Mixed Finite-Element Methods for Elliptic Convection-dominated Problems Arising in Semiconductor Physics

Dissertation
zur Erlangung des Grades
Doktor der Naturwissenschaften

Am Fachbereich Physik, Mathematik und Informatik
der Johannes Gutenberg-Universität Mainz

Stefan Holst,
geboren in Berlin

Mainz, 2005

Tag der mündlichen Prüfung: 13. Januar 2006

D77 - Mainzer Dissertation

Abstract

In dieser Arbeit wird ein adaptives numerisches Verfahren zur Simulation einer Klasse von makroskopischen Halbleitermodellen vorgestellt und analysiert. Dazu wird zunächst in die mathematische Modellierung von Halbleitern eingeführt. Dies dient zur Einordnung der im weiteren Verlauf numerisch in 2D genauer untersuchten Energie–Transport Modelle. Diese Modellklasse beschreibt den Fluß von geladenen Teilchen, d.h. von negativ geladenen Elektronen und sogenannten Löchern, das sind Pseudoteilchen mit positiver Ladung, und deren Energieverteilung in einem Halbleiterkristall anhand eines Systems von nichtlinearen gekoppelten partiellen Differentialgleichungen.
Eine wesentliche Schwierigkeit in der numerischen Behandlung dieser Gleichungen stellen einerseits die nichtlineare Kopplung und die nur teilweise durch die Daten abschätzbaren lokalen Phänomene, sogenannter "hot electron effects", dieser teils konvektionsdominanten Gleichungen dar. Die primären Größen der Modelle sind in der hier für die Simulationen verwendeten Formulierung Teilchen- und Energiedichten. Weiterhin entscheidend ist für den Anwender die Größe des Stromflusses durch Teile des Randes, sogenannte Kontakte. Das hier betrachtete numerische Verfahren verwendet gemischte Finite Elemente als Ansatzraum für die diskrete Lösung. Die stetige Diskretisierung der Normalkomponente der Stromdichte ist aus Sicht der Anwendung der entscheidende Vorteil dieser Elemente. Es wird gezeigt, daß im Laufe des Algorithmus unter bestimmten Bedingungen an die Triangulierung sichergestellt ist, daß die Teilchendichten positiv bleiben. In diesem Zusammenhang wird ebenfalls eine a priori Fehlerabschätzung für die diskrete Lösung einer linearen Konvektions-Diffusions-Gleichung bewiesen. Die lokalen Phänomene im Halbleiter werden durch adaptive Verfahren, die auf a posteriori Fehlerschätzern beruhen, geeignet aufgelöst. Es findet an dieser Stelle ein Vergleich verschiedener Fehlerschätzer statt.
Außerdem wird ein Verfahren zur Fehlerschätzung in von der Lösung abgeleiteten Größen, sogenannten 'functional outputs', auf die Diskretisierung mit gemischten Finiten Elementen übertragen. An einem Beispielproblem wird dargestellt, wie dieses Verfahren noch erfolgversprechend angewendet werden kann, wenn Standardfehlerschätzer keine Reduktion des Fehlers im Zuge iterativer Gitterverfeinerung erzielen.

Contents

Chapter 1

Introduction

Telecommunication was one of the major factors driving the development of modern societies in the last century. This development is closely related to the evolution of semiconductor technology. A technology that started in 1947 when Bardon, Brattain and Schockley presented the first semiconductor device (a germanium transistor) for which they were awarded the Nobel Prize in 1956.

Semiconductor devices as a replacement for the electronic devices used at that time were much smaller. The combination of several transistors and other structures in an electronic circuit on a single semiconductor crystal led to the so-called integrated circuits (J. Kilby of Texas Instruments, R. Noyce of Fairchild Semiconductor) which was the beginning of a process leading to todays chips containing millions of transistors on an area of about 1cm^2. In the industry the term "scaling" refers to the continued reduction of the size of the structures on a chip.

The characteristic length scales reached dimensions of 90nm in June 2004 in Intel's Pentium 4 and the process to build structures with 65nm is already in development. In these dimensions quantum effects cannot always be neglected. However, in the *energy-transport model* that we will use for the device simulation it is looked at the movement of electrons in a semiconductor in a continuum sense. Similar to fluid dynamics a hierarchy of models exists to describe this "electron gas". The quantum physical structure of the crystal is incorporated only via special predetermined parameters in the most detailed so called microscopic model in this hierarchy, the *semi–classical semiconductor Boltzmann equation*. The least complex

and best understood macroscopic model, the *drift-diffusion model*, consists of the mass conservation equation and a constitutive equation for the current density only. It has been derived from phenomenological considerations [**102**], a rigorous derivation is due to Poupaud [**97**]. We will clarify the position of the energy-transport model in the hierarchy of these semi–classical semiconductor models.

Numerical simulations of chips with microscopic models are impossible. Even macroscopic models are only applied to simulate very few connected devices. In the past the behaviour of a device in an integrated circuit was described with a rather small set of parameters that had been extracted from precedingly-performed device simulations. The scaling of device dimensions has led to a larger set of necessary parameters to describe the device reaction on external inputs. For techniques to automatically identify significant parameters see [**70**]. There is a demand for more detailed information on the interaction of the device with a connected circuit. Therefore methods to directly couple device and circuit simulation have been developed, see [**109**]. The active region of a device is modeled by macroscopic device equations and the remaining circuit is described by differential algebraic equations. A very interesting question in this coupling in present devices is for instance the device temperature.

In both of these fields the drift-diffusion model is not accurate enough to capture the additional features. Refined models were proposed by physicists and later mathematically analyzed, namely the hydrodynamic model introduced by Bløtekjær [**18**] and Baccarani and Wordemann [**7**] and the energy-transport model introduced by Stratton [**111**].

The energy-transport models are more complex than the drift-diffusion equations but keep their parabolic nature. For the numerical solution of the hydrodynamic model, which contains hyperbolic modes, special (efficient) algorithms are necessary (see, e.g.,[**61**]). This intermediate complexity makes the energy-transport system appropriate for fast and accurate semiconductor simulations. Extensions of recently developed algorithmic device design optimization techniques [**71**] for the drift-diffusion equations seem to be possible.

To solve this model with a flexible and robust numerical method that can automatically adapt to the local behaviour of solutions to semiconductor device simulations is the main goal of this work. In [**49**] it was shown that the energy-transport equations can be written in a drift-diffusion form. The unknowns in this formulation are the electron number and the electron energy density which should remain positive to be physically meaningful. Under this premise the discretization of drift dominated problems is not

straight forward, for instance, standard methods for convection-dominated problems, e.g. streamline diffusion methods, do not preserve the positivity of the numerical solution.

A main concern in stationary device simulation is the current voltage relation, which is of high importance to appraise the device effectivity. In order to generate such curves the current flowing through parts of the device boundary is quantified for a set of different values of an externally-applied voltage that is a difference in the electrostatic potential. In standard one-field finite-element methods the current is a derived quantity and therefore not of as high accuracy as the original unknown. To the contrary, mixed finite-elements introduce the current density as a second independent variable. Moreover, the current density will then be approximated in $H(\mathrm{div}, \Omega)$, which as a consequence, leads to the continuity of the current density's normal component, see [**99**], and thus to current conservation.

In the simulation we use an exponentially fitted mixed-hybrid finite-element method to discretize the stationary energy-transport model in two space dimensions. More precisely, we use the finite-element introduced by Marini and Pietra [**89**] adapted to energy-transport equations [**49, 79**] and an adaptive refinement of the elements based on an error estimator which is motivated by results of Hoppe and Wohlmuth [**72, 73, 117**]. We adopt a gradient recovery based estimator introduced by Carstensen and Bartels in [**31, 12**] and compare it with the above estimator. These estimators are used together to define a new estimator for the mixed finite-element method that estimates directly the error in a derived quantity of interest. This estimator is an extension of the dual weighted residual approach developed by Becker and Rannacher [**14, 13, 15**] to mixed finite-elements.

The method we develop and analyze is applied to some example device to validate it in comparison with one dimensional simulations of [**49**] and to show its numerical convergence and robustness when simulating a MESFET (metal semiconductor field effect transistor) device, in which areas with very low densities, so-called depletion regions, complicate the simulation. Finally, a simulation of a submicron MOSFET (metal-oxide semiconductor FET) device with a channel length of 70nm is performed.

In Section 1.1, the introduction will review the relationships between the different models to describe charge transport in a semiconductor, i.e. we present a hierarchy of models that can be seen as formal limiting equations derived from the Boltzmann equation, following the presentation in [**17, 50**]. It follows an introduction into adaptive finite-element methods especially concerning relations between the elements developed in [**89**] and standard mixed finite-elements in Section 1.2, before we start with analyz-

ing the numerical method for linear elliptic convection-diffusion problems in Chapter 2. Chapter 3 is devoted to the treatment of the nonlinear system of energy-transport equations. An iterative procedure is defined that extends the continuation method for the drift-diffusion model, the Gummel map [**69**], and finally we present the numerical examples.

1.1 Semiconductor Device Modeling

In this section we review different limiting procedures to derive energy-transport models from the semi–classical Boltzmann equation. Thus, we get more information about the simplifying assumptions that limit the accuracy of energy-transport models, including a display of its advantages over the drift-diffusion models. In physics literature the refined models that we will present are referred to as models for *hot electron* transport. Reference [**100**] is mentioned for the physical data that might occur in the formulas. For the Boltzmann picture in the theory of dilute gases, gas molecules obey Newton's laws

$$\partial_t x = v, \quad \partial_t v = \tfrac{1}{m} F(t), \qquad t > 0, x \in \mathbb{R}^d, \tag{1.1}$$

where v denotes the velocity and F describes a force field. Particles that do not interact with each other stay on a trajectory $(x(t), v(t))$. Letting $f(t, x, v)$ be their distribution function at a time t in position-velocity phase-space, we can express the last sentence by

$$\begin{aligned} 0 &= \frac{\mathrm{d}f}{\mathrm{d}t}(t, x(t), v(t)) = \partial_t f + \partial_t x(t) \cdot \nabla_x f + \partial_t v(t) \cdot \nabla_v f \\ &= \partial_t f + v \cdot \nabla_x f + \tfrac{1}{m} F \cdot \nabla_v f. \end{aligned}$$

This is the *Vlasov equation* or collisionless *Boltzmann transport equation*. Particle interactions that instantaneously change the particles' velocity but not their positions are called collisions. The classical Boltzmann equation states that the rate of change in particle distribution function along trajectories is only due to collisions, consequently we have, denoting the collision operator by $Q(f)$:

$$\partial_t f + v \cdot \nabla_x f + \tfrac{1}{m} F \cdot \nabla_v f = Q(f), \quad x, v \in \mathbb{R}^3, t > 0. \tag{1.2}$$

1.1.1 Semi-Classical Picture of Quantum Mechanics

The situation for electrons in a semiconductor crystal is considerably different. In the semi-classical picture of electron transport in a crystal, the laws

of classical mechanics (1.1) are substituted by

$$\partial_t x = v(\boldsymbol{k}) = \frac{1}{\hbar}\nabla_{\boldsymbol{k}} E(\boldsymbol{k}) \tag{1.3}$$

$$\partial_t v_c = (\boldsymbol{m}^*)^{-1}\partial_t(\hbar\boldsymbol{k}) = -(\boldsymbol{m}^*)^{-1}\frac{q}{\hbar}\nabla_x V(x,t), \tag{1.4}$$

where $E(\boldsymbol{k})$ is the electron energy in a certain energy band depending on the pseudo wave vector $\boldsymbol{k}$ and $V(x,t)$ is the electrostatic potential, $\hbar$ the reduced Planck constant and q the electron charge. In analogy to classical mechanics the product $p = \hbar\boldsymbol{k}$ is termed *crystal momentum* and $\boldsymbol{m}^*$ the *effective mass tensor*, motivated via the identification

$$\partial_t v(\boldsymbol{k}) = \frac{1}{\hbar}\underbrace{\frac{\mathrm{d}^2 E(\boldsymbol{k})}{\mathrm{d}\boldsymbol{k}^2}}_{=\hbar^2(\boldsymbol{m}^*)^{-1}}\partial_t\boldsymbol{k} = (\boldsymbol{m}^*)^{-1}\partial_t p.$$

To understand this relation we have to make a small excursion into quantum mechanics. At the smallest length scales an electron is quantum mechanically described as a *wave* that is a complex valued function $\psi(x,t)$ that solves the *Schrödinger equation* for a given potential V

$$i\hbar\partial_t\psi = -\frac{\hbar^2}{2m}\Delta\psi - qV\psi, \quad t>0, \quad \psi(x,0)=\psi_0(x), \quad x\in\mathbb{R}^3,$$

where i is the imaginary unit, m the free electron mass. The semi-classical picture tries to find a particle reinterpretation of these waves. We now want to give a short abridgement of this procedure. If we neglect time dependence in V the solutions to the Schrödinger equation can be solved by a separation ansatz, which leads to the eigenvalue problem of the stationary Schrödinger equation

$$-\frac{\hbar^2}{2m}\Delta\psi - qV\psi = E\psi \quad \text{ in } \mathbb{R}^3. \tag{1.5}$$

The simplest case is $V\equiv 0$, which permits solutions and eigenvalues of the form

$$\psi(x) = \exp(i\boldsymbol{k}\cdot x), \quad E = \frac{(\hbar|\boldsymbol{k}|)^2}{2m}, \quad \boldsymbol{k}\in\mathbb{R}^3.$$

These solutions are called *plane waves* for the wave vector $\boldsymbol{k}$ and it can be shown that $\hbar\boldsymbol{k}$ is their quantum mechanical momentum, which means that the eigenvalues represent the energy of these wave.
If V is periodic and describes the potential of the atomic nuclei in an infinite crystal, or more precisely, if $V(x) = V(x+y)$ for all $x\in\mathbb{R}^3$ and $y\in L$, L

being the crystal lattice, the structure of the solution is more complex. The *Bloch–Theorem* says that the solutions in this case have the form

$$\psi_B(\boldsymbol{k}, x) = \exp(i\boldsymbol{k} \cdot x)u(x),$$

where the *Bloch function* u is L–periodic and $\boldsymbol{k} \in B$. The Brillouin zone B can be understood as the largest area, so that an arbitrary plane wave $\exp(i\tilde{k}\cdot x)$ cannot be split into a product of a L-periodic function and a plane wave $\exp(i\boldsymbol{k}\cdot x)$ with $|\boldsymbol{k}| \leq |\tilde{k}|$. For a one-dimensional example of this decomposition see Figure 1.1.

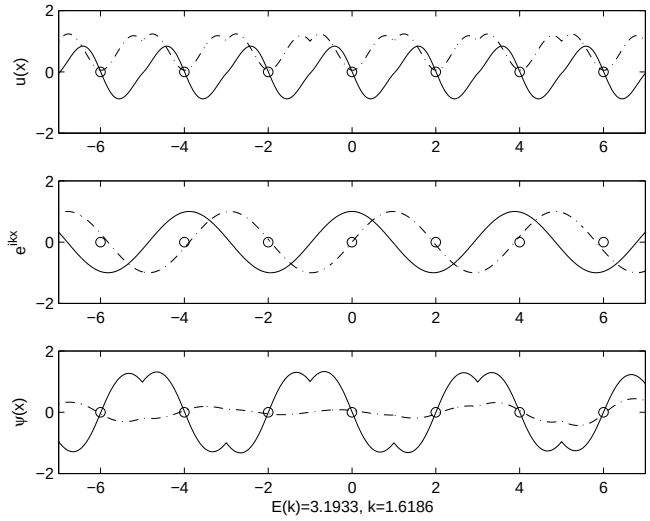

Figure 1.1: Bloch decomposition for a lattice(o) potential V. Solid lines display real parts, dash-dotted lines imaginary parts. The wave length of u and the plane wave are different. The state ψ is not necessarily periodic.

Since $\psi_B(\boldsymbol{k}, x)$ is not a plane wave, $\boldsymbol{k}$ is called *pseudo–wave vector*. Inserting ψ_B in (1.5) we obtain for any fixed $\boldsymbol{k} \in B$ an eigenvalue problem for u. The sequence of eigenvalue-eigenfunction pairs is denoted by $(E_n(\boldsymbol{k}), u_{n,\boldsymbol{k}})$. The function $k \mapsto E_n(k)$ is called *dispersion relation* or the n-th *energy band*. It shows how the energy of the n-th band depends on the (pseudo-)wave vector $\boldsymbol{k}$. The union of ranges of E_n over $n \in \mathbb{N}$ is not necessarily the whole real line $\mathbb{R}$. Energies that are not in the range are denoted as "forbidden" energies. They form intervals, the so-called energy gaps.
A crystal will always be of limited size, which leads to a limited number of available states in each energy band, or equivalently formulated, each state

occupies a certain volume in $\boldsymbol{k}$-space. The occupation of the available states characterize the material. The *Pauli principle* which holds for electrons says that each state can only be occupied by one electron of a specific spin. Hence, each state may be occupied by a fixed number of electrons, which we set to one, neglecting different spins.
In thermodynamic equilibrium at zero temperature the quantum mechanical system has minimal energy, which requires that the states are filled from the state of lowest energy up to a certain level. This energy is called *Fermi level* $E_F = q\mu$, and can be expressed by the *chemical potential* $q\mu$. For a semiconductor or insulator the Fermi level lies within a band gap and all the states in the bands below are occupied. The band below the gap is called the *valence band* in contrast to the band above the Fermi level which is called *conduction band*.
Conductance under an applied electric field can only start if an electron gains enough energy to reach an empty state in an upper band. The size of the energy gap therefore differs between insulators, semiconductors and conductors. For semiconductors it is approximately 0.1 to 0.5eV, for insulators it is larger than 1eV and for conductors the fully occupied energy bands and the empty bands overlap.
However, if the system gets excited thermally, the distribution of electrons is only given via a probability distribution function. This function is determined by the supposed state-changing mechanism and the *principle of detailed balance*, which roughly says that in thermal equilibrium state transitions from one state to an other happen equally in each direction. For electrons with a thermal energy k_BT in a semiconductor, k_B being the Boltzmann constant, this leads to the *Fermi Dirac distribution*

$$f(E) = \frac{1}{1 + \exp\left(\frac{E-E_F}{k_BT}\right)}. \tag{1.6}$$

States that are used for conduction are close to the Fermi level, since it deserves only little energy to free these states. For the intended derivation, it is therefore sufficient to include the energy bands at the gap, $E_v(\boldsymbol{k})$ and $E_c(\boldsymbol{k})$, the band edges, and to approximate them by Taylor-polynomials around their extrema $E_c(\boldsymbol{k}_c) = E_c^0$, and $E_v(\boldsymbol{k}_v) = E_v^0$.

$$\begin{aligned} E_c(\boldsymbol{k}-\boldsymbol{k}_c) &= E_c^0 + \frac{1}{2}(\boldsymbol{k}-\boldsymbol{k}_c)^\top \frac{\mathrm{d}^2 E_c(\boldsymbol{k})}{\mathrm{d}\boldsymbol{k}^2}(\boldsymbol{k}-\boldsymbol{k}_c) + O(|\boldsymbol{k}-\boldsymbol{k}_c|^3) \\ E_v(\boldsymbol{k}-\boldsymbol{k}_v) &= E_v^0 + \frac{1}{2}(\boldsymbol{k}-\boldsymbol{k}_v)^\top \frac{\mathrm{d}^2 E_v(\boldsymbol{k})}{\mathrm{d}\boldsymbol{k}^2}(\boldsymbol{k}-\boldsymbol{k}_v) + O(|\boldsymbol{k}-\boldsymbol{k}_v|^3) \end{aligned} \tag{1.7}$$

We now want to associate a velocity to an electron in an energy band. So far, we only have a description of the quantum mechanical state of one electron in a lattice-periodic potential. Measurable quantities of this state are operators on the wave function ψ. The important observables, the electron density $n(x,t)$ and the electron current density $J(x,t)$, are defined as

$$n(x,t) = |\psi(x,t)|^2 \quad \text{and} \quad J(x,t) = -\frac{\hbar q}{m}\,\mathrm{Im}((\overline{\psi}\nabla\psi)(x,t)).$$

A first idea is to use the classical particle velocity $J = -qn\hat{v}(x,t)$. But this does not give an idea of the motion of an electron as a wave packet. The velocity of a wave packet is the mean velocity over a primitive lattice cell D and the following relation can be deduced if no external potential is applied apart from the lattice potential:

$$v_b(\boldsymbol{k}) = \int_D \hat{v}_b(k) n_b(k) / \int_D n_b(k)\,\mathrm{d}x = \frac{1}{\hbar}\nabla_{\boldsymbol{k}} E_b(\boldsymbol{k}), \quad b \in \{c, v\}. \tag{1.8}$$

The notion of the group velocity is band specific. By introducing this velocity into the Boltzmann equation we model a priori only the flow of electrons within one band. Electron interaction and band interchange have to be modeled differently, see below.
We now consider the structure of the energy bands in more detail, starting with the conduction band. Close to the band minimum a second order approximation of the dispersion relation may be sufficient. The Hessian matrix at the band minimum is symmetric and positive definite, e.g., it can be written in its symmetry axis by:

$$\frac{1}{\hbar^2}\frac{\mathrm{d}^2 E_c(\boldsymbol{k})}{\mathrm{d}\boldsymbol{k}^2}\bigg|_{\boldsymbol{k}=\boldsymbol{k}_c} = (\boldsymbol{m}_c^*)^{-1} = \begin{pmatrix} 1/m_1^* & 0 & 0 \\ 0 & 1/m_2^* & 0 \\ 0 & 0 & 1/m_3^* \end{pmatrix}$$

This is regarded as the definition of the effective mass tensor. Often it is reduced further to an isotropic effective mass $\boldsymbol{m}_c^* = m_c^*\mathrm{Id}$. By shifting (1.7) so that the conduction band minimum is zero and lies at $\boldsymbol{k}_c = 0$ we obtain the *parabolic band approximation*:

$$E_c(\boldsymbol{k}) = \frac{\hbar^2}{2m_c^*}|\boldsymbol{k}|^2 \tag{1.9}$$

Compared to the dispersion relation of a free particle, electrons in the conduction band are regarded as free particles with the mass m_c^*.
In case of electrons in the valence band the Hessian matrix at the band maximum is symmetric and negative definite. Reduced to the isotropic

approximation this would lead to a negative effective electron mass. This can be avoided if we assume that these states are occupied by pseudo particles of an opposite charge. Physically this can be understood as vacancies in the valence band, generally termed *holes*. Substituting $\tilde{q} = -q$ in equation $J_b = -\tilde{q} n_b \hat{v}_v(x,t)$ we obtain $\tilde{v}_v(\boldsymbol{k}) = -v_v(\boldsymbol{k})$. It is therefore convention to use the $\tilde{E}_v(\boldsymbol{k}) = -E_v(\boldsymbol{k})$ as the dispersion relation for holes and to write

$$\tilde{v}_v(\boldsymbol{k}) = \frac{1}{\hbar}\nabla_{\boldsymbol{k}}\tilde{E}_v(\boldsymbol{k}),$$
$$(\boldsymbol{m}_v^*)^{-1} = \frac{1}{\hbar^2}\frac{\mathrm{d}^2\tilde{E}_v(\boldsymbol{k})}{\mathrm{d}\boldsymbol{k}^2}\bigg|_{\boldsymbol{k}=\boldsymbol{k}_c}$$

and then to use the isotropic approximation $\boldsymbol{m}_v^* = m_v^*\mathrm{Id}$.
If in addition to the lattice potential an external potential is applied, $V = V_L + \tilde{V}$, the band structure changes significantly and the bands are coupled. However, it is usually assumed that the variation of the external potential is small on the length scale of the lattice potential. So that by the process of homogenization [**98, 63**] to the larger length scale, the external potential $\tilde{V}$ becomes the driving potential in (1.4) only. Before summarizing this

(quasi-)particle	velocity	mass
free electron	$\frac{1}{\hbar}\nabla(\hbar\boldsymbol{k})^2/(2m)$	m=9.1·10^{-31}kg
conduction band el.	v_c=$\frac{1}{\hbar}\nabla E_c(\boldsymbol{k})$	$\boldsymbol{m}_c^*$=$\big(\frac{d^2E_c(\boldsymbol{k}_c)}{\mathrm{d}\boldsymbol{k}^2}\big)^{-1}$
valence band hole	v_v=$\frac{1}{\hbar}\nabla\tilde{E}_v(\boldsymbol{k})$, $\tilde{E}_v$=$-E_v$	$\boldsymbol{m}_v^*$=$\big(\frac{\mathrm{d}^2\tilde{E}_v(\boldsymbol{k}_v)}{\mathrm{d}\boldsymbol{k}^2}\big)^{-1}$

Table 1.1: Comparison of free electrons with the semi-classical interpretation of electrons in a crystal as *quasi particles with different masses*.

excursion into quantum mechanics in Table 1.1, we want to mention that higher order terms in the band diagram approximation cannot be neglected for larger fields. The following variation is frequently referred to as *non-parabolic* band approximation (in the sense of Kane[**82**]):

$$E_c(\boldsymbol{k})(1+\alpha E_c(\boldsymbol{k})) = \frac{\hbar^2}{2m^*}|\boldsymbol{k}|^2.$$

The validity of the semi-classical picture is restricted by the Heisenberg uncertainty principle. Position and momentum are conjugate variables and both of them cannot be measured sharply. It is assumed that the uncertainty of the momentum is small, so that the electron energy is defined sharply.

According to this, the uncertainty in the electron position should be small in the length scale of the variation of the external potential.
So far, the electrostatic potential V was regarded as given, and the Boltzmann equation models the motion of one particle in this potential. Electron and hole ensembles interact on short distances via collision mechanisms. These are commonly denoted as *scattering events* to emphasize the wave character of these quasi particles in the semi-classical picture. Additionally, they interact on long ranges via the *Coulomb force*

$$F(x,y) = -\frac{q}{4\pi\varepsilon_s}\frac{x-y}{|x-y|^3},$$

where ε_s is *electrical permittivity*, a material constant. If $\rho(x,t)$ is the total space charge density, the resulting electric field is

$$E_{\text{ef}}(x,t) = \frac{1}{4\pi\varepsilon_s}\int_{\mathbb{R}^3}\rho(y,t)\frac{x-y}{|x-y|^3}\,\mathrm{d}y.$$

The total space charge is the sum of the electron $n(x,t)$, the hole density $p(x,t)$ and the density of impurities ions, $C(x)$ that are inserted during the production process and are afterwards seen as fixed in the crystal. The electrostatic potential, defined via $E_{\text{ef}} = -\nabla V$, is the solution to the *Poisson equation*

$$\epsilon_s\Delta V = \rho = q(n-p-C) \quad \text{in } \mathbb{R}^3.$$

Remark 1.1. *The concentration of impurity ions, the so-called doping, $C(x)$, varies almost discontinuously over a large range of values. After scaling the equation in a dimensionless form this leads to a small diffusion coefficient, and locally to very large electric fields, also causing the main difficulty in simulating the macroscopic models for charge transport in semiconductor devices.*

Electron and hole densities are given via the Boltzmann equation leading to the coupled *Boltzmann-Poisson system* that is the starting point for the derivation of macroscopic models in the following:

$$\partial_t f_c + v_c(\boldsymbol{k})\cdot\nabla_x f_c + \frac{q}{\hbar}\nabla V\cdot\nabla_{\boldsymbol{k}} f_c = Q_c(f_c) + I_c(f_c,f_v), \tag{1.10}$$

$$\partial_t f_v + v_v(\boldsymbol{k})\cdot\nabla_x f_v - \frac{q}{\hbar}\nabla V\cdot\nabla_{\boldsymbol{k}} f_v = Q_v(f_v) + I_v(f_c,f_v), \tag{1.11}$$

$$\varepsilon_s\Delta v = q(n-p-C) \quad x\in\mathbb{R}^3, \boldsymbol{k}\in B, \tag{1.12}$$

where the densities are given through

$$n(x,t) = \int_B f_c(x,\boldsymbol{k},t)\,\mathrm{d}\boldsymbol{k}, \quad p(x,t) = \int_B f_v(x,\boldsymbol{k},t)\,\mathrm{d}\boldsymbol{k}. \tag{1.13}$$

Accompanied with initial values and periodic boundary conditions

$$f_i(x, \boldsymbol{k}, t) = f(x, -\boldsymbol{k}, t), \quad x \in \mathbb{R}^3,\ \boldsymbol{k} \in \partial B,\ t > 0 \tag{1.14}$$

$$f_i(x, \boldsymbol{k}, 0) = f_i^0(x, \boldsymbol{k}), \quad x \in \mathbb{R}^3,\ \boldsymbol{k} \in B,\ i \in \{c, v\}. \tag{1.15}$$

The terms on the right-hand sides $Q_i(f_i)$ and $I_i(\cdot, \cdot)$ model the scattering of particles. The terms $I_i(\cdot, \cdot)$ especially denote the effect of the generation of an electron-hole pair by absorbing energy and the counterpart of recombination of such a pair under the emission of energy. Before we detail the structure of these terms further, we note that the Boltzmann-Poisson system is a nonlinear integro-differential system in seven dimensions and therefore numerically only treatable with very high effort, e.g., employing Monte Carlo simulations. Often, they are used to fix parameters of previously phenomenological models with a higher accuracy and to investigate different, especially non-isotropic materials, see [**56, 115**]. In the following, we seek for simplifications through reducing the dimension.

We now have to specify the collision operator from which the macroscopic models are derived. As mentioned above, collisions are regarded as events changing the velocity instantaneously at a fixed point in space, and therefore concern mainly the pseudo wave vector $\boldsymbol{k}$, which is often called Bloch state in this context. For the moment, we neglect the possibility that a scattering event leads to a jump in the band index. We will therefore omit the band index as far as possible.

