a trial space $V_h \times W_h$ and write the finite volume scheme as a saddle point problem which has a solution in that trial space. We use theorem 3.1 from the article by Nicolaides[5]_to prove stability of the problem and existence of the solution. We then choose a projection $\Pi_h \times P_h$ of the solution $(\boldsymbol{\sigma}, u)$ of the continuous problem. We use the stability of the problem to give an upper bound on the global discretisation error in terms of the local discretisation error. We show that we may express the local discretisation error in terms of partial derivatives of $\boldsymbol{\sigma}$. Consistency follows immediately from the expression obtained. We then use the expression for the residual to construct a deferred correction scheme, based on the finite volume scheme in that form. We prove that, if the original scheme gives an $\mathcal{O}(h^k)$ accurate approximation, then this

Report NM-R9203
ISSN 0169-0388
CWI
P.O. Box 4079, 1009 AB Amsterdam, The Netherlands

deferred correction scheme gives an $\mathcal{O}(h^{k+1})$ accurate approximation.

Our analysis shows, that the discretisation error for the Scharfetter-Gummel scheme is second order in areas of constant cell size and slowly varying electrical potential (i.e. the jump in ψ - the scaled potential - between cell centres is smaller than 2). It also shows the scheme to be only first order accurate if the ratio of adjacent cell edges differs too much from one or if the number of boundary cells is a large fraction of the total number of cells.

In section 2 we formulate a model problem. Section 3 discusses the discretisation spaces to be used. We give a description of the discretisation in section 4. Section 5 gives conditions that imply existence of the solution and stability of the problem. In section 6 we describe a quadrature rule. Section 7 shows consistency and section 8 gives the a-posteriori error estimate. In section 9 we summarise our results and draw some conclusions.

2 The model equation.

In this paper we study a model equation for the semi-conductor continuity equation. For a discussion of both numerical and physical aspects of semi-conductor modelling, we refer to the books by Markowich[6], or Selberherr[7], or the papers by Polak et al. [8] or Engl et al. [9]. For a review of numerical aspects of such models, we refer to the articles by Bank et al. [3, 10, 11]. We consider a linearised model for the equations for one of the two charge carrier densities.

$$-a(\mathbf{grad}\,u+u\,\mathbf{grad}\,\psi) = \boldsymbol{\sigma} \text{ on } \Omega\,, \tag{1a}$$

$$\operatorname{div}\boldsymbol{\sigma} = f \text{ on } \Omega\,, \tag{1b}$$

with Dirichlet boundary conditions on some parts of the boundary

$$u\,|_{\Gamma_1} = g\,, \tag{1c}$$

and mixed or Robin boundary conditions on the remaining parts of the boundary

$$\mathbf{grad}\,u\cdot\mathbf{n}_{\partial\Omega}\,|_{\Gamma_2} + u\,\mathbf{grad}\,\psi\cdot\mathbf{n}_{\partial\Omega}\,|_{\Gamma_2} = 0\,, \tag{1d}$$

where the notation $\mathbf{n}_{\partial A}$ denotes the outward unit normal vector on the boundary of a domain A. In equation (1a), ψ corresponds to the electrical potential scaled by the thermal voltage. We place the following restrictions on the coefficients. We assume that the coefficients a and ψ are continuous and differentiable, $a,\psi \in C^1(\bar{\Omega})$, we also assume that a is bounded away from zero, $\exists\ a_0 > 0 \in \mathbb{R}: a \geqslant a_0$ on Ω, and ψ is piecewise bilinear on Ω. We assume that the function g is continuous and differentiable, $g \in C^1(\partial\Omega)$. Note that the connected subsets of the Dirichlet boundary generally correspond to the contacts of the device. We assume that the right hand side f is square integrable, i.e. $f \in \mathrm{L}^2(\Omega)$, and that $\Gamma_1 \bigcup \Gamma_2 = \partial\Omega$ and $\Gamma_1 \bigcap \Gamma_2 = \emptyset$. We assume that the shape of Ω, Γ_1 and Γ_2 and the conditions on a, ψ, f and g guarantee that $\boldsymbol{\sigma} \in \mathrm{H}^1(\Omega)^2$ and $u \in C(\Omega)$.

For later reference, we give an equivalent system of equations, obtained by a transformation of the dependent variable,

$$\boldsymbol{\sigma} = -a\exp(-\psi)\,\mathbf{grad}\,U\,, \tag{2a}$$

$$\operatorname{div}\boldsymbol{\sigma} = f\,, \tag{2b}$$

$$U\,|_{\Gamma_1} = \exp(\psi)g\,|_{\Gamma_1}\,, \tag{2c}$$

$$\boldsymbol{\sigma}\cdot\mathbf{n}_{\partial\Omega}\,|_{\Gamma_2} = 0\,, \tag{2d}$$

where $U = \exp(\psi)u$. Note that, in these variables, condition (2d) implies homogeneous Neumann boundary conditions for U on Γ_2.

We assume that U is square integrable and that $\boldsymbol{\sigma}$ lies in the space,

$$\mathrm{H}(\mathrm{div};\Omega) := \{\ \boldsymbol{\tau} \in \mathbf{L}^2(\Omega)\ \mid\ \operatorname{div}\boldsymbol{\tau} \in \mathrm{L}^2(\Omega)\ \}\,,$$

with the inner product,

$$(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)_{\mathrm{H(div;}\Omega)} = (\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)_{\mathbf{L}^2(\Omega)} + (\operatorname{div}\boldsymbol{\tau}_1, \operatorname{div}\boldsymbol{\tau}_2)_{\mathrm{L}^2(\Omega)} \quad \forall \;\; \boldsymbol{\tau}_1,\boldsymbol{\tau}_2 \in \mathbf{L}^2(\Omega) \; ,$$

where

$$\mathbf{L}^2(\Omega) = \{ \; \boldsymbol{\tau}:\Omega\rightarrow\mathbb{R}^2 \mid \int_\Omega \boldsymbol{\tau}\cdot\boldsymbol{\tau} \; d\mu < \infty \; \}$$

with the usual inner product,

$$(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)_{\mathbf{L}^2(\Omega)} = \int_\Omega \boldsymbol{\tau}_1\cdot\boldsymbol{\tau}_2 \; d\mu \quad \forall \;\; \boldsymbol{\tau}_1,\boldsymbol{\tau}_2 \in \mathbf{L}^2(\Omega) \; .$$

Properties of H(div;Ω) are found in Girault and Raviart[12]. We wish to define a subspace V of H(div;Ω) that contains all elements that satisfy the homogeneous Neumann boundary condition given in (2d). To do this properly, we define this subspace as the closure in H(div;Ω) of the space $\mathscr{V}(\Omega)$ of $C^\infty(\bar{\Omega})$ functions that satisfy the condition (2d),

$$\mathscr{V}(\Omega) := \{ \; \boldsymbol{\tau} \in C^\infty(\bar{\Omega})^2 \mid (\boldsymbol{\tau}\cdot\mathbf{n}_{\partial\Omega})|_{\partial\Omega} = 0 \;\text{ on }\; \Gamma_2 \; \} \, ,$$

where we assume that Γ_1 and Γ_2 are such that this definition makes sense. Now V is by definition a closed subspace of H(div;Ω) and a Hilbert space for the H(div;Ω) inner product.

3 The discretisation spaces.

The Scharfetter-Gummel discretisation can best be interpreted as a finite volume scheme, so we need an mesh of finite volumes, which we call the primary or finite volume mesh, and a dual mesh with the cell centres of the original mesh as vertices. In addition, the dual mesh needs vertices on the centres of those edges of finite volumes that lie on the boundary of the domain. For that purpose we add cells of zero thickness to the finite volume mesh, to avoid the need for special formulas that refer to dual mesh vertices on the domain edge. We restrict ourselves to rectangular domains and to partitions of Ω that are Cartesian products of partitions of the sides of the rectangle. We assume, that the boundaries between Γ_1 and Γ_2 coincide with vertices of the mesh. We assume, that ψ is piecewise bilinear on the cells of the dual mesh.

We use a Cartesian coordinate system and we position our rectangular domain Ω as follows,

$$\Omega = \,]0,L_1[\times]0,L_2[\; . \tag{3}$$

We use the following naming conventions. The horizontal unit vector is denoted by $\mathbf{e}_1$ and the vertical unit vector is denoted by $\mathbf{e}_2$. All lower case bold letters are vectors, the corresponding lower case italic letters with subscript 1 or 2 are the vector components in the horizontal or vertical direction.

3.1. The partition.

To introduce names for the vertices and cells we need to specify the partitions of the sides of our domain. We use the letter P for the partition of the horizontal axis and the letter Q for the partition of the vertical axis,

$$P = \{ \; 0 = p_{-1} = p_0 < p_1 < \cdots < p_{N_1} = p_{N_1+1} = L_1 \; \} \, , \tag{4}$$

$$Q = \{ \; 0 = q_{-1} = q_0 < q_1 < \cdots < q_{N_2} = q_{N_2+1} = L_2 \; \} \, , \tag{5}$$

where we added p_{-1}, p_{N_1+1}, q_{-1}, q_{N_2+1} to take into account the zero-width boundary cells. The partition of Ω is given by $P \times Q$. In the obvious way we introduce a notation for particular points in the primary and in the dual mesh. First, the vertices of the primary mesh,

$$\mathbf{x}_{i,j} = (p_i, q_j)^T \;\text{ for }\; i = -1,0,1,2,\ldots,N_1+1 \; , j = -1,0,1,2,\ldots,N_2+1 \; . \tag{6}$$

We denote the vertices of the dual cells by,

$$\mathbf{x}_{i-1/2,j-1/2} = \frac{\mathbf{x}_{i-1,j-1} + \mathbf{x}_{i,j}}{2} \;\text{ for }\; i=0,1,2,\ldots,N_1+1 \; , j=0,1,2,\ldots,N_2+1 \; . \tag{7}$$

Finally, we introduce,

$$\mathbf{x}_{i,j-\frac{1}{2}} = \frac{\mathbf{x}_{i,j-1} + \mathbf{x}_{i,j}}{2} \quad \text{for } i=0,1,2,\ldots,N_1\, , j=1,2,\ldots,N_2\, . \tag{8}$$

and

$$\mathbf{x}_{i-\frac{1}{2},j} = \frac{\mathbf{x}_{i-1,j} + \mathbf{x}_{i,j}}{2} \quad \text{for } i=1,2,\ldots,N_1\, , j=0,1,2,\ldots,N_2\, . \tag{9}$$

We denote the finite volumes, i.e. the cells of the partition $P \times Q$ by,

$$\Omega_{i-\frac{1}{2},j-\frac{1}{2}} = \{\, \mathbf{x} \,|\, \mathbf{x}_{i-1,j-1} < \mathbf{x} < \mathbf{x}_{i,j} \,\} \ \text{ for } i=1,2,\ldots,N_1\, , j=1,2,\ldots,N_2\, , \tag{10}$$

where the notation

$$\mathbf{a} < \mathbf{b}\, , \tag{11}$$

has its usual meaning, i.e.

$$\mathbf{a} < \mathbf{b} \Leftrightarrow a_1 < b_1 \ \text{ and } \ a_2 < b_2\, . \tag{12}$$

Similarly,

$$\Gamma_{i-\frac{1}{2},j} = \{\, \mathbf{x} \,|\, \mathbf{x}_{i-1,j} \leqslant \mathbf{x} \leqslant \mathbf{x}_{i,j} \,\} \ \text{ for } i=1,2,\ldots,N_1\, , j=0,1,2,\ldots,N_2\, , \tag{13}$$

and

$$\Gamma_{i,j-\frac{1}{2}} = \{\, \mathbf{x} \,|\, \mathbf{x}_{i,j-1} \leqslant \mathbf{x} \leqslant \mathbf{x}_{i,j} \,\} \ \text{ for } i=0,1,2,\ldots,N_1\, , j=1,2,\ldots,N_2\, . \tag{14}$$

In our error analysis in section 7, we also need to identify the cells and edges of the dual mesh, these are denoted by,

$$\hat{\Omega}_{i,j} = \{\, \mathbf{x} \,|\, \mathbf{x}_{i-\frac{1}{2},j-\frac{1}{2}} < \mathbf{x} < \mathbf{x}_{i+\frac{1}{2},j+\frac{1}{2}} \,\} \ \text{ for } i=0,1,2,\ldots,N_1\, , j=0,1,2,\ldots,N_2\, , \tag{15}$$

$$\hat{\Gamma}_{i-\frac{1}{2},j} = \{\, \mathbf{x} \,|\, \mathbf{x}_{i-\frac{1}{2},j-\frac{1}{2}} \leqslant \mathbf{x} \leqslant \mathbf{x}_{i-\frac{1}{2},j+\frac{1}{2}} \,\} \ \text{ for } i=1,2,\ldots,N_1\, , j=0,1,2,\ldots,N_2\, , \tag{16}$$

and

$$\hat{\Gamma}_{i,j-\frac{1}{2}} = \{\, \mathbf{x} \,|\, \mathbf{x}_{i-\frac{1}{2},j-\frac{1}{2}} \leqslant \mathbf{x} \leqslant \mathbf{x}_{i+\frac{1}{2},j-\frac{1}{2}} \,\} \ \text{ for } i=0,1,2,\ldots,N_1\, , j=1,2,\ldots,N_2\, . \tag{17}$$

Note that, at the start of this section, we assumed $\psi|_{\hat{\Omega}_{i,j}}$ to be bilinear. This implies that $\psi|_{\hat{\Gamma}_r}$ is linear for all $r \in \tilde{E}$, where $\tilde{E}$ is the collection of index tuples of edge centres,

$$\tilde{E} = \{\, e=(i,j-\tfrac{1}{2}) \,|\, i=0,1,2,\ldots,N_1\, , j=1,2,\ldots,N_2 \,\} \cup$$
$$\{\, e=(i-\tfrac{1}{2},j) \,|\, i=1,2,\ldots,N_1\, , j=0,1,2,\ldots,N_2 \,\}\, .$$

We indicate the set of indices of all edges that are not on the Neumann boundary by,

$$E = \{\, e \in \tilde{E} \,|\, \Gamma_e \subset \bar{\Omega} - \Gamma_2 \,\}\, .$$

Finally, we define the set of index tuples of cell centres, by

$$M = \{\, e=(i-\tfrac{1}{2},j-\tfrac{1}{2}) \,|\, i=1,2,\ldots,N_1\, , j=1,2,\ldots,N_2 \,\}\, ,$$

and we extend the definition of the Kronecker-δ to index tuples,

$$\delta_{rs} = \begin{cases} 1 & \text{if } r=s\, , \\ 0 & \text{if } r \neq s\, . \end{cases}$$

3.2. Local coordinates.

When we analyse the quadrature rules - in section 6 - and the discretisation - in section 7 - it is convenient to have at our disposal a local coordinate system with its origin at the intersection of a primary and a dual mesh line. We define this system as follows. Take a unit vector $\mathbf{e}_{x,r}$ parallel to $\hat{\Gamma}_r$, and let the direction of increasing coordinates correspond to the direction of increasing coordinates in the global coordinate system given at the start of section 3. Take a unit vector $\mathbf{e}_{y,r}$ parallel to Γ_r and directed to give a right hand coordinate system when combined with $\mathbf{e}_{x,r}$. I.e. $\mathbf{e}_{x,r}$ is a normal vector on Γ_r and $\mathbf{e}_{y,r}$ is a normal vector on $\hat{\Gamma}_r$. We shall use the letters x and y for local coordinates, so if $\mathbf{x}$ is an arbitrary position vector in the global coordinate system and $\mathbf{x}_r$ is the global coordinate vector of the intersection of Γ_r and $\hat{\Gamma}_r$, then

$$\mathbf{x} = \mathbf{x}_r + x\mathbf{e}_{x,r} + y\mathbf{e}_{y,r} .$$

When we use the terms left and right, we shall mean left and right with respect to the local coordinate system. We denote the length of Γ_r by $h_{r,y} = \lambda(\Gamma_r)$. So, the highest local coordinate on Γ_r is $y = ½h_{r,y}$. We denote the width of the cell to the left of Γ_r - i.e. the cell to the left of the origin of the local coordinate system - by $h_{r,L}$, we denote the width of the cell to the right of Γ_r by $h_{r,R}$. So, the highest local coordinate on $\hat{\Gamma}_r$ is $x = ½h_{r,R}$. If $\mathbf{x}_r$ lies on the boundary of Ω where the global coordinate along $\hat{\Gamma}_r$ is highest, then we have a cell with width $h_{r,R}=0$ to the right of Γ_r. The same holds at the other boundary.

We construct a function ψ_r on each $\hat{\Gamma}_r$,

$$\psi_r(x) = \psi(\mathbf{x}_r + x\mathbf{e}_{x,r}) . \tag{18a}$$

By linearity, we can write this as,

$$\psi_r(x) = \beta_r x + \gamma_r \, , \; x \in [-½h_{r,L}, ½h_{r,R}] . \tag{18b}$$

We define ϕ_r to be the difference between the values of ψ in the two cell centres,

$$\phi_r(x) = \beta_r \frac{h_{r,L}+h_{r,R}}{2} . \tag{18c}$$

To have a convenient notation, we introduce a special notation σ_r for the $\mathbf{e}_{x,r}$ component of a continuous vector valued function $\boldsymbol{\sigma}$ given as a function of local coordinates, i.e.

$$\sigma_r(x,y) = \boldsymbol{\sigma}(\mathbf{x}_r + x\mathbf{e}_{x,r} + y\mathbf{e}_{y,r})\cdot\mathbf{e}_{x,r} . \tag{19}$$

3.3. Some local projections.

In this paper we need several projections that are mesh dependent. To simplify their definition, we introduce a notation for the average of a function over a given area or a given line segment. We denote the average over an area A by,

$$P[A](f) = \frac{1}{\mu(A)}\int_A f \, d\mu , \tag{20}$$

for all measurable and bounded $A \subset \Omega$ with $\mu(A) > 0$ and all f, integrable over A, where μ is the Lebesgue measure on $\mathbb{R}^2$. We denote the average over a line segment Γ by,

$$P[\Gamma](f) = \frac{1}{\lambda(\Gamma)}\int_\Gamma f \, d\lambda , \tag{21}$$

for all measurable finite line segments Γ with $\lambda(\Gamma) > 0$ and all f, integrable over Γ, that lie in $\bar{\Omega}$. Here λ is the Lebesgue measure on $\mathbb{R}$.

3.4. Some global projections and the trial spaces.

In this paper we examine the difference between the solution of (1) and a discrete approximation of that solution. To do this we need to compare a known discrete solution with an unknown continuous solution. We simplify the problem by using a projection $\Pi_h \times \bar{\mathrm{P}}_h$ of the continuous solution onto the trial space $V_h \times W_h$. The problem then reduces to the study of the interpolation error - i.e. the difference between the continuous solution and its projection - and the discretisation error - i.e difference between this projection and our discrete solution -. In general the projection can not be calculated numerically, but its properties and accuracy are known, so the problem reduces to finding a measure for the distance - in the trial space - between the projection and the discrete solution. This approach differs from the standard approach in Hilbert spaces, because the chosen projection is not necessarily orthogonal. However, the approach can also be found in Douglas and

Roberts[13]. In this section we describe the trial space $V_h \times W_h$. Using the local projections defined earlier we then construct the global projection $\Pi_h \times \overline{\mathrm{P}}_h : \mathrm{H}^1(\Omega)^2 \times \mathrm{L}^2(\Omega) \rightarrow V_h \times W_h$.

We use the lowest order Raviart-Thomas space[14] for our trial space. The subspace V_h of the trial space is spanned by vector valued functions that satisfy the homogeneous Neumann boundary conditions,

$$V_h = Span(\{ \boldsymbol{\eta}_r \mid r \in E \}) , \tag{22}$$

where the basis vectors $\boldsymbol{\eta}_r$ have a triangular prism shaped components (tent functions),

$$\boldsymbol{\eta}_r(\mathbf{x}_r + x\mathbf{e}_{x,r} + y\mathbf{e}_{y,r}) = \begin{cases} \dfrac{h_{r,L}+x}{h_{r,L}}\mathbf{e}_{x,r} \ \text{ if } \ (x,y) \in [-h_{r,L},0] \times [-\tfrac{1}{2}h_{r,y}, \tfrac{1}{2}h_{r,y}] \\ \dfrac{h_{r,R}-x}{h_{r,R}}\mathbf{e}_{x,r} \ \text{ if } \ (x,y) \in [0,h_{r,R}] \times [-\tfrac{1}{2}h_{r,y}, \tfrac{1}{2}h_{r,y}] \\ \qquad 0 \ \text{ elsewhere} \end{cases} \tag{23}$$

for all $r \in E$. For W_h, we use a space of piece wise constant functions,

$$W_h = Span(\{ \chi_{\Omega_{i-\frac{1}{2},j-\frac{1}{2}}} \mid i=1,2, \ldots ,N_1 \, , j=1,2, \ldots ,N_2 \}) , \tag{24}$$

where χ_A is the characteristic function of $A \subset \Omega$, i.e.

$$\chi_A(\mathbf{x})=1 \ \text{ if } \ \mathbf{x} \in A \ , \chi_A(\mathbf{x})=0 \ \text{ if } \ \mathbf{x} \in \Omega - A \ . \tag{25}$$

Next we define the projection $\Pi_h \times \overline{\mathrm{P}}_h$. The map $\overline{\mathrm{P}}_h : C(\Omega) \rightarrow W_h$ is a projection such that the function and its image coincide at cell centres:

$$\overline{\mathrm{P}}_h(f) = \sum_{s \in M} f(\mathbf{x}_s)\chi_{\Omega_s} \ . \tag{26}$$

and the mapping $\Pi_h : \mathrm{H}^1(\Omega)^2 \rightarrow V_h$, taken from [14], is given by,

$$\Pi_h(\mathbf{f}) = \sum_{r \in E} P[\Gamma_r](\boldsymbol{f} \cdot \mathbf{e}_{x,r})\boldsymbol{\eta}_r \ . \tag{27}$$

These are the basic ingredients for our calculations. However, we still need some other definitions associated with cell edges. We define the space E_h spanned by the characteristic functions of dual cell edges,

$$E_h = Span(\{ \chi_{\hat{\Gamma}_r} \mid r \in E \}) ,$$

and the space G_h,

$$G_h = Span(\{ \chi_{\Gamma_r} \mid \Gamma_r \subset \Gamma_1 \}) ,$$

and we introduce a map $Q_h : V_h \rightarrow E_h$, similar to $\overline{\mathrm{P}}_h$,

$$Q_h(\mathbf{f}) = \sum_{r \in E} \mathbf{f}(\mathbf{x}_r) \cdot \mathbf{e}_{x,r}\chi_{\hat{\Gamma}_r} \ . \tag{28}$$

Finally, we define the additional global projection, $\mathrm{P}_\mathrm{h} : \mathrm{L}^2(\Omega) \rightarrow W_h$,

$$\mathrm{P}_\mathrm{h}(f) = \sum_{s \in M} P[\Omega_s](f)\chi_{\Omega_s} \ . \tag{29}$$

and we notice that the pair of projections Π_h and P_h are those discussed by Raviart and Thomas[14].

4 Discretisation of the system.

We construct a scheme for the approximation of the solution (σ,u) of (1). We proceed as follows. We formulate the set of integral equations that hold for the solution of (1) and that correspond to the classical finite volume equations. We then write this set of equations as a saddle-point problem. Finally we replace exact integration by quadrature rules where appropriate.

Given a (horizontal) dual mesh edge $\hat{\Gamma}_{i,j-\frac{1}{2}}$, the following formula holds for the solution $(\boldsymbol{\sigma},u)$ of (1):

$$(\exp(\psi)u)\left[\frac{p_{i+1}+p_i}{2},\frac{q_{j-1}+q_j}{2}\right] - (\exp(\psi)u)\left[\frac{p_i+p_{i-1}}{2},\frac{q_{j-1}+q_j}{2}\right] = \qquad (A)$$

$$- \int_{x_1=\frac{p_{i-1}+p_i}{2}}^{\frac{p_i+p_{i+1}}{2}} \left[\frac{\exp(\psi)}{a}\boldsymbol{\sigma}\cdot\mathbf{e}_1\right]\left[x_1,\frac{q_{j-1}+q_j}{2}\right] dx_1 \,.$$

This follows immediately from (2a). An analogous formula is derived for a vertical dual mesh edge. In this way we find one equation for each dual mesh edge. Note that, if $\mathbf{x}_r$ lies on the Dirichlet part of the boundary, then one of the endpoints of the integration coincides with $\mathbf{x}_r$ and $u(\mathbf{x}_r)$ is given by $g(\mathbf{x}_r)$. For each cell Ω_m of the primary mesh, (2b) implies that

$$\int_{\partial\Omega_m} \boldsymbol{\sigma}\cdot\mathbf{n}_{\partial\Omega_m}\, d\lambda = \int_{\Omega_m} f\, d\mu \,. \qquad (B)$$

This gives us an equation for each cell Ω_m. The set of equations given above is the starting point for our derivation of a finite volume version of the Scharfetter-Gummel scheme. Our derivation is a variation on the derivation of a finite volume scheme as given in [15].