The number of electrons in a specific Bloch state $\boldsymbol{k}$ is changed by two different state transitions types. Either an electron is scattered "out" of its initial state $\boldsymbol{k}$ to the state $\boldsymbol{k}'$, or an electron is scattered "in" the state $\boldsymbol{k}$ from its former state $\boldsymbol{k}'$. At first, we describe the modeling of the "out" type transition. The rate of electrons at a point (x, t) changing its state from $\boldsymbol{k}$ to $\boldsymbol{k}'$ is proportional to the number of electrons available in the state $\boldsymbol{k}$ that is $f(x, k, t)$. The Pauli exclusion principle allows scattering into a specific state only, if this state is free. The proportion of unoccupied states $\boldsymbol{k}'$ is given by $(1 - f(x, \boldsymbol{k}', t))$. The probability that a particle changes its Bloch state from $\boldsymbol{k}$ to $\boldsymbol{k}'$ is denoted by $s(x, \boldsymbol{k}, \boldsymbol{k}')$ and called *scattering rate*. Altogether the rate at which particles change the state at the point (x, t) from $\boldsymbol{k} \to \boldsymbol{k}'$ is

$$s(x, \boldsymbol{k}, \boldsymbol{k}') f(x, \boldsymbol{k}, t)(1 - f(x, \boldsymbol{k}', t)).$$

The scattering "in" type transition is obtained by exchanging $\boldsymbol{k}$ and $\boldsymbol{k}'$ thus leading to the collision operator:

$$Q(f)(x, \boldsymbol{k}, t) = \int_B s(x, \boldsymbol{k}', \boldsymbol{k}) f'(1 - f) - s(x, \boldsymbol{k}, \boldsymbol{k}') f(1 - f') \, \mathrm{d}\boldsymbol{k}', \tag{1.16}$$

abbreviating $f = f(x, \boldsymbol{k}, t)$ and $f' = f(x, \boldsymbol{k}', t)$. Before classifying different collision operators, we also want to describe the recombination generation process phenomenologically. The generation of an electron-hole pair is only possible, if the final states are empty, thus, with the generation rate $g(x, \boldsymbol{k}, \boldsymbol{k}') \geq 0$, leading to

$$g(x, \boldsymbol{k}, \boldsymbol{k}')(1 - f_c)(1 - f_v').$$

The recombination of an electron in state $\boldsymbol{k}$ and a hole in state $\boldsymbol{k}'$ is given analogously by

$$r(x, \boldsymbol{k}, \boldsymbol{k}')f_c f_v',$$

where $r(x, \boldsymbol{k}, \boldsymbol{k}') \geq 0$. The physical principle of detailed balance at equilibrium relates r and g by $r(x, \boldsymbol{k}, \boldsymbol{k}') = \exp((E_c(\boldsymbol{k}) - E_v(\boldsymbol{k}'))/(k_B T))g(x, \boldsymbol{k}, \boldsymbol{k}')$. We obtain:

$$I_c(f_c, f_h)(x, \boldsymbol{k}, t) = \int_B g(x, \boldsymbol{k}, \boldsymbol{k}') \left(1 - e^{(E_c(\boldsymbol{k}) - E_v(\boldsymbol{k}'))/(k_B T)}\right) f_c f_p' \, \mathrm{d}\boldsymbol{k}',$$
$$I_v(f_c, f_h)(x, \boldsymbol{k}, t) = \int_B g(x, \boldsymbol{k}, \boldsymbol{k}') \left(1 - e^{(E_c(\boldsymbol{k}') - E_v(\boldsymbol{k}))/(k_B T)}\right) f_c' f_p \, \mathrm{d}\boldsymbol{k}'.$$

For more advanced models for recombination generation effects and the derivation of macroscopic limits, we refer to [**41**]. In the following, we will neglect these terms.

Different physical scattering mechanisms can be modelled via (1.16). We will not describe the underlying physical processes in a detailed way, see e.g. [**86**], as for the following derivation, it is more important to distinguish between elastic and inelastic collisions. For many scattering mechanisms, it is possible to separate a predominant elastic contribution from an inelastic correction. This is a key point in the derivation of macroscopic models we review. In the process of elastic collision the two states before and after the collision $\boldsymbol{k}$ and $\boldsymbol{k}'$ have equal energies $E(\boldsymbol{k}) = E(\boldsymbol{k}')$, i.e. the state of the particle changes only within surfaces of equal energy. These collision operators may be summarized by the formula:

$$Q_{\mathrm{el}}(f) = \int_B \phi_{\mathrm{el}}(\boldsymbol{k}, \boldsymbol{k}')\delta(E(\boldsymbol{k}) - E(\boldsymbol{k}'))(f' - f) \, \mathrm{d}\boldsymbol{k}, \tag{1.17}$$

where δ is the Dirac delta 'function', and ϕ_{el} is a cross section of the scattering rates of the underlying physical scattering mechanisms. The evaluation of the Dirac delta function in (1.17) is performed by means of the coarea formula, [**55, 88**]:

Lemma 1.1 (Coarea formula). *For any smooth function f, and periodic C^1 function $E(k) : B \to R \subset \mathbb{R}$ with nondegenerate critical points, it holds:*

$$\int_B f(\boldsymbol{k})\,\mathrm{d}\boldsymbol{k} = \int_R \int_{E^{-1}(\varepsilon)} f(k) \frac{\mathrm{d}F(\boldsymbol{k})}{|\nabla_{\boldsymbol{k}} E(\boldsymbol{k})|}\,\mathrm{d}\varepsilon, \tag{1.18}$$

where $\mathrm{d}F(\boldsymbol{k})$ is the two dimensional hypersurface element, and $E^{-1}(\varepsilon) = \{\boldsymbol{k} \in B : E(\boldsymbol{k}) = \varepsilon\}$.

Binary scattering of electrons will play an important role in the derivation of energy-transport models; we therefore report the collision term:

$$Q_{\mathrm{ee}}(f)(\boldsymbol{k}) = \int_{B^3} \phi_{\mathrm{ee}}(\boldsymbol{k}, \boldsymbol{k}', \boldsymbol{k}_1, \boldsymbol{k}_1')\delta(E'+E_1'-E-E_1)\delta_p(\boldsymbol{k}'+\boldsymbol{k}_1'-\boldsymbol{k}-\boldsymbol{k}_1) \\ [f'f_1'(1-f)(1-f_1)-ff_1(1-f')(1-f_1')]\,\mathrm{d}\boldsymbol{k}_1\,\mathrm{d}\boldsymbol{k}'\,\mathrm{d}\boldsymbol{k}_1'. \tag{1.19}$$

The mapping of the collision operator should result in a function whose domain within the $\boldsymbol{k}$ variable remains B, δ_p is a periodic Delta function and accounts for projecting $\boldsymbol{k}$ into B if $\boldsymbol{k}' + \boldsymbol{k}_1' - \boldsymbol{k}_1$ is not in B. Additionally, it is required by the principle of detailed balance that

$$\phi_{\mathrm{ee}}(\boldsymbol{k}, \boldsymbol{k}', \boldsymbol{k}_1, \boldsymbol{k}_1') = \phi_{\mathrm{ee}}(\boldsymbol{k}', \boldsymbol{k}, \boldsymbol{k}_1, \boldsymbol{k}_1') = \phi_{\mathrm{ee}}(\boldsymbol{k}_1, \boldsymbol{k}_1', \boldsymbol{k}, \boldsymbol{k}').$$

1.1.2 Macroscopic Models

In the scaled form the semiconductor Boltzmann equation for those electrons in the conduction band can be written as

$$\alpha^2 \partial_t f + \alpha(\nabla_{\boldsymbol{k}} E(\boldsymbol{k}) \cdot \nabla_x f + \nabla_x V \cdot \nabla_{\boldsymbol{k}} f) = Q(f). \tag{1.20}$$

Let λ_{el} be the length of the mean free path between two consecutive *elastic* collisions and λ_{inel} the corresponding length for *inelastic* collisions. The scaling parameter is then given by

$$\alpha^2 = \frac{\lambda_{\mathrm{el}}}{\lambda_{\mathrm{inel}}}.$$

It is assumed that elastic collisions dominate the collision operator, i.e. $0 < \alpha \ll 1$. For the comparison of limiting procedures it is convenient to define $\beta = \lambda_{\mathrm{ee}}/\lambda_{\mathrm{el}}$, where λ_{ee} is the mean free path length between two binary *electron-electron* collisions. The collision operator is assumed to have the form

$$Q(f) = Q_{\mathrm{el}} + \beta Q_{\mathrm{ee}} + \alpha^2 Q_{\mathrm{inel}}$$

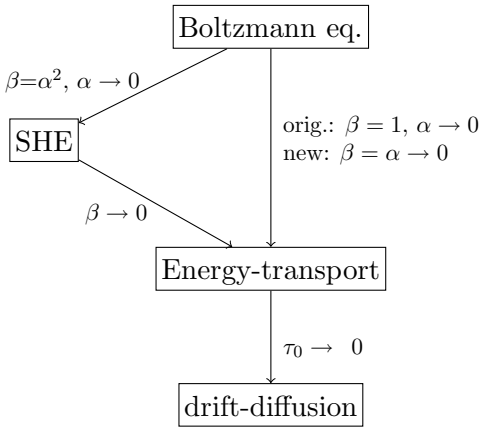

Figure 1.2: The scaling parameter α relates time between two consecutive collisions and the macroscopic time, whereas β relates the distances between consecutive collisions of specific types, τ_0 is the energy relaxation time.

The models that are derived from this scaling are depicted in Figure 1.2, including the limiting procedures connecting them. The direct limit from the Boltzmann equation to the energy transport equations was investigated in [**16**]. It has the drawback that the coefficients in the resulting model are given via operator equations that require additional numerical effort. For this limit elastic and electron-electron collisions are set to be of equal order, i.e. $\beta = \alpha^0$. In [**17**] a different procedure via an intermediate model was proposed that allowed for much simpler and in a reduced case even explicit expressions for the coefficients in the resulting energy transport model. The intermediate model, the SHE model, is often referred to as a mesoscopic model, since it describes the distribution function via levels of kinetic energy, which is much sharper than describing the width of the Fermi-Dirac distribution via temperature. More recently, it was found that it is possible to retrieve the same energy transport model as in the two step procedure via a different direct scaling limit, see [**50**]. In this limit electron-electron collisions lie in the intermediate scale between elastic and inelastic collisions, $\beta = \alpha^1$.

On a formal level we now sketch the derivation via the SHE model and subsequently the limit to the energy-transport model. We follow the presentation in [**50**]. Mathematically more rigorous proofs of the results may be found there and in the references therein. The first step is to insert the Hilbert expansion

$$f = f_0 + \alpha f_1 + \alpha^2 f_2 + \dots$$

into (1.20) for $\beta = \alpha^2$ and to identify terms of equal order in α to obtain

$$\overset{\alpha^0}{\Longrightarrow} \quad Q_{\mathrm{el}}(f_0) = 0, \tag{1.21}$$

$$\overset{\alpha^1}{\Longrightarrow} \quad Q_{\mathrm{el}}(f_1) = \nabla_{\boldsymbol{k}} E(\boldsymbol{k}) \cdot \nabla_x f_0 + \nabla V \cdot \nabla_{\boldsymbol{k}} f_0, \tag{1.22}$$

$$\overset{\alpha^2}{\Longrightarrow} \quad Q_{\mathrm{el}}(f_2) = \partial_t f_0 + \nabla_{\boldsymbol{k}} E(\boldsymbol{k}) \cdot \nabla_x f_1 + \nabla_x V \cdot \nabla_{\boldsymbol{k}} f_1 - (Q_{\mathrm{inel}} + Q_{\mathrm{ee}})(f_0). \tag{1.23}$$

The first equation requires $f_0 \in N(Q_{\mathrm{el}})$, the kernel of Q_{el} which is given by functions only depending on $E(\boldsymbol{k})$, thus

$$f_0(x, \boldsymbol{k}, t) = F(x, E(\boldsymbol{k}), t).$$

The solvability conditions for (1.22)-(1.23) lead to the equations that govern the evolution of F.

Proposition 1.1. [17]*Formally, for $\beta = \alpha^2$ the solution to (1.20) tends to a function $F = F(x, E(k), t)$ as $\alpha \to 0$ where $F(x, \varepsilon, t)$ satisfies the following* **SHE model***:*

$$N(\varepsilon)\partial_t F + \nabla_* J = S_{ee,inel}(F), \tag{1.24}$$

$$J(x, \varepsilon, t) = -D(x, \varepsilon)\nabla_* F, \tag{1.25}$$

with

$$\nabla_* = \nabla_x + \nabla_x V \partial_\varepsilon, \quad N(\varepsilon) = \int_B \delta(E(k) - \varepsilon)\, \mathrm{d}\boldsymbol{k}, \tag{1.26}$$

$$S_{ee,inel}(F)(\varepsilon) = \int_B (Q_{ee} + Q_{inel})(F(x, E(\boldsymbol{k}), t))\delta(E(\boldsymbol{k}) - \varepsilon)\, \mathrm{d}\boldsymbol{k}, \tag{1.27}$$

$$D(x, \varepsilon) = -\int_B \nabla_{\boldsymbol{k}} E(\boldsymbol{k}) Q_{el}^{-1}(\nabla_{\boldsymbol{k}} E(\boldsymbol{k}))\delta(E(\boldsymbol{k}) - \varepsilon)\, \mathrm{d}\boldsymbol{k}. \tag{1.28}$$

The term $N(\epsilon)$ is called the density of states of energy ε.

The computation of the diffusion matrix $D(x, \varepsilon)$ requires the inversion of the elastic collision operator. For a number of cases of physical interest this can be done explicitly, see below. The macroscopic quantities, electron density and electron energy density, are given by

$$\begin{pmatrix} n(x,t) \\ \mathcal{E}(x,t) \end{pmatrix} = \int_R \begin{pmatrix} 1 \\ \varepsilon \end{pmatrix} N(\varepsilon) F(x, \varepsilon, t)\, \mathrm{d}\varepsilon.$$

These quantities implicitly depend on the distribution function F. If we assume for a moment that F obeys a certain distribution parameterized by

macroscopic quantities, the multiplication of (1.24) with $(1,\varepsilon)^\top$ together with this F leads to a purely macroscopic *energy-transport* model. Indeed, the next asymptotic limit will provide such a distribution function for F. We assume that electron-electron collisions dominate the physical regime and we set afresh the right-hand side in equation (1.24) to

$$S(F_\beta) = \int_B (Q_{\text{inel}} + \frac{1}{\beta} Q_{\text{el}})(F_\beta(x, E(\boldsymbol{k}), t))\delta(E(\boldsymbol{k}) - \varepsilon)\,\mathrm{d}\boldsymbol{k}.$$

As in the previous step in the limit $\beta \to 0$ the function F_0 is required to be in the kernel of the dominant collision operator S_{ee}, which is obtained from (1.27) by omitting the inelastic collisions. The dominant collision operator has the following properties:

Lemma 1.2. [17]

(i) Entropy inequality: $\displaystyle\int_R S_{ee}(F)\ln(F/(1-F))\,\mathrm{d}\varepsilon \le 0.$

(ii) Mass and energy conservation: $\displaystyle\int_R S_{ee}(F)(1,\varepsilon)^\top\,\mathrm{d}\varepsilon = (0,0)^\top.$

(iii) Equilibrium states: $S_{ee}(F)=0 \iff \exists(\mu,T)\in\mathbb{R}\times[0,\infty)$ *s.t.*

$$F = \mathcal{F}_{\mu,T}(\varepsilon) = \frac{1}{1+\exp(\frac{\varepsilon-\mu}{T})}$$

The equilibrium distribution is the Fermi-Dirac distribution. The inelastic collision operator S_{inel}, which is defined by dropping Q_{ee} in (1.27) is assumed to be mass conservative, $\int_R S_{\text{inel}}\,\mathrm{d}\varepsilon = 0$. Equipped with the Fermi-Dirac parameterization of F we perform the above mentioned multiplication of (1.24) and (1.25) by $(1,\varepsilon)^\top$ and obtain the following result. The potential V may be given by a self-consistent coupling with the Poisson equation.

Proposition 1.2. [17] *Formally, F_β tends to the Fermi-Dirac distribution function $\mathcal{F}_{\mu,T}(\varepsilon)$ where the position-time-dependent chemical potential $\mu(x,t)$ and temperature $T(x,t)$ satisfy the* **energy-transport model***:*

$$\frac{\partial}{\partial t}\begin{pmatrix} n(\mu,T) \\ \mathcal{E}(\mu,T)\end{pmatrix} - \begin{pmatrix}\operatorname{div}_x J_1 \\ \operatorname{div}_x J_2\end{pmatrix} = \begin{pmatrix}0 \\ J_1\cdot\nabla_x V\end{pmatrix} + \begin{pmatrix}0 \\ W\end{pmatrix}, \tag{1.29}$$

$$\begin{pmatrix}J_1 \\ J_2\end{pmatrix} = \begin{pmatrix}\mathcal{D}_{11} & \mathcal{D}_{12} \\ \mathcal{D}_{21} & \mathcal{D}_{22}\end{pmatrix}\begin{pmatrix}\left(\nabla_x\left(\frac{\mu}{T}\right) - \frac{\nabla_x V}{T}\right) \\ \nabla_x\left(-\frac{1}{T}\right)\end{pmatrix}. \tag{1.30}$$

The entries of the diffusion matrix $\mathcal{D}$ are given by

$$\mathcal{D}_{ij}(\mu,T)=\int_R D(x,\varepsilon)\mathcal{F}_{\mu,T}(1-\mathcal{F}_{\mu,T})\varepsilon^{i+j-2}\,\mathrm{d}\varepsilon,\quad i,j\in\{1,2\}. \tag{1.31}$$

The term W is called energy relaxation term and depends on the inelastic scattering processes:

$$W(\mu,T)=\int_R S_{inel}(\mathcal{F}_{\mu,T})\varepsilon\,\mathrm{d}\varepsilon. \tag{1.32}$$

The last limit is to recover the drift-diffusion model. It is assumed that the energy loss due to inelastic collision is large and that the electron temperature returns to the lattice temperature. We therefore substitute the last term in (1.29) by $(0,\sigma^{-1}W)^\top$. Following the lines in [**17**] we perform a Hilbert expansion of the form

$$\mu^\sigma=\mu_0+\sigma\mu_1+\ldots,\quad T^\sigma=T_0+\sigma T_1,\quad J_i^\sigma=J_{i0}+\sigma J_{i1}.$$

Identifying the equations of the order σ^{-1} and σ^0, we obtain:

$$0=W(\mu_0,T_0,T_L), \tag{1.33}$$

$$\partial_t n(\mu_0,T_0)+\operatorname{div} J_{10}=0, \tag{1.34}$$

where J_{10} is obtained from the first row of (1.30) by substituting μ^σ,T^σ with μ_0,T_0. It was shown in [**17**] that the solution to (1.33) and (1.34) for a constant lattice temperature T_L is $T_0=T_L$ and

$$J_{10}=\mathcal{D}_{11}(x,\mu_0,T_L)\frac{\nabla(\mu_0-V)}{T_L}. \tag{1.35}$$

Assuming for simplicity that $D(x,\varepsilon)$ is a constant, $T_L=1$, a parabolic band structure and $n(\mu_0,T_L)=ce_0^\mu$, for some constant c, simple calculations lead to $\mathcal{D}_{11}=e^\mu=n/c$. Introduced in (1.35) we obtain for the current

$$J_{10}=\frac{n}{c}\nabla(\log n-c-V))=\frac{1}{c}(\nabla n-n\nabla V)$$

which is the classical drift-diffusion current density relation.
Summarizing this subsection, we point out that due to the understanding of limiting procedures all parameters in the macroscopic models are based on microscopic mechanisms. It is now possible to use refined parameters obtained by Monte Carlo simulations, for adjusting the parameters in the macroscopic models, enabling the simulation of advanced materials. The

energy-transport models have to be solved on a domain with fewer dimensions than the mesoscopic or even the microscopic models, and they still incorporate the temperature which is important for simulating "hot electron" effects. This makes energy-transport models altogether very interesting for simulations in the design process of new devices. Other so-called hydrodynamic models can be derived via moment methods from the Boltzmann equation. The resulting macroscopic model contains hyperbolic modes. Special numerical methods have been developed for those models as well, see [**18**, **38**, **60**, **61**]. The energy-transport model is of parabolic type. This makes the discretization in the transient case a little easier. In the physical literature, energy-transport equations have been derived from hydrodynamic models usually by neglecting certain convection terms (see, e.g., [**103**] and references therein). This approach can be made rigorously by considering a diffusion time scaling [**62**], which gives an additional connection in the hierarchy of macroscopic semiconductor models, see also [**81**] for a numerical comparison.
In the next section, we give a short overview on the state of the art in energy-transport models. Furthermore, we specify the parameters on the microscopic level that lead to the class of stationary energy-transport models which be the subject for the simulations in this work.

1.1.3 Energy-transport Models – An Overview

The analysis of the class of models consisting of the equations (1.29)-(1.32), coupled to the Poisson equation (1.12) were performed under the assumption of uniformly bounded diffusion coefficients, $\mathcal{D}$, in [**47**, **48**]. The existence and uniqueness of weak solutions to both the stationary and the time-dependent (initial) boundary-value problems were proved. Existence results with different assumptions (for instance, near-equilibrium situations) were shown in [**3**, **68**, **78**, **54**, **40**, **39**].
The numerical discretization of energy-transport models has been investigated in the physical literature for quite some years [**5**, **36**, **57**, **37**, **110**, **114**]. Mathematicians started to pay attention to these models in the 1990s, using ENO (essentially non-oscillatory) numerical schemes [**77**], finite-difference methods [**58**, **59**, **101**], mixed finite-volume schemes [**20**], mixed finite-element methods [**85**, **92**], or mixed finite-elements for the formulation (1.36) [**49**] (see also [**24**] for an overview), but always for a *fixed* discretization mesh.
In this work, we consider only *stationary* energy-transport models. The structure of the class of models that we will use in the simulations needs

to be elaborated. In order to find constitutive relations for $\mathcal{D}$ and W, we formulate the following three main hypotheses:

- The energy band diagram is spherical symmetric. The Brillouin zone is therefore replaced by $\mathbb{R}^3$. We continue to denote $E(k) = E(\boldsymbol{k})$, for some $\boldsymbol{k}$ with $|\boldsymbol{k}| = k$, and assume that $E(k)$ is strictly monotone.
- The Fermi-Dirac distribution is substituted by the Maxwell-Boltzmann distribution given by $M(\mu, T) = \exp(-(\varepsilon - \mu)/T)$. Assuming that ε is large enough, we approximate $\mathcal{F}(1 - \mathcal{F}) \approx M$ as well. In the physics literature, this assumption is referred to as a semiconductor with non-degenerate statistics.
- The elastic collision operator is a *relaxation time* operator. These operators follow from (1.17) under the assumption that the scattering rate is energy and position dependent only $\phi_x(\boldsymbol{k}, \boldsymbol{k}') = \phi(x, E(\boldsymbol{k})) = \phi_0(x)E^\beta$, $\beta > -2$.

The key point is that under these assumptions the constitutive relations of both currents (1.30) obey the same convection diffusion form, which was discovered in [**49**]. Indeed, it is possible to define the new variables g_1, g_2 such that the constitutive relations for the current densities have a drift-diffusion form:

$$J_i = \nabla g_1 - \frac{\nabla V}{T} g_1, \quad i \in \{1, 2\}, \tag{1.36}$$

where the temperature is given by $T = T(g_1, g_2)$.(In fact, $g_1 = \mathcal{D}_{11}$ and $g_2 = \mathcal{D}_{21}$; see [**49**] for details.)
In literature, the values $\beta = 1/2$ (used by Chen et al. [**36**]) and $\beta = 0$ (used by Lyumkis et al. [**87**]) have been employed. In case $\beta = 1/2$ the diffusion matrix for the parabolic band approximation has the form

$$\mathcal{D}(x, \mu, T) = \frac{n(\mu, T)}{6\sqrt{2}\pi\phi_0(x)} \begin{pmatrix} \mathrm{Id} & \frac{3}{2}T\mathrm{Id} \\ \frac{3}{2}T\mathrm{Id} & \frac{15}{4}T^2\mathrm{Id} \end{pmatrix}.$$

We call the resulting model with this diffusion matrix the *Chen model*. In case $\beta = 0$, we obtain

$$\mathcal{D}(x, \mu, T) = \frac{2T^{1/2}n(\mu, T)}{3\phi_0(x)} \begin{pmatrix} \mathrm{Id} & 2T\mathrm{Id} \\ 2\mathrm{Id} & 6T^2\mathrm{Id} \end{pmatrix}.$$

The corresponding energy-transport equations are called the *Lyumkis model*. If we use the Fokker-Planck approximation[**106**] of the energy relaxation term (1.32), it can be written in those new variables as $W = c_1 g_1 - c_2 g_2$

with $c_i = c_i(g_1, g_2) \geq 0$. Using the current densities (1.36) we thus write the stationary energy-transport model in a scaled form: (See Chapter 3 for the precise constitutive relations.)

$$-\operatorname{div} J_1 = 0, \tag{1.37}$$

$$-\operatorname{div} J_2 + c_2 g_2 = -J_1 \cdot \nabla V + c_1 g_1, \tag{1.38}$$

$$\lambda^2 \Delta V = n(g_1, g_2) - C. \tag{1.39}$$

The above equations have to be solved in a bounded domain $\Omega \subset \mathbb{R}^2$. We now (implicitly) complement the equations with physically motivated mixed Dirichlet-Neumann boundary conditions:

$$n = n_D,\ T = T_D,\ V = V_D \qquad \text{on } \Gamma_D, \tag{1.40}$$

$$J_1 \cdot \nu = J_2 \cdot \nu = \nabla V \cdot \nu = 0 \qquad \text{on } \Gamma_N, \tag{1.41}$$

modelling the (Ohmic or Schottky) contacts Γ_D and the insulating boundary parts Γ_N. We have assumed that $\partial\Omega = \Gamma_D \cup \Gamma_N$ and $\Gamma_D \cap \Gamma_N = \varnothing$ holds and that the exterior normal unit vector ν exists almost everywhere on $\partial\Omega$. By freezing the coefficients in c_1, c_2 and $\nabla V T^{-1}$ (obtained from the previous iteration in the global iteration procedure), the energy-transport model can be written (in each iteration step) as a system of convection-diffusion equations.

As we pointed out in Remark 1.1 a very high electric field might be present in parts of the domain, which leads to a convection dominated problem. Therefore, an adaptive refinement strategy is crucial to resolve accurately and efficiently very sharp gradients occurring in the device simulations.

Standard mixed finite-element families like *Raviart–Thomas* (RT) or *Brezzi–Douglas–Marini* (BDM) elements of polynomial degree p are extensively used in various applications where the currents or fluxes are the more important quantities. However, in the discretization of a symmetric elliptic problem with zeroth order term they might not preserve the positivity of the solution for given positive boundary data, which is crucial in our simulation. Another class of mixed finite-elements was proposed by Marini and Pietra in [**89**] to overcome this deficiency in the simulation of drift-diffusion equations. Interestingly, the *Marini–Pietra* elements (MP) can be regarded as an extension of the RT–elements. In the case of a non-homogeneous Dirichlet problem without internal load, the discrete solutions are the same. The discretization of the drift-diffusion model takes advantage of the existence of a transformation on the continuous level which allows to rewrite the convection-diffusion problem as a symmetric elliptic problem. Due to

the non-constant temperature in the energy-transport model, this transformation can only be made on the discrete level. In Chapter 2, we detail this transformation and prove an a priori error estimate for the discrete solution of a general convection-diffusion-reaction problem.
In the next section, we present the relations between the above mentioned mixed finite-element classes, and, in the last part of this section, we focus on the mesh refinement process in an adaptive solution process.

1.2 A Mixed Finite–Element Framework

In this section we sketch an abstract error analysis framework, in which standard mixed finite-elements and the more specialized *Marini-Pietra* (MP) elements can jointly be analyzed. This framework was originally published in [**89**] and in a more streamlined form in [**24**, Chapters 3.1 to 3.6]. The main result of this comparison is given in Proposition 1.3 and Remark 1.2 on page 29.
A simplified linear model problem suits to elaborate the similarities and differences of different mixed finite-element methods. We shall use standard notation for Sobolev spaces, [**1**], and the respective norms that will occur. Let Ω be a polygonal domain in $\mathbb{R}^2$, and a, d, g and f be sufficiently regular functions, then we consider the problem: Find $u \in H^1(\Omega)$ so that

$$-\operatorname{div}(a\nabla u) + du = f \quad \text{in } \Omega, \qquad u = g \quad \text{on } \partial\Omega.$$

Assume that d is non-negative and bounded, a is bounded; the lower bound is given by a constant $a_0 > 0$. The central step for deriving a mixed formulation from the above equation is to introduce the flux $\sigma = a\nabla u$ as an independent variable, thus we have

$$\begin{aligned} a^{-1}\sigma - \nabla u &= 0 && \text{in } \Omega, \\ -\operatorname{div}\sigma + du &= f && \text{in } \Omega, \\ u &= g && \text{on } \Gamma = \partial\Omega. \end{aligned} \tag{1.42}$$

We abbreviate the notation of the normed function spaces by

$$\begin{aligned} W &= L^2(\Omega), & \|w\|_W &= \|w\|_0 = \|w\|_{L^2(\Omega)}, \\ V &= H(\operatorname{div},\Omega), & \|\boldsymbol{q}\|_V^2 &= \|\boldsymbol{q}\|_0^2 + \|\operatorname{div}\boldsymbol{q}\|_0^2. \end{aligned}$$

Let us introduce the bilinear forms $a(\cdot,\cdot)\!:\!V\times V \to \mathbb{R}$, $b(\cdot,\cdot)\!:\!W\times V \to \mathbb{R}$, $d(\cdot,\cdot)\!:\!W\times W \to \mathbb{R}$

$$a(\sigma,\tau) = \int_\Omega a^{-1}\sigma\tau \,\mathrm{d}x, \quad b(v,\tau) = \int_\Omega v \operatorname{div}\tau \,\mathrm{d}x, \quad d(u,w) = \int_\Omega duv \,\mathrm{d}x.$$

The weak mixed formulation then reads:

$$\begin{aligned} &\text{Find } (\sigma, u) \in V \times W \text{ such that} \\ &a(\sigma, \tau) + b(u, \tau) = \langle g, \tau \cdot \nu \rangle_{|\Gamma} \quad \forall \tau \in V, \\ &b(w, \sigma) - d(u, w) = -(f, w) \quad \forall w \in W, \end{aligned} \tag{1.43}$$

where $\langle \cdot, \cdot \rangle_{|\Gamma}$ is the duality pairing between $H^{1/2}(\Gamma)$ and its dual space $H^{-1/2}(\Gamma)$, and $(\cdot, \cdot)$ is the inner product in $L^2(\Omega)$. The following properties ensure well-posedness of the continuous problem (1.43), see [**24**, Theorem 3.1]. There exist constants $\alpha > 0$, $\gamma > 0$ so that:

(P1) $a(\cdot, \cdot)$, $b(\cdot, \cdot)$, and $d(\cdot, \cdot)$ are continuous bilinear forms.