We introduce some notation in order to write this in the form of a saddle point problem. We define two operators $\mathscr{E}:W_h\rightarrow W_h$ and $\mathscr{E}_{\partial\Omega}:C(\partial\Omega)\rightarrow G_h$, and two bilinear forms, $\alpha_{SG}:V\times E_h\rightarrow\mathbb{R}$ and $b:V_h\times W_h\rightarrow\mathbb{R}$:

$$\mathscr{E}t_h = \sum_{s\in M} \exp(\psi(\mathbf{x}_s))t_h(\mathbf{x}_s)\chi_{\Omega_s} \quad \forall\ t\in W_h\,, \qquad (30)$$

$$(\mathscr{E}_{\partial\Omega}g)\,|_{\Gamma_r} = \exp(\psi(\mathbf{x}_r))g(\mathbf{x}_r)\chi_{\Gamma_r} \quad \forall\ \Gamma_r\subset\Gamma_1\,,$$

$$\alpha_{SG}(\boldsymbol{\sigma},Q_h\boldsymbol{\eta}_r) := \lambda(\Gamma_r)\int_{\hat{\Gamma}_r} \exp(\psi)\,\boldsymbol{\sigma}\cdot\mathbf{e}_{x,r}\,d\lambda \quad \forall\ r\in E \quad \forall\ \boldsymbol{\sigma}\in \mathrm{H}^1(\Omega)^2\,, \qquad (31)$$

and

$$b(\boldsymbol{\tau}_h,t_h) := \int_\Omega \operatorname{div}\boldsymbol{\tau}_h\ t_h\, d\mu \quad \forall\ \boldsymbol{\tau}_h\in V_h\,,\ t_h\in W_h\,.$$

Using these definitions, we can write the equations as follows,

$$\alpha_{SG}(\boldsymbol{\sigma},Q_h\boldsymbol{\eta}_r) - b(\boldsymbol{\eta}_r,\mathscr{E}\bar{\mathrm{P}}_h u) = -<\mathscr{E}_{\partial\Omega}g,\boldsymbol{\eta}_r\cdot\mathbf{n}_{\partial\Omega}> \quad \forall\ r\in E\,, \qquad (32a)$$

and

$$b(\boldsymbol{\sigma},t_h) = (f,t_h) \quad \forall\ t_h\in W_h\,. \qquad (32b)$$

Equation (32a) corresponds with (A) and gives a relation between the current along an edge and the value of u at the endpoints of that edge. Equation (32b) corresponds with (B) and gives a relation between the currents through the different edges of a given cell. Note that in the form α_{SG} the basis vector $\boldsymbol{\eta}_r$ just serves to indicate the edge over which the integration takes place. We shall use the same convention in the quadrature rule for α_{SG}. We obtain our discrete system by replacing $\boldsymbol{\sigma}$ by $\boldsymbol{\sigma}_h$, $\bar{\mathrm{P}}_h u$ by u_h and α_{SG} by a quadrature rule α_h. The discrete system has the form,

$$(\boldsymbol{\sigma}_h,u_h)\in V_h\times W_h\,,$$

$$\alpha_h(\boldsymbol{\sigma}_h,Q_h\boldsymbol{\tau}_h) - b(\boldsymbol{\tau}_h,\mathscr{E}u_h) = -<\mathscr{E}_{\partial\Omega}g,\boldsymbol{\tau}_h\cdot\mathbf{n}_{\partial\Omega}> \quad \forall\ \boldsymbol{\tau}_h\in V_h\,, \qquad (33a)$$

$$b(\boldsymbol{\sigma}_h,t_h) = (f,t_h) \quad \forall\ t_h\in W_h\,. \qquad (33b)$$

We discuss a specific quadrature rule α_h for α_{SG} in section 6. To facilitate the study of the properties of different versions of α_h, we introduce a bilinear form $(.,.)_h:V_h\times E_h\rightarrow\mathbb{R}$,

$$(\boldsymbol{\sigma}_h,Q_h\boldsymbol{\tau}_h)_h := \sum_{r\in E} \mu_r\Pi[\Gamma_r](\boldsymbol{\sigma}_h)\cdot\mathbf{e}_{x,r}\ \Pi[\Gamma_r](\boldsymbol{\tau}_h)\cdot\mathbf{e}_{x,r} \quad \forall\ \boldsymbol{\sigma}_h\,,\ \boldsymbol{\tau}_h\in V_h\,,$$

where μ_r is

$$\mu_r = \lambda(\Gamma_r)\,\lambda(\hat{\Gamma}_r) \quad \forall \; r \in E \, ,$$

and λ is the Lebesgue measure on $\mathbb{R}$. The bilinear form $(.,.)_h$ is a weighted version of the Euclidean inner product on V_h. We prove that in V_h the norm derived from this inner product and the $\mathrm{L}^2(\Omega)$-norm are equivalent.

Lemma 1.

$$\| \boldsymbol{\sigma}_h \|^2_{\mathbf{L}^2(\Omega)} \leqslant (\boldsymbol{\sigma}_h, Q_h \boldsymbol{\sigma}_h)_h \leqslant 3 \| \boldsymbol{\sigma}_h \|^2_{\mathbf{L}^2(\Omega)} \, ,$$

where $\mathbf{L}^2(\Omega) = \mathrm{L}^2(\Omega)^2$.

Proof.

We start by determining the value of $(\boldsymbol{\sigma}_h, \boldsymbol{\sigma}_h)_{\mathbf{L}^2(\Omega)}$. To simplify matters, we introduce coordinates s_r for $\boldsymbol{\sigma}_h$ with respect to the basis $\boldsymbol{\eta}_r$ given in (23) and we split $\boldsymbol{\sigma}_h$ into mutually orthogonal $\mathbf{e}_i$ parts $(i = 1,2)$,

$$\boldsymbol{\sigma}_{1,j} = \sum_{i=0}^{N_1} \boldsymbol{\eta}_{i,j-½} s_{i,j-½} \, ,$$

and

$$\boldsymbol{\sigma}_{2,i} = \sum_{j=0}^{N_2} \boldsymbol{\eta}_{i-½,j} s_{i-½,j} \, .$$

Now,

$$(\boldsymbol{\sigma}_h, \boldsymbol{\sigma}_h)_{\mathbf{L}^2(\Omega)} = \sum_{i=1}^{N_1} (\boldsymbol{\sigma}_{2,i}, \boldsymbol{\sigma}_{2,i})_{\mathbf{L}^2(\Omega)} + \sum_{j=1}^{N_2} (\boldsymbol{\sigma}_{1,j}, \boldsymbol{\sigma}_{1,j})_{\mathbf{L}^2(\Omega)} \, .$$

We see immediately, that

$$(\boldsymbol{\sigma}_{1,j}, \boldsymbol{\sigma}_{1,j})_{\mathbf{L}^2(\Omega)} = \sum_{i=1}^{N_1} \frac{1}{3} (s^2_{i-1,j-½} + s^2_{i,j-½} + s_{i-1,j-½} s_{i,j-½}) \mu(\Omega_{i-½,j-½}) \, ,$$

so

$$\frac{1}{6} \sum_{i=1}^{N_1} (s^2_{i-1,j-½} + s^2_{i,j-½}) \mu(\Omega_{i-½,j-½}) \leqslant (\boldsymbol{\sigma}_{1,j}, \boldsymbol{\sigma}_{1,j})_{\mathbf{L}^2(\Omega)} \leqslant \frac{1}{2} \sum_{i=1}^{N_1} (s^2_{i-1,j-½} + s^2_{i,j-½}) \mu(\Omega_{i-½,j-½}) \, .$$

Furthermore,

$$(\boldsymbol{\sigma}_{1,j}, \boldsymbol{\sigma}_{1,j})_h = \frac{1}{2} \sum_{i=0}^{N_1-1} \mu(\Omega_{i+½,j-½}) s^2_{i,j-½} + \frac{1}{2} \sum_{i=1}^{N_1} \mu(\Omega_{i-½,j-½}) s^2_{i,j-½} \, .$$

□

5 Existence and uniqueness of the solution.

In this section we give sufficient conditions for the existence and uniqueness of the solution of the discrete system.

We plan to use theorem 3.1 from Nicolaides[5] to prove existence, uniqueness and stability for the discrete scheme. To apply the theorem, we need to define norms on our discrete spaces and to verify the conditions (2.1, 2,2, 3.1 and 3.2) given in [5]. We shall use the norms associated with the following inner products on V_h and W_h,

$$(\boldsymbol{\sigma}_h, \boldsymbol{\tau}_h)_{V_h} = (\boldsymbol{\sigma}_h, \boldsymbol{\tau}_h)_h + (\operatorname{div} \boldsymbol{\sigma}_h, \operatorname{div} \boldsymbol{\tau}_h)_{\mathrm{L}^2(\Omega)} \quad \forall \; \boldsymbol{\sigma}, \boldsymbol{\tau} \in V_h \, , \tag{34a}$$

and

$$(u_h,t_h)_{W_h} = (u_h,t_h)_{L^2(\Omega)} \quad \forall \ u_h,t_h \in W_h \ . \tag{34b}$$

The conditions 1, 2 and 3 that follow are equivalent to the conditions (2.1, 2,2, 3.1 and 3.2) imposed by Nicolaides.

Condition 1.

The bilinear form α_h is bounded, i.e. there is a $0 < A \in \mathbb{R}$, independent of the mesh, such that

$$\alpha_h(\sigma_h,\tau_h) \leqslant A \, \| \sigma_h \|_{V_h} \| \tau_h \|_{V_h} ,$$

and α_h is coercive on the kernel of the divergence operator in V_h, i.e. there exists a $0 < \delta \in \mathbb{R}$, independent of the mesh, such that

$$\delta \, (\sigma_h,\sigma_h)_h \leqslant \alpha_h(\sigma_h,\sigma_h) ,$$

for all $\sigma_h \in V_h \bigcap \mathscr{N}(\mathrm{div})$. Our condition 1 corresponds to conditions (3.1) and (3.2) in the paper by Nicolaides[5].

Condition 2

The bilinear form b is bounded and there exists a $0 < \gamma' \in \mathbb{R}$, independent of the mesh, such that

$$\sup_{0 \neq \tau_h \in V_h} \frac{| b(\tau_h,t_h) |}{\| \tau_h \|_{V_h}} \geqslant \gamma' \| t_h \|_{L^2(\Omega)} \ .$$

Condition 3

There exists a $0 < \gamma \in \mathbb{R}$, independent of the mesh, such that

$$\sup_{0 \neq w \in W_h} \frac{| b(\sigma,w) |}{\| w \|_{W_h}} \geqslant \gamma \inf_{z \in (\mathscr{N}(\mathrm{div}) \bigcap V_h)} \| \sigma - z \|_{V_h} \quad \forall \ \sigma \in V_h \ .$$

Our conditions 2 and 3 correspond to conditions (2.1) and (2.2) in [5]. We can now give a version of theorem 3.1 of Nicolaides.

Theorem 1.

If the conditions 1, 2 and 3 are satisfied, then the discrete system (33) has a unique solution and the norm of the solution is bounded by,

$$\| \sigma_h \|_{V_h} \leqslant \frac{1}{\delta} \| \mathscr{E} g \|_{V_h'} + \frac{1}{\gamma} \left(1 + \frac{A}{\delta} \right) \| f \|_{L^2(\Omega)} ,$$

$$\| \mathscr{E} u_h \|_{L^2(\Omega)} \leqslant \frac{1}{\gamma'} \left(A \| \sigma_h \|_{V_h} + \| \mathscr{E} g \|_{V_h'} \right) .$$

Proof.

The proof is a direct application of theorem 3.1 in [5].

□

Now, we have to ask ourselves when these conditions are satisfied. In section 6, we shall introduce an α_h that satisfies condition 1. Because the remaining two conditions are not easily verified in the form given here, we give alternative conditions 2a and 3a that are easier to verify. Lemma 2 shows that 2a and 3a imply 2 and 3.

Condition 2a.

The corresponding Poisson problem is regular, i.e. Ω, $\Gamma_1, \Gamma_2 \subset \partial\Omega$ are such that there exists a $C > 0$, $C \in \mathbb{R}$ such that

$$\begin{gathered}
\forall\ f \in \mathrm{L}^2(\Omega)\ \exists!\ u \in \mathrm{H}^2(\Omega) : \\
\Delta u = f \ \text{ on } \Omega , \\
u = 0 \ \text{ on } \Gamma_1 , \\
\mathbf{grad}\, u \cdot \mathbf{n}_{\partial\Omega} = 0 \ \text{ on } \Gamma_2 , \\
\| u \|_{\mathrm{H}^2(\Omega)} \leqslant C \| f \|_{\mathrm{L}^2(\Omega)} ,
\end{gathered}$$

Condition 3a

The map Π_h has the following approximation property, there is a $K > 0$, $K \in \mathbb{R}$, independent of the mesh, such that

$$\| \mathbf{grad}\, u - \Pi_h \mathbf{grad}\, u \|_{\mathrm{L}^2(\Omega)} \leqslant KH \| u \|_{\mathrm{H}^2(\Omega)} ,$$

where H is the maximum mesh diameter.

Assuming that 2a and 3a do indeed imply 2 and 3, it remains to see when 2a and 3a are satisfied. For condition 2a we refer to Grisvard[16]. Condition 3a follows almost immediately from the assumption that $\boldsymbol{\sigma}$ has components in $\mathrm{H}^1(\Omega)$. To illustrate this, we prove Lemma 3. This lemma proves that Π_h has the approximation property.