(P2) $a(\tau, \tau) \geq \alpha \|\tau\|_0^2, \ \forall \tau \in V.$

(P3) $\sup\limits_{\tau \in V} \dfrac{b(w, \tau)}{\|\tau\|_V} \geq \gamma \|w\|_W \quad \forall w \in W, \quad$ an *inf-sup* condition.

Furthermore, the solution (σ, u) lies in a more regular space $V^* \times W$, see [**24**, Theorem 3.2], with

$$V^* = \{\tau \in (L^p(\Omega))^2 | \, p > 2, \ \operatorname{div} \tau \in L^2(\Omega)\}.$$

and an inf-sup condition in a stronger norm $\|\cdot\|_{V^*}$ holds as well. For the discrete problem, we assume the Ω is exactly covered by a regular simplicial triangulation $\mathcal{T}_h$ in the sense of Ciarlet [**42**]. The discrete spaces V_h and W_h are assumed to consist of piecewise polynomial functions and the element-wise restrictions are denoted as $V_h(K)$ and $W_h(K)$. Here we treat only conforming approximations of the vector valued part of the solution, i.e. $V_h \subset V$. However, to state a discrete problem this requirement is not mandatory and non-conforming approximation spaces fitting in a slightly extended framework have been proposed as well, see [**89**, Example 5].
The discrete analogue of the inf-sup condition has some implication for the possible combination of approximation spaces V_h and W_h. It is therefore useful to define the normed space $\hat{V}$ related to a triangulation

$$\hat{V} = L^2(\Omega) \cap \prod_{K \in \mathcal{T}_h} H(\operatorname{div}, K), \quad \|\tau\|_{\hat{V}}^2 = \|\tau\|_{0,\Omega}^2 + \sum_{K \in \mathcal{T}_h} \|\operatorname{div} \tau\|_{0,K}^2.$$

The bilinear form b is meaningless on the space $\hat{V}$. It is substituted by

$$b_h(w, \tau) = \sum_{K \in \mathcal{T}_h} b_h^K(w, \tau), \quad \text{with } b_h^K(w, \tau) = \int_K w \operatorname{div} \tau \, \mathrm{d}x,$$

and we notice that $b(w,\tau) \equiv b_h(w,\tau)$ for all $\tau \in V$ and $w \in W$. We associate an operator $B : \hat{V} \to W'$ with the bilinear form $b_h(\cdot,\cdot)$ and distinguish between:

$$\begin{aligned} \mathrm{Ker}B_h &:= \big\{\tau_h \in \textstyle\prod_K V_h(K) |\ b_h^K(w_h,\tau_h) = 0\ \forall w_h \in W_h\big\}, \quad \text{and} \\ \mathrm{Ker}\hat{B} &:= \big\{\tau \in \hat{V} |\ b_h(w,\tau) = 0\ \forall w \in W\big\}. \end{aligned}$$

The discrete formulation of (1.43) is now:

$$\begin{aligned} &\text{Find } (\sigma_h, u_h) \in V_h \times W_h \text{ such that} \\ &a(\sigma_h,\tau_h) + b(u_h,\tau_h) = \langle g, \tau_h \cdot \nu\rangle_{|\Gamma} \quad \forall \tau_h \in V_h, \\ &b(w_h,\sigma_h) - d(u_h,w_h) = -(f,w_h) \quad \forall w_h \in W_h. \end{aligned} \tag{1.44}$$

In this framework the proof of existence of discrete solutions and optimal error bounds is based on the following additional assumptions on the discrete spaces:

(P4) $\mathrm{Ker}B_h \subset \mathrm{Ker}\hat{B}$.

(P5) There exists an interpolation operator $\Pi_h : V^* \to V_h$ such that

$$b_h(w_h, \tau - \Pi_h\tau) = 0\ \forall w_h \in W_h, \text{ and} \qquad \|\Pi_h\tau\|_{\hat{V}} \le C\|\tau\|_{V^*}\ \forall \tau \in V^*$$

with a mesh-independent constant C.

These assumptions have the following consequences. Firstly, they imply that a discrete inf-sup condition holds:

$$\exists\, \gamma > 0 : \quad \sup_{\tau_h \in V_h} \frac{b_h(w_h,\tau_h)}{\|\tau_h\|_{\hat{V}}} \ge \gamma \|w_h\|_W. \tag{1.45}$$

This inter-relates the dimension of the approximation spaces:

$$dim(\mathrm{div}\, V_h) \equiv dim(W_h).$$

Moreover, for any choice of basis functions $\{w_1,\ldots,w_r\}$ and $\{d_1,\ldots,d_r\}$ in $W_h(K)$ and $\mathrm{div}\, V_h(K)$ respectively the matrix

$$\int_K d_i w_j \,\mathrm{d}x \quad i,j \in \{1,\ldots,r\} \tag{1.46}$$

is nonsingular. To prove (1.45) we use the inf-sup condition given for the continuous problem in the form that for any $w_h \in W_h \subset W$, there exists $\tau \in V^*$ such that

$$\frac{b(w_h,\tau)}{\|\tau\|_{V^*}} \le \gamma^* \|w_h\|_W$$

Now we infer

$$\frac{b_h(w_h, \Pi\tau)}{\|\Pi\tau\|_{\hat{V}}} \geq \frac{b_h(w_h, \tau)}{C\|\tau\|_{\hat{V}}} \geq \frac{b(w_h, \tau)}{C\|\tau\|_{\hat{V}}} \geq \frac{\gamma^*}{C}\|w_h\|_W$$

hence (1.45) holds with $\gamma = \gamma^*/C$. The assumption **(P4)** and (1.45) render the second consequence. We now need to define an interpolation operator on the primal field. Because of the invertibility of the matrix in equation (1.46) this can be done by requiring that

$$\int_K (w - P_h w) \operatorname{div} \tau_h \, \mathrm{d}x = 0 \quad \forall \tau_h \in V_h(K). \tag{1.47}$$

Now we are able to write the abstract error estimate:

Proposition 1.3. [**24**] *Let (σ, u) be a solution to (1.43) and (σ_h, u_h) be a solution of (1.44). Under the assumptions* **(P1)**–**(P5)** *there exists a constant C that is independent of the mesh size h, such that the following estimate holds.*

$$\|\Pi_h\sigma - \sigma_h\|_{\hat{V}} + \|P_h u - u_h\|_W \leq C(\|\sigma - \Pi_h\sigma\|_0 + \|u - P_h u\|_W)$$

Proof. Due to the conformity of the finite dimensional spaces, we obtain a discrete difference problem by subtracting (1.43) from (1.44) and adding and subtracting projections of σ and u into the discrete spaces:

$$\begin{cases} a(\Pi_h\sigma - \sigma_h, \tau_h) + b_h(P_h u - u_h, \tau_h) = a(\Pi_h\sigma - \sigma, \tau_h) & \forall \tau_h \in V_h \\ b_h(w_h, \Pi_h\sigma - \sigma_h) - d(P_h u - u_h, w_h) = d(u - P_h u, w_h) & \forall w_h \in W_h \end{cases}$$

Well-posedness of the discrete problem yields the error estimate. The constant C depends on the continuity properties of the underlying bilinear forms, the coercivity of $a(\cdot,\cdot)$ and the inf-sup condition for $b_h(\cdot,\cdot)$. □

An estimate on the difference $\|\sigma - \sigma_h\|$ in a certain norm $\|\cdot\|$ may be deduced from Proposition 1.3 via the triangle inequality

$$\|\sigma - \sigma_h\| \leq \|\sigma - \Pi_h\sigma\| + \|\Pi_h\sigma - \sigma_h\|.$$

If both terms on the right-hand side can be bounded in a specific norm we directly deduce an estimate on the error in this norm. Taking, for instance, $\|\cdot\| = \|\cdot\|_0$ the second term is bounded due to $\|\cdot\|_0 \leq \|\cdot\|_{\hat{V}}$. An interpolation estimate in the L^2–norm is at hand for standard mixed finite-elements as well as for the *MP*–elements. A bound in the stronger norm on $\hat{V}$ requires an estimate for $\|\Pi_h\sigma - \sigma\|_{\hat{V}}$ which is not available for the *MP*–elements. We

recall that $\Pi_h\sigma$ is only defined via $b_h(w_h, \sigma - \Pi_h\sigma) = 0$, which depends on the actual pair (V_h, W_h) of approximation spaces.

Remark 1.2. [24, equations (3.110)-(3.111), (3.124)] *The Raviart–Thomas element of lowest order RT_0 as well as the conforming Marini–Pietra element are accompanied by $W_h(K) = P_0(K)$, the element-wise constant functions. For both elements we have:*

$$\|\sigma - \sigma_h\|_0 + \|u - u_h\|_W \leq Ch(|\sigma|_1 + |u|_1).$$

For the RT_0 it holds $\|\sigma - \Pi_h\sigma\|_{\hat{V}} \leq Ch(|\sigma|_1 + |\operatorname{div}\sigma|_1)$, see [23, 42]*, consequently we deduce*

$$\|\sigma - \sigma_h\|_{\hat{V}} + \|u - u_h\|_W \leq Ch(|\sigma|_1 + |\operatorname{div}\sigma|_1 + |u|_1).$$

The advantages of the *MP*–elements will become apparent on the implementational side. Furthermore, these advantages enable the a priori analysis for a general convection–diffusion problem that we will carry out in Chapter 2. Here we explain further similarities and the fundamental difference between RT_0– and *MP*–elements. The dimension of the two spaces is the same, they are spanned locally by

$$\begin{aligned} RT_0(K) &= \operatorname{span}\{(0,1)^\top, (1,0)^\top, \boldsymbol{q}_{RT_0}\}, \quad \boldsymbol{q}_{RT_0} = (x,y)^\top \\ MP(K) &= \operatorname{span}\{(0,1)^\top, (1,0)^\top, \boldsymbol{q}_{MP}\}, \quad \boldsymbol{q}_{MP} \in P_2(K). \end{aligned}$$

The precise form of $\boldsymbol{q}_{MP}$ will be specified later, at this stage we only note that it is a special second order polynomial in each component. Let τ_i denote the basis elements. By construction (see Section 2.1) the matrix

$$\int_{e_j(K)} \tau_i \cdot n_j \,\mathrm{d}s \qquad \forall K \in \mathcal{T}_h,$$

is invertible for both elements, where n_j is the outward normal vector of the edge e_j. Consequently the local projections $\Pi_h^{RT_0}\tau$ and $\Pi_h^{MP}\tau$ are given by

$$\int_{e_j(K)} (\tau - \Pi_h^S\tau)\cdot n_j \,\mathrm{d}s = 0 \quad \forall j \in \{1,2,3\}, S \in \{MP, RT_0\}.$$

For $\tau_h \in RT_0(K)$ it is obvious that $\operatorname{div}\tau_h \in P_0(K)$. The projection P_h is therefore the usual L^2-projection

$$\int_K (v - P_h v)w = 0 \qquad \forall w \in W.$$

This is not true for the *MP*–elements since $\operatorname{div} \boldsymbol{q}_{MP} \in P_1(K)$, but an interpolation estimate still exists. On the other hand, if we equip $\hat{V}$ with the norm

$$\|\tau\|_h^2 = \|\tau\|_0^2 + \sum_{K \in \mathcal{T}_h} {h_K}^2 \|\operatorname{div} \tau\|_0^2,$$

it is possible to prove $\|\Pi_h^{MP}\tau\|_h \leq C\|\tau\|_{V^*}$. It is not necessary for the *RT*-elements to exchanging the norm on $\hat{V}$, i.e. we directly have $\|\Pi_h^{RT_0}\tau\|_{\hat{V}} \leq C\|\tau\|_{V^*}$. Accordingly, the properties **(P4)**, **(P5)** are fulfilled and the abstract framework applies.
As a final statement of this section, we point out that the main requirement in the target application is the conforming approximation in $H(\operatorname{div}, \Omega)$ which leads to conservation of the total particle current. This requirement is fulfilled by both elements.

1.2.1 Hybridization

The algebraic system that corresponds to (1.44) may be indefinite, for instance if $\gamma \equiv 0$. Compared with standard one-field finite-element methods the linear system is much larger and it would be require to construct conforming basis functions in $H(\operatorname{div}, \Omega)$. A different way to solve the discrete problem (1.44) avoiding all of the above deficiencies is to drop the continuity property in the definition of the discrete spaces, to enforce it through an additional constraint and to solve the new problem by means of a Lagrange multiplier method. This procedure is called *hybridization*. We closely follow in this subsection the lines of [**24**, Section 3.6].
The space of multipliers is directly attached to $\mathcal{E}_h$, the set of edges of the triangulation. For the lowest order method it is precisely:

$$\Lambda_h = \{\mu_h \in L^2(\mathcal{E}_h) |\, \mu_h|_e \in P_0(e) \forall e \in \mathcal{E}_h\}.$$

In order to impose boundary conditions, we write for a given function g:

$$\Lambda_{h,g} = \{\mu_h \in \Lambda_h |\, \int_e \mu_h - g \,\mathrm{d}s = 0 \;\forall e \in \mathcal{E}_h \cap \Gamma\}.$$

The constraint will be based on the bilinear form

$$c_h(\mu_h, \tau_h) = \sum_{e \in \mathcal{E}_h} \int_e \mu_h [\tau \cdot n] \,\mathrm{d}s,$$

where $[\tau_h \cdot n]|_e$ denotes the jump of $\tau \cdot n$ over the edge e. For a function $\tau \in \hat{V}$ the constraint that τ is actually in V is given by the requirement

$$c_h(\mu_h, \tau) = 0 \qquad \forall \mu_h \in \Lambda_{h,0}.$$

The hybridized discrete problem then is:

$$\begin{aligned}
&\text{Find } (\sigma_h, u_h, \lambda_h) \in (\textstyle\prod V_h(K)) \times W_h \times \Lambda_{h,g} \text{ such that} \\
&a(\sigma_h, \tau_h) + b(u_h, \tau_h) - c_h(\lambda_h, \tau_h) = 0 \qquad \forall \tau_h \in V_h, \\
&b(w_h, \sigma_h) - d(u_h, w_h) = -(f, w_h) \qquad \forall w_h \in W_h, \\
&c_h(\mu_h, \sigma_h) = 0 \qquad \forall \mu_h \in \Lambda_{h,0}.
\end{aligned} \tag{1.48}$$

Existence and uniqueness follows directly from an inf-sup condition on the new bilinear form c_h. Once more we stress that through enforcing the continuity of the normal component of the vector-valued variable in the last of the above equations force the solution to be equal to that of problem (1.44). The Lagrange multipliers λ_h are an approximation of the solution's primal variable u that is only defined on the skeleton of the triangulation. Error estimates are derived utilizing the results of Proposition 1.3 in the following form:

Proposition 1.4. *If we assume that the triangulation is quasi-uniform, and the vector valued variable is approximated either by RT_0– or* MP*–elements, the following estimate holds (see* [**24**, equation (3.200)]*):*

$$\|\lambda_h - \mathcal{P}u\|_{0,\mathcal{E}_h} \leq C(h^{1/2}\|\sigma - \sigma_h\|_0 + h^{-1/2}\|P_h u - u_h\|_0) \leq Ch^{1/2}, \tag{1.49}$$

where $\mathcal{P}$ is the L^2-projection onto Λ_h.

So far, the introduction of the Lagrange multipliers has enlarged the corresponding linear system even further. Setting aside any inter-element continuity property in the approximation spaces for (σ_h, u_h) allows for an element-by-element elimination of these variables expressing them in terms of λ_h. By that means the linear system is reduced to the size of the number of internal edges of the triangulation. Here lies the main advantage of the *MP*–elements, since even in case $\gamma \not\equiv 0$ the resulting matrix is a *M-matrix*, and therefore a discrete maximum principle holds, ensuring that particle densities do not become negative. This is not true for the RT_0–element, and has negative effects, as shown in [**90**].
The procedure of eliminating the original unknowns is called *static condensation* and it has a major impact on the analysis that we carry out in Section 2.2, where we prove a new estimate for the Lagrange multipliers for a general convection-diffusion problem. The new technique to derive this estimate does not rely on independently-obtained estimates on the other variables.
Furthermore, the Lagrange multipliers λ_h allow for a postprocessing in which new higher order approximations $\hat{u}_h$ to the primal variable are generated.

Although this higher order estimate was only proven for symmetric problems approximated by standard mixed elements, it holds numerically also in the case of drift dominated simulations. This effect will be exploited to estimate *a posteriori* the error of a computed solution and to have an adaptive control on the refinement of the mesh in Section 2.3. The next section will give a brief introduction in this field.

1.2.2 Basic Adaptive Algorithm

Numerical errors are intrinsic to the simulation of the partial differential equations their exact solutions lie in infinite dimensional function spaces which are substituted by finite dimensional approximations in order to simulate them in the computer. The question is, how the error can be effectively minimized. The finite-element method is often also called a projection method, since it produces the optimal approximate solution for a given discrete space, with respect to a specific problem dependent norm, cf. *Galerkin orthogonality*.

The question, how the error can be effectively minimized, can be rephrased into how can appropriate discrete spaces be defined. The a priori estimation of numerical errors delivers only the asymptotic behaviour of a simulation technique. A discrete space that has been designed by means of an a priori bound to guarantee a prescribed accuracy in the solution is often far too complex.

If a less complex space is used instead, how can the accuracy of the calculated solution be measured? It is the objective of *a posteriori error analysis* to answer this question. In case of an insufficient accuracy, the discretization may be refined. An a posteriori error estimator $\eta(u_h)$ – computed from the approximate solution u_h – may therefore be broken up into local element-wise contributions η_K. These local indicators form the basis for adaptive refinement algorithms.

Different strategies can be pursued to manipulate the spaces. For a fixed triangulation the polynomial degree p can be altered, which is referred to as p–refinement. On the other hand, if the polynomial degree is fixed, the triangulation can be refined, thereby changing the element diameter h, accordingly denoted as h–refinement. A variation in both directions is called hp-refinement, although it is hard to decide, whether one should increase the polynomial degree or refine the mesh, see [**107, 75**] and the references therein. The approximation of functions with a large gradient by high order polynomials tend to be oscillating. This prohibits the use of high order elements, since, in regions with positive but small densities, these oscillations

might range into negative, thus unphysical, values. In this work we only consider the h–refinement for elements of lowest order.
The main concepts that distinguish the quality in estimating the error $\|u - u_h\|$ are reliability, efficiency and asymptotic exactness. *Reliability* guarantees that the error is below a certain bound

$$\|u - u_h\| \leq C_r \eta(u_h),$$

where C_r is a mesh independent constant. The true error might be much smaller, which would again lead to over-refinement, as in the case of mesh design by an a priori bound. This can be avoided if a lower bound is available. Therefore, an estimator is called *efficient* if a bound from below with a mesh independent constant C_e is given

$$C_e \eta(u_h) \geq \|u - u_h\|.$$

An estimator is said to be *asymptotic exact* if both constants C_r, and C_e tend to one. Two different a posteriori estimators will be assessed numerically in Section 2.3. Under certain assumptions on the approximation spaces and the regularity of the solution, reliability and efficiency estimates hold. These two estimators are used together to construct a new goal oriented estimator. These estimators are even more efficient in the sense that they directly consider the error in a (local) quantity of interest, whereas standard estimators may have only indirect control on these quantities through the norm that they control. The procedure is illustrated in Section 2.4.
The adaptive algorithm for all these estimators does not change besides the substitution of the error indicator. It is depicted in Algorithm 1. The actual mesh refinement is based on the PDEToolbox of `MATLAB`® .

Algorithmus 1 : Adaptive mesh refinement

Input : Initial mesh $\mathcal{T}_0$, tolerance $\delta > 0$, max. triangle number $K_{\max}$
Output : Refined mesh $\mathcal{T}'$, discrete solution u_h
$k \leftarrow 0$
repeat
  compute discrete solution u_h on the mesh $\mathcal{T}_k$
  compute global error estimator $\eta(u_h)$ and local error indicator η_K
  collect a set RK of elements have to be refined
  generate a refined mesh $\mathcal{T}_{k+1}$
  $k \leftarrow k + 1$
until $\eta(u_h) < \delta$ *or* $|T_k| > K_{\max}$
$\mathcal{T}' \leftarrow \mathcal{T}_{k-1}$

Chapter 2

A Hybridized Mixed Finite–Element Method for Convection-Diffusion Problems

This chapter consists of three main parts. In the first section the discretization is delineated leading to the main important property that the matrix of the algebraic system is an M-matrix, which is proven in Lemma 2.1. In Section 2.2 we describe the new interpretation of the discrete system after hybridization, that can be understood as a discrete bilinear form acting on the Lagrange multipliers only. We will obtain a new *a priori* estimate on the Lagrange multiplier that requires only minimal regularity of the solution to the continuous problem in Theorem 2.3. The last two parts of this chapter are devoted to different kinds of *a posteriori* error estimation techniques combined in a new dual-weighted-residual estimator[**15**] in Section 2.4.

2.1 Discretization

Basically, the method treats the convection term with a local exponential fitting discretization. These methods, often also referred to as Scharfetter and Gummel (SG) type discretization, have been used in many ways to solve convection–diffusion problems. The originally one-dimensional idea of Scharfetter and Gummel [**105**] has been extended to higher dimensions to

the so called SG-Box Method [9, 10, 25]. Several other generalizations of this idea can be found, especially if monotone schemes are desired, cf. e.g. [118, 96] and more recently [116]. In contrast to the edge-averaged finite-element method presented in [118] our method preserves additionally to the maximum principal also the continuity of the normal component of the flux density. In this section we state the precise problem assumption before elaborating on the connection of the final discretization given through problem (2.10) to the method introduced by Van Nooyen in [96]. We will therefore include an intermediate discretization, problem (2.4), that, additionally, has important utility for the analysis in Section 2.2.
We state the convection–diffusion problem directly as a system of first order equations, which is closely related to the mixed discretization of the problem. Let $\Omega \subset \mathbb{R}^2$ be a bounded polygonal domain and consider

$$\begin{aligned} \sigma &= -a(\nabla u - \boldsymbol{b}u) && \text{in } \Omega, \\ \operatorname{div}\sigma + du &= f && \text{in } \Omega, \\ u = g \text{ on } \Gamma_D,\ \sigma\cdot\nu &= 0 && \text{on } \Gamma_N \end{aligned} \tag{2.1}$$

We impose the following assumptions on the data.

Assumption 1. *The diffusion coefficient $a \in C^0(\bar{\Omega})$ with $0 < a_{min} \leq a(x) \leq a_{max}$ and $\boldsymbol{b} \in W^{1,\infty}(\bar{\Omega})$, $d \in C^0(\Omega), 0 \leq d(x) + \frac{1}{2}\operatorname{div}(a\boldsymbol{b})(x)$ almost everywhere in Ω, $f \in L^2(\Omega), \bar{\Gamma}_D \cup \Gamma_N = \partial\Omega$, $\Gamma_D \cap \Gamma_N = \varnothing$ and $g \in H^{1/2}(\Gamma_D)$.*

Under even weaker assumptions it is known that a maximum principle holds for this operator [64]. Let Ω be exactly covered by a regular simplicial triangulation $\mathcal{T}_h$ in the sense of [42]. We denote by $\mathcal{E}_h$ the set of edges e of the triangulation and by $\mathcal{N}_h$ the set of all nodes or vertices. May also Γ_D and Γ_N always be connected via a vertex of the triangulation. By $P_k(\mathcal{O})$ we denote the space of polynomials over the domain $\mathcal{O}$ of degree less than or equal to k. We denote the norms on spaces $H^k(\mathcal{O})$ by $\|\cdot\|_{k,\mathcal{O}}$. Associated with $\mathcal{T}_h$ let $W_h \subset L^2(\Omega)$ be the space of piecewise constant functions, let Λ_h be the space of edgewise constant functions, that may incorporate given boundary data in the following form

$$\Lambda_{h,u_D} = \Big\{\lambda \in \prod_{e\in\mathcal{E}_h} P_0(e); \int_e u_D - \lambda \,\mathrm{d}s = 0 \;\forall e \in \Gamma_D\Big\}.$$

In the mixed finite element discretization of (2.1) σ is approximated directly by a discrete function σ_h. This vector valued approximation requires in the hybridized setting only local regularity, for instance, $\sigma_h|_K \in V_h(K) \subset$

$H^1(K)^2$, $\forall K \in \mathcal{T}_h$. We may set $V_h = \prod_{K\in\mathcal{T}_h} V_h(K)$ composed by the local Marini-Pietra elements defined through

$$V_h(K) := \operatorname{span}\{(1,0)^\top, (0,1)^\top, \boldsymbol{q}\}.$$

Where $\boldsymbol{q} = (\phi_1, \phi_2)^\top \in P_2(K)$ is characterized by

$$\begin{cases} \int_K \phi_i \,\mathrm{d}x = 0 \quad \text{for } i = 1,2, \\ (\phi_1,\phi_2)^\top|_{e_i} \cdot \nu_{e_i} = \delta_{1i} \quad \text{for } e_i \in \partial K, \\ (\phi_1,\phi_2)^\top \cdot t(m_1) = 0, \end{cases} \tag{2.2}$$

where m_1 is the midpoint of the edge e_1 and t the respective tangent vector. The local edge numbering will be uniquely determined through the drift term as we will specify below. This specific choice of $\boldsymbol{q}$ will confine the influence of the zeroth order term to diagonal of the final algebraic system, see (2.13)-(2.14) below. The space of Raviart–Thomas finite-elements of lowest order can be equipped with local nodal basis functions of similar structure, where we denote by x_B the barycenter of K:

$$RT(K) := \operatorname{span}\{(1,0)^\top, (0,1)^\top, x - x_B\}.$$

The final discretization will be an inconsistent approximation of the original problem. Our aim is now to state an intermediate mesh dependent but consistent problem, which illustrates the connection of the final discretizations to other approaches to solve the convection-diffusion problem and it will serve as a tool in the later proofs. We abbreviate by $\bar{f} \in W_h$ the L^2-projection of a function f onto W_h. The important point is that we associate to $\boldsymbol{b}$ a function $\psi \in L^2(\Omega)$ through the local requirement

$$\nabla\bar{\boldsymbol{b}}|_K = \nabla\psi|_K \qquad \forall K \in \mathcal{T}_h, \tag{2.3}$$

so that the extension of $\psi|_K$ to $\overline{K}$ attains its largest value zero at one vertex of K. Let $V_{h,\psi} = \prod_{K\in\mathcal{T}_h} e^{\psi} V_h(K)$ be the space of exponentially fitted test functions. We also use the notation $\tilde{\tau} = e^{\psi}\tau$. We state the first discrete problem.