Lemma 2.
If all mesh edges have a length that is bounded above by a constant H_0, then the conditions 2a and 3a imply conditions 2 and 3 with $\gamma = \gamma' = \dfrac{1}{3C(1+KH_0)}$.

Proof.
Assume that (2a) and (3a) hold and take a fixed $w \in W_h$. If we solve the Poisson problem for $f = w$ then, according to (2a), the solution u_w satisfies,

$$\| \mathbf{grad}\, u_w \|_{\mathrm{H}(\mathrm{div};\Omega)} \leqslant C \| w \|_{\mathrm{L}^2(\Omega)} ,$$

and

$$(\mathrm{div}\, \mathbf{grad}\, u_w , w) = \| w \|^2_{\mathrm{L}^2(\Omega)} .$$

Furthermore, (3a) implies that for all $u \in \mathrm{H}^2(\Omega)$

$$\| \mathbf{grad}\, u - \Pi_h \mathbf{grad}\, u \|_{\mathrm{L}^2(\Omega)} \leqslant KH_0 \| u \|_{\mathrm{H}^2(\Omega)} .$$

So we find, that

$$\mathrm{P_h}\, \mathrm{div}\, \mathbf{grad}\, u_w = \mathrm{P_h} w ,$$

and

$$\| \Pi_h \mathbf{grad}\, u_w \|_{\mathrm{H}(\mathrm{div};\Omega)} \leqslant H_0 K \| u_w \|_{\mathrm{H}^2(\Omega)} + \| \mathbf{grad}\, u_w \|_{\mathrm{L}^2(\Omega)} + \| \mathrm{div}\, \Pi_h \mathbf{grad}\, u_w \|_{\mathrm{L}^2(\Omega)} .$$

Our w lies in W_h, so special properties of $\mathrm{P_h}$ and Π_h imply,

$$\mathrm{div}\, \Pi_h \mathbf{grad}\, u_w = \mathrm{P_h} w = w .$$

We combine this with condition (3a) to find,

$$\begin{gathered}
\| \Pi_h \mathbf{grad}\, u_w \|_{\mathrm{H}(\mathrm{div};\Omega)} \leqslant \\
\| u_w \|_{\mathrm{H}^2(\Omega)} + H_0 K \| u_w \|_{\mathrm{H}^2(\Omega)} \leqslant C(1 + H_0 K) \| w \|_{\mathrm{L}^2(\Omega)} .
\end{gathered}$$

We see immediately, that

$$\| \mathrm{div} \|_{\mathscr{L}(V_h, W_h)} \leqslant 1 ,$$

and

$$\forall\ w \in W_h \ \exists\ \boldsymbol{\tau}_h \in V_h : \operatorname{div}\boldsymbol{\tau} = w \quad\text{and}\quad \|\boldsymbol{\tau}_h\|_{\mathrm{H(div;\Omega)}} \leqslant C(1+H_0K)\|w\|_{\mathrm{L}^2(\Omega)} .$$

Lemma 1 implies, that

$$\|\boldsymbol{\tau}_h\|_{\mathrm{H(div;\Omega)}} \leqslant \|\boldsymbol{\tau}_h\|_{V_h} \leqslant \sqrt{3}\,\|\boldsymbol{\tau}_h\|_{\mathrm{H(div;\Omega)}} \quad \forall\ \boldsymbol{\tau}_h \in V_h .$$

We find, that

$$\frac{|b(\Pi_h \,\mathbf{grad}\, u_w, w)|}{\|\Pi_h\,\mathbf{grad}\, u_w\|_{V_h}} \geqslant \frac{1}{3C(1+KH_0)}\|w\|_{\mathrm{L}^2(\Omega)} .$$

Now suppose $\boldsymbol{\sigma}_h \in V_h$. Then the above derivation implies, that there is a $\boldsymbol{\tau}_h \in V_h$ such that $\operatorname{div}\boldsymbol{\tau}_h = \operatorname{div}\boldsymbol{\sigma}_h$ and $\|\boldsymbol{\tau}_h\|_{\mathrm{H(div;\Omega)}} \leqslant C(1+H_0K)\|\operatorname{div}\boldsymbol{\sigma}_h\|_{\mathrm{L}^2(\Omega)}$. Moreover, $\boldsymbol{\tau}_h - \boldsymbol{\sigma}_h \in \mathscr{N}(\operatorname{div})$, so

$$\inf_{z_h \in V_h \bigcap \mathscr{N}(\operatorname{div})} \|\boldsymbol{\sigma}_h + z_h\|_{V_h} \leqslant \|\boldsymbol{\tau}_h\|_{V_h} \leqslant 3C(1+H_0K)\|\operatorname{div}\boldsymbol{\sigma}_h\|_{\mathrm{L}^2(\Omega)} .$$

So we find,

$$\frac{(\operatorname{div}\boldsymbol{\sigma}_h, \operatorname{div}\boldsymbol{\sigma}_h)_{\mathrm{L}^2(\Omega)}}{\|\operatorname{div}\boldsymbol{\sigma}_h\|_{\mathrm{L}^2(\Omega)}} = \|\operatorname{div}\boldsymbol{\sigma}\|_{\mathrm{L}^2(\Omega)} \geqslant \frac{1}{3C(1+H_0K)} \inf_{z_h \in (\mathscr{N}(\operatorname{div}) \bigcap V_h)} \|\boldsymbol{\sigma}_h - z_h\|_{V_h} \quad \forall\ \boldsymbol{\sigma} \in {}_h V_h .$$

This implies,

$$\sup_{0 \neq w \in W_h} \frac{|b(\boldsymbol{\sigma},w)|}{\|w\|_{W_h}} \geqslant \frac{1}{3C(1+H_0K)} \inf_{\boldsymbol{z} \in (\mathscr{N}(\operatorname{div}) \bigcap V_h)} \|\boldsymbol{\sigma} - \boldsymbol{z}\|_{V_h} \quad \forall\ \boldsymbol{\sigma} \in V_h .$$

□

Lemma 3.
If f is a square integrable function with square integrable derivatives on a rectangle $\Omega=[0,h_1]\times[0,h_2]$ with sides $\Gamma_{1,1}=\{\,h_1\,\}\times[0,h_2]$, $\Gamma_{2,1}=[0,h_1]\times\{\,h_2\,\}$, $\Gamma_{1,0}=\{\,0\,\}\times[0,h_2]$ and $\Gamma_{2.0}=[0,h_1]\times\{\,0\,\}$, then the following inequality holds for all $s \in \mathrm{L}^\infty([0,h_1])$ and $\mathscr{R}(s) \subset [0,1]$,

$$\|f-(1-s)\Pi[\Gamma_{1,0}]f - s\Pi[\Gamma_{1,1}]f\|_{\mathrm{L}^2(\Omega)} \leqslant \sqrt{2}(h_1^2+h_2^2)^{1/2}\|\mathbf{grad} f\|_{\mathrm{L}^2(\Omega)} .$$

Proof.
We start by proving the above inequality for $f \in C^1(\Omega)$. Then we can extend the inequality by density to $\mathrm{H}^1(\Omega)$. We see that,

$$\|f-(1-s)\Pi[\Gamma_{1,0}]f - s\Pi[\Gamma_{1,1}]f\|_{\mathrm{L}^2(\Omega)}^2 =$$

$$\int_{x=0}^{h_1}\int_{y=0}^{h_2}\left(\frac{1}{h_2}\int_{z=0}^{h_2}\Big[(1-s(x))(f(x,y)-f(0,z)) + s(x)(f(x,y)-f(h_1,z))\Big]dz\right)^2 dxdy .$$

We use partial derivatives to rewrite the expression,

$$\|f-(1-s)\Pi[\Gamma_{1,0}]f - s\Pi[\Gamma_{1,1}]f\|_{\mathrm{L}^2(\Omega)}^2 =$$

$$\int_{x=0}^{h_1}\int_{y=0}^{h_2}\left(\frac{1}{h_2}\int_{z=0}^{h_2}\left[(1-s(x))\left\{\int_{a=0}^{x}\frac{\partial f}{\partial a}(a,z)da + \int_{b=z}^{y}\frac{\partial f}{\partial b}(x,b)db\right\} + \right.\right.$$

$$\left.\left. s(x)\left\{\int_{a=h_1}^{x}\frac{\partial f}{\partial a}(a,z)da + \int_{b=z}^{y}\frac{\partial f}{\partial b}(x,b)db\right\}\right]dz\right)^2 dxdy .$$

We use Hölder and extend the integrals where appropriate,

$$\|f-(1-s)\Pi[\Gamma_{1,0}]f-s\Pi[\Gamma_{1,1}]f\|_{L^2(\Omega)}^2 \leqslant$$
$$\int_{x=0}^{h_1}\int_{y=0}^{h_2}\left[\frac{h_1^{1/2}}{h_2^{1/2}}\|\partial f/\partial x_1\|_{L^2(\Omega)} + h_2^{1/2}\left[\int_{b=0}^{h_2}\left[\frac{\partial f}{\partial b}(x,b)\right]^2 db\right]^{1/2}\right]^2 dxdy\ .$$

We use $(|A|+|B|)^2 \leqslant 2(A^2+B^2)$ to write this as,

$$\|f-(1-s)\Pi[\Gamma_{1,0}]f-s\Pi[\Gamma_{1,1}]f\|_{L^2(\Omega)}^2 \leqslant$$
$$2\int_{x=0}^{h_1}\int_{y=0}^{h_2}\left[\frac{h_1}{h_2}\|\partial f/\partial x_1\|_{L^2(\Omega)}^2 + h_2\int_{b=0}^{h_2}\left[\frac{\partial f}{\partial b}(x,b)\right]^2 db\right] dxdy\ .$$

This reduces to,

$$\|f-(1-s)\Pi[\Gamma_{1,0}]f-s\Pi[\Gamma_{1,1}]f\|_{L^2(\Omega)}^2 \leqslant$$
$$2h_1^2\|\partial f/\partial x_1\|_{L^2(\Omega)}^2 + 2h_2^2\|\partial f/\partial x_2\|_{L^2(\Omega)}^2\ .$$

□

6 The quadrature rule.

In the previous section we left open the choice of the quadrature rule for the computation of α_h. In this section we select a quadrature rule and we check whether it meets condition 1 from section 5. If it is to satisfy the normal addition rules for integrals, the quadrature rule must respect the local support and vector character of the basis vector functions given in (24), so it must satisfy,

$$\alpha_h(\boldsymbol{\eta}_{i,j-1/2}, Q_h\boldsymbol{\eta}_{k-1/2,l}) \equiv 0\ ,\ j\neq l \Rightarrow \alpha_h(\boldsymbol{\eta}_{i,j-1/2}, Q_h\boldsymbol{\eta}_{k,l-1/2}) \equiv 0\ ,\ i\neq k \Rightarrow \alpha_h(\boldsymbol{\eta}_{i-1/2,j}, Q_h\boldsymbol{\eta}_{k-1/2,l}) \equiv 0\ .$$

On $\hat{\Gamma}_r$ we use a one-point rule with $\mathbf{x}_r$ as nodal point. We choose the weight at the node in such a way, that

$$\alpha_{h,1}(\mathbf{e}_{x,r}, Q_h\boldsymbol{\eta}_r) = \alpha_{SG}(\mathbf{e}_{x,r}, Q_h\boldsymbol{\eta}_r)\ , \tag{35a}$$

i.e. the rule is exact for all $\boldsymbol{\sigma}$ that have a constant component along $\hat{\Gamma}_r$.

The discretisation obtained in this way can be derived in several other ways, see e.g. the papers by Bank et al.[3, 4], the discretisation is closely related to the method given by MacNeal[15]. If we use the quadrature rule given above, we find the following formula for $\alpha_{h,1}(\boldsymbol{\eta}_r, Q_h\boldsymbol{\eta}_s)$,

$$\alpha_{h,1}(\boldsymbol{\eta}_r, Q_h\boldsymbol{\eta}_s) = \delta_{rs}\lambda(\Gamma_r)\int_{\hat{\Gamma}_r}\frac{\exp(\psi)}{a}\,d\lambda\ , \tag{35b}$$

this shows that the corresponding matrix is a diagonal matrix. It is clear that this rule corresponds to the use of a Scharfetter-Gummel scheme for each of the two directions $\mathbf{e}_1$ and $\mathbf{e}_2$ separately.

Lemma 4.
If ψ is piecewise linear, then

$$\|\boldsymbol{\sigma}_h\|_{L^2(\Omega)}^2 \min_{r\in E} P[\hat{\Gamma}_r](\exp(\psi)/a) \leqslant \alpha_{h,1}(\boldsymbol{\sigma}_h,\boldsymbol{\sigma}_h) \leqslant 3\|\boldsymbol{\sigma}_h\|_{L^2(\Omega)}^2 \max_{r\in E} P[\hat{\Gamma}_r](\exp(\psi)/a)\ .$$

Proof.
This follows immediately from (35b), the definition of $(.,.)_h$ and lemma 1.