Find $(\sigma_h, u_h, \lambda_h) \in V_h \times W_h \times \Lambda_{h,g}$ so that

$$\sum_{K\in\mathcal{T}_h} \int_K c\sigma_h\tilde{\tau}_h - u_h \operatorname{div}\tilde{\tau}_h - u_h\boldsymbol{b}\tilde{\tau}_h \,\mathrm{d}x + \int_{\partial K\cap\Omega} \lambda_h\tilde{\tau}_h\cdot\nu \,\mathrm{d}s = -\int_{\Gamma_D} g\tilde{\tau}_h\cdot\nu \,\mathrm{d}s, \tag{2.4a}$$

$$\sum_{K\in\mathcal{T}_h} \int_K w_h \operatorname{div}\sigma_h + dw_h u_h \,\mathrm{d}x = \int_\Omega f w_h \,\mathrm{d}x, \tag{2.4b}$$

$$\sum_{K\in\mathcal{T}_h} \int_{\partial K} \mu_h\sigma_h\cdot\nu \,\mathrm{d}s = 0, \tag{2.4c}$$

holds for every $(\tilde{\tau}_h, w_h, \mu_h) \in V_{h,\psi} \times W_h \times \Lambda_{h,0}$ with $c = a^{-1}$.
This scheme can be understood as a kind of Petrov-Galerkin finite-element method. The space $V_{h,\psi}$ is in some sense the nonconforming analogue of the conforming exponential test space for rectangular meshes that was introduced in [**96**]. There Van Nooyen derived a monotone mixed Petrov-Galerkin scheme for a fixed rectangular grid. A rectangular grid allows to separate the components of the vector valued test functions. Therefore a conforming construction of the test space is possible as in one dimension. We circumvent this loss of conformity of the test function by applying a static condensation procedure and obtain finally also a monotone scheme of first order.
The final scheme is now derived by substituting the terms in (2.4) with suitable approximations. The generated inconsistencies will be analyzed in the following section. At first we note that the choice of $V_{h,\psi}$ does not completely symmetrize the method and we get the principal inconsistency of the scheme, since we drop in the discretization the last term of

$$\int_K -u \operatorname{div} \tilde{\tau}_h - u\boldsymbol{b}\tilde{\tau}_h \,\mathrm{d}x = \int_K -ue^{\psi} \operatorname{div} \tau_h - u(\boldsymbol{b} - \bar{\boldsymbol{b}})e^{\psi}\tau_h \,\mathrm{d}x. \tag{2.5}$$

Due to the local nature of the requirement $\nabla_K \psi = -\bar{\boldsymbol{b}}_{|_K}$ we can further ensure that $\psi \leq 0$ and thereby bound $\|e^{\psi}\|_{L^\infty(\Omega)} \leq 1$. This prevents error amplification for instance in the last term in (2.5). The exponential functions in the integrals will be substituted by carefully chosen integral mean values, given by

$$R_K = \frac{1}{|K|}\int_K e^{\psi} \,\mathrm{d}x,\ S_{e_i(K)} = \frac{1}{|e_i(K)|}\int_{e_i(K)} e^{\psi} \,\mathrm{d}s,\ S_K = \max_{e\in\partial K} S_{e(K)}, \tag{2.6}$$

where $|K|$ and $|e|$ denote the area of a triangle K and the length of an edge e, respectively, and we define the function $R, S \in W_h$ through $R|_K = R_K$ and $S|_K = S_K$. We will also approximate the coefficients c, d in the equations (2.4a) and (2.4b) in W_h. In order to define a matrix description of the actual numerical scheme that we will analyze, we introduce the operators $A : V_h \to V_h$, $B : V_h \to W_h$, $C : V_h \to \Lambda_h$, $D : W_h \to W_h$

$$(Ap, q)_\Omega = \int_\Omega R\bar{c}p \cdot q \,\mathrm{d}x, \qquad (Bp, v)_\Omega = \sum_{K\in\mathcal{T}_h}\int_K v \operatorname{div} p \,\mathrm{d}x \tag{2.7}$$

$$(Dw, v)_\Omega = \int_\Omega \bar{d}wv \,\mathrm{d}x, \qquad (Cp, \mu)_{\mathcal{E}_h} = -\sum_{K\in\mathcal{T}_h}\int_{\partial K} \mu p\cdot\nu \,\mathrm{d}s, \tag{2.8}$$

and the adjoints adjusted to the nonsymmetric operator, defined locally by

$$(\widetilde{B}^\top u, q)_K = \int_K S_K u \operatorname{div} q \,\mathrm{d}x, \quad (\widetilde{C}^\top \lambda, q)_K = -\int_{\partial K} S_{e_i(K)} \lambda q \cdot \nu \,\mathrm{d}s. \quad (2.9)$$

Here and in the following $(\cdot,\cdot)_{\mathcal{O}}$ denotes the $L^2(\mathcal{O})$ inner product. The algebraic system corresponding to the hybridized mixed method is then defined as follows: Find $(\sigma_h, u_h, \lambda_h) \in V_h \times W_h \times \Lambda_{h,g}$ so that

$$\begin{pmatrix} A & -\widetilde{B}^\top & -\widetilde{C}^\top \\ B & D & 0 \\ C & 0 & 0 \end{pmatrix} \begin{pmatrix} \sigma_h \\ u_h \\ \lambda_h \end{pmatrix} = \begin{pmatrix} 0 \\ F \\ 0 \end{pmatrix}, \quad (2.10)$$

where F is the vector of nodal values of $\bar{f}$, and with respective bases for V_h,W_h, and $\Lambda_{h,g}$ the matrix parts are given by the corresponding operators defined by the equations (2.7)-(2.9).
In applications the use of this algebraic system is limited due to its size, but it is the basis of our analysis in the following section. The number of unknowns can be reduced by eliminating the unknowns σ_h and u_h from (2.10). The elementary matrices associated with the element K will be denoted with the superscript K. Then it holds $\widetilde{B}^K = S_K B^K \in \mathbb{R}^3$. For $\widetilde{C}^K$ we get

$$\widetilde{C}^K = C^K \operatorname{diag}(S_{e_1}, S_{e_2}, S_{e_3}).$$

In view of the definition of V_h (see (2.2)) the matrix corresponding to A has a diagonal structure and can easily be inverted, i.e. σ_h can be removed from the system by static condensation. Similar arguments for $-B^T A^{-1}\widetilde{B} - D$ allow to eliminate u_h.
We obtain the final linear system $M\boldsymbol{\lambda}_h = G$ acting on the vector of nodal values $\boldsymbol{\lambda}_h$ of the Lagrange multipliers only, with

$$M = C^\top A^{-1}\widetilde{C} - C^\top A^{-1}\widetilde{B}(B^\top A^{-1}\widetilde{B} + D)^{-1}B^\top A^{-1}\widetilde{C},$$
$$G = -C^\top A^{-1}\widetilde{B}(B^\top A^{-1}\widetilde{B} + D)^{-1}F.$$

For the current density σ_h and for u_h we get

$$\sigma_h = A^{-1}[\widetilde{B}(B^\top A^{-1}\widetilde{B} + D)^{-1}(-F - B^\top A^{-1}\widetilde{C}\boldsymbol{\lambda}_h) + \widetilde{C}\boldsymbol{\lambda}_h], \quad (2.11)$$
$$u_h = (B^\top A^{-1}\widetilde{B} + D)^{-1}(\bar{f} - BA^{-1}\tilde{C}^\top \boldsymbol{\lambda}_h). \quad (2.12)$$

To emphasize the structure of the equation $M\boldsymbol{\lambda}_h = G$ we detail the contribution of the element matrix M^K. Here $\boldsymbol{\lambda}_h$ denotes the vector of nodal

values according to the basis $\{\chi_{e_i} : \forall e_i \in \mathcal{E}_h\}$, where χ_e is the characteristic function of the edge e. With the local edge numbering so that $S_K = S_{e_1}$ and $n_i = |e_i|\nu_i$, ν_i being the outward normal of the respective edge e_i, the local element contribution to the global stiffness matrix M has the form:

$$M_{i,j}^K = \begin{cases} \dfrac{S_K}{\bar{c}R_K}\dfrac{|e_1|^2}{|K|} + \bar{d}|K|\gamma(\bar{d}) \quad \text{for } i = j = 1, \\ \dfrac{S_{e_j}}{\bar{c}R_K}\dfrac{n_i \cdot n_j}{|K|}, \end{cases} \tag{2.13}$$

$$\gamma(\bar{d}) = |e_1|^2\Big(|e_1|^2 + \bar{d}|K|\|\tau_3\|^2_{L^2(K)}\frac{\bar{c}R_K}{S_K}\Big)^{-1}. \tag{2.14}$$

Remark 2.1. (i) In the implementational practice possible cancellation effects in the computation of the quotients S_{e_j}/R_K have to be dealt with. The resulting numerical technique is now transfered into the requirement $e^{\psi} \leq 1$, which prevents the error amplification in our analysis. Since only local quotients occur, the method is still equivalent to the one proposed in [**25, 26, 90**] for the special case $\boldsymbol{b} = \nabla\psi$, where ψ is a piecewise linear function. Additionally the scheme used in [**80**] is also covered by the above method.

(ii) A scaling argument shows that even in the hyperbolic limit the approximation of the zeroth order term $\bar{d}\gamma(\bar{d})$ is of the correct order of magnitude compared to the drift term (cf. [**90**]).

Remark 2.2. Since

$$\tilde{M}_{ij}^K = \frac{n_i \cdot n_j}{|K|}, \quad i, j = 1, 2, 3,$$

is the elementary stiffness matrix corresponding to a P_1 non-conforming finite-element discretization of the Laplace operator and since for constant potential it holds $S_K/R_K = 1$, a treatment of the lower order terms with a non-conforming P_1 elements is motivated. In Chapter 3 we perform for the first device example a comparison of the different treatments.

Lemma 2.1. *If the triangulation $\mathcal{T}_h$ is weakly acute then the global stiffness matrix M is a M-Matrix for every given data satisfying Assumption 1.*

Proof. Each column of the matrix M corresponds, due to the chosen basis of Λ_h, to one edge of the triangulation. With the global edge numbering as shown in Figure 2.1 the contributions to the column of the edge e_3 stem from the two adjacent triangles K_1, K_2 with edges e_1, e_2, e_3 and e_3, e_4, e_5

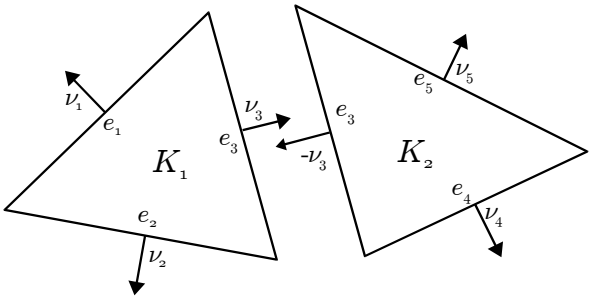

Figure 2.1: Degrees of freedom that contribute to the column corresponding to the edge e_3

respectively. Regarding now (2.13) this leads to at most five nonzero entries. The coefficients R_{K_l}, $S_{e_m(K_l)}$ and $\bar{c}$ are positive and $\bar{d}\gamma(\bar{d})$ is nonnegative. This leads to positive diagonal entries in M. The off diagonal entries' sign is determined by the sign of the term $|e_i||e_j|\nu_i \cdot \nu_j$ which is negative due to the fact that the triangles are non obtuse.
It remains to show that the matrix is column diagonally dominant. Each matrix M^{K_l} is column diagonally dominant, see (2.13), and this property is invariant under summation. Elimination of boundary edges completes the proof. □

We complete this section with the component description of the reconstruction of the current density σ_h and the piecewise constant approximation of the primal variable u_h. Due to the block structure in (2.11) and (2.12) σ_h and u_h are reconstructed element per element. Let $\lambda_i, i = 1, 2, 3$ be the values $\boldsymbol{\lambda}_h$ corresponding to the edges e_i of a triangle K, then we have:

$$\sigma_h^K = -\sum_{j=1}^{3} S_{e_j(K)} R_K^{-1} \lambda_j n_j - \frac{\gamma(\bar{d})|K|}{|e_1|}(\bar{d}\lambda_1 - \bar{f})\boldsymbol{q}^K \quad \text{and} \tag{2.15}$$

$$u_h^K = \gamma(\bar{d})\Big(\lambda_1 + \frac{\bar{f}|K|\bar{c}R_k\|\boldsymbol{q}\|^2}{S_{e_1}|e_1|^2}\Big). \tag{2.16}$$

The reconstruction of the current density is clearly very important, whereas the benefit of the reconstruction of u_h is questionable. The space W_h contains less degrees of freedom compared to the space of the Lagrange multiplier Λ_h, which indicates a loss of approximation quality. Actually there are other reconstructions that show numerically a higher convergence order and that will be considered in Section 2.3.

2.2 A priori Analysis

The basis of the new error analysis is to reinterpret the result of the static condensation procedure $M\boldsymbol{\lambda} = G$ as a variational problem $a(\lambda, \mu) = b(\mu)$. The bilinear form $a(\cdot, \cdot)$ acting on the Lagrange multiplier consists of certain lifting operators that represent the original problem. This general technique, developed by Cockburn and Gopalakrishnan[**43, 44**], has been applied to several approximation spaces like Raviart–Thomas and Brezzi–Douglas–Marini[**43**] elements and later to a variable degree Raviart–Thomas space[**44**] for second order selfadjoint elliptic problems. The error estimate on the Lagrange multiplier is deduced in the norm induced by the bilinear form $a(\cdot, \cdot)$. Already in the symmetric case the discretization is not consistent in the sense that for the exact solution u

$$a(\mathcal{P}u, \mu) - b(\mu) \neq 0,$$

for any $\mu \in \Lambda_{h,0}$. This lack of consistency has to be estimated carefully to derive the error estimates. In the nonsymmetric case of the convection-diffusion problem considered here, we have to recover in a first step the reinterpretation result, see Theorem 2.1, by defining an extended set of lifting operators. The seconds step comprises of identifying the additional inconsistencies in Theorem 2.2 and analyzing their convergence behaviour, yielding the main result in Theorem 2.3. This is followed by a subsection containing proofs of the more technical details.
Inspired by the ideas of [**44**] we now start with the new interpretation of (2.13) by introducing the lifting operators. The most important change is how the local liftings need to be adopted to represent the nonsymmetric nature of the continuous problem. We define therefore two classes of lifting operators, starting with the definition of $\widetilde{\boldsymbol{Q}} : \Lambda(K) \to V_h(K)$, $\widetilde{\boldsymbol{\mathcal{Q}}} : V_h(K) \to V_h(K)$, $\widetilde{\boldsymbol{\mathcal{Q}}} : W_h(K) \to V_h(K)$ and $\widetilde{\boldsymbol{U}} : \Lambda(K) \to W_h(K)$, $\widetilde{\boldsymbol{\mathcal{U}}} : V_h(K) \to W_h(K)$, $\widetilde{\boldsymbol{\mathcal{U}}} : W_h(K) \to W_h(K)$ through

$$\begin{pmatrix} A & -\widetilde{B}^\top \\ B & D \end{pmatrix} \begin{pmatrix} \widetilde{\boldsymbol{Q}}\lambda & \widetilde{\boldsymbol{\mathcal{Q}}}\alpha & \widetilde{\boldsymbol{\mathcal{Q}}}f \\ \widetilde{\boldsymbol{U}}\lambda & \widetilde{\boldsymbol{\mathcal{U}}}\alpha & \widetilde{\boldsymbol{\mathcal{U}}}f \end{pmatrix} = \begin{pmatrix} \widetilde{C}^\top\lambda & \alpha & 0 \\ 0 & 0 & f \end{pmatrix} \tag{2.17}$$

Additionally we define for a "transposed" problem similar operator pairs $\boldsymbol{Q}, \boldsymbol{U}$ and $\boldsymbol{\mathcal{Q}}, \boldsymbol{\mathcal{U}}$ through

$$\begin{pmatrix} A & -B^\top \\ B & S_K^{-1}D \end{pmatrix} \begin{pmatrix} \boldsymbol{Q}\lambda & \boldsymbol{\mathcal{Q}}f \\ \boldsymbol{U}\lambda & \boldsymbol{\mathcal{U}}f \end{pmatrix} = \begin{pmatrix} C^\top\lambda & 0 \\ 0 & f \end{pmatrix}. \tag{2.18}$$

We see that $(\boldsymbol{Q}\lambda, \boldsymbol{U}\lambda)$ can be regarded as a discrete solution to

$$-\operatorname{div}(ae^{-\psi}\nabla u) + e^{-\psi}du = 0 \text{ in } K, \quad u = \lambda \text{ on } \partial K,$$

in the sense that $(\boldsymbol{Q}\lambda, \boldsymbol{U}\lambda)$ is an approximation of $(-ae^{-\psi}\nabla u, u)$. By this means we obtain, similar to the symmetrization in Petrov-Galerkin methods [94], the reinterpretation result.

Theorem 2.1. *Assume that $(\sigma, u, \lambda) \in V_h \times W_h \times \Lambda_{h,g}$ is a solution of (2.10) for a right hand side $(\alpha, \delta, 0)^\top$ with $\alpha \in V_h$ and $\delta \in W_h$. Then λ is the unique solution to*

$$a(\lambda, \mu) = b(\mu) \quad \text{for all } \mu \in \Lambda_{h,0}, \tag{2.19}$$

where

$$a(\lambda, \mu) = \sum_{K \in \mathcal{T}_h} \int_K \bar{c} R_k \, \widetilde{\boldsymbol{Q}}\lambda \, \boldsymbol{Q}\mu \, dx + \int_K \bar{d} \, \widetilde{\boldsymbol{U}}\lambda \boldsymbol{U}\mu \, \mathrm{d}x, \tag{2.20}$$

$$b(\mu) = (\delta, \boldsymbol{U}\mu) - (\alpha, \, \boldsymbol{Q}\mu). \tag{2.21}$$

Proof. The proof can be done elementwise. It follows mainly the ideas of [44]. The differences lie in the nonsymmetry of the operators in (2.17) and (2.18) as we will now demonstrate. Comparing the definition of the operators with (2.10) it directly follows that

$$\sigma = \widetilde{\boldsymbol{Q}}\lambda + \widetilde{\boldsymbol{Q}}\alpha + \widetilde{\boldsymbol{\mathcal{Q}}}\delta, \tag{2.22}$$

$$u = \widetilde{\boldsymbol{U}}\lambda + \widetilde{\boldsymbol{U}}\alpha + \widetilde{\boldsymbol{\mathcal{U}}}\delta. \tag{2.23}$$

Inserting the first expression into the last equation of (2.10) and multiplying with a test function $\mu \in \Lambda(\partial K)$ we obtain

$$0 = (\widetilde{\boldsymbol{Q}}\lambda, C^\top \mu) + (\widetilde{\boldsymbol{Q}}\alpha, C^\top \mu) + (\widetilde{\boldsymbol{\mathcal{Q}}}\delta, C^\top \mu)$$

The first term will give the definition of $a_K(\lambda, \mu)$ whereas the two last terms define $-b(\mu)$. By the definition of the lifting operator $\boldsymbol{Q}$ in (2.18) it follows from the second equation of (2.17) that

$$\begin{aligned}(\widetilde{\boldsymbol{Q}}\lambda, C^\top \mu) &= (\widetilde{\boldsymbol{Q}}\lambda, A\,\boldsymbol{Q}\mu) + (\widetilde{\boldsymbol{Q}}\lambda, -B^\top \boldsymbol{U}\mu) \\ &= (\widetilde{\boldsymbol{Q}}\lambda, A\,\boldsymbol{Q}\mu) + (D\,\widetilde{\boldsymbol{U}}\lambda, \boldsymbol{U}\mu) =: a_K(\lambda, \mu)\end{aligned}$$

Next we will show $(\widetilde{\boldsymbol{Q}}\alpha, C^\top \mu) = -(\alpha, \, \boldsymbol{Q}\mu)$. Since $S_K \boldsymbol{U}\lambda \in W_h(K)$ we get from the second equation in (2.18)

$$\begin{aligned}(\widetilde{\boldsymbol{Q}}\alpha, B^\top \boldsymbol{U}\mu) &= (B\,\widetilde{\boldsymbol{Q}}\alpha, \boldsymbol{U}\mu) = -(D\,\widetilde{\boldsymbol{U}}\alpha, \boldsymbol{U}\mu) \\ &= -(S_K\,\widetilde{\boldsymbol{U}}\alpha, S_K^{-1} D\boldsymbol{U}\mu) = (S_K\,\widetilde{\boldsymbol{U}}\alpha, B\,\boldsymbol{Q}\mu) \\ &= (\widetilde{B}^\top \widetilde{\boldsymbol{U}}\alpha, \, \boldsymbol{Q}\mu).\end{aligned}$$

Thus finally

$$(\widetilde{\boldsymbol{Q}}\alpha, C^\top\mu) = (A\,\widetilde{\boldsymbol{Q}}\alpha - \widetilde{B}^\top\,\widetilde{\boldsymbol{U}}\alpha,\, \boldsymbol{Q}\mu) = -(\alpha,\, \boldsymbol{Q}\mu).$$

Similar manipulations lead to $(\widetilde{\boldsymbol{\mathcal{Q}}}\delta, C^\top\mu) = -(\delta, \boldsymbol{U}\mu)$ which completes the proof. □

If we identify again an element $\mu \in \Lambda_h$ with its corresponding vector of nodal values $\boldsymbol{\mu}$, we can express the bilinear form a in (2.20) in terms of the matrix M with

$$a(\lambda, \mu) = \boldsymbol{\mu}^\top M \boldsymbol{\lambda}$$

From Theorem 2.1 we deduce an a priori error estimate for the Lagrange multiplier in the "energy norm"

$$\|\lambda\|_a = \sup_{\|\boldsymbol{Q}\mu\|+\|\boldsymbol{U}\mu\|\neq 0} \frac{|a(\lambda,\mu)|}{\|\boldsymbol{Q}\mu\| + \|\boldsymbol{U}\mu\|}.$$

The definiteness even in the case $\boldsymbol{b} \neq 0$ follows then directly from Lemma 2.1 since for all $\mu \neq 0$ we have

$$a(\mu,\mu) = \tfrac{1}{2}\boldsymbol{\mu}^\top(M + M^\top)\boldsymbol{\mu} > 0$$

Remark 2.3. *If $|\boldsymbol{b}| \to 0$ this norm is equivalent to the norm $|\!|\!| \cdot |\!|\!|_a$ which was studied for the Raviart-Thomas element in* [44].

Before stating the main result we present some properties of the operators defined in (2.17) and (2.18) that elucidate in which way the structure of the continuous problem is recovered. Therefore, an element $\boldsymbol{v} \in V_h(K)$ is decomposed into $\boldsymbol{v} = \boldsymbol{v}_0 \oplus \boldsymbol{v}_1$ so that $(\boldsymbol{v}_0, \boldsymbol{q})_K = 0$ (see (2.2)) and denote likewise for example by $\widetilde{\boldsymbol{Q}}_0$ and $\widetilde{\boldsymbol{Q}}_1$ the respective parts of the operator $\widetilde{\boldsymbol{Q}}$. The proofs of the following assertions will be given in the Section 2.2.1.

Lemma 2.2. *For $|\boldsymbol{b}| \neq 0$ and $d = 0$ we have the equivalences*

(i) $\boldsymbol{Q}_0\mu|_K = 0 \Longleftrightarrow \mu$ *is constant on* ∂K.

(ii) $\widetilde{\boldsymbol{Q}}_0\lambda^*|_K = 0 \Longleftrightarrow \lambda^*|_{e_i(K)} = lS_{e_i}^{-1}$ *for a constant* $l \in \mathbb{R}$.

(iii) If λ is constant on ∂K, then $\widetilde{\boldsymbol{Q}}_0\lambda = -\bar{c}^{-1}\bar{\boldsymbol{b}}\lambda$.

Another important observation is that the part that $\boldsymbol{Q}_1$ contributes to the L^2-norm decreases with the mesh size.

Lemma 2.3. *There exists a mesh-independent constant $C > 0$ such that for $d_M = \max_K d$, $\mu \in \Lambda_h$, and $f \in W_h(K)$,*

$$\| \boldsymbol{Q}_1 \mu\|_{0,K} \leq C h_K \sqrt{d_M} \| \boldsymbol{U} \mu\|_{0,K}, \tag{2.24}$$

$$\| \boldsymbol{\mathcal{Q}} f\|_{0,K} \leq C h_K \| f \|_{0,K}. \tag{2.25}$$

We define the projection $\mathcal{P}$ onto $\Lambda_{h,g}$ in the usual L^2-sense by requiring locally on each edge e

$$\int_e (\mathcal{P}u - u)\mu \, ds = 0 \quad \forall \mu \in P_0(e).$$

Similarly we denote by P the L^2-orthogonal projection onto W_h, and recall that the projection Π onto V_h is given by properties **(P4)** and **(P5)** from the abstract framework of Section 1.2.

Theorem 2.2. *Let (σ, u) denote the exact solution to (2.1), then the inconsistency of the method can be described by*

$$a(\mathcal{P}u - \lambda, \mu) = (\epsilon_{e^\psi} + \epsilon_{\bar{\boldsymbol{b}}} + \epsilon_{\bar{c}} + \epsilon_\Pi, \boldsymbol{Q}\mu) + (\epsilon_P, \boldsymbol{U}\mu), \tag{2.26}$$

where

$$\epsilon_{e^\psi} = \bar{c}(R - e^\psi)\sigma + C^\top \mathcal{P}(e^{\psi_K}(\mathcal{P}u - u)) + B^\top u(S - e^\psi), \tag{2.27}$$

$$\epsilon_{\bar{\boldsymbol{b}}} = u(\boldsymbol{b} - \bar{\boldsymbol{b}})e^\psi, \quad \epsilon_{\bar{c}} = (\bar{c} - c)e^\psi \sigma, \quad \epsilon_P = P(d(Pu - u)), \tag{2.28}$$

$$\epsilon_\Pi = \bar{c} R(\Pi\sigma - \sigma). \tag{2.29}$$

Proof. The exact solution satisfies (2.4). In order to pass into a discrete formulation we substitute (σ, u, u) by $(\Pi\sigma, Pu, \mathcal{P}u)$ and transform the equation to obey the form of equation (2.10). With $(\tilde{\tau}, w, \mu) \in \widetilde{V}_h \times W_h \times \Lambda_{h,0}$, $\tilde{\tau} = e^\psi \tau$, $\tau \in V_h$ we obtain from the first term in (2.4a)

$$\begin{aligned}\int_K c\sigma\tilde{\tau} \, \mathrm{d}x = &\int_K \bar{c} R \Pi\sigma\tau \, \mathrm{d}x + \int_K \bar{c} R(\sigma - \Pi\sigma)\tau \, \mathrm{d}x \\ &+ \int_K \bar{c}(e^\psi - R)\sigma\tau + (c - \bar{c})\sigma e^\psi \tau \, \mathrm{d}x.\end{aligned}$$

Since the function S lies in W_h, we have $\int_K u S_K \operatorname{div} \tau \, dx = \int_K P u S_K \operatorname{div} \tau \, dx$

and we obtain for the second and third terms in (2.4a), recalling (2.5),

$$\begin{aligned}\int_K -u\operatorname{div}\tilde{\tau} - u\boldsymbol{b}\tilde{\tau}\,\mathrm{d}x &= \int_K -PuS_K\operatorname{div}\tau\,dx - \int_K u(\boldsymbol{b}-\bar{\boldsymbol{b}})e^{\psi}\tau\,\mathrm{d}x \\ &\quad + \int_\Omega u(e^{\psi}-S_K)\operatorname{div}\tau\,dx, \\ \int_{\partial K} u\tilde{\tau}\cdot\nu\,ds &= \int_{\partial K}\mathcal{P}ue^{\psi_K}\tau\cdot\nu\,ds + \int_{\partial K} e^{\psi_K}(u-\mathcal{P}u)\tau\cdot\nu\,ds\end{aligned}$$

This can be rephrased in the discrete system of equations that are satisfied by $(\Pi\sigma, Pu, \mathcal{P}u)$

$$\begin{pmatrix} A & -\widetilde{B}^\top & -\widetilde{C}^\top \\ B & D & 0 \\ C & 0 & 0 \end{pmatrix}\begin{pmatrix}\Pi\sigma \\ Pu_h \\ \mathcal{P}u\end{pmatrix} = \begin{pmatrix}\alpha \\ Pf+\epsilon_P \\ 0\end{pmatrix}, \tag{2.30}$$

with $\alpha \in V_h$ and given through

$$\begin{aligned}(\alpha_K,\tau) = &\int_{\partial K} e^{\psi_K}(u-\mathcal{P}u)\tau\cdot\nu\,\mathrm{d}s + \int_K u(e^{\psi}-S_K)\operatorname{div}\tau\,\mathrm{d}x \\ &+ (\bar{c}\sigma(e^{\psi}-R) + \epsilon_\Pi + \epsilon_{\bar{\boldsymbol{b}}} + \epsilon_{\bar{c}},\tau)_K\end{aligned}$$

By subtracting from (2.30) the equation (2.10) for the discrete solution and with Theorem 2.1 the assertion is proven. □

Theorem 2.3. *Suppose that $u \in H^1(\Omega)$ and $\sigma = -a(\nabla u - \boldsymbol{b}u) \in H^1(\Omega)^2$, and further that $c \in W^{1,\infty}(\Omega)$, $\boldsymbol{b} \in W^{1,\infty}(\Omega)^2$ then there exists a mesh-independent constant C depending only on the $W^{1,\infty}$-norms of c and $\boldsymbol{b}$ so that*

$$\|\lambda - \mathcal{P}u\|_a \le Ch(\|u\|_{H^1(\Omega)} + \|\sigma\|_{H^1(\Omega)^2}).$$

Remark 2.4. *Depending on the regularity of the solution, expressed by $u \in H^s(\Omega)$ (cf. [67]), we have for $s = 1$ or $s = 3/2$ that the drift dependent part of the constant C is of order $\|\boldsymbol{b}\|_{L^\infty(\Omega)}h^{s-1/2}$. In the case $s = 3/2$, the extra regularity is only needed for the last term in the definition of ϵ_{e^ψ}, see equations (2.33)-(2.35) below.*

Proof. The main difficulty in proving Theorem 2.3 lies in estimating ϵ_{e^ψ}. The rest mainly follows from standard results in approximation theory together with $\|e^{\psi}\|_{L^\infty(\Omega)} \le 1$.

Before we start estimating these terms we state two lemmas. The first one is concerned with the boundedness of the operators $B^\top$ and $C^\top$, and the second lemma relies on the properties of $B^\top$. The proofs will be given in Section 2.2.1.