□

So, formally $\alpha_{h,1}$ satisfies condition 1 from section 5 and we can apply theorem 1 from section 5 to the discrete scheme based on this quadrature rule. We use the word "formally" to indicate that the constant A in condition 1 may need to be very large. This is due to the appearance of the exponential weighting function in the nodal weight for the quadrature rule. In general we shall use the words "formal" and "formally" to indicate that certain statements hold, but only for very small h.

7 Consistency.

As discussed in section 3.4, we use a projection onto the trial space to split the difference between the solution of (1) and its discrete approximation into an interpolation error and a discretisation error as follows,

$$\|\boldsymbol{\sigma}-\boldsymbol{\sigma}_h\|_{\mathrm{H(div;\Omega)}} \leqslant \|\boldsymbol{\sigma}-\Pi_h\boldsymbol{\sigma}\|_{\mathrm{H(div;\Omega)}} + 3\|\Pi_h\boldsymbol{\sigma}-\boldsymbol{\sigma}_h\|_{V_h} ,$$

$$\|U-\mathscr{E}u_h\|_{\mathrm{L}^2(\Omega)} \leqslant \|U-\bar{\mathrm{P}}_h U\|_{\mathrm{L}^2(\Omega)} + \|\bar{\mathrm{P}}_h U-\mathscr{E}u_h\|_{W_h} .$$

The interpolation error can be estimated by standard approximation theory. Here we study the discretisation error.

7.1. Effects of piecewise bilinear interpolation for ψ.

At the start of section 3 we assumed that ψ was piecewise bilinear. If this does not hold then we can estimate the error caused by approximation of ψ with the aid of the following lemma.

Lemma 5.
If $\psi \in C^2([0,h])$ and we replace ψ in

$$\int_0^h \exp(\psi)\, d\lambda ,$$

by ψ_I, defined as

$$\psi_I(x) = \frac{h-x}{h}\psi(0) + \frac{x}{h}\psi(h) \quad \forall \;\; x \in [0,h] ,$$

then

$$\left|\int_0^h \exp(\psi)\, d\lambda - \int_0^h \exp(\psi_I)\, d\lambda\right| \leqslant \left|\int_0^h \exp(\psi_I)\, d\lambda\right| \left[\exp(h^2\,\|d^2\psi/dx^2\|_{\mathrm{L}^\infty([0,h])}) - 1\right] .$$

Proof.
We start by giving an estimate for

$$\|\psi-\psi_I\|_{\mathrm{L}^\infty([0,h])} .$$

If $\psi \in C^2([0,h])$, then

$$\psi(x)-\psi_I(x) = \psi(0)+\int_{y=0}^{x} \frac{d\psi}{dx}(y)dy - \psi(0) - \frac{x}{h}\int_{z=0}^{h} \frac{d\psi}{dx}(z)dz =$$

$$\frac{1}{h}\int_{z=0}^{h}\int_{y=0}^{x} \frac{d\psi}{dx}(y) - \frac{d\psi}{dx}(z)dydz = \frac{1}{h}\int_{z=0}^{h}\int_{y=0}^{x}\int_{w=z}^{y} \frac{d^2\psi}{dx^2}(w)dwdydz =$$

$$\frac{1}{h}\int_{z=0}^{h}\int_{y=0}^{x}\int_{w=0}^{y} \frac{d^2\psi}{dx^2}(w)dwdydz - \frac{1}{h}\int_{z=0}^{h}\int_{y=0}^{x}\int_{w=0}^{z} \frac{d^2\psi}{dx^2}(w)dwdydz$$

$$= \int_{y=0}^{x} (x-y)\,\frac{d^2\psi}{dx^2}(y)dy - \frac{x}{h}\int_{z=0}^{h} (h-z)\frac{d^2\psi}{dx^2}(z)dz .$$

This implies that

$$\|\psi-\psi_I\|_{\mathrm{L}^\infty([0,h])} \leqslant h^2\,\|d^2\psi/dx^2\|_{\mathrm{L}^\infty([0,h])} .$$

We combine integrals and reorder terms to find,

$$\left|\int_0^h \exp(\psi)\, d\lambda - \int_0^h \exp(\psi_I)\, d\lambda\right| = \left|\int_0^h \exp(\psi_I)\,[\exp(\psi-\psi_I)-1]\, d\lambda\right| .$$

We use our estimate for

$$||\psi - \psi_I||_{L^\infty([0,h])} \cdot$$

and move the resulting constant term out of the integral to find the desired estimate.

□

This implies that the approximation of ψ by its bilinear interpolator causes a relative error in the coefficients of our quadrature rules that is formally of order $\mathcal{O}(h^2)$, where h is the maximum edge length and the error constant is dependent on $\partial^2\psi/\partial x_1^2$ and $\partial^2\psi/\partial x_2^2$. As we shall see in section 7.2, this is comparable in order to the local error resulting from the use of the quadrature rule $\alpha_{h,1}$.

7.2. The discretisation error.

In section 5, theorem 1, we gave an expression for the norm of the solution of a saddle-point system in terms of the right hand side. If we insert of the difference between the projection $(\Pi_h\sigma, \bar{P}_h u)$ of the solution of (1) and the solution (σ_h, u_h) of (33) into the saddle-point problem corresponding to the discrete system, the norm of the right hand side is given by

$$\sup_{0 \neq \tau_h \in V_h} \frac{|\alpha_h(\Pi_h\sigma, Q_h\tau_h) - \alpha_{SG}(\sigma, Q_h\tau_h)|}{||\tau_h||_{V_h}} , \tag{36}$$

for (33a) and 0 for the (33b). In this section, we consider this expression for $\alpha_h = \alpha_{h,1}$ and $a \equiv 1$.

We consider the above expression for $\tau_h = \eta_r$ and express it in the local coordinates and local functions defined in section 3.2. As we consider the expression for one fixed edge $\hat{\Gamma}_r$, we may omit the subscript r. The two bilinear forms of interest take the following form,

$$\alpha_{SG}(\sigma, Q_h\eta_r) = h_y \int_{x=-\frac{1}{2}h_L}^{\frac{1}{2}h_R} \sigma(x,0) \exp(\beta x + \gamma)\, dx , \tag{37}$$

$$\alpha_{h,1}(\sigma, Q_h\eta_r) = \left[\int_{y=-\frac{1}{2}h_y}^{\frac{1}{2}h_y} \sigma(0,y)\, dy \right] \int_{x=-\frac{1}{2}h_L}^{\frac{1}{2}h_R} \exp(\beta x + \gamma)\, dx . \tag{38}$$

We assume, that σ_1, σ_2 are elements of $C^4(\bar{\Omega})$. In the following lemma we give a formula for the difference between (37) and (38) for an arbitrary vector-valued function $\sigma \in C^4(\bar{\Omega})^2$. To simplify notation, we introduce the moments of $\exp(\psi)$ on all dual mesh edges,

$$\tilde{L}_{r,n} = \int_{x=-\frac{1}{2}h_{r,L}}^{\frac{1}{2}h_{r,R}} x^n \exp(\psi_r)\, dx ,$$

we see immediately, that

$$\alpha_{h,1}(\eta_r, Q_h\eta_r) = \lambda(\Gamma_r)\tilde{L}_{r,0} .$$

We also introduce scaled versions of these moments,

$$L_{r,n} = \frac{\tilde{L}_{r,n}}{\tilde{L}_{r,0}} = \frac{\displaystyle\int_{x=-\frac{1}{2}h_{r,L}}^{\frac{1}{2}h_{r,R}} x^n \exp(\beta_r x)\, dx}{\displaystyle\int_{x=-\frac{1}{2}h_{r,L}}^{\frac{1}{2}h_{r,R}} \exp(\beta_r x)\, dx} .$$

Lemma 6.

We consider a vector-valued function σ in local coordinates around $\mathbf{x}_r$. Let $H = \max(\lambda(\Gamma_r), \lambda(\hat{\Gamma}_r))$. If we assume that $\sigma_r \in C^4(\mathbb{R}^2)$, then we can expand σ_r in a Taylor series around the origin of the local coordinate system, as r is fixed we omit the subscript r on σ, x and y. We write σ_x, σ_y for the partial derivatives in the local x and y directions.

$$\alpha_{SG}(\boldsymbol{\sigma},Q_h\boldsymbol{\eta}_r) - \alpha_{h,1}(\boldsymbol{\sigma},Q_h\boldsymbol{\eta}_r) = \tag{39}$$

$$\alpha_{h,1}(\boldsymbol{\eta}_r,Q_h\boldsymbol{\eta}_r)\left[\sigma_x(0,0)L_{r,1} + ½\sigma_{xx}(0,0)L_{r,2} - \right.$$

$$\left. ½f_{yy}(0,0)\frac{h_y{}^2}{12} + \frac{1}{6}\sigma_{xxx}(0,0)L_{r,3} + \frac{1}{24}\sigma_{xxxx}(\mu_x,0)L_{r,4} - \frac{1}{24}\sigma_{yyyy}(0,\mu_y)\frac{h_y{}^4}{80}\right] ,$$

with $\mu_x \in [-½h_L, ½h_R]$, $\mu_y \in [-½h_y, ½h_y]$.
Proof.
to verify this, we subtract (38) from (37), expand all occurrences of σ in Taylor series around the local origin and carry out all integrations over y. After integration, we are left with the above expression for $\alpha_{SG} - \alpha_{h,1}$,

□

The form of this expression and the earlier expression for the error due to bilinear approximation of ψ suggest that the formal order behaviour is best studied by dividing the part of the error corresponding to a given dual edge $\hat{\Gamma}_r$ by $\alpha_{h,1}(\boldsymbol{\eta}_r,Q_h\boldsymbol{\eta}_r)$.

The use of a one point rule for α_h can affect accuracy. We specify three cases where

$$\frac{(\alpha_{SG}-\alpha_{h,1})(\boldsymbol{\sigma},Q_h\boldsymbol{\eta}_r)}{\alpha_{h,1}(\boldsymbol{\eta}_r,Q_h\boldsymbol{\eta}_r)} ,$$

may be $\mathcal{O}(h)$ in stead of $\mathcal{O}(h^2)$.

Case I.

If $\Gamma_r \subset \Gamma_1$, i.e. we are dealing with an edge on the Dirichlet boundary, then there are $\mu \in (0,1)$, $k_r \in [-K,K]$, where K depends only on the $L^\infty(\Omega)$ norm of derivatives of $\boldsymbol{\sigma}$, such that

$$\frac{|\alpha_{SG}(\boldsymbol{\sigma},Q_h\boldsymbol{\eta}_r) - \alpha_{h,1}(\Pi_h\boldsymbol{\sigma},Q_h\boldsymbol{\eta}_r)|}{\alpha_{h,1}(\boldsymbol{\eta}_r,Q_h\boldsymbol{\eta}_r)} = \mu H\sigma_x(0,0) + k_rH^2 ,$$

this contains a first order error term in the right hand side.

Case II.

If a vertical edge Γ_r lies in the interior on Ω and the width of the cell on the left side of he edge differs from that of the cell on the right side by more than a factor of order $\mathcal{O}(H^2)$, then there are $\mu \in (0,1)$, $k_r \in [-K,K]$, where K depends only on the $L^\infty(\Omega)$ norm of derivatives of $\boldsymbol{\sigma}$, such that

$$\frac{|\alpha_{SG}(\boldsymbol{\sigma},Q_h\boldsymbol{\eta}_r) - \alpha_{h,1}(\Pi_h\boldsymbol{\sigma},Q_h\boldsymbol{\eta}_r)|}{\alpha_{h,1}(\boldsymbol{\eta}_r,Q_h\boldsymbol{\eta}_r)} = \mu H\sigma_x(0,0) + k_rH^2 ,$$

because the first order error terms for these cells do not cancel even when $\beta=0$.

Case III.

Lastly, if an interval vertical edge Γ_r lies in the interior on Ω and the jump in ψ over the edge is larger than 2, then there is a $k_r \in [-K,K]$, where K depends only on the $L^\infty(\Omega)$ norm of derivatives of $\boldsymbol{\sigma}$, such that

$$\frac{|\alpha_{SG}(\boldsymbol{\sigma},Q_h\boldsymbol{\eta}_r) - \alpha_{h,1}(\Pi_h\boldsymbol{\sigma},Q_h\boldsymbol{\eta}_r)|}{\alpha_{h,1}(\boldsymbol{\eta}_r,Q_h\boldsymbol{\eta}_r)} = C\sigma_x(0,0) + k_rH^2 ,$$

because of the asymmetry of $\exp(\psi)$. The coefficient C of $\sigma_x(0,0)$ is given by

$$C = L_{r,1} ,$$

this is equivalent to

$$L_{r,1} = \tfrac{1}{2}\left[\tfrac{1}{2}(h_R-h_L) + \tfrac{1}{2}(h_R+h_L)G\left[\beta\frac{h_R+h_L}{4}\right]\right] ,$$

with

$$G(z) = \frac{z\cosh z - \sinh z}{z\sinh z} .$$

We see that,

$$\frac{dG}{dz}(z) = \frac{(\sinh z)^2 - z^2}{z^2(\sinh z)^2} .$$

The behaviour of G is as follows,

$$G(z) = -G(-z) ,$$

$$\lim_{z\to\infty} G(z) = 1 ,$$

$$0 \leqslant \frac{dG}{dz}(z) < 1 ,$$

so

$$-1 \leqslant G(z) \leqslant 1 .$$

If we assume that $h_R = h_L$, then

$$L_{r,1} = \tfrac{1}{2}(h_R+h_L)G\left[\beta\frac{h_R+h_L}{4}\right] .$$

So the order behaviour of $L_{r,1}$ is determined by the order behaviour of G. Assume that $h = h_L + h_R < 1$. The order behaviour of G is as follows. If $\beta h < 2$, then

$$G(\tfrac{1}{2}\beta h) \leqslant \frac{\tfrac{1}{2}\beta h(1+\tfrac{1}{2}(\tfrac{1}{2}\beta h)^2\cosh(1)) - \tfrac{1}{2}\beta h}{(\tfrac{1}{2}\beta h)^2} = \frac{\beta h\cosh(1)}{2} < \frac{\beta\cosh(1)}{2}h ,$$

so $G(\tfrac{1}{2}\beta h)$ is $\mathcal{O}(h)$. On the other hand, as long as $\beta h > 20$,

$$G(\tfrac{1}{2}\beta h) \geqslant G(10) = \frac{10\cosh(10) - \sinh(10)}{10\sinh(10)} > \frac{9}{10} ,$$

so for all meshes with $h > 20/\beta$, we have $G(\tfrac{1}{2}\beta h)$ is $\mathcal{O}(1)$. We see that $L_{r,1}$ is at worst $\mathcal{O}(h)$ and at best $\mathcal{O}(h^2)$. If $h < |2/\beta|$ then $L_{r,1}$ is $\mathcal{O}(H^2)$. If $h > |20/\beta|$ then is $\mathcal{O}(H)$.

8 An a-posteriori error estimator.

In this section we study an a-posteriori error estimate for the discretisation based on $\alpha_h = \alpha_{h,1}$. We calculate a correction to an initial solution and use this to improve the order of approximation, this method is related to the deferred correction scheme as described by Fox and Mayers in chapter 6 of [17].

8.1. A derivation of a deferred correction scheme.

In this section we give a deferred correction scheme. The discussion takes into account formal order only, i.e. it assumes that h is "small enough".

In equation (33) we take $\alpha_{h,1}$, given in (35b), as our α_h. If we insert $(\Pi_h\sigma - \sigma_h, \bar{\mathrm{P}}_h u - u_h)$ in (33) to determine an expression for the right hand side, then we find

$$\alpha_{h,1}(\Pi_h\sigma - \sigma_h, Q_h\tau_h) - (\mathrm{div}\,\tau_h, \mathcal{E}(\bar{\mathrm{P}}_h u - u_h)) = \qquad (40a)$$

$$\alpha_{h,1}(\Pi_h\sigma, Q_h\tau_h) - \alpha_{SG}(\sigma, Q_h\tau_h) \quad \forall \;\; \tau_h \in V_h ,$$

$$(\mathrm{div}\,(\Pi_h\sigma - \sigma_h), t_h) = 0 \quad \forall \;\; t_h \in W_h . \qquad (40b)$$

Our approach is the following. We assume that $\sigma_1,\sigma_2 \in C^4(\bar{\Omega})$ and we assume that we have Dirichlet boundary conditions on the entire boundary of our domain. We see from (39) that we can approximate the right hand side of (40a) by an expression in the partial derivatives of $\boldsymbol{\sigma}$. If we can justify the use of finite difference approximations based on $\boldsymbol{\sigma}_h$ for these derivatives, then we can solve (33) with a adjusted right hand side and obtain a better solution. First we show that we can approximate partial derivatives of $\boldsymbol{\sigma}$ of first or second order in a given direction by finite differences of $\Pi_h\boldsymbol{\sigma}$. Next we show that we can use finite differences of $\boldsymbol{\sigma}_h$ to approximate the finite differences of $\Pi_h\boldsymbol{\sigma}$. We introduce the following special notation.

$$\partial_{h,\kappa}f(\mathrm{x}) = \frac{f(\mathrm{x}+h\mathrm{e}_\kappa)-f(\mathrm{x})}{h} ,$$

$$\partial^2_{h,\kappa}f(\mathrm{x}) = \frac{f(\mathrm{x}+h\mathrm{e}_\kappa)-2f(\mathrm{x})+f(\mathrm{x}-h\mathrm{e}_\kappa)}{h^2} .$$

Lemma 7.
If $f \in C^3([0,1]\times[0,1])$, $h \in (0,1/4)$,

$$\Gamma(\mathrm{x}) = \{\ (x,y) \mid x_1=x\ ,y \in [x_2-h/2,x_2+h/2]\ \} ,$$

and $\mathrm{x} \in [h,1-h]\times[h,1-h]$, then

$$|\,\partial_{h,\kappa}f(\mathrm{x}) - \frac{\partial f}{\partial x_\kappa}(\mathrm{x}+\tfrac{1}{2}h\mathrm{e}_\kappa)\,| = \mathcal{O}(h^2) , \tag{41a}$$

$$|\,\partial^2_{h,\kappa}f(\mathrm{x}) - \frac{\partial^2 f}{\partial x_\kappa^2}(\mathrm{x}+\tfrac{1}{2}h\mathrm{e}_\kappa)\,| = \mathcal{O}(h^2) , \tag{41b}$$

$$|\,\partial_{h,1}(f-P[\Gamma(\mathrm{x})](f))(\mathrm{x})\,| = \mathcal{O}(h^2) , \tag{41c}$$

$$|\,\partial_{h,2}(f-P[\Gamma(\mathrm{x})](f))(\mathrm{x})\,| = \mathcal{O}(h^2) , \tag{41d}$$

$$|\,\partial^2_{h,1}(f-P[\Gamma(\mathrm{x})](f))(\mathrm{x})\,| = \mathcal{O}(h^2) , \tag{41e}$$

$$|\,\partial^2_{h,2}(f-P[\Gamma(\mathrm{x})](f))(\mathrm{x})\,| = \mathcal{O}(h^2) . \tag{41f}$$

Proof.
The above statements are easily verified through the use of Taylor expansions.

□

For a special case we justify the use of finite differences of $\boldsymbol{\sigma}_h$ to approximate the finite differences of $\Pi_h\boldsymbol{\sigma}$. We assume that the mesh is uniform, i.e. $\lambda(\Gamma_{i-\frac{1}{2},j})=h$ and $\lambda(\Gamma_{i,j-\frac{1}{2}})=h$ for a fixed $h \in \mathbb{R}$ for all edges. We also assume that ψ is linear and increasing on the entire domain, i.e.

$$\psi(\mathrm{x}) = \beta_1 x_1+\beta_2 x_2+\gamma ,$$

with fixed $\beta_1,\beta_2,\gamma \in \mathbb{R}$ and $\beta_1,\beta_2 > 0$. We introduce two vectors $\boldsymbol{R}_h,\boldsymbol{S}_h$ in V_h. The bilinear form $\alpha_{h,1}$ acting on the sum of these vectors generates the right hand side of (40a) up to third order. We define

$$L_n = \frac{\int_{x=-\frac{1}{2}h}^{\frac{1}{2}h} x^n\exp(x\beta_\kappa)\,dx}{\int_{x=-\frac{1}{2}h}^{\frac{1}{2}h} \exp(x\beta_\kappa)\,dx} ,$$

this is equal to $L_{r,n}$ if Γ_r does not lie on the Dirichlet boundary Γ_1. We define the vectors by giving their value for each edge $\Gamma_r, r \in E$, we express this value in terms of local coordinates,

$$R_{h,r} = L_1\sigma_{r,x}(0,0) + \tfrac{1}{2}L_2\sigma_{r,xx}(0,0) - \frac{h^2}{24}\sigma_{r,yy}(0,0) . \tag{42}$$

$$S_{h,r} = \left[(L_{r,1}-L_1)\sigma_{r,x}(0,0) + (L_{r,2}-L_2)\tfrac{1}{2}\sigma_{r,xx}(0,0)\right] , \tag{43}$$

note that $\boldsymbol{S}_h$ is non-zero only on the Dirichlet part Γ_1 of the boundary. If we compare (42) and (43) with (39), then we see immediately that,

$$\alpha_{SG}(\boldsymbol{\sigma},Q_h\boldsymbol{\eta}_r) - \alpha_{h,1}(\boldsymbol{\sigma},Q_h\boldsymbol{\eta}_r) = \alpha_{h,1}(\boldsymbol{R}_h + S_h,Q_h\boldsymbol{\eta}_r) + \mathcal{O}\left[h^3\alpha_{h,1}(\boldsymbol{\eta}_r,Q_h\boldsymbol{\eta}_r)\right] \quad \forall\ r \in E\ . \tag{44a}$$

We wish to approximate $\alpha_{h,1}(\boldsymbol{R}_h+S_h,Q_h\boldsymbol{\eta}_r)$ by $\alpha_{h,1}(\Pi_h\boldsymbol{R},Q_h\boldsymbol{\eta}_r)\ +\ <\ \mathscr{E}_{\partial\Omega}S,\boldsymbol{\eta}_r\cdot\mathbf{n}_{\partial\Omega}\ >$, where $\boldsymbol{R}$ and S are continuous functions on Ω and the Dirichlet part of the boundary (Γ_1) respectively, we define,

$$R_\kappa(\mathrm{x}) = L_1\frac{\partial\boldsymbol{\sigma}\cdot\mathbf{e}_\kappa}{\partial x_\kappa}(\mathrm{x}) + \tfrac{1}{2}L_2\frac{\partial^2\boldsymbol{\sigma}\cdot\mathbf{e}_\kappa}{\partial x_\kappa^2}(\mathrm{x}) - \frac{h^2}{24}\frac{\partial^2\boldsymbol{\sigma}\cdot\mathbf{e}_\kappa}{\partial x_{3-\kappa}^2}(\mathrm{x})\ ,$$

$$S(\mathrm{x}_r)|_{\Gamma_1} =$$

$$\frac{1}{h}\exp(-\psi_r(0))\alpha_{h,1}(\boldsymbol{\eta}_r,Q_h\boldsymbol{\eta}_r)\left[(L_{r,1}-L_1)\frac{\partial\boldsymbol{\sigma}\cdot\mathbf{e}_\kappa}{\partial x_\kappa}(\mathrm{x}_r) + (L_{r,2}-L_2)\tfrac{1}{2}\frac{\partial^2\boldsymbol{\sigma}\cdot\mathbf{e}_\kappa}{\partial x_\kappa^2}(\mathrm{x}_r)\right]$$

$$\forall\ \mathrm{x}_r \in \Gamma_1\ .$$

On each straight part of the Dirichlet boundary, we can extend S to a C^1 function on that part of the Dirichlet boundary by replacing x_r by x. We see immediately that,

$$\alpha_{h,1}(\Pi_h\boldsymbol{R}-\boldsymbol{R}_h,Q_h\boldsymbol{\eta}_r) = \mathcal{O}\left[h^2(L_1 + L_2 + L_3)\alpha_{h,1}(\boldsymbol{\eta}_r,Q_h\boldsymbol{\eta}_r)\right] , \tag{44b}$$

and

$$<\ \mathscr{E}_{\partial\Omega}S,\boldsymbol{\eta}_r\cdot\mathbf{n}_{\Gamma_1}\ >\ = \alpha_{h,1}(\boldsymbol{S}_h,Q_h\boldsymbol{\eta}_r)\ , \tag{44c}$$

for all r such that Γ_r is a part of the Dirichlet boundary. We see that, if problem (1) is solvable for all $(f{=}F,g{=}G)$, then , according to (33), the solution of (1) $(\boldsymbol{\rho},v)$ for $F = \mathrm{div}\boldsymbol{R}$ on Ω, $G{=}S$ on Γ_1 will satisfy the equations,

$$\alpha_{SG}(\boldsymbol{\rho},Q_h\boldsymbol{\eta}_r) - b(\boldsymbol{\eta}_r,\mathscr{E}\bar{\mathrm{P}}_h v) = -\ <\ \mathscr{E}_{\partial\Omega}S,\boldsymbol{\eta}_r\cdot\mathbf{n}_{\Gamma_1}\ > \quad \forall\ r \in E\ , \tag{45a}$$

$$b(\boldsymbol{\rho},t_h) = (\mathrm{div}\boldsymbol{R},t_h) \quad \forall\ \underline{t_h} \in W_h\ . \tag{45b}$$