Lemma 2.4. *The operators $B^\top{:}L^2(K)\to V_h(K)$ and $C^\top{:}L^2(\partial K)\to V_h(K)$ are defined for all $v\in L^2(K)$, $\mu\in L^2(\mathcal{E}_h)$ and $q\in V_h(K)$ by*

$$(B^\top v,q)_K=\int_K v\operatorname{div} q\,\mathrm{d}x \quad \text{and} \quad (C^\top\mu,q)_K=\int_{\partial K}\mu q\cdot n\,\mathrm{d}s. \tag{2.31}$$

There are positive mesh-independent constants c_1,c_2 and c_3 such that

$$c_1h_K^{-1}\|v\|_{0,K}\le\|B^\top v\|_{0,K}\le c_2\|\hat v\|_{0,\hat K} \quad \text{and} \quad \|C^\top\mu\|_{0,K}\le c_3h_K^{-1/2}\|\mu\|_{0,\partial K}. \tag{2.32}$$

Remark 2.5. *We denote by $\hat v$ in equation (2.32) and in the following the usual transform to the reference triangle K of the function v.*

Let us begin the estimation of e_ψ with the last term in (2.27). The range of $B^\top$ lies is the subspace spanned by $\boldsymbol{q}$. Recall that $\boldsymbol{Q}_1\mu$ denotes the part of $\boldsymbol{Q}\mu$ which lies in the subspace spanned by $\boldsymbol{q}$. With Lemma 2.4 and the supposed maximum principle yielding $u\in L^\infty(\Omega)$ we have

$$\begin{aligned}(B^\top u(e^\psi-S),\,\boldsymbol{Q}\mu)_K&\le(B^\top u(e^\psi-S),\,\boldsymbol{Q}_1\mu)_K\le\|u(e^{\hat\psi}-S)\|_{0,\hat K}\|\,\boldsymbol{Q}_1\mu\|_{0,K}\\&\le C\|u\|_{L^\infty(K)}\|e^{\hat\psi}-S\|_{0,\hat K}h_K\sqrt{d_M}\|\boldsymbol{U}\mu\|_{0,K}.\end{aligned}$$

We see here that this part of the error disappears if $d\equiv 0$. For a nonvanishing zeroth-order part we need to estimate $\|e^{\hat\psi}-S\|_{0,\hat K}$. Although S_K is only the mean value of e^{ψ_K} along one edge, we show now that this term is of order $O(\boldsymbol{b}h)$. This holds true since S_K is the largest edge mean value. The transformation to the reference triangle may be such that $\hat\psi$ is zero in the origin and such that e_1 is mapped to the horizontal edge of $\hat K$. In other words let

$$\hat\psi(x,y)=e^{-bh_K(x+ay)}\text{ with }b\ge0,a\ge1\text{ and }S_K=\frac{1-e^{-bh_K}}{bh_K}.$$

Then we have

$$\|e^{\hat\psi}-S_K\|_{0,\hat K}^2\le\frac{1}{24}(2a^2-2a+1)b^2h_K^2+O(h_K^3)\le C\|\boldsymbol{b}\|_{\infty,K}^2h_K^2.$$

For the second term in equation (2.27) we want to use Lemma 2.4. Since $\boldsymbol{v}\cdot\nu|_{e(K)}$ is constant for all $\boldsymbol{v}\in V_h(K)$ and $e\in\partial K$, we have

$$\begin{aligned}(C^\top e^\psi(u-\mathcal{P}u),\boldsymbol{v})_K &= \int_{\partial K} e^\psi(u-\mathcal{P}u)\boldsymbol{v}\cdot\nu\,\mathrm{d}s = \sum_{e\in\partial K}\int_e (e^\psi - S_e)(u-\mathcal{P}u)\boldsymbol{v}\cdot\nu\,\mathrm{d}s\\ &= (C^\top\mathcal{P}((e^\psi - S_{\partial K})(u-\mathcal{P}u)),\boldsymbol{v})_K,\end{aligned}$$

where we denote by $S_{\partial K}$ an element of Λ_h that takes on the edges $e_i(K)$ the value of $S_{e_i(K)}$. Due to $e^{\psi_K}|_e$ being smooth and $\|e^\psi\|_{L^\infty(\Omega)}\le 1$ we obtain

$$\|e^\psi - S_{e(K)}\|_{0,e(K)} \le |e(K)||e^\psi|_{1,e} \le |\bar{\boldsymbol{b}}|h\|e^\psi\|_{0,e} \le |\bar{\boldsymbol{b}}|h^{3/2}.$$

By requiring only $u\in H^1(\Omega)$ and remembering $\mathcal{P}u|_e$ being the integral mean over the edge e it follows with a mesh-independent constant c (see eg. [**52**, Lemma 3.32])

$$\|u-\mathcal{P}u\|_{0,e(K)} \le ch^{1/2}|u|_{1,K}, \quad \text{that leads to} \tag{2.33}$$

$$\|C^\top e^\psi(u-\mathcal{P}u)\|_{0,K} \le ch^{3/2}\|\boldsymbol{b}\|_{\infty,K}|u|_{1,K}. \tag{2.34}$$

If the exact solution has traces in $H^1(\mathcal{E}_h)$, (2.34) may be improved by substituting (2.33) with

$$\|u-\mathcal{P}u\|_{0,e(K)} \le ch|u|_{1,\partial K}. \tag{2.35}$$

The first term in ϵ_{e^ψ} in (2.27) is bounded and thus Theorem 2.3 follows. □

2.2.1 Proofs of technical lemmas

Proofs of Lemma 2.2. The first assertion has been proven already in [**65**], since $V(K)$ differs from the lowest order Raviart-Thomas space only in the non-constant part. If we only regard $\widetilde{\boldsymbol{Q}}_0$ and $\boldsymbol{Q}_0$, i.e the piecewise constant parts, the equations (2.17) and (2.18) simplify to

$$(\bar{c}R_K\,\boldsymbol{Q}_0\lambda - C^\top\lambda,\boldsymbol{v}) = 0 \;\text{ and }\; (\bar{c}R_K\,\widetilde{\boldsymbol{Q}}_0\lambda - \widetilde{C}^\top\lambda,\boldsymbol{v}) = 0 \quad \forall\boldsymbol{v}\in P_0(K)^2.$$

With the canonical basis for $V_h(K)_0$ and $\Lambda_h(K)$, we obtain the matrix $\mathcal{C}$ of maximal rank so that $(C^\top\lambda,\boldsymbol{v}) = \boldsymbol{v}^\top\mathcal{C}\boldsymbol{\lambda}$. Defining $\mathcal{S} = \mathrm{diag}(S_{e_1},S_{e_2},S_{e_3})$ leads on the other hand to $(\widetilde{C}^\top\lambda,\boldsymbol{v}) = \boldsymbol{v}^\top\mathcal{C}\mathcal{S}\boldsymbol{\lambda}$. This leads finally to $\mathrm{Ker}\,\widetilde{\boldsymbol{Q}} = \mathcal{S}^{-1}\,\mathrm{Ker}\,\mathcal{C}$. The kernel of $\mathcal{C}$ are the piecewise constant functions and $\boldsymbol{\lambda}^* = \mathcal{S}^{-1}\boldsymbol{l}$, for some constant $\boldsymbol{l}$, which proves the second assertion.

The assertion (*iii*) simply follows by

$$(\widetilde{C}^\top \lambda, \boldsymbol{v})_K = \sum_{e_i \in \partial K} \int_{e_i} \lambda S_{e_i} \boldsymbol{v} \cdot \nu \, \mathrm{d}s = \lambda \int_{\partial K} e^{\psi} \boldsymbol{v} \cdot \nu \, \mathrm{d}s$$
$$= \lambda \int_K \operatorname{div}(e^{\psi} \boldsymbol{v}) \, \mathrm{d}x = -\lambda \int_K e^{\psi} \bar{\boldsymbol{b}} \cdot \boldsymbol{v} \, \mathrm{d}x = -(R_K \bar{\boldsymbol{b}} \lambda, \boldsymbol{v}). \quad (2.36)$$

□

Proof of Lemma 2.4. Although the elementwise divergence does not map $V_h(K)$ onto $W_h(K)$, there exist for each $w \in W_h(K)$ a unique $\boldsymbol{v}_w \in V_h(K)$ so that $\|w\|_{0,K}^2 = (w, \operatorname{div} \boldsymbol{v}_w)_K$. Therefore we have with appropriate spaces $V_h(\hat{K})$ and $W_h(\hat{K})$ on the reference triangle $\hat{K}$ the inequality

$$\|\hat{w}\|_{0,\hat{K}} \leq \hat{C} \sup_{\boldsymbol{v} \in V_h(\hat{K})} \frac{(\hat{w}, \operatorname{div} \boldsymbol{v})_{\hat{K}}}{\|\boldsymbol{v}\|_{0,\hat{K}}} \quad \forall \hat{w} \in W_h(\hat{K}).$$

Employing the Piola-map for $\boldsymbol{v}$ (see [**23**, chapter III.1.3]) when returning to the element K yields

$$C h_K^{-1} \|w\|_{0,K} \leq \hat{C} \sup_{\boldsymbol{v} \in V_h(K)} \frac{(w, \operatorname{div} \boldsymbol{v})_K}{C h_K^{2/2-1} \|\boldsymbol{v}\|_{0,K}} \quad \forall w \in W_h(K),$$

which proves the first inequality. We notice that $B^\top$ maps onto the subspace spanned by $\boldsymbol{q} = (\phi_1, \phi_2)^\top$ since $\boldsymbol{q}$ is orthogonal to $(P_0(K))^2$ in the sense of L^2. A simple scaling argument therefore leads to

$$\|B^\top w\|_{0,K} = \frac{(w, \operatorname{div} \boldsymbol{q})_K}{\|\boldsymbol{q}\|_{0,K}} \leq C \|\hat{w}\|_{0,\hat{K}}, \quad \forall w \in L^2(K).$$

In the proof of the second inequality we denote by $\mathcal{B} = \{(1,0)^\top, (0,1)^\top, \boldsymbol{q}\}$ an orthogonal basis of $V_h(K)$. For each $\boldsymbol{v}_i \in \mathcal{B}$ we have $|\boldsymbol{v}_i \cdot \nu_{|K}| \leq 1$. Together with $C^\top f = \sum_{i=1}^3 \tau_i \boldsymbol{v}_i$ we get

$$|(C^\top f, \boldsymbol{v}_i)| = |\int_{\partial K} f \boldsymbol{v}_i \cdot \nu \, \mathrm{d}s| \leq \int_{\partial K} |f| \, 1 \, \mathrm{d}s \leq h^{1/2} \|f\|_{0,\partial K}$$

together with $\|\boldsymbol{v}_i\|_{0,K} \geq c h^{2/2}$

$$\|C^\top f\|_{0,K} \leq c h^{-1/2} \|f\|_{0,\partial K}.$$

□

Proof of Lemma 3.2. The proof is a slight variant of the results given in Lemma 4.1 of [44]. We report them since it is interesting to see how the constants at the end depend on the convection term $\boldsymbol{b}$. We begin with equation (2.25). The matrix in equation (2.18) is skew symmetric which leads together with $\boldsymbol{\mathcal{Q}}f = A^{-1}B^\top \boldsymbol{\mathcal{U}}f$ to $(B^\top \boldsymbol{\mathcal{U}}f, A^{-1}B^\top \boldsymbol{\mathcal{U}}f) + (S_K^{-1} D\boldsymbol{\mathcal{U}}f, \boldsymbol{\mathcal{U}}f) = (\boldsymbol{\mathcal{U}}f, f)$. Therefore we also deduce for Marini-Pietra elements with Lemma 2.4

$$\|\boldsymbol{\mathcal{U}}f\|_{0,K}^2 \leq h^2 \|B^\top \boldsymbol{\mathcal{U}}f\|_{0,K}^2 \leq \bar{c} R_K h^2 \|\boldsymbol{\mathcal{U}}f\|_{0,K} \|f\|_{0,K}. \tag{2.37}$$

We observe that the stability constant is slightly better than that of a purely diffusive problem with the same diffusion coefficient $a = c^{-1}$ since $R_K \leq 1$. Using now the first equation of (2.18) with $\boldsymbol{\mathcal{Q}}f$ as test function, we obtain in analogy to [44]

$$\begin{aligned}(\bar{c}R_K)^{-1} \|\boldsymbol{\mathcal{Q}}f\|_{0,K}^2 &\leq (B^\top \boldsymbol{\mathcal{U}}f, \boldsymbol{\mathcal{Q}}f)_K = (\boldsymbol{\mathcal{U}}f, B\boldsymbol{\mathcal{Q}}f)_K \\ &= (\boldsymbol{\mathcal{U}}f, f)_K - (S_K^{-1} D\boldsymbol{\mathcal{U}}f, \boldsymbol{\mathcal{U}}f)_K \leq \|\boldsymbol{\mathcal{U}}f\|_{0,K} \|f\|_{0,K} \\ &\leq \bar{c} R_K h^2 \|f\|_{0,K}^2 \quad \text{by (2.37)}.\end{aligned}$$

Regarding the second equation of (2.18) we obtain from $\boldsymbol{Q}_1\mu = l_\mu \boldsymbol{q}$ and $\boldsymbol{U}\mu \in P_0(K)$ that

$$\begin{aligned} l_\mu \int_K \operatorname{div} \boldsymbol{q} \, \mathrm{d}x &= S_K^{-1} \bar{d} \, \|\boldsymbol{U}\mu\|_{0,K} = S_K^{-1} \bar{d} \, \|\boldsymbol{U}\mu\|_{0,K} \\ \|\boldsymbol{Q}_1\mu\|_{0,K} &= S_K^{-1} \bar{d} \frac{\|\boldsymbol{q}\|_{0,K}}{|e_1|} \|\boldsymbol{U}\mu\|_{0,K} \quad \text{by (2.2)}. \end{aligned}$$

From here the result follows by a usual scaling argument since $\boldsymbol{q} \in P_2(K)$. For the variable degree Raviart–Thomas case the inequality (2.24) also holds for the non-solenoidal part of $\boldsymbol{Q}$ (see [44, Lemma 5.3]). □

2.3 A Posteriori Error Estimation

In this section we introduce two a posteriori estimators, that on the one hand are directly applied to the semiconductor application, and on the other hand they build the basic tools for the estimation methodology that is presented in Section 2.4. It is well known that in semiconductor modeling, the equations (2.1) are usually convection dominated due to high electric fields. An adaptive refinement strategy, therefore, is crucial to resolve accurately and efficiently very sharp gradients occurring in the device simulations. Reliable and efficient a posteriori error estimators are an indispensable tool for efficient adaptive algorithms.

Although we will combine different error estimators in the device simulation for the continuity equations and the Poisson equation, we will present here in detail only techniques to estimate the error in the hybridized mixed method to discretize the continuity equations. In Chapter 3 we will describe how to apply these techniques to the nonlinear system of the application.

We will distinguish two different types of estimators. One way to obtain an a posteriori estimate is to compare different approximate solutions. The rough idea is to take the solution in the richer approximation space as a substitute for the exact solution and to calculate the norm of the difference of these two approximations as an a posteriori estimate. These higher order approximation can be obtained in different ways, and the methods might be distinguished by the accuracy that is obtained and the effort that it takes to calculate them, [**2**]. If one of the approximations is obtained with reduced computational cost, these estimators are often referred to embedded estimators.

A different class of a posteriori error estimates is based on evaluating the dual norm of the residual. The development of residual based error estimates for mixed finite elements starts with the work of Braess and Verfürth [**21**], who required a saturation assumption to prove reliability and efficiency of their estimator for the RT-element. An overview of different approaches for the RT-element is given in [**73**] and the work of Carstensen [**28**] provides residual based error estimator for spaces with a structural assumption fulfilled by many typical examples, like RT- or BDM-elements.

In this section we will focus at the beginning to the first class of estimators. In Section 2.4 we return to residual based estimators with the extension of the so-called dual–weighted–residual (DWR) estimators. These estimators have been developed for standard finite elements by Becker, Rannacher, et.al., see [**14, 15**] and the references therein. We will extend this methodology to mixed finite-element methods. Let us point out, that at least for

symmetric problems all what follows also holds for RT-elements, especially the new estimator in Section 2.4 will also be applied to RT-elements.

2.3.1 An Embedded Estimator Controlling the L^2-Error

The error estimator is based on a comparison of the two approximations $\lambda_h \in \Lambda_{h,u_D}$, the Lagrange multiplier and $u_h \in W_h$ of the primal variable $u \in H^1(\Omega)$ of the exact solution to continuous problem (2.1). The first approximation λ_h is only defined on the edges of the triangles, whereas the second approximation u_h is piecewise constant on the triangles. Following [6], we introduce a suitable lifting of λ_h and we compare then functions in $L^2(\Omega)$. More precisely, we introduce the Crouzeix-Raviart finite-element space of lowest order [45],

$$CR_{h,\xi} = \{v \in L^2(\Omega) : v|_K \in P_1(K)\ \forall K \in \mathcal{T}_h; \tag{2.38}$$

$$v \text{ is continuous in } m_e\ \forall e \in \mathcal{E}_h \backslash \Gamma_h; \tag{2.39}$$

$$v|_K(m_e) = \textstyle\int_e \xi\, \mathrm{d}s/|e|\ \forall e \in \Gamma_h\}, \tag{2.40}$$

and we define the lifting $\hat{u}_h \in CR_{h,u_d}$ of λ_h and the interpolation $P_{CR}u$ of u in CR_{h,u_D} by

$$\int_e (\lambda_h - \hat{u}_h)\, \mathrm{d}s = \int_e (P_{CR}u - u)\, \mathrm{d}s = 0 \quad \forall e \in \mathcal{E}_h.$$

In the above definition of $CR_{h,\xi}$, $\Gamma_h \subset \mathcal{E}_h$ denotes the set of all edges on the boundary of Ω, and we recall that m_e is the midpoint of an edge $e \in \mathcal{E}_h$. Clearly, the nodal values of λ_h and $\hat{u}_h$ are the same.
We introduce the discretization errors, and abbreviate again the L^2–norm by $\|\cdot\|_0$

$$e_1 = \|\hat{u}_h - P_{CR}u\|_0 \quad \text{and} \quad e_2 = \|u_h - P_{CR}u\|_0. \tag{2.41}$$

The error estimator is based on the assumption that there exists a constant $0 \le \gamma < 1$ so that

$$e_1 \le \gamma e_2. \tag{2.42}$$

This saturation assumption gives rise to an upper and lower bound for the discretization error u_h since from

$$\|\hat{u}_h - u_h\|_0 \le \|\hat{u}_h - P_{CR}u\|_0 + \|u_h - P_{CR}u\|_0 \le (1+\gamma)\|u_h - P_{CR}u\|_0,$$
$$\|\hat{u}_h - u_h\|_0 \ge \|\hat{u}_h - P_{CR}u\|_0 - \|u_h - P_{CR}u\|_0 \ge (1-\gamma)\|u_h - P_{CR}u\|_0$$

it follows

$$(1+\gamma)^{-1}\|\hat{u}_h - u_h\|_0 \le \|u_h - P_{CR}u\|_0 \le (1-\gamma)^{-1}\|\hat{u}_h - u_h\|_0.$$

The first inequality expresses the efficiency and the second one the reliability of the error estimator defined by

$$\eta_{CR} = \|\hat{u}_h - u_h\|_0.$$

Introducing the local contributions

$$\eta^2_{CR,K}(\hat{u}_h, u_h) = \frac{|K|}{3}\sum_{i=1}^{3}(\hat{u}_h(m_{e_i}) - u_h|_K)^2, \tag{2.43}$$

we can see that the error estimator now writes

$$\eta^2_{CR} = \sum_{K\in\mathcal{T}_h} \eta^2_{CR,K}. \tag{2.44}$$

Moreover, as the functions λ_h and $\hat{u}_h$ have the same values on the nodes,

$$\eta^2_{CR,K}(\lambda_h, u_h) = \frac{|K|}{3}\sum_{i=1}^{3}(\lambda_h|_{e_i} - u_h|_K)^2.$$

2.3.2 Benchmark Problems

The saturation assumption (2.42) will be numerically verified for the above elements in two test problems. We take test problems that have been used in the literature in order to study different schemes for convection-diffusion equations [**27, 93, 104**]. The main feature of these problems is that the solutions exhibit very large gradients for strong convective data, which is typical in semiconductor device simulations. In both cases, we set $\Omega = (0,1)^2$, consider the problem (2.1) and denote the space variables by x and y. The first problem is constructed from the explicitly given solution

$$u(x,y) = xy(1 - e^{(x-1)/\varepsilon})(1 - e^{(y-1)/\varepsilon}),$$

which forms sharp layers at the boundaries $\{x = 1\}$ and $\{y = 1\}$ for small $\varepsilon > 0$ (see Figure 2.2 (left)). The right-hand side f is constructed according to the data diffusion coefficient $a = \varepsilon$, drift $\beta = (1,1)^T$, and zeroth order term $d = 2$ in Ω, see (2.1). We choose the boundary condition $u = 0$ on $\partial\Omega$. The errors e_1 and e_2, defined in (2.41), for $\varepsilon = 2^{-j}$ $(j = 2, 4, 6, 8)$ are shown in Figure 2.3. The error e_2 decreases with a rate approximately equal to 1. Generally, the rate of convergence of e_1 is larger than that of e_2. In Table 2.1, the ratios e_1/e_2 are listed. For solutions with small gradients (or equivalently, large ε), the ratios decrease with smaller mesh sizes. For

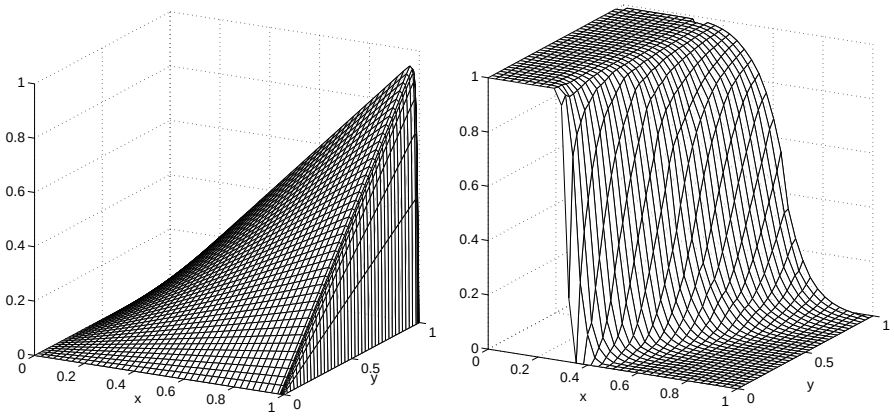

Figure 2.2: Solutions of the first test problem for $\varepsilon = 2^{-8}$ (left) and of the second test problem for $h_{\max} = 0.02$ (right).

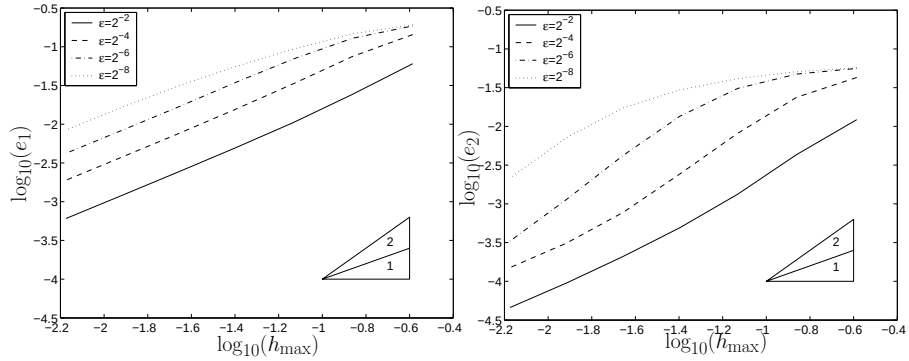

Figure 2.3: Logarithmic plot of e_1 (left) and e_2 (right) for $\varepsilon = 2^{-j}$, $j = 2, 4, 6, 8$ (from bottom to top). For ε not too small the rate of convergence for e_1 is larger than 1, which is approximately the rate for e_2.

smaller ε the ratios decrease only for finer meshes. Hence, the saturation assumption (2.42) with $\gamma < 1$ seems to hold if the gradients of the solution are not too large or if the mesh is fine enough to resolve the shape of the solution.

The second problem focuses on the crosswind dissipation along a transported discontinuity. We take the same parameters as in [**104**], $\varepsilon = 10^{-6}$, $\beta =$

	ε					
$h_{\max}$	2^{-2}	2^{-4}	2^{-6}	2^{-8}	2^{-10}	2^{-12}
1/4	**0.2279**	0.3511	0.3656	0.3559	0.3538	0.3537
1/8	0.2187	**0.3829**	0.4462	0.4120	0.4001	0.3995
1/16	0.1708	0.3213	**0.5710**	0.5579	0.5039	0.4949
1/32	0.1412	0.2191	0.5397	0.7374	0.6026	0.5264
1/64	0.1252	0.1483	0.3658	**0.8053**	0.8375	0.6261
1/128	0.1150	0.1207	0.2021	0.6432	**1.0440**	0.8658
1/256	0.1110	0.1167	0.1165	0.3919	0.9372	**1.1904**

Table 2.1: Ratio e_1/e_2 for different diffusion coefficients ε and maximum mesh sizes $h_{\max}$.

$(1,3)^T$, $f = d = 0$ and

$$u_D(x,y) = \begin{cases} 1 & \text{for } x = 0,\ y \in [0,1] \text{ or } x \in [0,1/3],\ y = 0 \\ 0 & \text{else.} \end{cases}$$

The solution for $h_{\max} = 0.02$ is shown in Figure 2.2 (right). In addition to the interior layer there is also a layer at the domain boundary. Since an explicit solution is not available, the errors are computed by using a reference solution from a very fine mesh with $h_{\max} = 1/384$. From the results in Table 2.2 we see that even for this quite small diffusion coefficient the quotient e_1/e_2 is smaller than one for all used meshes. The quotients e_i/η_{CR} $(i = 1, 2)$, defined in (2.44) and depicted in Table 2.2, are called the efficiency indices. Roughly speaking, the closer this index is to one the more accurate is the error indicator.

$h_{\max}$	e_1	e_2	e_1/e_2	e_1/η_{CR}	e_2/η_{CR}
1/3	0.1796	0.2444	0.73	1.08	1.47
1/6	0.1389	0.1753	0.79	1.29	1.63
1/12	0.1089	0.1324	0.82	1.43	1.75
1/24	0.0827	0.0984	0.84	1.56	1.85
1/48	0.0584	0.0694	0.84	1.57	1.87
1/96	0.0365	0.0449	0.81	1.40	1.72

Table 2.2: Global errors e_1 and e_2 for the second test problem.

2.3.3 Error Control Based on the Current Density

In the semiconductor device application we will see that the error control based on the L^2-error in the primal variable u leads to a rather rough refinement control. Since users of semiconductor device simulation are mainly interested in the current through parts of the boundary of the device an estimate of the error in the current density is likely to give a better refinement control. In the application of the MOSFET device we will numerically verify that the estimator which we will present now offers a significantly better mesh refinement if one looks at the terminal current.
The error estimator targets at the error $\|\sigma - \sigma_h\|_{0,\Omega}$. It is constructed by substituting σ by higher order reconstruction based on local averaging of σ_h. Such kind of estimators are often referred to as gradient recovery or Zienkiewich-Zhu type [**120**] error indicator. The estimator used here has been analyzed in various model problems for second-order elliptic boundary value problems [**31**, **12**, **32**, **33**]. We will follow in our presentation mainly the lines in [**29**] and adopt it to the mixed discretization of a convection–diffusion equation.
At first we introduce an averaging space V_h^* which carefully adopts the boundary conditions and in the simplest case it contains element–wise linear functions which are globally continuous. One error indicator can be defined by

$$\eta_M := \min\{\|\sigma_h - v_h\|_{0,\Omega} : \forall v_h \in V_h^*\}.$$

Provided that the exact solution σ offers sufficient regularity a standard interpolation estimate directly leads to a proof of the efficiency of this estimator up to a remainder term (h.o.t.) of higher order than the error itself.

$$\begin{aligned}\eta_M &= \min_{v_h \in V_h^*} \|\sigma_h - \sigma + \sigma - v_h\|_{0,\Omega} \\ &\leq \|\sigma - \sigma_h\|_{0,\Omega} + \min_{v_h \in V_h^*} \|\sigma - v_h\|_{0,\Omega} \\ &= \|\sigma - \sigma_h\|_{0,\Omega} + \text{h.o.t.}\end{aligned}$$

We note that the efficiency coefficient is one, which means that η_M is a direct lower bound of the error in the current density up to higher order terms. Unfortunately this minimum might not be easy to calculate and is therefore replaced by an upper bound

$$\eta_M \leq \eta_{ZZ} := \|\sigma_h - \mathcal{A}\sigma_h\|_{0,\Omega},$$

where $\mathcal{A}\sigma_h \in V_h^*$ is computed by a local averaging operator $\mathcal{A}$ that is adjusted to the given boundary data according to the equation for the current

density. If it is possible to prove the reliability $\|\sigma - \sigma_h\|_{0,\Omega} \leq C\eta_M$ of η_M then clearly the reliability of η_{ZZ} holds with the same multiplicative constant. The efficiency of the error indicator η_{ZZ} is proven once an upper bound

$$C_{\text{eff}}\, \eta_M \geq \eta_{ZZ}$$

would give the equivalence of the error indicators η_{ZZ} and η_M. Such kind of problems have been investigated in [29] and bounds that depend only on the shape of the triangulation and not on the mesh-size have been obtained. Numerically the efficiency index of η_{ZZ} behaves better than the value of the upper bound C_{eff} suggests.
We will now define the averaging operator $\mathcal{A}$ and the space V_h^* and finally illustrate the indicator efficiency for the numerical example of the previous section. The construction of the operator $\mathcal{A}$ consists of two parts. In the first step an averaging operator $\mathcal{M} : V_h \to S^2_{1,h}$ maps σ_h onto $\sigma_h^* \in (\mathcal{S}_{1,h})^2$ and in a second step σ_h^* is projected onto V_h^*. The spaces $\mathcal{S}_{k,h}$ consist of globally continuous functions that are element–wise polynomials of degree not more than k, for later reference we write

$$\mathcal{S}_{k,h} := \{v \in C^0(\Omega) : v|_K \in P_k(K),\, \forall K \in \mathcal{T}_h\}. \tag{2.45}$$

By choosing the standard nodal basis in $\mathcal{S}_{1,h}$ the characterization of V_h^* amounts to define at each vertex p of the triangulation a subspace $V_{h,p}^* \subset \mathbb{R}^2$ of nodal values. In the interior this space will be unrestricted but for boundary nodes this may be an affine subspace, according to the boundary data. In order to characterize these restrictions we introduce some notation. Let $\mathcal{E}_N$, $\mathcal{E}_D$ be the set of edges along the boundary part Γ_N and Γ_D, respectively. The presentation exploits the simplification of the two dimensional setting but can be easily be extended to higher dimensions. For a given edge $e \in \partial K$ we denote by ν_e the outward unit normal vector, and t_e the unit tangent vector. For any $p \in \mathcal{E}_{\partial\Omega}$ we define for given Dirichlet boundary data u_D and homogeneous Neumann data the affine subspace

$$\begin{aligned} V_{h,p}^* = \{\boldsymbol{v}^* \in \mathbb{R}^2 : \boldsymbol{v}^* \cdot \nu_e = 0\ \forall e \in \mathcal{E}_N \text{ and } \text{ with } p \in e \\ \boldsymbol{v}^* \cdot t_e = a(\partial_{t_e} u_D + \beta \cdot t_e u_D)\ \forall e \in \mathcal{E}_D, p \in e\} \end{aligned}$$

In three dimensions this might possibly lead to an overdetermined linear system, whereas in two dimensions a vertex $p \in \partial\Omega$ is always the intersection of two neighboring boundary edges $\mathcal{E}_p = e_1 \cup e_2$. However we need to require the compatibility conditions that are $u_D \in C^1(\mathcal{E}_p)$ for $\mathcal{E}_p \subset \Gamma_D$ and $0 = a(\partial_{t_e} u_D + \beta \cdot t_e u_D)|_{\Gamma_D}$ if $p \in \Gamma_N \cap \bar{\Gamma}_D$.

This leads to the definition, denoting the set of vertices of the triangulation by $\mathcal{N}_h$

$$V_h^* = \{\boldsymbol{q}_h \in \mathcal{S}_{1,h}^2 : \ \forall p \in \mathcal{N}_h \cap \partial\Omega, \boldsymbol{q}_h(p) \in V_{h,p}^*\}.$$

For the purpose of unifying the definition of the averaging operator for the boundary and the interior nodes we set $V_{h,p}^* := \mathbb{R}^2$ for all $p \in \mathcal{N}_h \cap \Omega$ and denote by $\pi_p : \mathbb{R}^2 \to \mathbb{R}^2$ the orthogonal projection onto $V_{h,p}^*$. We set

$$(\mathcal{A}\sigma)(p) = \pi_p \circ \mathcal{M}_p(\sigma)$$

where $\mathcal{M}_p(\sigma)$ is the average of σ over a certain domain, here $\Omega_p = \{K \in \mathcal{T}_h : p \in \overline{K}\}$ the patch of elements K having the node p in common, more precisely

$$\mathcal{M}_p\sigma = \frac{1}{|\Omega_p|} \int_{\Omega_p} \sigma \, \mathrm{d}x.$$

Remark 2.6. *The actual choice of the averaging operator and the specific properties of the Marini-Pietra elements lead to a very simple implementation of the averaging operator. It is precisely the local property $\boldsymbol{q} \perp P_0(K)$ that allows the simplification: Let $\sigma \in V_h$ with $\sigma|_K = (\sigma_1, \sigma_2)^\top + \sigma_3 \boldsymbol{q}$, $\sigma_i(K) \in \mathbb{R}$.*

$$\begin{aligned}\mathcal{M}_p\sigma &= \frac{1}{|\Omega_p|} \sum_{K \in \Omega_p} \int_K \sigma \, \mathrm{d}x = \frac{1}{|\Omega_p|} \sum_{K \in \Omega_p} \int_K \begin{pmatrix}\sigma_1 \\ \sigma_2\end{pmatrix} \mathrm{d}x \\ &= \frac{1}{|\Omega_p|} \sum_{K \in \Omega_p} |K| \begin{pmatrix}\sigma_1 \\ \sigma_2\end{pmatrix}.\end{aligned}$$

Proofs of the reliability of this estimator usually show the equivalence to a residual type estimator for which a reliability proof involves a global stability constant that tends to be large for convection dominant problems. We are therefore much more interested in the actual performance for model problems as for the previous estimator. The ideas of a residual type estimator will be presented in Section 2.4.