If we subtract $\boldsymbol{R}$ from $\boldsymbol{\rho}$ in these equations, then we find,

$$\alpha_{SG}(\boldsymbol{\rho}-\boldsymbol{R},Q_h\boldsymbol{\eta}_r) - b(\boldsymbol{\eta}_r,\mathscr{E}\bar{\mathrm{P}}_h v) = -\alpha_{SG}(\boldsymbol{R},Q_h\boldsymbol{\eta}_r) -\ <\ \mathscr{E}_{\partial\Omega}S,\boldsymbol{\eta}_r\cdot\mathbf{n}_{\Gamma_1}\ > \quad \forall\ r \in E\ , \tag{46a}$$

$$b(\boldsymbol{\rho}-\boldsymbol{R},t_h) = 0 \quad \forall\ t_h \in W_h\ . \tag{46b}$$

We can write (46a) as follows,

$$\alpha_{h,1}(\Pi_h(\boldsymbol{\rho}-\boldsymbol{R}),Q_h\boldsymbol{\eta}_r) - b(\boldsymbol{\eta}_r,\mathscr{E}\bar{\mathrm{P}}_h v) = \tag{47a}$$

$$\alpha_{h,1}(\Pi_h(\boldsymbol{\rho}-\boldsymbol{R}),Q_h\boldsymbol{\eta}_r) - \alpha_{SG}(\boldsymbol{\rho}-\boldsymbol{R},Q_h\boldsymbol{\eta}_r) - \alpha_{SG}(\boldsymbol{R},Q_h\boldsymbol{\eta}_r) -\ <\ \mathscr{E}_{\partial\Omega}\boldsymbol{S}\cdot\mathbf{n}_{\Gamma_1},\boldsymbol{\eta}_r\cdot\mathbf{n}_{\Gamma_1}\ > \quad \forall\ r \in E\ ,$$

for $r \in E$, the right hand side of this equation can be written as follows,

$$\alpha_{h,1}(\Pi_h\boldsymbol{\rho},Q_h\boldsymbol{\eta}_r) - \alpha_{SG}(\boldsymbol{\rho},Q_h\boldsymbol{\eta}_r) - \alpha_{h,1}(\Pi_h\boldsymbol{R}-\boldsymbol{R}_h,Q_h\boldsymbol{\eta}_r) - \alpha_{h,1}(\boldsymbol{R}_h+\boldsymbol{S}_h,Q_h\boldsymbol{\eta}_r)\ . \tag{48a}$$

According to 47a, if we subtract 46 from 40 and we use (44,a,b,c) then we get,

$$\alpha_{h,1}(\Pi_h\boldsymbol{\sigma} - \boldsymbol{\sigma}_h - \Pi_h(R-\boldsymbol{\rho}),Q_h\boldsymbol{\eta}_r) - (\mathrm{div}\boldsymbol{\eta}_r,\mathscr{E}(\bar{\mathrm{P}}_h u - u_h - \bar{\mathrm{P}}_h v)) = \tag{49a}$$

$$\alpha_{SG}(\boldsymbol{\rho},Q_h\boldsymbol{\eta}_r) - \alpha_{h,1}(\Pi_h\boldsymbol{\rho},Q_h\boldsymbol{\eta}_r) + \mathcal{O}\left[(L_1+L_2+L_3)h^2\alpha_{h,1}(\boldsymbol{\eta}_r,Q_h\boldsymbol{\eta}_r)\right] + \mathcal{O}\left[h^3\alpha_{h,1}(\boldsymbol{\eta}_r,Q_h\boldsymbol{\eta}_r)\right] \quad \forall\ r \in E\ ,$$

$$(\mathrm{div}(\Pi_h\boldsymbol{\sigma}-\boldsymbol{\sigma}_h) - \Pi_h(R-\boldsymbol{\rho}),t_h) = 0 \quad \forall\ t_h \in W_h\ . \tag{49b}$$

If we assume that L_1 is $\mathcal{O}(h)$ but not $\mathcal{O}(h^2)$, - this holds if e.g. $\beta_1,\beta_2\ >\ 20/h$ - and that problem (1) satisfies the following regularity condition for all $f \in \mathrm{L}^2(\Omega), g \in \mathrm{H}^{3/2}(\partial\Omega)$,

$$\|u\|_{L^2(\Omega)} + \|\sigma_1\|_{H^1(\Omega)} + \|\sigma_2\|_{H^1(\Omega)} \leqslant C(\|f\|_{L^2(\Omega)} + \|g\|_{H^{3/2}(\partial\Omega)}) ,$$

then we can derive an estimate for

$$|\alpha_{SG}(\boldsymbol{\rho}, Q_h\boldsymbol{\eta}_r) - \alpha_{h,1}(\Pi_h\boldsymbol{\rho}, Q_h\boldsymbol{\eta}_r)| .$$

To prove that the problem has this regularity, we use theorem 5.2.2. by P. Grisvard[16], for our problem, the theorem states that, if we have Dirichlet boundary conditions everywhere and (1) has a unique solution, then operator (1) - with $\Gamma_1 = \partial\Omega$ is a bijective continuous mapping from $H^2(\Omega)$ to $L^2(\Omega) \times H^{3/2}(\partial\Omega)$. From the equivalence of (1) and (2) and the ellipticity of (2), we see that - for Dirichlet boundary conditions - equation (1) has a unique solution. Now the above mentioned theorem states that the operator is bounded, so bounded inverse theorem (Schechter[18], theorem 4.1) implies that the operator has a bounded inverse. This in turn implies that the above regularity condition holds.

According to our assumption that $\sigma_1, \sigma_2 \in C^4(\overline{\Omega})$ and the fact that L_1 is $\mathcal{O}(h)$, we have $\boldsymbol{R} = h\boldsymbol{F}$ with $F_1, F_2 \in H^4(\overline{\Omega})$, $\operatorname{div} \boldsymbol{R} = hf$ with $f \in H^3(\Omega)$ and $S = h^2 g$ with $g \in H^3(\partial\Omega)$. According to lemma 6, this implies that,

$$|\alpha_{SG}(\boldsymbol{\rho}, Q_h\boldsymbol{\eta}_r) - \alpha_{h,1}(\Pi_h\boldsymbol{\rho}, Q_h\boldsymbol{\eta}_r)| \leqslant$$
$$hCL_1(\|f\|_{L^2(\Omega)} + \|g\|_{H^{3/2}(\partial\Omega)})|\alpha_{h,1}(\boldsymbol{\eta}_r, Q_h\boldsymbol{\eta}_r)| = \mathcal{O}(h^2\alpha_{h,1}(\boldsymbol{\eta}_r, Q_h\boldsymbol{\eta}_r)) \quad \forall \ r \in E .$$

This implies that $(\boldsymbol{\sigma}_h + \Pi_h(\boldsymbol{R} - \boldsymbol{\rho}), u_h + \overline{P}_h v)$ considered as an approximation to $(\Pi_h\boldsymbol{\sigma}, \overline{P}_h u)$ is one order of h more accurate than $(\boldsymbol{\sigma}_h, u_h)$, i.e. it is $\mathcal{O}(h^2)$.

Now assume, that L_1 *is* $\mathcal{O}(h^2)$ - this holds if e.g. $\beta_1, \beta_2 < 2/h$ - and problem (1) satisfies the above regularity condition. Now according to our assumption that $\sigma_1, \sigma_2 \in C^4(\overline{\Omega})$ and the fact that L_1 is $\mathcal{O}(h^2)$, we have $\boldsymbol{R} = h^2\boldsymbol{F}$ with $F_1, F_2 \in H^4(\overline{\Omega})$, $\operatorname{div} \boldsymbol{R} = h^2 f$ with $f \in H^3(\Omega)$ and $S = h^2 g$ with $g \in H^3(\partial\Omega)$. Note that $L_{r,1}$ is $\mathcal{O}(h)$ if Γ_r is a part of the Dirichlet edge. According to lemma 6, away from the Dirichlet edge,

$$|\alpha_{SG}(\boldsymbol{\rho}, Q_h\boldsymbol{\eta}_r) - \alpha_{h,1}(\Pi_h\boldsymbol{\rho}, Q_h\boldsymbol{\eta}_r)| \leqslant$$
$$C(L_1 + L_2)h(\|f\|_{L^2(\Omega)} + \|g\|_{H^{3/2}(\partial\Omega)})|\alpha_{h,1}(\boldsymbol{\eta}_r, Q_h\boldsymbol{\eta}_r)| = \mathcal{O}(h^3\alpha_{h,1}(\boldsymbol{\eta}_r, Q_h\boldsymbol{\eta}_r)) \quad \forall \ r \in E ,$$

and on the Dirichlet edge,

$$|\alpha_{SG}(\boldsymbol{\rho}, Q_h\boldsymbol{\eta}_r) - \alpha_{h,1}(\Pi_h\boldsymbol{\rho}, Q_h\boldsymbol{\eta}_r)| \leqslant$$
$$CL_{r,1}h(\|f\|_{L^2(\Omega)} + \|g\|_{H^{3/2}(\partial\Omega)})|\alpha_{h,1}(\boldsymbol{\eta}_r, Q_h\boldsymbol{\eta}_r)| = \mathcal{O}(h^2\alpha_{h,1}(\boldsymbol{\eta}_r, Q_h\boldsymbol{\eta}_r)) \quad \forall \ r \in E .$$

This implies that

$$\|\Pi_h\boldsymbol{\sigma} - \boldsymbol{\sigma}_h - \Pi_h(\boldsymbol{R} - \boldsymbol{\rho})\|_{L^2(\Omega)} = \mathcal{O}(h^3) ,$$

because expression (36) for the above case is bounded by

$$CN_1N_2h^5 + 2D(N_1 + N_2)h^4 .$$

We can summarise the two results given above as follows,

$$\|\Pi_h\boldsymbol{\sigma} - \boldsymbol{\sigma}_h - \Pi_h(\boldsymbol{R} - \boldsymbol{\rho})\|_{L^2(\Omega)} = \mathcal{O}(h^{k+1}) ,$$

where k is the order of L_1, i.e. $L_1 = \mathcal{O}(h^k)$.

This in turn implies that,

$$\|\Pi_h\boldsymbol{\sigma} - \boldsymbol{\sigma}_h - \Pi_h(\boldsymbol{R} - \boldsymbol{\rho})\|_{L^\infty(\Omega)} = \mathcal{O}(h^k) ,$$

on at most a $\mathcal{O}(h)$ part of Ω and

$$\|\Pi_h\boldsymbol{\sigma} - \boldsymbol{\sigma}_h - \Pi_h(\boldsymbol{R} - \boldsymbol{\rho})\|_{L^\infty(\Omega)} = \mathcal{O}(h^{k+1})$$

elsewhere.

We use this to justify the approximation of the partial derivatives $\partial^n / \partial x_\kappa^n$ of $\boldsymbol{\sigma}$ in $\boldsymbol{R}_h$ by divided differences $\partial_{h,\kappa}^n$ of $\boldsymbol{\sigma}_h$. As

$$\Pi_h\boldsymbol{\sigma}-\boldsymbol{\sigma}_h = \Pi_h(\boldsymbol{R}-\boldsymbol{\rho}) + \mathcal{O}(h^{k+1}),$$

and $\boldsymbol{R}-\boldsymbol{\rho}=h^k\boldsymbol{t}$ with $\boldsymbol{t} \in C^2(\overline{\Omega})^2$, we find,

$$\partial^n_{h,\kappa}\left[\Pi_h\sigma-\boldsymbol{\sigma}_h\right](\mathrm{x}_r) = \partial^n_{h,\kappa}\left[\Pi_h(\boldsymbol{R}-\boldsymbol{\rho})\right](\mathrm{x}_r) + \mathcal{O}(h^{k-n}) = h^k\frac{\partial^n}{\partial x^n_\kappa}\boldsymbol{t}(\mathrm{x}_r) + \mathcal{O}(h^{k-n}),$$

for $\kappa=1,2$ on a $\mathcal{O}(h)$ part of the domain Ω and

$$\partial^n_{h,\kappa}\left[\Pi_h\sigma-\boldsymbol{\sigma}_h\right](\mathrm{x}_r) = \partial^n_{h,\kappa}\left[\Pi_h(\boldsymbol{R}-\boldsymbol{\rho})\right](\mathrm{x}_r) + \mathcal{O}(h^{k+1-n}) = h^k\frac{\partial^n}{\partial x^n_\kappa}\boldsymbol{t}(\mathrm{x}_r) + \mathcal{O}(h^{k+1-n}),$$

for $\kappa=1,2$ elsewhere. Combined with lemma 7 we find that for $\kappa=1,2$ and $n=1,2$,

$$|\frac{\partial^n\boldsymbol{\sigma}}{\partial x_\kappa}(\mathrm{x}_r) - \partial_{h,\kappa}\boldsymbol{\sigma}(\mathrm{x}_r)| = \mathcal{O}(h^{k-n}),$$

on a $\mathcal{O}(h)$ part of the domain and

$$|\frac{\partial^n\boldsymbol{\sigma}}{\partial x_\kappa}(\mathrm{x}_r) - \partial_{h,\kappa}\boldsymbol{\sigma}(\mathrm{x}_r)| = \mathcal{O}(h^{k+1-n}),$$

elsewhere.

Let us denote by $\tilde{\boldsymbol{R}}_h$ the approximation of $\boldsymbol{R}_h$ and by $\tilde{\boldsymbol{S}}_h$ the approximation of $\boldsymbol{S}_h$, obtained by substituting $\partial^n_{h,\kappa}\boldsymbol{\sigma}_h$ for $\partial^n/\partial x^n_\kappa$ with $n=1,2$. We see that

$$\tilde{\boldsymbol{R}}_h+\tilde{\boldsymbol{S}}_h-\boldsymbol{R}-\boldsymbol{S} = L_1\,\mathcal{O}(h^{k-1}) + L_2\,\mathcal{O}(h^{k-2}),$$

on $\mathcal{O}(h)$ of all cells and

$$\tilde{\boldsymbol{R}}_h+\tilde{\boldsymbol{S}}_h-\boldsymbol{R}-\boldsymbol{S} = L_1\,\mathcal{O}(h^{k}) + L_2\,\mathcal{O}(h^{k-1}),$$

elsewhere. Let $(\tilde{\boldsymbol{\sigma}}_h,\tilde{u}_h)$ be the solution of

$$\alpha_{h,1}(\tilde{\boldsymbol{\sigma}}_h,Q_h\boldsymbol{\tau}_h) - (\operatorname{div}\boldsymbol{\tau}_h,\mathscr{E}(\tilde{u}_h)) = \tag{50a}$$

$$\alpha_{h,1}(\tilde{\boldsymbol{R}}_h+\tilde{\boldsymbol{S}}_h,Q_h\boldsymbol{\tau}_h) - \ < \mathscr{E}_{\partial\Omega}g,\boldsymbol{\eta}_r\cdot\mathbf{n}_{\Gamma_1} > \quad \forall\ \boldsymbol{\tau}_h \in V_h\,,$$

$$(\operatorname{div}(\tilde{\boldsymbol{\sigma}}_h),t_h) = (f,t_h) \quad \forall\ t_h \in W_h\,, \tag{50b}$$

then

$$\alpha_{h,1}(\Pi_h\boldsymbol{\sigma} - \tilde{\boldsymbol{\sigma}}_h,Q_h\boldsymbol{\tau}_h) - (\operatorname{div}\boldsymbol{\tau}_h,\mathscr{E}(\overline{\mathrm{P}}_hu - \tilde{u}_h)) = \tag{51a}$$

$$\alpha_{h,1}(\boldsymbol{R}_h+\boldsymbol{S}_h,Q_h\boldsymbol{\tau}_h) - \alpha_{h,1}(\tilde{\boldsymbol{R}}_h+\tilde{\boldsymbol{S}}_h,Q_h\boldsymbol{\tau}_h)$$

$$\forall\ \boldsymbol{\tau}_h \in V_h\,,$$

$$(\operatorname{div}(\Pi_h\boldsymbol{\sigma}-\tilde{\boldsymbol{\sigma}}_h),t_h) = 0 \quad \forall\ t_h \in W_h\,, \tag{51b}$$

so - in $\mathrm{L}^2(\Omega)$ norm - $(\tilde{\boldsymbol{\sigma}}_h,\tilde{u}_h)$ is formally one order of h closer to $(\Pi_h\boldsymbol{\sigma},\overline{\mathrm{P}}_hu)$ than $(\boldsymbol{\sigma}_h,u_h)$.

We can derive an a-posteriori error estimate by calculating the difference between the discrete solution with and without a tilde. It may be possible to derive a mesh-refinement criterion from $\tilde{\boldsymbol{R}}_h$.

8.2. Numerical results.

In this section we show how the deferred correction method works in practice. We consider problem (1) with Dirichlet boundary conditions on the entire boundary,

$$\Gamma_1 = \partial\Omega\,,\ a = 0.01 \ \text{ and } \ \psi = 100(x_1+x_2)\,,$$

and data derived from a known solution,

$$u = \tanh(8(x_1-x_2))\,.$$

It follows that,

$$g = u\,|_{\partial\Omega}\ ,$$

$$f = -\frac{\operatorname{div}(\mathbf{grad}\,u + u\,\mathbf{grad}\,\psi)}{100}\ .$$

We show results for the Scharfetter-Gummel version of the discretisation and the results obtained after applying the correction discussed in section 8.1 once, twice, thrice or ten times.

We take the unit square for Ω. On a mesh of $n \times n$ cells, with mesh width $h = 1/n$, we have $4n - 4$ Dirichlet edge cells and a total of n^2 cells. We use the 2-norm as norm for the error vectors,

$$\|v\| = \left[\frac{1}{|I|} \sum_{i \in I} v_i^2 \right] ,$$

where $|I|$ is the number of elements in the index set.
All experiments satisfied the expected symmetry relation

$$\log_2 \|(\Pi_h\sigma - \sigma_h)\cdot \mathbf{e}_1\| = \log_2 \|(\Pi_h\sigma - \sigma_h)\cdot \mathbf{e}_2\| \ ,$$

for the accuracy given in the tables.

the $\log_2$ of the errors for $\alpha_h = \alpha_{h,1}$.		
meshwidth	$\log_2 \|\mathrm{P}_h u - u_h\|$	$\log_2 \|(\Pi_h\sigma - \sigma_h)\cdot \mathbf{e}_1\|$
1 / 2	-1.6	-0.9
1 / 4	-1.5	-1.4
1 / 8	-1.9	-1.9
1 / 16	-2.6	-2.6
1 / 32	-3.8	-3.8
1 / 64	-5.5	-5.4
1 / 128	-7.3	-7.3

We see that the large jump in ψ per cell on the coarsest meshes, combined with the large gradient of the solution relative to the coarsest meshes result in convergence slower than $\mathcal{O}(h)$. For a fine mesh, $h < 1/32$, we see that the convergence behaviour tends to $\mathcal{O}(h^2)$. For intermediate meshes intermediate convergence rates are found.

the $\log_2$ of the errors after one correction.		
meshwidth	$\log_2 \|\mathrm{P}_h u - u_h\|$	$\log_2 \|(\Pi_h\sigma - \sigma_h)\cdot \mathbf{e}_1\|$
1 / 2	-2.2	-1.6
1 / 4	-2.1	-2.0
1 / 8	-2.7	-2.8
1 / 16	-3.9	-3.9
1 / 32	-6.0	-6.0
1 / 64	-9.1	-9.1
1 / 128	-12.9	-12.8

We still see slow convergence rates at the coarsest meshes, probably due to the relative steepness of the solution on that mesh. Convergence speed on the finer meshes is improved by the correction. We see that - as predicted below equation (4.51b) in section 8.1 - where the previous table shows first order behaviour between meshes, we now find second order convergence. And where the previous table shows second order behaviour between meshes, we now find third order convergence.

the $\log_2$ of the errors after two corrections		
meshwidth	$\log_2 \| P_h u - u_h \|$	$\log_2 \| (\Pi_h \sigma - \sigma_h)\cdot e_1 \|$
1 / 2	-2.6	-1.8
1 / 4	-2.4	-2.4
1 / 8	-3.1	-3.3
1 / 16	-4.7	-4.8
1 / 32	-7.6	-7.6
1 / 64	-11.8	-11.7
1 / 128	-15.4	-15.3

We still see slow convergence at the coarsest meshes. Again we find $k+1$-th order behaviour between meshes where the previous table shows k-th order order behaviour between meshes.

the $\log_2$ of the errors after ten corrections.		
meshwidth	$\log_2 \| P_h u - u_h \|$	$\log_2 \| (\Pi_h \sigma - \sigma_h)\cdot e_1 \|$
1 / 2	-2.8	-2.0
1 / 4	-2.8	-2.7
1 / 8	-3.7	-3.9
1 / 16	-5.9	-6.1
1 / 32	-9.2	-9.2
1 / 64	-12.6	-12.6
1 / 128	-15.4	-15.4

After ten iterations no further significant changes occurred. We see that we have third order behaviour from $h = 1/16$ onward.

9 Conclusions.

In section 4 and 6 we have seen that the Scharfetter-Gummel discretisation in two dimensions can be written as a saddle point problem. We can use theorem 3.1 by Nicolaides[5] to show that this discretisation is at least formally stable and consistent. We then showed consistency. In section 8 we presented a technique to obtain a local error indicator and we gave numerical results.

The results on a posteriori error estimates can be summarised as follows. We show that it gives an approximation of the error that is an $\mathcal{O}(h^{k+1})$ accurate approximation to the true error, when the true error is $\mathcal{O}(h^k)$. This can also be seen in the numerical results for this method.

We see that the two dimensional Scharfetter-Gummel scheme for the current continuity equation is stable and consistent. Our error analysis in section 7 yields the following information on the order of the error. For small enough h, he error is order two only if a cell is not adjacent to the boundary and has a size that differs at most $\mathcal{O}(h^2)$ from its neighbours. If these conditions do not hold the error is of order $\mathcal{O}(h)$. To be certain that the global order of the error is $\mathcal{O}(h^2)$ the change in ψ between cell centres must be smaller than 2. For semiconductors this means that the change in the voltage scaled by the thermal voltage must be smaller than 2. In Section 8 it is shown that is possible to calculate a correction to the solution of the Scharfetter-Gummel scheme. From this we can derive an a-posteriori error estimator.

A search of the literature shows that papers on a posteriori error estimates for finite volume or mixed finite element discretisations - other than for fluid dynamics - .are rare. There are papers that deal with a posteriori error estimates for the mixed discretisation of the Navier-Stokes equations, see e.g. the paper by Verfürth, [19] but the techniques used there are geared to that type of problem.

References

1. D. L. Scharfetter and H. K. Gummel, "Large-Signal Analysis of a Silicon Read Diode Oscillator," *IEEE Transactions on Electron Devices*, vol. ED-16, no. 1, pp. 64-77, 1969.
2. A. M. Il'in, "Differencing scheme for a differential equation with a small parameter affecting the highest derivative," *Mathematical Notes of the Academy of Sciences of the USSR*, vol. 6, no.

1-2, pp. 596-602, 1969.

3. Wolfgang Fichtner, Donald J. Rose, and Randolph E. Bank, "Numerical methods for semiconductor device simulation," *SIAM journal of scientific and statistical computing*, vol. 4, no. 3, pp. 416-435, 1983.

4. Wolfgang Fichtner, Donald J. Rose, and Randolph E. Bank, "Semiconductor device simulation," *SIAM journal of scientific and statistical computing*, vol. 4, no. 3, pp. 391-415, 1983.

5. R. A. Nicolaides, "Existence, uniqueness and approximation for generalized saddle point problems," *SIAM J. Numer. Anal.*, vol. 19, no. 2, pp. 349-357, 1982.

6. Peter A. Markowich, *The Stationary Semiconductor Device Equations*, Springer-Verlag, Wien New York, 1986.

7. Siegfried Selberherr, *Analysis and simulation of semiconductor devices*, Springer-verlag, Wien New York, 1984.

8. S. J. Polak, C. den Heijer, H. A. Schilders, and P. Markowich, "Semiconductor device modelling from the numerical point of view," *International Journal for Numerical Methods in Engineering*, vol. 24, pp. 763-838, 1987.

9. Walter L. Engl, Heinz K. Dirks, and Bernd Meinerzhagen, "Device Modeling," *Proceedings of the IEEE*, vol. 71, no. 1, pp. 10-33, January 1983.

10. R. E. Bank, W. Fichtner, D. J. Rose, and R. K. Smith, "Algorithms for semiconductor device simulation," in *Large Scale Scientific Computation*, ed. B. Engquist, Progress in Scientific Computing, Birkhäuser, 1987.

11. Randolph E. Bank, Joseph W. Jerome, and Donald J. Rose, "Analytical and numerical aspects of semiconductor device modelling," in *Computing Methods in Applied Sciences and Engineering*, ed. J. L. Lions, vol. V, pp. 593-597, North-Holland, 1982.

12. V. Girault and P. Raviart, *Finite Element Methods for Navier-Stokes Equations*, Springer series in computational mathematics, 5, Springer-Verlag, 1986.

13. J. Douglas, Jr. and J. E. Roberts, "Global estimates for mixed methods for second order elliptic equations," *Mathematics of computation*, vol. 44, no. 169, pp. 39-52, 1985.

14. P. A. Raviart and J. M. Thomas, "A mixed finite element method for 2-nd order elliptic problems," in *Mathematical aspects of the finite element method*, Lecture Notes in Mathematics, vol. 606, pp. 292-315, Springer, 1977.

15. R. H. MacNeal, "An asymmetrical finite difference network," *Quart. Appl. Math.*, vol. XI, no. 3, pp. 295-310, 1953.

16. P. Grisvard, *Elliptic Problems in Nonsmooth Domains*, Pitman, 1985.

17. L . Fox and D. F. Mayers, *Numerical Solution of Ordinary Differential Equations for scientists and engineers*, Chapman and Hall, 1987.

18. M. Schechter, *Principles of Functional Analysis*, Academic Press, 1971.

19. R. Verfürth, "A Posteriori Error Estimators for the Stokes Equations," *Numerische Mathematik*, vol. 55, pp. 309-325, 1989.