As for the estimator η_{CR} in Section 2.3.2 we compare now the indicator value η_{ZZ} with the estimated error $\|\sigma - \sigma_h\|_0$. For notational convenience we denote $e_3 = \|\sigma - \sigma_h\|_0$ and report the values of the efficiency index e_3/η_{ZZ} in the Tables 2.3 and 2.4. In a second step we compare the adaptive mesh refinement controlled by the two estimators η_{CR} and η_{ZZ} in Figure 2.4.

The efficiency indexes in Table 2.3 are calculated for meshes, that are created starting from a Delauney triangulation by refining each triangle of the previous refinement step into four congruent triangles and thereby halving

the mesh size. For relatively large diffusion coefficients ε the efficiency indexes remain almost constant and strictly below one. This indicates some kind of super convergence phenomenon in the numerical solution which often happens for congruent meshes.The adaptive refinement procedure in [**12**] incorporates a mesh distortion procedure, which moves the internal mesh points randomly out of the barycenter. This technique should prevent the super convergence and an improvement of the estimator efficiency may be achieved. However, mesh distortion would increase the number of obtuse triangles and thereby reduce the numerical stability of our method as discussed in Section 2.2. Here we can see that the dependence on the mesh congruency is reduced if convection becomes dominant. We do not incorporate a mesh distortion in the adaptive refinement in the application. In Table 2.4 the triangulation of each row is a Delauney triangulation for the stated maximal element size $h_{\max}$. We see that even for these typically completely unstructured meshes the efficiency index remains bounded although it strongly varies for the finest triangulation. This numerical example does not allow for further conclusions, but it seems that at least in the large diffusion case the efficiency index tends towards asymptotic exactness.
The Delauney triangulation may contain obtuse triangles which may lead to these pollution effects. We refer to [**29**] where precise bounds are analyzed for the first time.
Locally refined meshes that are generated by an adaptive refinement algorithm do not bear as many local element similarities as the meshes of the first example. In this sense the result of the second mesh category gives a more realistic view on the estimation accuracy in the application.

	ε						
$h_{\max}$	2^0	2^{-2}	2^{-4}	2^{-6}	2^{-8}	2^{-10}	2^{-12}
1/4	0.6367	0.7479	0.8668	0.939	1.03	1.057	1.06
1/8	0.6372	0.7527	0.8582	0.7667	0.7914	0.8265	0.8331
1/16	0.6364	0.7471	0.863	0.7422	0.5943	0.592	0.6046
1/32	0.6362	0.7456	0.8718	0.8179	0.5723	0.4497	0.4474
1/64	0.6362	0.7449	0.8745	0.8932	0.6714	0.4142	0.3331
1/128	0.6361	0.7442	0.8748	0.9331	0.8007	0.5051	0.2871
1/256	0.6339	0.7401	0.8733	0.9478	0.9007	0.6422	0.3361

Table 2.3: Efficiency index values e_3/η_{ZZ} for nested meshes with locally congruent triangles for the first benchmark problem.

Let us recall that the second benchmark problem focuses on the crosswind

	ε						
$h_{\max}$	2^0	2^{-2}	2^{-4}	2^{-6}	2^{-8}	2^{-10}	2^{-12}
1/4	0.6367	0.7479	0.8668	0.939	1.03	1.057	1.06
1/8	0.6417	0.7635	0.8551	0.732	0.7323	0.7611	0.7679
1/16	0.6373	0.7814	0.9424	0.7297	0.5651	0.5679	0.5828
1/32	0.6442	0.8563	1.211	0.9759	0.6309	0.5103	0.5082
1/64	0.6565	1.086	1.914	1.553	0.8946	0.6064	0.5449
1/128	0.7183	1.838	3.622	2.924	1.533	0.9086	0.7601
1/256	0.7438	2.159	4.30	3.52	1.951	1.009	0.7026

Table 2.4: Efficiency index values e_3/η_{ZZ} for the Delauney triangulations for a given maximal edge length $h_{\max}$.

dissipation along a transported discontinuity. Unstructured grids that are not aligned to the direction of the transport in the problem diffuse the solution too much. It is therefore very important that an adaptive refinement procedure automatically refines at the onset of the discontinuity. The comparison in Figure 2.4 is done for $\varepsilon = 10^{-6}$. For both estimators the top 5% of the triangles K with largest $\eta^K_{CR,ZZ}$ are refined in each step and we present the result after 14 steps. Both estimators detect the discontinuity and we obtain very similar contour plots. The mesh produced under the control of η_{CR} contains a large area of refined triangles, this leads with 7580 elements to a higher number of triangles compared to 4040 elements in the refined mesh produced under the estimator η_{ZZ}. Focusing the area around the onset of the discontinuity it is clearly visible that the estimator η_{ZZ} is capable to refine closer to the discontinuity and therefore resolves the solution with far less triangles. The difference in the element number is also due to the process of removal of hanging nodes.
These results clearly indicate, that a refinement based on or aiming at controlling the current density could be advantageous. On the other hand is the computational effort to compute the estimator is higher than for η_{CR}, due to the larger averaging regions in evaluating the η_{ZZ}. For symmetric problems the estimator η_{CR} may be sufficient. In Section 3.4 we will assess the estimator in the semiconductor application. Another important application will be given in the next section where the goal is to refine the mesh in order to approximate a certain linear functional acting on the solution with best possible quality. In the construction of this new estimator both η_{CR} and η_{ZZ} will be incorporated.

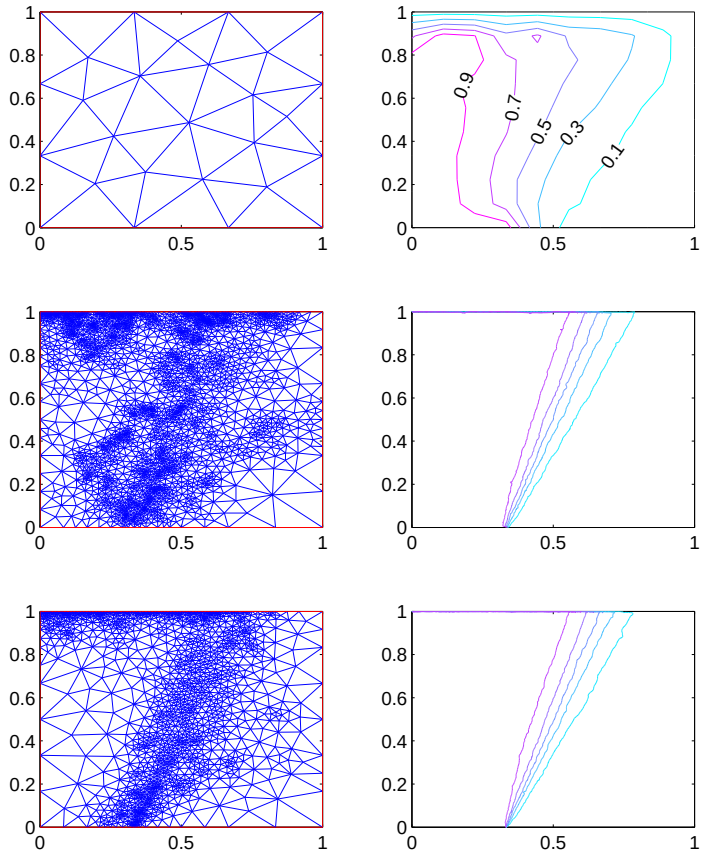

Figure 2.4: Comparing the resulting meshes after 14 adaptive mesh refinements, where only 5% of the triangles have been refined in each step. The first row shows the initial triangulation. In the second row the refinement is controlled by η_{CR} and in the last row by η_{ZZ}.

2.4 The Dual–Weighted–Residual Estimation for Efficient Mesh Refinement Control

Residual-based a posteriori error estimates have been used successfully for many finite element methods and various applications. Introductions to this field might be found in any modern textbook concerning finite element analysis (cf. [**52, 22**]). Typically these estimates bound the error in the energy norm and involve a global stability constant of a dual problem, whose deter-

mination might be viewed as a starting point for the following investigations. The analytical methods to quantify this constant (see eg. [53]) are limited. In more complex situations one could determine this constant to a certain extend by using benchmark calculations before beginning the simulation of the original problem.
But controlling the error in a global norm might not lead to a good control of local quantities that are very often the main objective of the simulation. The dual problem may provide further information to control the error in these quantities, especially if local phenomena are the main sources of the error. The key idea of the approach we present here is to compute the solution to a dual problem before refining the mesh. The dual solution is used to calculate weighting factors that emphasize parts of the local residuals in the indicator.
The main advantage is that the functional in the dual problem may be chosen freely, which means that there is only little effort necessary to switch to the estimation between several quantities of interest or even a global norm. The only demand is that this quantity must be defined as, or at least approximated by, a linear functional acting on the space in which the solution to the primal problem exists.
The development of this methodology started in standard conforming finite-element schemes and lead to the survey article by Becker and Rannacher [15]. The method has been extended also to stabilized methods for transport problems, but to our knowledge this is the first time that an extension of the output oriented estimation approach to mixed finite-elements is performed. For a residual-based estimator for mixed finite-elements that evaluate the residual on carefully chosen test functions, cf. [28]. This procedure is not extendable to the DWR–framework, where the residual is evaluated on a dual problem's solution. We will see how to circumvent this deficiency.
The rest of this section is organized as follows. We briefly summarize the DWR–approach for standard finite-element discretizations before we present the new estimator for the mixed finite-element method, and finish with a benchmark problem that shows the practical applicability of the DWR–method.

2.4.1 Methodology for Standard Finite–Element Methods

Consider the model problem

$$-\Delta u = f \ \text{ in } \Omega, \quad u = 0 \ \text{ on } \partial\Omega, \tag{2.46}$$

for a polygonal domain $\Omega \subset \mathbb{R}^2$. The function space used in the weak formulation will be $V := H_0^1(\Omega)$ and in the variational formulation of (2.46)

we will look for $u \in V$, so that

$$a(u,v) = (f,v), \qquad \text{for all } v \in V, \tag{2.47}$$

with $a(u,v) := (\nabla u, \nabla v)$, and $(\cdot,\cdot)$ being the L^2-scalar product. To simplify the presentation of the estimator's structure we choose the approximation space $V_h := \mathcal{S}_{2,h} \cap H_0^1(\Omega)$ of continuous second order polynomials, cf. (2.45) for a definition. The discrete version of (2.46) reads as: Find $u_h \in V_h$ so that

$$a(u_h, v_h) = (f, v_h), \qquad \text{for all } v_h \in V_h.$$

Following the notation in [**15**] we denote the residual $\rho_h(\cdot) := (f,\cdot) - a(u_h,\cdot)$ of the discrete solution $u_h \in V_h$ and the error $e = u - u_h$. One important property is the 'Galerkin orthogonality'

$$0 = \rho_h(v_h) = a(e, v_h), \qquad \text{for all } v_h \in V_h.$$

The users interest, which might be the value of the solution at a fixed point in the domain or at the boundary $\delta_{x_0}(u)$ or boundary integrals for instance, is denoted as a functional $J(\cdot) \in H^{-1}(\Omega)$. The DWR–method consists of estimating the functional error $J(u) - J(u_h)$. Therefore we introduce the dual problem: Find $z \in V$ so that

$$J(v) = a(v,z), \qquad \text{for all } v \in V. \tag{2.48}$$

We observe that the functional error can be expressed by the residual functional applied to $z \in V$ the solution of the dual problem. We employ the Galerkin orthogonality for an arbitrary $v_h \in V_h$ to deduce

$$J(e) = a(u - u_h, z) = \rho_h(z) = \rho_h(z - v_h).$$

With an element–wise integration by parts we get the expression

$$J(e) = \sum_{K \in \mathcal{T}_h} (f - \Delta u_h, z - v_h)_K - (\nu \cdot \nabla u_h, z - v_h)_{\partial K}.$$

This leads to the error estimate: $|J(e)| \le \eta_\omega^{(1/2)}(u_h)$ with (cf. [**14, 15**])

$$\eta_\omega^{(1)}(u_h) = \sum_{K \in \mathcal{T}_h} |(f - \Delta u_h, z - v_h) - (\tfrac{1}{2}\nu \cdot [\nabla u_h], z - v_h)_{\partial K}|, \tag{2.49}$$

or in emphasizing the structure of element residuals ρ_K and weights ω_K with

$$\eta_\omega^{(2)}(u_h) = \sum_{K \in \mathcal{T}_h} \rho_K \omega_K, \quad \begin{cases} \rho_K := \|f - \Delta u_h\|_K + h^{-1/2} \|\tfrac{1}{2}\nu \cdot [\nabla u_h]\|_{\partial K}, \\ \omega_K := \|z - v_h\|_K + h^{1/2} \|z - v_h\|_{\partial K}. \end{cases} \tag{2.50}$$

The term $[\nabla u_h]$ denotes the jump of ∇u_h for an interior edge and is extended by 0 for boundary edges. The first part of ρ_K measures the local error production due to data approximation, whereas the second part measures the local smoothness of the solution.
Clearly, ρ_K is computable for each element once the approximate solution u_h is obtained, but the evaluation of the weights ω_K remains open. Three approaches to evaluate the weighting factors have been presented and assessed in [**15**, Section 5], that we briefly describe here to motivate the choice taken later for the new estimator.

(i) Starting with the first approach in the literature, [**13**], where we simply set $v_h = I_h z$, a suitable interpolant is chosen to derive

$$\omega_K \approx \|z - I_h z\|_{0,K} \leq c_i h^2 |z|_{2,K}.$$

For simplicity z_h the discrete dual solution on the same mesh is employed as an approximation to z. The quality of these weights is suboptimal.

Compared to separating the norms in $\eta_\omega^{(2)}$ the estimation by evaluation the formula (2.49) is more accurate. The remaining two alternatives substitute v_h in (2.49) by the discrete dual solution z_h on the same mesh and choose a more accurate approximation as a substitute for z.

(ii) This more precise approximation can be an interpolation of z_h on a space of higher order polynomials. In the test examples this method behaved more accurately compared to **(i)**. The interpolation for one element amounts to evaluating z_h on patches of elements and is therefore more costly than the first approach.

(iii) This method consists in computing globally $\hat{z}_h \in \hat{V}_h$ on a richer approximation space and set $v_h = I_h \hat{z}_h$. Enriching is either possible by refining the mesh or by increasing the polynomial degree. In both cases this is much more costly since a much larger dual problem needs to be solved. On the other hand this is the most accurate way to estimate the functional error. Another extension is to use completely different meshes for the primal and the dual problem as performed in [**74**] for a stabilized method applied to transport problems.

In the mixed finite-element method the postprocessing techniques of the previous section directly provide these higher order approximations without changing the triangulation, which makes them very attractive for the use in this methodology.

2.4.2 DWR-Estimator for Mixed Finite–Element Methods

We derive the estimator for a symmetric second order elliptic equation with mixed Dirichlet-Neumann boundary data. For notational simplicity we assume that the triangulation covers $\Omega \subset \mathbb{R}^2$ completely and that the problem data is sufficiently regular. The continuous problem reads

$$\begin{aligned} \sigma = a\nabla u, \quad -\operatorname{div}\sigma + du = f \quad \text{in } \Omega \\ \sigma\cdot\nu = 0 \ \text{ on } \Gamma_N, \quad u = g \quad \text{on } \Gamma_D \end{aligned} \tag{2.51}$$

Compared to the convection–diffusion problem that we were treating before there is no additional drift term in the first equation in (2.51) and we remark that we changed the sign there too. In order to treat convection–diffusion problems discretized with the method of the previous sections we would need to deal with the inconsistencies like $u(\boldsymbol{b}-\bar{\boldsymbol{b}})$ as well, which would require some additional thoughts.
One idea to recover the structure of the setting in standard finite-element methods could be to work with the bilinear $a(\cdot,\cdot)$ form of Section 2.2. This bilinear form acts only on discrete quantities and we do not have Galerkin orthogonality there, which would reduce the estimation accuracy further. Therefore we resort to the more standard bilinear form on the product space $X = H_N(\operatorname{div},\Omega)\times L^2(\Omega)$, where

$$H_N(\operatorname{div},\Omega) := \{\tau \in H(\operatorname{div},\Omega) |\, \langle \tau\cdot\nu, v\rangle = 0, \forall v \in H^1_{0,\Gamma_D}(\Omega)\},$$

and $H^1_{0,\Gamma_D}(\Omega)$ contains the functions whose traces on Γ_D are zero. We obtain the weak form: Find $[\sigma,u]\in X$ so that

$$A([\sigma,u],[\tau,v]) = B([\tau,v]) \quad \forall [\tau,v]\in X, \text{ with} \tag{2.52a}$$

$$A([\sigma,u],[\tau,v]) = \int_\Omega c\sigma\cdot\tau + u\operatorname{div}\tau - v\operatorname{div}\sigma + duv \,\mathrm{d}x, \tag{2.52b}$$

$$B([\tau,v]) = \int_\Omega fv\,\mathrm{d}x + \langle \tau\cdot\nu, g\rangle_{(H^{1/2}_{00}(\Gamma_D))',H^{1/2}_{00}(\Gamma_D)}. \tag{2.52c}$$

The coefficient $c = a^{-1}$ is again the inverse of the diffusion coefficient. We will not make particular use of properties of the spaces occurring in the dual pairing on the Dirichlet boundary. We mention them here as an example for a functional in X'. Exhaustive discussions on these spaces especially on nonsmooth domains may be found in [**66, 67**] although we adopt the more intuitive notation of [**8**], see also [**95**] for polygonal domains.
The first order system formulation is easily retrieved from (2.52) by choosing specific test functions. The equation for the current density is expressed by

$$A([\sigma,u],[\tau,0]) = B([\tau,0]) \quad \forall [\tau,0]\in X \tag{2.53}$$

and the second equation in (2.51) respectively by

$$A([\sigma, u], [0, v]) = B([0, v]) \quad \forall [0, v] \in X. \tag{2.54}$$

With conforming discrete spaces, eg. *RT*-elements of lowest order or *MP*-element for the current density, and the piecewise constant functions W_h here combined in $X_h = V_h \times W_h$ we formulate the discrete mixed finite-element problem: Find $[\sigma_h, u_h] \in X_h$ so that

$$A([\sigma_h, u_h], [\tau_h, v_h]) = B([\tau_h, v_h]) \quad \forall [\tau_h, v_h] \in X_h. \tag{2.55}$$

We abbreviate $[\epsilon, e] = [\sigma - \sigma_h, u - u_h]$. As before we obtain the residual $\rho_{[\sigma_h, u_h]} \in X'$ and due to $X_h \subset X$ we directly have Galerkin orthogonality

$$\rho_{[\sigma_h, u_h]}([\tau, v]) = A([\epsilon, e], [\tau, v]) \quad \forall [\tau, v] \in X. \tag{2.56}$$

$$0 = A([\epsilon, e], [\tau_h, v_h]) \quad \forall [\tau_h, v_h] \in X_h. \tag{2.57}$$

If the quantity of interest can be formulated as a functional $F \in X'$ the dual problem has the following form: Find $[\zeta, z] \in X$ so that

$$F([\vartheta, w]) = A([\vartheta, w], [\zeta, z]) \quad \forall [\vartheta, w] \in X. \tag{2.58}$$

Inserting here the error $[\epsilon, e]$ as test function and in view of (2.53), (2.54) and (2.57) we obtain

$$\begin{aligned} F([\epsilon, e]) =& A([\epsilon, e], [\zeta, z]) = \rho_{[\sigma_h, u_h]}([\zeta - \zeta_h, z - z_h]) \quad \forall [\zeta_h, z_h] \in X_h \\ =& B([\zeta - \zeta_h, 0]) - \int_\Omega c\sigma_h(\zeta - \zeta_h) + u_h \operatorname{div}(\zeta - \zeta_h) \,\mathrm{d}x + \\ & \int_\Omega \underbrace{(f + \operatorname{div} \sigma_h - du_h)}_{=:\rho_V}(z - z_h) \,\mathrm{d}x. \end{aligned} \tag{2.59}$$

The last term contains the volumetric residual ρ_V that we have already seen in the standard finite-element case, representing the error due to data approximation. We still deserve a measure of the local smoothness of the solution. In standard finite-element methods this was measured by $\nu \cdot [\nabla u]_e$, the jump of the gradient's normal component over internal edges. Now, due to the current density's approximation in $H_N(\operatorname{div}, \Omega)$, these jumps would be zero. The question is how to combine the remaining parts of (2.59) in a locally balanced form, that can be regarded as a measure of the solution's local smoothness. The tool of choice would be element-wise partial integration in the second last term in (2.59), but applied directly we end up

with element boundary integrals of jumps of u_h, which do not balance with any other term. To cure this problem we choose some $\tilde{u} \in H^1(\Omega)$, add and subtract to (2.59) $\tilde{u}\,\mathrm{div}(\zeta - \zeta_h)$, and by partial integration we get

$$F([\epsilon, e]) = \int_\Omega \Big\{ (\nabla \tilde{u} - c\sigma_h)(\zeta - \zeta_h) + (\tilde{u} - u_h)\,\mathrm{div}(\zeta - \zeta_h) + \rho_V(z - z_h) \Big\}\,\mathrm{d}x + B([\zeta - \zeta_h, 0]). \tag{2.60}$$

Our objective is to define $\tilde{u} \in H^1(\Omega)$ in an efficiently computable way. In the implementation of the mixed method u_h itself is derived from the Lagrange multipliers λ_h. The interpretation of the Lagrange multipliers as nodal values of $\hat{u}_h \in CR_{h,g}$ does not serve directly as a substitute, since $CR_{h,g} \not\subset H^1(\Omega)$, cf. (2.38) for a definition. Therefore we use the projection $P_S : L^2(\Omega) \to \mathcal{S}_{1,h}$ introduced in [**117**, Subsection 4.2]. If we restrict the arguments to $v \in CR_{h,g}$, the projection $P_S v \in \mathcal{S}_{1,h}$ is described through the values at the vertices p

$$P_S v(p) = \frac{1}{n_p} \sum_{K \in \omega_p} v|_K(p),$$

where n_p is the number of elements K_i so that p is a vertex of K_i and $\omega_p = \bigcup_i K_i$. The averaging that enters by defining $\tilde{u} := P_S \hat{u}_h$ clearly connects in (2.60) local solution values $[\sigma_h, u_h]|_K$ with the values in the patch of elements containing K, and therefore measures the smoothness of the solution. This enables us to present a DWR-estimator in the mixed finite-element setting.

Proposition 2.1. *Let $[\sigma, u] \in X$ be the solution to (2.52) and $[\sigma_h, u_h] \in V_h \times W_h = X_h \subset X$ the discrete solution in a conforming approximation space and $\hat{u}_h \in CR_h$ the reinterpretation of the Lagrange multipliers occurring in solving the matrix problem. Let $F \in X'$ describe the quantity of interest, and abbreviate the error by $[\epsilon, e] = [\sigma - \sigma_h, u - u_h]$. The functional error $F([\epsilon, e])$ may be bounded by*

$$|F(\epsilon, e)|^2 \le \sum_{K \in \mathcal{T}_h} \eta_{V,K}\omega_{z,K} + \eta_{S_1,K}\omega_{\zeta,K} + \eta_{S_2,K}\omega_{\mathrm{div}\,\zeta,K} =: \eta_\omega, \tag{2.61}$$

where

$$\eta_{V,K} = \|f + \mathrm{div}\,\sigma_h - du_h\|_{K,0}, \qquad \omega_{z,K} = \|\hat{z}_h - z_h\|_{K,0}, \tag{2.62}$$

$$\eta_{S_1,K} = \|\nabla(P_S\hat{u}) - c\sigma_h\|_{K,0}, \qquad \omega_{\zeta,K} = \|\mathcal{A}\zeta_h - \zeta_h\|_{K,0}, \tag{2.63}$$

$$\eta_{S_2,K} = \|(P_S\hat{u}) - u_h\|_{K,0}, \qquad \omega_{\mathrm{div}\,\zeta,K} = \|\,\mathrm{div}(\mathcal{A}\zeta_h - \zeta_h)\|_{K,0}. \tag{2.64}$$

Here $[\zeta_h, z_h]$ *denotes the solution to the discrete dual problem, and again* $\hat{z}_h \in CR_h$ *denotes the reinterpretation of the Lagrange multipliers in the dual matrix problem. The operator* $\mathcal{A}$ *is the averaging operator defined in Subsection 2.3.3 adjusted to the boundary data of the dual problem.*

Remark 2.7. *Carstensen presented residual estimators for standard mixed finite-elements in* [28] *and in a unifying framework very recently in* [30]. *In* [30] *it is shown in the case* $c \equiv 1$ *for RT-elements:*

$$\min_{w \in H^1(\Omega)} \|\sigma_h - \nabla w\|_0 \approx \|h_{\mathcal{E}_h}^{1/2} [\sigma_h \cdot t_{\mathcal{E}_h}]\|_{L^2(\bigcup \mathcal{E}_h)}.$$

Although the proof delivers no construction of the minimizer, this result supports the argument that contributions $\|\nabla \tilde{u} - c\sigma_h\|_{0,K}$ *measure the local smoothness of the solution. The estimates of Carstensen are focusing global norms. The technique to derive the estimate is based on the evaluation of the residual* $\rho_{[\sigma_h, u_h]}$ *on test functions of the form*

$$\boldsymbol{q}_h = (-\partial_y v_h, \partial_x v_h)^\top = \mathrm{Curl}_h\, v \in H(\mathrm{div}, \Omega),$$

where $v \in S_{1,h}$. *If we aim at functional error estimation the test function is not at our disposal, since it is a solution* $[\zeta, z]$ *to a dual problem, which is the main difference between this approach and the estimation in global norms.*

Remark 2.8. *We remind the reader that error estimates for* $\|\mathrm{div}(\zeta - \zeta_h)\|_0$ *are not available for* MP*–elements. This is caused by the missing commuting diagram property for these elements, cf.* [24]. *This lack might cause an overrefinement. In the following we perform simulations for the* RT*– and* MP*–finite-elements, and there is no difference in the error decay that could be interpreted as an overrefinement.*

2.4.3 A Problem of the SIAM 100-Digit Challenge

The challenge consisted of ten numerical problems posted in the *SIAM News* issue January/February 2002 and later also in *Science*. The solution to each of the problems was a real number and the contestants could get a point for every of the first ten correctly computed digits of these numbers. This challenge was a great succuss, in the sense that 94 teams from 24 countries entered the competition, many of them solved all the problems and published there results, finally all this culminated in a book [19]. Folkmar Bornemann, investigated the last of these ten problems further and finally found that the solution can be written in a closed form, that can be evaluated with arbitrary precision, see below. This together with the fact that the problem

can be solved by finite-elements, and a remark Folkmar Bornemann made in a talk, that there is a lack of a good error estimator for this problem in the finite-element package that he used, attracted the attention of the author. Additionally to the performance demonstration, this problem serves as an example of how to apply a DWR-estimator in cases where the quantity of interest cannot be written as an element of X'. The problem was originally stated in the form:

> A particle at the center of a 10x1 rectangle undergoes a Brownian motion (i.e. 2D random walk with infinitesimal step length) till it hits the boundary. What is the probability that it hits on of the ends rather than at the sides?

Translated into an elliptic boundary value problem this reads: If $u(x_1, x_2)$ describes the probability that a particle starting at the point $\boldsymbol{x} = (x_1, x_2)^\top$ hits one of the ends, then u is the solution to

$$\begin{aligned} -\Delta u = 0 \text{ in } (-L, L) \times (-0.5, 0.5),\ L = 5 \\ u|_{\{|x_2|=L\}} = 1, \quad u|_{\{|x_2|=0.5\}} = 0. \end{aligned}$$

The domain is chosen such that the center of the rectangle, our starting point, lies in the origin. This problem may be solved by separation of variables, which describes the solution as an infinite series. Solutions in closed form may be obtained by methods of complex analysis or the theory of elliptic functions, see [19]. For our objective, to solve the problem numerically, already the separation ansatz delivers sufficiently precise values. Before this

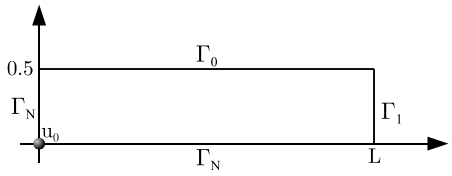

Figure 2.5: Computational domain

problem is solved numerically the symmetries are exploited, see Figure 2.5, to yield

$$\begin{aligned} -\Delta u = 0 \text{ in } \Omega = (0, L) \times (0, 0.5), \\ u|_{\Gamma_1} = 1, \quad u|_{\Gamma_0} = 0, \quad \nu \cdot \nabla u = 0 \text{ on } \Gamma_N, \\ \Gamma_N = \{(\boldsymbol{x} |\ x_1 = 0 \vee x_2 = 0\}, \Gamma_1 = \{x_1 = L\},\ \Gamma_0 = \{x_2 = 0.5\}. \end{aligned} \tag{2.65}$$

The solution to the original problem is $u_0 = 3.8375\,89792\,51226\ldots \cdot 10^{-7}$. Interestingly, if $L = \sqrt{3}/2$, the probability to reach Γ_1 with a random walk starting in the origin is $u_0 = 1/6$. We perform simulations for both aspect ratios. Transfered into the mixed finite-elements setting we have: Find $[\sigma, u] \in X = H_N(\mathrm{div}, \Omega) \times L^2(\Omega)$:

$$A([\sigma,u],[\tau,v]) = \int_\Omega \sigma\cdot\tau + u\,\mathrm{div}\,\tau - v\,\mathrm{div}\,\sigma\,\mathrm{d}x = \int_{\Gamma_1} \tau\cdot\nu\,\mathrm{d}s = B([\tau,v])$$

for all $[\tau, v] \in X$. The functional output clearly is δ_0, the Dirac measure of the origin. Here we encounter the first difficulty, since δ_0 is not an element of X' and neither of X_h'. We need to find an approximation of δ_0 in W_h. Functions in W_h are only defined inside a triangle. For an adaptively refined mesh the origin is generally a vertex that is connected to several triangles. In the computation we chose therefore $u_0{:=}(P_S u_h)(0)$, which simplifies to the integral mean value of u_h over Ω_0, denoting by Ω_0 the patch of elements connected to zero. This corresponds to the approximation of $\delta_{0,h} \in W_h$ of the form

$$\delta_{0,h}|_K = \begin{cases} |\Omega_0|^{-1}, & \text{if } K \in \Omega_0 \\ 0, & \text{elsewhere.} \end{cases}$$

The evaluation of the functional output can be reformulated by

$$\delta_{0,h}(u) = F([\sigma,u]) = A([\sigma,u],[\zeta,z]) = B([\zeta,z]) = \int_{\Gamma_1} \zeta\cdot\nu\,\mathrm{d}s.$$

For a function $\zeta \in V_h \subset H(\mathrm{div}, \Omega)$ the jumps $[\zeta\cdot\nu]$ over inter-element edges vanish, therefore the last boundary integral might be a better way to evaluate the functional output in the end. Neglecting the connection between computing $u_{0,h}$ and the functional output, the source term in the dual problem $F([\vartheta, w]) = A([\vartheta, w], [\zeta, z])$ can be any discrete version of δ_0. In the computation we tried also the following approximation; let K_0 be the triangle whose barycenter has the closest distance to the origin and set

$$\tilde{\delta}_{0,h}|_K = \begin{cases} 1/|K_0|, & \text{if } K = K_0, \\ 0, & \text{otherwise.} \end{cases}$$

Regarding again the primal problem, the jumping boundary condition makes a refinement in the top left corner together with a low order finite-element approximation necessary. To the contrary the smoothness of the solution close to the origin would allow a higher order polynomial approximation. We will not be able to obtain ten digits with our method. The aim of

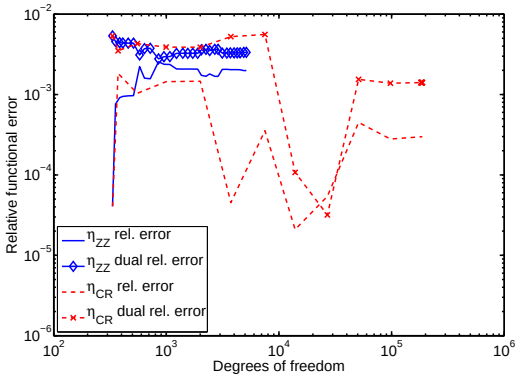

Figure 2.6: Refinement controlled by the estimators η_{ZZ} and η_{CR} in the case $L = \sqrt{3}/2$. For both estimators the functional error does not decrease. Refinement under η_{ZZ} only refines near the top right corner and even after many refinement steps the error does not decrease.

the following simulation is that starting from a coarse mesh the estimator should control the refinement automatically so that the estimated functional error decreases. For a comparison the functional error is displayed in Figure 2.6 if the refinement is controlled by η_{CR} or η_{ZZ} in the case $L = \sqrt{3}/2$. Although the jump in the boundary data is closer to the point of interest the refinement does not lead to a steadily increased accuracy in the output. Next compare the influence of different approximation of the functional output $\delta_{0,h}$ and $\tilde{\delta}_{0,h}$. The estimated error in the functional output decays for both approximation with the same rate, see Figure 2.7. If the exact functional output is regarded as unknown, this simulation can assure the user that the computed quantity is calculated with a higher than the estimated accuracy. To improve the global functional error estimate one could substitute $[\zeta, z] := [\mathcal{A}\zeta_h, \hat{z}_h]$ directly in evaluating the right-hand side of equation (2.60). The example on the right in Figure 2.7, shows that after some refinement steps controlled by η_ω the evaluation of $u_{0,h}$ does not differ from evaluating $F([\zeta_h, z_h])$. This simulation was performed with a maximum element number of 40000, and by refining elements that sum up to a fixed fraction of the total estimator value.

This example problem permits a comparison of the approximation by *MP*– or *RT*–elements with very little changes. Considering the question of a possible overrefinement for the *MP*–elements the next test gives a first positive answer, in the sense that the error decays for both elements with the same approximate rate. The similarities in the discretizations are due to the fact

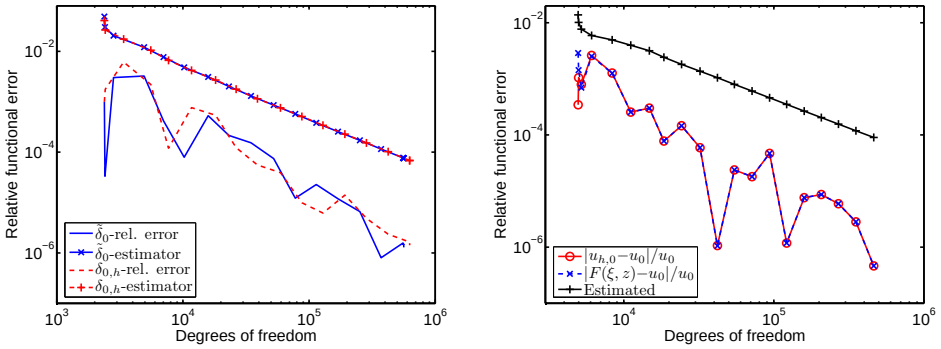

Figure 2.7: On the left a comparison of refinements for approximate dual functionals $\delta_{0,h}$ and $\tilde{\delta}_{0,h}$. The good agreement of the final approximate values supports the argument that the final values are of a higher accuracy than estimated. On the right the difference in evaluating $u_{0,h}$ or $F([\zeta_h, z_h])$ is almost invisible after some refinement steps.

that there is no zeroth order term d in (2.65). The matrix entries in the Lagrange multiplier formulation $M\boldsymbol{\lambda} = G$ for the RT–elements, given as an example in [**43**, Section 3.2], are the same as in the linear system for the MP–elements. The parts in G representing the Dirichlet boundary data are the same, too. Together this leads in the primal problem to $\lambda^{RT} \equiv \lambda^{MP}$. The reconstruction of $[\sigma_h, u_h]$ (cf. proof of Theorem 2.1 and [**44**]) simplifies due to the symmetry of the present problem to

$$\sigma_h^{RT,MP} = \boldsymbol{Q}\lambda, \quad u_h^{RT} = \boldsymbol{U}^{RT}\lambda, \quad u_h^{MP} = \boldsymbol{U}\lambda.$$

For the dual problem the approximation of $\delta_{0,h}$ leads to different right-hand sides G^{RT} and G^{MP}, but the reconstructions of the current density still bear some similarities:

$$\begin{aligned} \zeta_h^{RT} &= \boldsymbol{Q}\lambda^{RT} + \boldsymbol{\mathcal{Q}}^{RT}\delta_{0,h}, & z_h^{RT} &= \boldsymbol{U}^{RT}\lambda^{RT} + \boldsymbol{\mathcal{U}}^{RT}\delta_{0,h}, \\ \zeta_h^{MP} &= \boldsymbol{Q}\lambda^{MP} + \boldsymbol{\mathcal{Q}}^{MP}\delta_{0,h}, & z_h^{MP} &= \boldsymbol{U}^{MP}\lambda^{MP} + \boldsymbol{\mathcal{U}}^{MP}\delta_{0,h}. \end{aligned}$$

The averaging operator $\mathcal{A}$ used to compute the higher order approximation $\mathcal{A}\zeta_h$ does not change if we define the basis of $RT(K)$ as:

$$RT(K) = \operatorname{span}\{(1,0)^\top, (0,1)^\top, \boldsymbol{x} - \boldsymbol{x}_B\}, \quad \boldsymbol{x}_B \text{ the barycenter of } K.$$

In Figure 2.8 we display for both aspect ratios the error decay under the new estimator applied to the RT– and MP–elements implementation. Clearly the

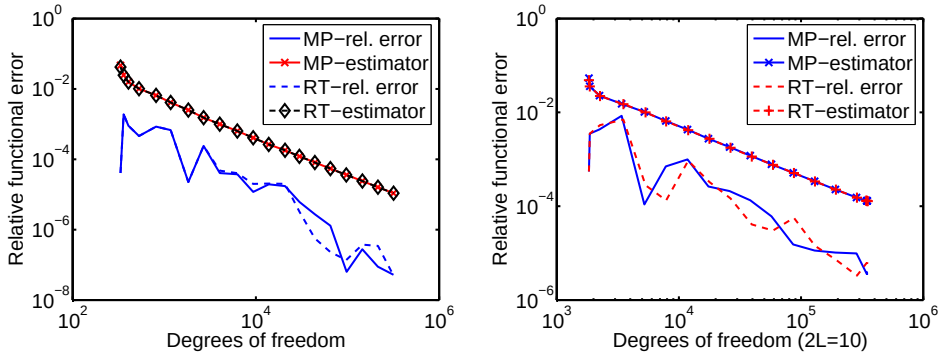

Figure 2.8: Comparison of RT– and MP– elements for the current density variable. On the left for $L = \sqrt{3}/2$, on the right $L = 5$.

problem for $L = 5$ is harder to solve, but for both elements we see a very similar refinement behavior and the difference in the estimated values of u_0 is very small, too.

We point out that standard estimators fail to control the refinement process as we see from Figure 2.6. In contrast to that we see in all the examples of the new estimator that the estimated error decreases with each refinement step. The true error remains always much smaller than the estimated value although it does not behave with the same monotonicity.

Chapter 3

The Semiconductor Application

This chapter will explain how the mixed finite-element method is applied to the device simulation problem. As a first step we present the complete physical model and scrutinize the scaling of the equation. This is followed by a characterization of the state of thermal equilibrium. This description will provide the boundary data as well as a motivation for the iterative procedure that is used to solve the nonlinear system, which can be regarded as an approximate Newton method.

3.1 The Complete Physical Model

The introduction was meant to show the physical background leading to the convection-diffusion equations, we will now specify the precise constitutive relations used in the device simulation. We start with defining the complete model in the physical variables, detail the scaling to a dimensionless form in order to recover the precise coefficients in the final system of convection-diffusion equations.
All the following device examples are majority carrier devices which means that the terminal current depends mainly on the current induced by the majority carriers. The flow of the majority carriers, the electrons, is modeled by the system of energy transport equations. We neglect temperature effects for the minority carriers thus couple the drift-diffusion equation for holes to the problem. The carrier transport equations are coupled through the generation recombination effects. These effects are accounted for in the term $R(n, p)$. For the Shockley-Read-Hall recombination generation effect

this term has the form:

$$R(n,p) = \frac{np - n_i^2}{\tau_p(n+n_i) + \tau_n(p+n_i)},$$

with the intrinsic density n_i and the electron and hole life times τ_n, τ_p. The energy relaxation term is given in the Fokker-Planck approximation [**106**, **49**] by

$$W(n,t) = \frac{3}{2}\frac{nk_B(T-T_0)}{\tau_\beta(T/T_0)},$$

where k_B is the Boltzmann constant, T_0 the lattice temperature. The index $\beta > -1$ determines the energy dependence of the inelastic scattering cross section, thus the temperature dependent energy relaxation time τ_β will be specified for the Chen model ($\beta = 1/2$) and the Lyumkis model ($\beta = 0$) separately. The unscaled complete problem has the form

$$-\operatorname{div} J_1 = -qR(n,p), \tag{3.1}$$

$$-\operatorname{div} J_2 = -J_1 \cdot \nabla V + W(n,T) - \frac{3}{2}k_B T R(n,p), \tag{3.2}$$

$$-\operatorname{div} J_p = qR(n,p) \tag{3.3}$$

$$\varepsilon_0\varepsilon_r \Delta V = q(n - p - C(x)), \tag{3.4}$$

with current density relations

$$J_1 = q\left(\nabla\left(\mu_\beta^{(1)}\left(\frac{T}{T_0}\right)\frac{k_B T}{q}n\right) - \mu_\beta^{(1)}\left(\frac{T}{T_0}\right)n\nabla V\right), \tag{3.5}$$

$$J_2 = \nabla\left(\mu_\beta^{(2)}\left(\frac{T}{T_0}\right)\frac{(k_B T)^2}{q}n\right) - \mu_\beta^{(2)}\left(\frac{T}{T_0}\right)nk_B T\nabla V, \tag{3.6}$$

$$J_p = -q\left(\mu_{p,0}\frac{k_B T_0}{q}\nabla p + \mu_{p,0}p\nabla V\right), \tag{3.7}$$

where q is the elementary charge. The terms $\mu_\beta^{(i)}$ are the electron and electron energy mobility relations, which each depend on the inelastic scattering process, too, and therefore have different forms depending on the parameter β, see equation (3.8) below. Instead of adjusting the material related parameters in the scattering cross section and the effective mass, the material properties are expressed by the low field mobilities $\mu_{n,0}, \mu_{p,0}$ and a typical energy relaxation time τ_0 which can be quantified by measurements. For the precise dependence of τ_0 and $\mu_{n,0}, \mu_{p,0}$ on the microscopic quantities, the effective mass of electrons in the conduction band m_c^* and the scattering cross section $\phi_0(x)$, we refer to [**49**, Section 2.4].

Let additionally C_m be the maximal value of the doping profile, ℓ^* the diameter of the device, and $U_0 = k_B T_0/q$ the thermal voltage. Using the scaling

$$\begin{gathered} n \to C_m n, \quad p \to C_m p, \quad C \to C_m C, \quad T \to T_0 T, \quad V \to U_0 V, \quad x \to \ell^* x, \\ J_1 \to j_0 J_1, \quad J_2 \to (U_0 j_0) J_2, \quad J_p \to (q\mu_{p,0} U_0 C_m/\ell^*) J_p, \\ R \to (j_0/(\ell^* q))R, \quad n_i \to C_m n_i, \quad W \to (U_0 j_0/\ell^*) W, \end{gathered}$$

with $j_0 = (q\mu_{n,0} U_0 C_m/\ell^*)$ the scaling factor of the electron current density, we can introduce the new variables g_i for $i = 1, 2$, based on the scaled version of the mobilities $\mu_\beta^{(i)}$:

$$g_i(n,T) = \mu_\beta^{(i)}(T) T^i n, \quad \mu_\beta^{(i)}(T) = \frac{2\Gamma(i+1-\beta)}{\sqrt{\pi}} T^{-\frac{1}{2}-\beta}. \tag{3.8}$$

The symbol Γ denotes the Gamma function defined by

$$\Gamma(s) = \int_0^\infty u^{s-1} e^{-u} du, \quad s > 0.$$

In the scaled equations below we find various products of mobilities and relaxation times $\mu_{r,0}\tau_s$, $r \in \{n, p\}$, $s \in \{0, n, p\}$ for which the above scaling leads to the dimensionless form

$$\mu_{r,0}\tau_s \to \ell^2/U_0 \mu_{r,0}\tau_s.$$

In the new variables the energy relaxation term has the form

$$W(n,T) = \frac{3}{2\tau_\beta(T)T}\left(\frac{g_1}{\mu_\beta^{(1)}(T)} - \frac{g_2}{\mu_\beta^{(2)}(T)}\right), \quad \text{with}$$
$$\tau_\beta(T) = \frac{3\sqrt{\pi}\mu_{n,0}\tau_0}{4\Gamma(\beta+2)} T^{\frac{1}{2}-\beta}.$$

The scaled equations in the variables (g_1, g_2, p, V) have the form

$$-\operatorname{div} J_i + c_i g_i = f_i, \quad J_i = \nabla g_i - g_i \frac{\nabla V}{T(g_1, g_2)}, \tag{3.9}$$
$$-\operatorname{div} J_p + c_p p = f_p, \quad J_p = -\nabla p - p\nabla V, \tag{3.10}$$
$$\lambda^2 \Delta V = n(g_1, g_2) - p - C(x) \qquad \text{in } \Omega, \tag{3.11}$$
$$g_i = g_{D,i}, \quad p = p_D, \quad V = V_D \qquad \text{on } \Gamma_D,$$
$$J_1 \cdot \nu = J_2 \cdot \nu = J_p \cdot \nu \nabla V \cdot \nu = 0 \qquad \text{on } \Gamma_N,$$

where $g_{D,i} = g_i(n_D, T_D)$ for $i = 1, 2$ and $\lambda = \sqrt{\varepsilon_0 \varepsilon_r U_T/(qC_m \ell^{*2})}$ is the scaled Debye length. The coefficients c_i and right-hand sides f_i are now specified for the models employed in the simulations. For the Chen model ($\beta = 1/2$) we have $\Gamma(3/2) = \sqrt{\pi}/2$ and $\Gamma(5/2) = 3\sqrt{\pi}/4$ and therefore

$$c_1 = \frac{r}{\mu_{n,0}} p, \quad f_1 = \frac{r}{\mu_{n,0}} n_i^2, \quad c_p = \frac{r}{\mu_{p,0}} n, \quad f_p = \frac{r}{\mu_{p,0}} n_i^2 \tag{3.12}$$

$$c_2 = \frac{1}{\mu_{n,0}} \left(\frac{1}{\tau_0} + rp\right) \quad f_2 = -J_1 \cdot \nabla V + \frac{3}{2\mu_{n,0}} \left(\frac{g_1}{\tau_0} + T n_i^2 r\right) \tag{3.13}$$

$$T(g_1, g_2) = \frac{2g_2}{3g_1}, \quad n(g_1, g_2) = g_1, \tag{3.14}$$

where $r(g_1, g_2, p) = (\tau_p(n(g_1, g_2) + n_i) + \tau_n(p + n_i))^{-1}$. In many situations, the doping is used to create very high electron concentrations, for example, the minority carrier concentration is then several orders of magnitude below that of the majority carrier and is therefore neglected completely in the simulation. Thus for unipolar simulations the equation (3.10) is omitted and in equation (3.11) it is set $p = 0$ and $r = 0$, which leads to $c_1 = f_1 = 0$. The coefficients in (3.9) for unipolar simulations with the Lyumkis model ($\beta = 0$) in the variables (g_1, g_2, V) have the form

$$c_2(g_1, g_2) = \frac{1}{\tau_0 \mu_{n,0} T(g_1, g_2)}, \quad f_2(g_1, g_2) = -J_1 \cdot \nabla V + \frac{1}{2} c_2, \tag{3.15}$$

$$T(g_1, g_2) = \frac{g_2}{2g_1}, \quad n(g_1, g_2) = \left(\frac{\pi}{2} \frac{g_1^3}{g_2}\right)^{1/2}. \tag{3.16}$$

Name	Description	Value
$\tau_p \,/\, \tau_n$	carrier life times	$10^{-5} \,/\, 10^{-6}$s
$\mu_{p,0} \,/\, \mu_{n,0}$	low-field carrier mobilities	$450 \,/\, 1500$ cm^2/Vs
n_i	intrinsic density	$1.4 \cdot 10^{10}$ cm^{-3}
τ_0	energy relaxation time	0.4 ps
ε_0	permittivity of vacuum	$8.85{\cdot}10^{-14}$ As/(Vcm)
$\varepsilon_r(Si)$	relative permittivity of silicon	11.7
$\varepsilon_r(SiO_2)$	relative permittivity of the oxide	3.8
T_0	ambient temperature	300K

Table 3.1: Material and model parameters.

The precise Dirichlet boundary will be specified in the next section, since they are derived from the relations at thermal equilibrium. To conclude

this section we collect the material parameters of Silicon in Table 3.1. The relative permittivity of SiO_2 will be used in the simulation of a MOSFET device in Subsection 3.4.4.

3.2 Thermal equilibrium

The state of thermal equilibrium plays an important role also in the device simulation. In this situation the system of energy-transport equations reduces to a semilinear Poisson equation, which is clearly much simpler to solve numerically than the full nonlinear system. When generating current-voltage curves the solution at thermal equilibrium is used to compute the starting values for the nonlinear solution mechanism to solve the complete system.
The stationary state of thermal equilibrium is characterized by zero current flow, and an electron temperature that equals the lattice temperature. If the lattice temperature is constant, the energy-transport system reduces to the drift-diffusion model. Together this implies that $R(n,p) = 0$ or $np = n_i^2$, see (3.1). The current density relation J_n and J_p reduce to

$$J_n = \nabla n - n\frac{\nabla V}{T_0}, \quad J_p = \nabla p + p\frac{\nabla V}{T_0},.$$

The equation $J_n = J_p = 0$ together with $np = {n_i}^2$ yields, for details see [**91**],

$$n = n_i e^V, \quad p = n_i e^{-V}. \tag{3.17}$$

Inserting this into the Poisson equation we obtain the semilinear elliptic equation

$$\begin{aligned} \lambda^2 \Delta V &= n - p - C = n_i(e^V - e^{-V}) - C \\ &= 2n_i \sinh V - C \quad \text{in } \Omega. \end{aligned} \tag{3.18}$$

In order to avoid boundary layers the built-in potential V_{el} is the unique solution to boundary value problem that is obtained by requiring a vanishing right-hand side of (3.18)

$$V_{\text{el},D} = \operatorname{arcsinh}\left(\frac{C}{2n_i}\right) \text{ on } \Gamma_D, \quad \nabla V_{\text{el}} \cdot \nu = 0 \text{ on } \Gamma_N.$$

Once the built-in potential has been computed, the equilibrium particle densities are directly given by (3.17). A damped Newton algorithm to solve this problem would have the form:

(i) Initialize V_0 obeying the boundary conditions.

(ii) Do
  (a) Let $n = n_i e^{V_0}$, $p = n_i e^{-V_0}$.
  (b) Set $V := V_0$ and solve

$$\begin{cases} \lambda^2 \Delta\phi - (p+n)\phi = -\lambda^2 \Delta V + n - p - C & \text{in } \Omega \\ \phi = 0 \quad \text{on } \Gamma_D, \quad \nabla\phi \cdot \nu = 0 \quad \text{on } \Gamma_N. \end{cases} \tag{3.19}$$

  (c) Set $V_0 := V + t\phi$, with some $t \in (0, 1]$.
  Until convergence.
(iii) Set $V_{\text{el}} := V_0$.

The built-in potential characterizes the device in thermal equilibrium. If we want to solve the macroscopic semiconductor device equation in a non-equilibrium situations, we need a to accommodate the iterative method to solve the coupled nonlinear system. A damped Newton method would be much more complex compared to the equilibrium case described above. Instead, often a different class of iteration methods is used, the so-called *Gummel maps*. We now give a brief idea, starting with the boundary conditions, of how to define a Gummel mapping.
Non-equilibrium situations are induced by adding a potential difference at the Dirichlet contacts. It is assumed that the particle densities remain in thermal equilibrium at the ohmic contacts. This leads to Dirichlet boundary data in the application [**91**, Chapter 4]

$$\begin{aligned} n_D &:= \frac{1}{2}\big(\sqrt{C^2 + 4n_i^2} + C\big), \quad & p_D &:= \frac{1}{2}\big(\sqrt{C^2 + 4n_i^2} - C\big), \\ V_D &:= V_{\text{el}} + V_{\text{appl.}}, & T_D &:= 1. \end{aligned} \tag{3.20}$$

The Neumann boundary conditions remain unchanged, see (1.41).
In order to illustrate the iterative procedure to solve the nonlinear system, we may define new variables, ρ_n and ρ_p from

$$n = n_i \rho_n e^{V}, \quad p = n_i \rho_p e^{-V}.$$

Gummel-type mappings consist of variants of the above algorithm for the equilibrium case. They are obtained by substituting the equilibrium relations in step (ii)(a) with drift-diffusion equations for the given potential V_0 to compute updates ρ_n and ρ_p or equivalently to obtain n and p. These new values for the densities n and p are then used in step (ii)(b) to update the potential until convergence is reached. Clearly ρ_n and ρ_p are equal to one in the state of thermal equilibrium. But they vary largely if an external potential is applied and (n, p, V) is a solution to the complete drift-diffusion

model. Proofs of convergence of such approximate Newton methods rely on $\rho_n - 1$ and $\rho_p - 1$ being small in suitable norms.
Therefore convergence results for Gummel mappings were only derived for small applied voltage differences and $R = 0$, see [**76**] and the references therein. To the contrary, the extension of this idea to the energy-transport system still leads to a flexible and stable method that converges even in situations where of a full Newton method did not converge due to the large condition number of the Jacobian matrix. We point out that in the Gummel iteration the information about the strong coupling of the unknowns is incorporated into the Poisson equation only. This allows a flexible extension to more advanced simulations including multi band methods where equations for each band are coupled through extended recombination generation effects, see Kerkhoven [**83**]. This procedure of step-wise solving linear elliptic subproblems reduces the memory consumption which would allow to increase the number of degrees of freedom in the finite-element space so that extensions to finite-element approximations of three-dimensional domains are possible. An extension of the mixed finite-element method of Section 2.1 to three dimensions has not been derived yet, and only very few exponentially fitted methods have been proposed for three dimensions, see [**4**] and the references therein.

3.3 Global Iteration

For the classical drift-diffusion model, several iterative procedures for solving the coupled system have been proposed in the literature (see e.g., [**76, 84, 9**] and the references therein). We have used a Gummel-type [**69**] method for the numerical examples in the following section, combined with an iteration procedure for the temperature. An adaptive mesh refinement is integrated in a second step.
This decoupling procedure allows us to apply the mixed scheme of Chapter 2 to the corresponding subproblems (see below). We present the iteration in case of the energy-transport-drift-diffusion system, when the Chen model is used for the constitutive relations in the energy-transport part.
We recall that in the Chen model, $g_1 = n$ and $g_2 = (3/2)nT$ holds but in general, n is a function of g_1 and T (see [**49**]). We assume that a set of functions $(g_1^{(l)}, g_2^{(l)}, p^{(l)}, V^{(l)}, T^{(l)})$, some constant $\delta T^{(l)}$ and the piecewise constant function $\overline{T}$ are given. These functions may not only stem from a foregoing iteration step but can also stem from an approximate solution on a coarser mesh which has been interpolated. The algorithm to solve the

system (called algorithm (A)) is as follows:

1. Let $g_1^* = g_1^{(l)}$, $p^* = p^{(l)}$, $V = V^{(l)}$.
2. Do
   (a) find g_1 such that

$$\begin{cases} -\operatorname{div} J_1 + \frac{r(g_1^*, T^{(l)}, p^*)}{\mu_{n,0}} p^* g_1 = n_i^2 \frac{r(g_1^*, T^{(l)}, p^*)}{\mu_{n,0}} & \text{in } \Omega, \\ J_1 = \nabla g_1 - \nabla V \overline{T}^{-1} g_1 & \text{in } \Omega, \\ g_1 = n_D \quad \text{on } \Gamma_D, \quad J_1 \cdot \nu = 0 \quad \text{on } \Gamma_N; \end{cases} \tag{3.21}$$

   (b) find p such that

$$\begin{cases} -\operatorname{div} J_p + \frac{r(g_1, T^{(l)}, p^*)}{\mu_{p,0}} g_1 p = n_i^2 \frac{r(g_1, T^{(l)}, p^*)}{\mu_{p,0}} & \text{in } \Omega, \\ J_p = \nabla p + \nabla V p & \text{in } \Omega, \\ p = p_D \quad \text{on } \Gamma_D, \quad J_p \cdot \nu = 0 \quad \text{on } \Gamma_N; \end{cases} \tag{3.22}$$

   (c) set $n = g_1$ and $V_1 = V + \delta V$, where δV is the solution of

$$\begin{cases} \lambda^2 \Delta(\delta V) - (p+n)\delta V = -\lambda^2 \Delta V + n - p - C & \text{in } \Omega, \\ \delta V = 0 \quad \text{on } \Gamma_D, \quad \nabla(\delta V) \cdot \nu = 0 \quad \text{on } \Gamma_N; \end{cases} \tag{3.23}$$

   (d) set $g_1^* := g_1$, $p^* := p$, $V := V_1$;

   until $\|\delta V\|_{L^2} < \varepsilon(\delta T^{(l)})$.
3. Find $g_2^{(l+1)}$ such that for $r = r(g_1, T^{(l)}, p)$

$$\begin{cases} -\operatorname{div} J_2^{(l+1)} + \frac{3}{2\mu_{n,0}}(r + \tau_0^{-1}) g_2^{(l+1)} \\ \qquad = -J_1 \cdot \nabla V + \frac{3}{2\mu_{n,0}}(T_0 \tau_0^{-1} g_1 + n_i^2 T^{(l)} r) & \text{in } \Omega, \\ J_2^{(l+1)} = \nabla g_2^{(l+1)} - \nabla V \overline{T}^{-1} g_2^{(l+1)} & \text{in } \Omega, \\ g_2^{(l+1)} = \frac{3}{2} n_D T_D \quad \text{on } \Gamma_D, \quad J_2^{(l+1)} \cdot \nu = 0 \quad \text{on } \Gamma_N. \end{cases} \tag{3.24}$$

4. Set $g_1^{(l+1)} := g_1$, $p^{(l+1)} := p$.
5. Compute $T^{(l+1)} = \frac{2}{3} g_2^{(l+1)} / g_1^{(l+1)}$, a piecewise constant approximation $\overline{T}$, and let $\delta T^{(l+1)} = \|T^{(l+1)} - T^{(l)}\|_{L^\infty}$.
6. Define $V^{(l+1)} = V + \delta V$ where δV is the solution of (3.23).

As a stopping criterion we use the L^∞-norm of two consecutive iteration functions $(g_1^{(l)}, g_2^{(l)}, T^{(l)}, p^{(l)}, V^{(l)})$.
Steps 2(a), 2(b), and 3 are performed by means to the mixed scheme described in Section 2.1. The subtle part is the definition of the drift coefficient $\boldsymbol{b}$ according to equation (2.1) on the discrete level, which we now describe. We therefore denote the finite-element approximations of $g_i^{(l+1)}, V^{(l+1)}$ by g_i^h and V^h respectively. In the iteration process an electron temperature T^h is defined via (3.14) as a function which is piecewise constant on the edges. The linearized Poisson equation in step 2(c) is solved by using a nonconforming method approximating the potential by $V^h \in CR_{h,V_D}$, with $CR_{h,\xi}$ defined in (2.38). For the discretization of the hole current equation (3.22) in step 2(b) we can directly use $\bar{\boldsymbol{b}} = \nabla V^h$ since the hole temperature is constant. This is equivalent to setting in equation (2.3) $\psi = V^h$, cf. Remark 2.1(i). In the electron particle and energy flux equations (3.21) and (3.24) of steps 2(a) and 3, we substitute the temperature by the local mean temperature $\overline{T}|_K$ which is defined by the (arithmetic) mean value of T^h on the edges $e \in \partial K$. Observing that ∇V^h is piecewise constant for $V^h \in CR_{h,V_D}$, we can define the piecewise constant drift coefficient $\boldsymbol{b} = \bar{\boldsymbol{b}} = \nabla V^h \overline{T}^{-1}$ and accordingly ψ through the equation (2.3).
After convergence of the iterative scheme, the current densities are computed using formula (2.15), originating from the mixed scheme.

3.3.1 Refinement Strategy

For the adaptive procedure we need to define an indicator for the whole system of equations. The estimator η_{CR} of Subsection 2.3.1 with its local contributions $\eta_{CR,K}$ defined in (2.43) is employed for the particle and energy flux equations. To simplify the notation we drop the index CR in this section. An estimator for the error of the Poisson equation is taken from the literature. Several estimators for elliptic problems discretized by Crouzeix-Raviart elements can be found in the literature, see, e.g., [**34, 46, 117**]. We use the estimator derived in [**117**]. More precisely, the error estimator for the electrostatic potential has the form

$$(\eta^V)^2 = \sum_{K \in \mathcal{T}_h} (\eta_K^V)^2$$

with the local contributions

$$(\eta_K^V)^2 = \Big(\frac{h_T}{\lambda}\Big)^2 \|n - p - C\|_{0;T}^2 + \sum_{i=1}^{3} \frac{1}{2}\Big(h_{e_i} \int_{e_i} [\nabla V \cdot \nu]^2 ds + \frac{\lambda^2}{h_{e_i}} \int_{e_i} [V]^2 ds\Big),$$

where $[u]$ is the jump of u across an edge e_i.
We also add the heuristic term $\eta_K(g_2^h/g_1^h, \bar{g}_2^h/\bar{g}_1^h)$ to the indicator of the system, with g_i^h and $\bar{g}_i^h$ the approximations of g_i^h in the spaces Λ_h and W_h respectively, see (2.43). The idea is to monitor directly changes of the quotient g_2^h/g_1^h. In the present situation, the quotient equals the temperature (multiplied by 3/2). In general, in particular for nonparabolic band diagrams, the temperature depends nonlinearly on that quotient; see [**49**]. We observed that the estimators for g_1^h or g_2^h alone are not able to resolve strong changes of the quotient in regions with very low densities. The error estimator that we used in our simulations reads as follows (with $\bar{g}_i^h, \bar{p}^h \in W_h$):

$$\eta_{K,ET}^2 = \eta_K^2(g_1^h, \bar{g}_1^h) + \eta_K^2(g_2^h, \bar{g}_2^h) + (\eta_K^V)^2 + \eta_K^2(g_2^h/g_1^h, \bar{g}_2^h/\bar{g}_1^h), \quad (3.25)$$

$$\eta_{K,ET\&DD}^2 = \eta_{K,ET}^2 + \eta_K^2(p^h, \bar{p}^h), \quad (3.26)$$

with $\eta_K(p^h, \bar{p}^h)$ defined through equation (2.43).
The estimator η_{ZZ} can be used as a substitute for the first two summands in (3.25) as indicated in Subsection 2.3.3. For the MOSFET device we will compare simulation results obtained under the refinement control of η_{CR} with those obtained by means of η_{ZZ}.
In the adaptive strategy we refine the triangles belonging to the set

$$RK = \{K \in \mathcal{T}_h : \eta_{K,ET\&DD} \geq 0.7 \max_{K \in \mathcal{T}_h} \eta_{K,ET\&DD}\}.$$

More precisely, the triangles of this set are first refined into four congruent triangles. Possible hanging nodes are removed by using the algorithm of Bank [**11**]. This algorithm may introduce triangles with obtuse angles. The number of these triangles are diminished by the so-called "barycentric mesh regularization" technique. The idea is to move any interior vertex P of the triangulation $\mathcal{T}_h$ to the barycenter of the set $D_P = \{K \in \mathcal{T}_h : P \in \partial K\}$ of neighboring triangles (for details see [**113**]).
The direct interpolation of the nonconforming P_1 discretization of V^h given on $\mathcal{T}_h$ might be very "oscillating" in regions where the potential has steep gradients. Therefore we construct a conforming P_1 interpolation on $\mathcal{T}_{h/2} \supset \mathcal{T}_h$. For the vertices which are the midpoints of edges in $\mathcal{T}_h$, we use the corresponding values of V^h. The values at the remaining vertices are determined by the mean value of V^h taken over the midpoints of the edges that are connected to the corresponding vertex. The variables g_1^h, g_2^h and p^h, which are defined at the midpoints of the edges, are treated analogously.
For the global iteration we assume that additionally to the data needed for the algorithm (A) of the previous subsection, the estimator η is initialized. The adaptive procedure then reads as follows:

Do

1. Perform the steps of the algorithm (A) until $\|(g_1^{(l+1)}, g_2^{(l+1)}, p^{(l+1)}, V^{(l+1)}, T^{(l+1)}) - (g_1^{(l)}, g_2^{(l)}, p^{(l)}, V^{(l)}, T^{(l)})\|_{L^\infty} < 10^{-4} \cdot \eta$;
2. Compute $\eta^2 = \sum_{K \in \mathcal{T}_h} \eta_{K,ET\&DD}$;
3. Let $RK = \{K \in \mathcal{T}_h : \eta_{K,ET\&DD} \geq 0.7 \max_{K \in \mathcal{T}_h} \eta_{K,ET\&DD}\}$;
4. Generate a new mesh $\mathcal{T}_h$ by refining all $K \in RK$ regularly, removing any hanging nodes and regularizing the mesh;
5. Generate the new vectors of nodal values by interpolating the solution from the previous mesh;

until $\eta < 10^{-5}$ or $\#K > N_{max}$.

3.4 Semiconductor Devices

In this section we present the numerical simulations for a set of device examples. The first device is used to assess the numerical convergence of the scheme, whereas the following examples represent configurations of more realistic devices. These include some specialties for the boundary data, so-called Schottky contacts and Oxide layers, which have to be explained in the respective sections.

3.4.1 A Ballistic Diode

We consider a two-dimensional diode which is uniform in one dimension. The semiconductor domain is $\Omega = (0, l_x) \times (0, l_y)$, where $l_x = 0.6\mu$m, $l_y = 0.2\mu$m, and the length of the channel equals 0.4μm. The n^+ doping regions are defined in $(0, 0.1\mu\text{m}) \times (0, l_y)$ and $(0.5\mu\text{m}, l_x) \times (0, l_y)$, the n region (or channel region) in $(0.1\mu\text{m}, 0.5\mu\text{m}) \times (0, l_y)$. The diode has two Ohmic contacts on both sides of the domain whereas the remaining boundary parts are insulating. We choose:

$$\Gamma_{D1} = \{x = 0\}, \quad \Gamma_{D2} = \{x = l_x\}, \quad \Gamma_N = \{y = 0\} \cup \{y = l_y\}.$$

On Γ_N homogeneous Neumann boundary data are imposed and on Γ_D the Dirichlet boundary data are given as follows:

$$\begin{aligned} n = C_m, \quad T = T_0, \quad V = V_{\text{el}} \qquad & \text{on } \Gamma_{D1}, \\ n = C_m, \quad T = T_0, \quad V = U + V_{\text{el}} \qquad & \text{on } \Gamma_{D2}. \end{aligned}$$

Here, the ambient temperature is $T_0 = 300\text{K}$, $U = 1.5\text{V}$ is the applied voltage, and the built-in potential V_{el} is given by

$$V_{\text{el}} = U_0 \ln(n/n_i). \tag{3.27}$$

The doping concentration in the n^+ region is $C_m = 5{\cdot}10^{17}\text{cm}^{-3}$, in the channel region we take a doping density of $2{\cdot}10^{15}\text{cm}^{-3}$. These values are the same as in [**36, 49, 92, 110**] and allow for a comparison with the results presented in these papers.
First we report the relative errors (RE) of the computed solutions in the L^2-norm in Table 3.2 and Table 3.3, using the Chen model (see Section 3.1). The mesh on which the reference solution was computed has 20480 triangles and maximal edge length $h = 1/160$. As mentioned in Remark 2.2 we compare the proposed scheme with a P_1 nonconforming treatment of the zeroth-order terms, cf. Remark 2.2. The convergence rates of the mixed discretization for T_h, n_h, $g_{2,h}$ are 1.55, 1.85, and 1.84, respectively. For a constant potential one would expect the P_1 nonconforming scheme to give better results, since there the lower-order terms have a stronger effect on the resulting matrix. For the present situation the errors do almost not differ. This shows that the convection parts play here an important role. We finally mention that similar results have been obtained using the Lyumkis model.

h	RE for Ψ_h	RE for T_h	RE for n_h	RE for $g_{2,h}$
0.1	$5.4 \cdot 10^{-3}$	0.0122	0.100	0.100
0.05	$1.9 \cdot 10^{-3}$	$4.96 \cdot 10^{-3}$	0.029	0.029
0.025	$5.1 \cdot 10^{-4}$	$1.4 \cdot 10^{-3}$	$7.7 \cdot 10^{-3}$	$7.7 \cdot 10^{-3}$

Table 3.2: Relative errors for the mixed scheme (Chen model).

h	RE for Ψ_h	RE for T_h	RE for n_h	RE for $g_{2,h}$
0.1	$5.4 \cdot 10^{-3}$	0.0122	0.100	0.100
0.05	$1.9 \cdot 10^{-3}$	$3.4 \cdot 10^{-3}$	0.029	0.029
0.025	$4.8 \cdot 10^{-4}$	$9.9 \cdot 10^{-4}$	$7.8 \cdot 10^{-3}$	$7.8 \cdot 10^{-3}$

Table 3.3: Relative errors for the P_1 nonconforming scheme (Chen model).

Now we present the numerical results for a non-uniform mesh with 900 triangles. In Figure 3.1 the (unscaled) electron temperature for the Lyumkis

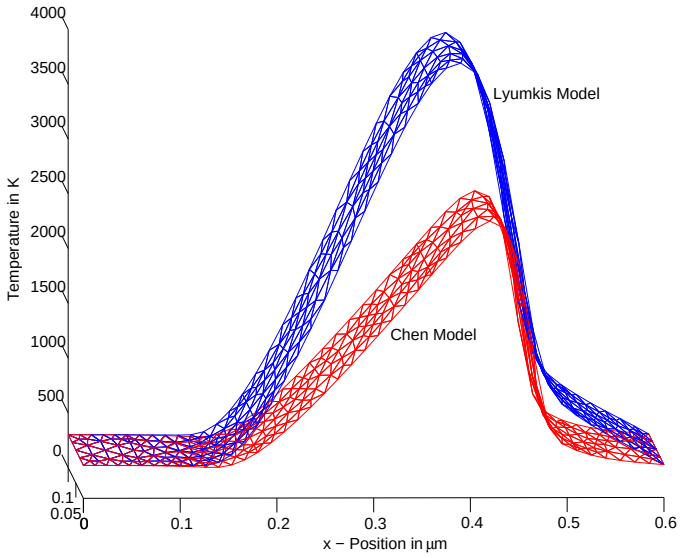

Figure 3.1: Electron temperature versus position in a ballistic diode.

and the Chen model are shown. As expected, the temperature profiles are almost uniform in one space direction. Moreover, they coincide with the values computed in [**49**].
In Figure 3.2 we show the (unscaled) electron mean velocity $u = J_1/(qn)$. Again, the profiles coincide almost with the corresponding results in [**49**]. The maximum values for the Lyumkis and the Chen model are $3.03 \cdot 10^7$cm/s and $1.44 \cdot 10^7$cm/s, respectively. In [**49**], the maximum values $2.92 \cdot 10^7$cm/s (Lyumkis model) and $1.44 \cdot 10^7$cm/s (Chen model) are reported.

3.4.2 A MESFET Device

The MESFET (metal-semiconductor field-effect transistor) device is used as a switch or amplifier [**112**]. We present two examples of this device type. For the first example we use data from the literature, see below. The second example in the next section will be the first adaptive simulation and presents the different states of a switch.
The device behavior is mainly governed by the size of the depletion region (i.e., a region with very low electron density) that develops around the Schottky contact at the gate, see Figures 3.3 and 3.9. This depletion region enlarges if the gate voltage is decreased, and therefore diminishes the channel width which leads to a reduced current for a fixed applied drain voltage

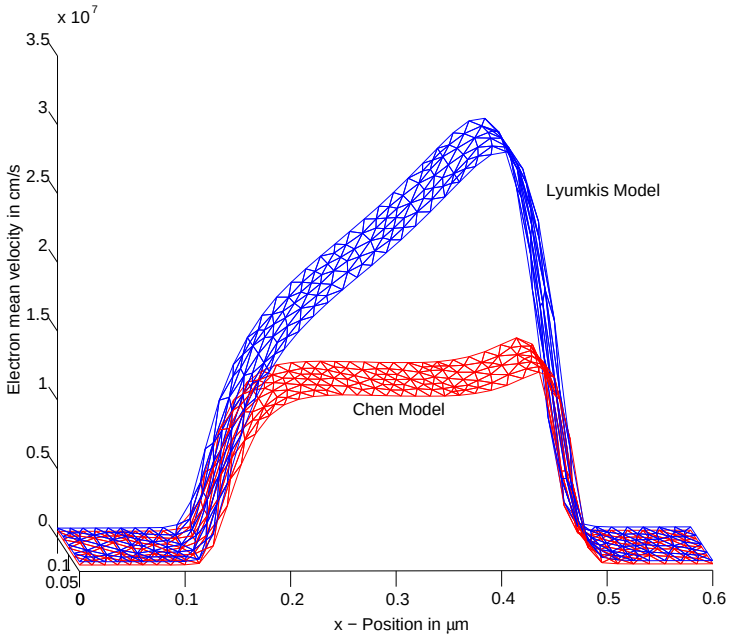

Figure 3.2: Electron mean velocity versus position in a ballistic diode.

(closed state). For larger gate voltage, the depletion region becomes smaller and a significant current can flow (open state).

The device geometry of the first example is as follows. The device consists of two high-doped n^+ regions near the Ohmic contacts (called source and drain) and an n region with a Schottky contact (called gate); see Figure 3.3. The source and drain contact lengths are 0.1μm; the gate contact length is 0.2μm. For the doping profile we use the smoothed function presented in Figure 3.4. It holds (in μm):

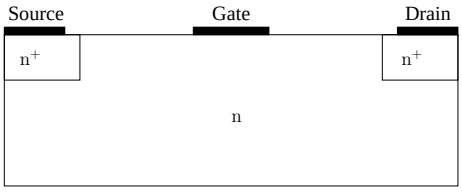

Figure 3.3: Geometry of the MESFET.

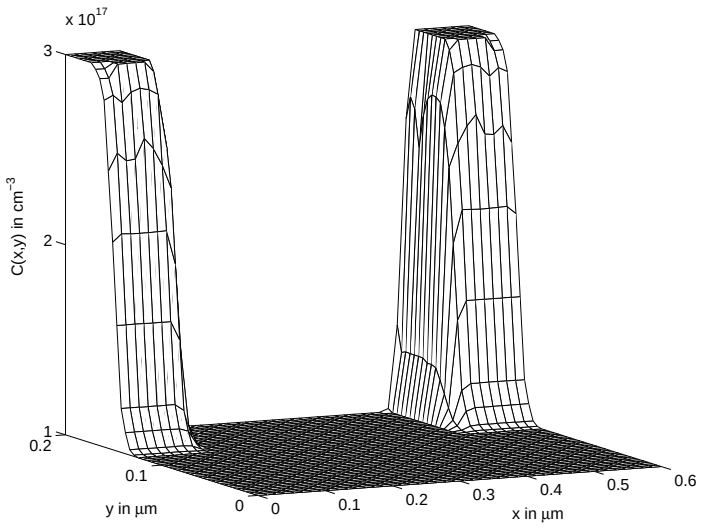

Figure 3.4: Doping profile of the MESFET.

$$C(x,y) = \begin{cases} 3 \cdot 10^{17}\mathrm{cm}^{-3} & : (x,y) \in [0,0.1] \times [0.15,0.2] \cup [0.5,0.6] \times [0.15,0.2], \\ 1 \cdot 10^{17}\mathrm{cm}^{-3} & : \text{else.} \end{cases}$$

The boundary data is given as follows (with V_{el} as defined in (3.27)):

- at the source: $n = 3 \cdot 10^{17}\mathrm{cm}^{-3}$, $T = 300\mathrm{K}$, $V = V_{\mathrm{el}}$;
- at the drain: $n = 3 \cdot 10^{17}\mathrm{cm}^{-3}$, $T = 300\mathrm{K}$, $V = V_{\mathrm{el}} + 2\mathrm{V}$;
- at the gate: $n = 3.9 \cdot 10^{5}\mathrm{cm}^{-3}$, $T = 300\mathrm{K}$, $V = V_{\mathrm{el}} - 0.8\mathrm{V}$;
- for the remaining boundary segments, homogeneous Neumann boundary conditions for J_1, J_2, and V are used.

These values are the same as in [**77, 38**], but we do not prescribe the velocity on the contact parts. The value for n at the gate contact has been computed from the formula (5.1-19) from [**108**]. We use the definition of the temperature and the energy relaxation term of the Chen model.
For this device geometry, the Gummel-type iteration presented in Section 3.3 converges very slowly. The reason may be the complex structure of the temperature profile, in particular the large gradients near the gate contact. For this reason, a full Newton scheme has been used.

The discrete version of equation (3.14) for $T = f(g_1, g_2)$ is of algebraic nature and therefore, it is different from the discrete versions of (3.9)-(3.11) since it does not stem from a variational formulation. Using it directly in the derivation of the Newton scheme leads to a badly conditioned Jacobian matrix. To avoid this we assume that T_h and f_h, given by $T_h = f_h = f(g_1^h, g_2^h)$, are a P_1 nonconforming approximation of the temperature and the right-hand side of (3.14). The "variational" formulation of the equation $T_h = f_h$, with basis functions φ_i, is as follows:

$$\int_\Omega T_{hj}\varphi_j\varphi_i dxdy = \int_\Omega f_{hj}\varphi_j\varphi_i dxdy.$$

Now, since $\int \varphi_j\varphi_i dxdy = |\mathrm{supp}(\varphi_j)|/3 \cdot \delta_{ij}$ we get

$$T_{hj}|\mathrm{supp}(\varphi_j)|/3 = f_{hj}|\mathrm{supp}(\varphi_j)|/3, \tag{3.28}$$

which is of the same order of magnitude as the differential equations. This leads especially for bad initial values or high increments in the bias continuation to much faster convergence compared to a Newton scheme that does not incorporate the scaling factor $|\mathrm{supp}(\varphi_j)|/3$ in equation (3.28).
Figures 3.5-3.8 show the computational results for a mesh with 1564 nodes. It has been a priori refined near the Dirichlet boundaries and the junctions. In Figure 3.5 the electron density is shown. The depletion area around the gate contact is clearly seen.
The electron temperature is presented in Figure 3.6. The maximal temperature is 3152K. As expected, the temperature is large near the drain contact. Near the gate, the gradient of the temperature is very large, which may indicate that the fixed boundary temperature of 300K is physically not appropriate.
In Figure 3.7 the electrostatic potential is shown. Since the electrons are moving in the direction of positive potential, we expect a significant current flow from the source to the drain. This is confirmed by Figure 3.8 where the vector of the particle current density is depicted. In fact, with the above data, the MESFET device is in an "open" state. In a "closed" state the depletion region is much larger than shown in Figure 3.5 and the current density much smaller as we will see in the next example.

3.4.3 A Double-Gate MESFET

To improve the control of the channel we add a second gate contact at the lower boundary, more precisely the device consists of two high-doped n^+

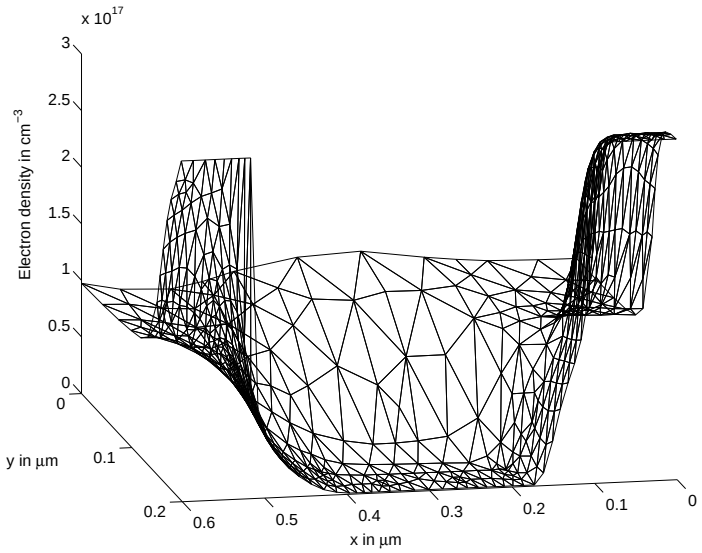

Figure 3.5: Electron density in the MESFET (Chen model).

regions near the Ohmic contacts (called source and drain) and an n region with an upper and a lower Schottky contact (called gate) in a sandwich configuration (see Figure 3.9).
The model parameters of the MESFET of size $0.6\mu\text{m}\times 0.24\mu\text{m}$ are as follows. The source and drain contact lengths are 0.24μm; the gate contact length is 0.2μm. Moreover, the length of the low-doped (channel) region is 0.36μm. The discontinuous doping profile is given by

$$C(x,y) = \begin{cases} 1\cdot 10^{17}\text{cm}^{-3} & : x \in [0.12\mu\text{m}, 0.48\mu\text{m}],\ y \in [0, 0.24\mu\text{m}] \\ 3\cdot 10^{17}\text{cm}^{-3} & : \text{else.} \end{cases}$$

At the source and drain contacts the data for the particle density are equal to the equilibrium values, see equation (3.20). Apart from the Schottky contact the boundary values are as before. At the Schottky contact the boundary data for the potential consists of the usual parts, the built-in potential and the applied voltage, but additionally the so-called Schottky barrier height is subtracted, i.e. $V|_G = V_{\text{applied}} + V_{\text{el}} - V_{\text{barrier}}$. A barrier height of 0.8V is used as a typical value for a n-type silicon/metal contact. The temperature at all contacts is equal to the ambient temperature T_0. The particle density at the Schottky gates is computed from formula (5.1-19) in [**108**]. In the following we summarize the boundary conditions:

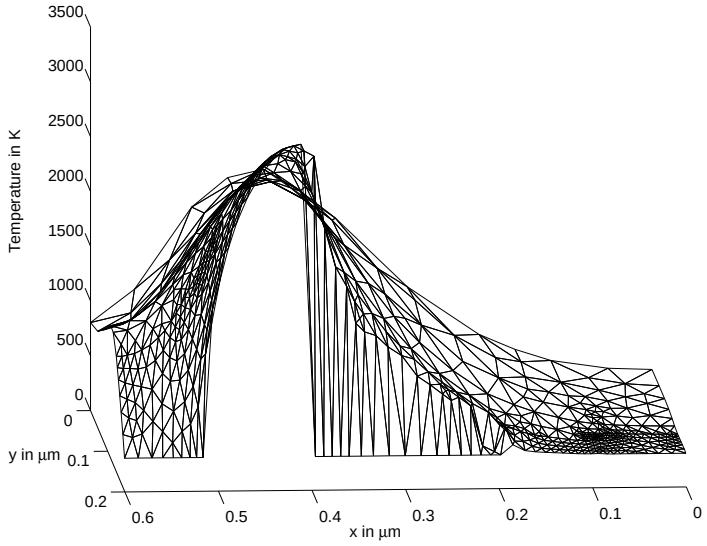

Figure 3.6: Electron temperature in the MESFET (Chen model).

- at the source: $n = 3 \cdot 10^{17}\text{cm}^{-3}$, $V = \Phi_0$;
- at the drain: $n = 3 \cdot 10^{17}\text{cm}^{-3}$, $V = \Phi_0 + 2\text{V}$;
- at the gates:

  open state: $n = 3.9 \cdot 10^{5}\text{cm}^{-3}$, $V = 0\text{V} + \Phi_0 - 0.8\text{V}$;
  close state: $n = 2.4 \cdot 10^{5}\text{cm}^{-3}$, $V = -1.2\text{V} + \Phi_0 - 0.8\text{V}$;

- for the remaining boundary segments, homogeneous Neumann boundary conditions for J_1, J_2, and V are used.

In Figure 3.10, we present the electron and the energy density in the open and closed state, respectively. In the open state we can observe clearly the channel between the depletion regions at the gate contacts, built by the electrons. As expected, the energy density is much larger in the open state than in the closed state.
The electrostatic potential and the electron temperature are depicted in Figure 3.11. Although the electron density is very low in the channel and at the drain junction, the electron temperature is much higher in the closed state since the electric field is larger here than in the open state.
In Figures 3.12 and 3.13 we present two current-voltage curves: the drain current I_D depending on the drain voltage U_D with no applied gate voltage

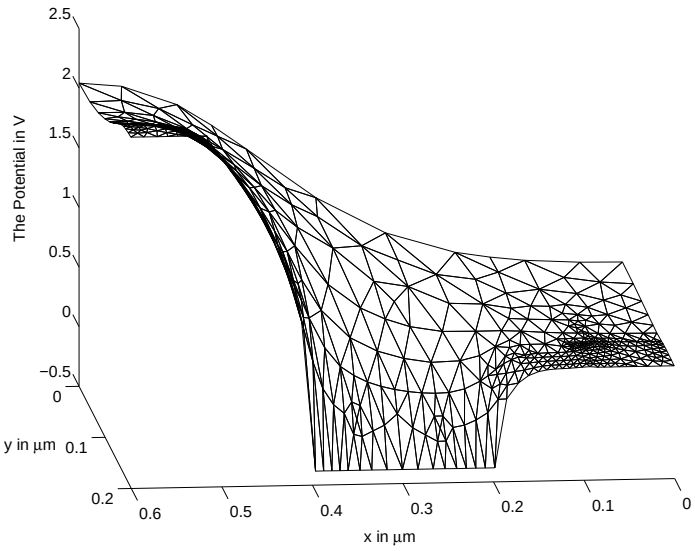

Figure 3.7: Electrostatic potential in the MESFET (Chen model).

and the drain current I_D depending on the gate voltage V_G with a drain voltage of 2V. The dependence of $I_D(V_G)$ is approximately quadratic which confirms the results mentioned in [**112**]. Moreover, it can be seen that the value of the current density is not strongly affected by the number of triangles. This situation changes when simulating a MOSFET device (see the next section).

3.4.4 A Deep Submicron MOSFET

A MOSFET (metal-oxide semiconductor field-effect transistor) device can be used as a voltage-driven switch and is the most used device in computer technology. We simulate a transistor of size 420nm×210nm with an effective channel length of 70nm and an oxide thickness of 1.5nm. The length of the source and drain contacts is 30nm (see Fig. 3.14). The doping profile is given by a step function with values 10^{19}cm^{-3} in the n^+-region and -10^{17}cm^{-3} in the p bulk region. The geometry and data of this device are adapted from the work of Cassan et al. [**35**]. Cassan et al. compare Monte-Carlo simulations of the Boltzmann transport equation with numerical results from drift-diffusion and energy-transport models. Under the assumption that direct tunneling through the oxide is the dominant mechanism producing gate currents, they develop a post processing procedure to calculate these

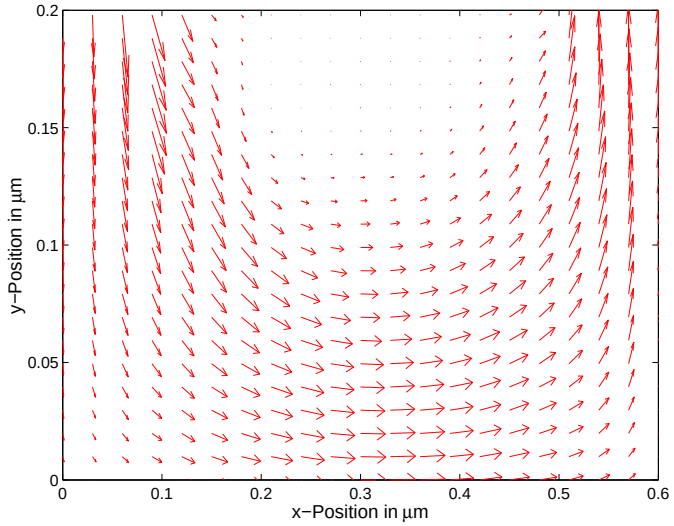

Figure 3.8: Electron current density in the MESFET (Chen model).

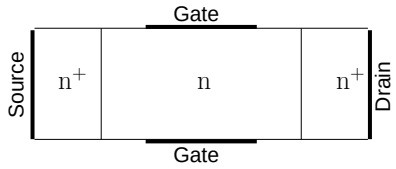

Figure 3.9: Geometry of the MESFET with two coupled gate contacts.

currents from the simulations using the energy-transport and drift-diffusion models. They observed that gate currents for n-MOSFET devices with an oxide thickness less than 2nm, calculated by post processing, are in good agreement with gate currents calculated from Monte-Carlo simulations. This justifies in some sense the use of the energy-transport model for the above MOSFET geometry.

The current-voltage characteristics of the device are mainly influenced by the electric field at the semiconductor oxide junction. To model the influence of the oxide we assume that the particles do not penetrate the oxide region. We denote the semiconductor region by Ω_S, Ω_O is the oxide region, $\Gamma_{S/O} = \partial\Omega_S \cap \partial\Omega_O$ is the silicon/silicon oxide interface, Γ_G is the gate contact part of $\partial\Omega_O$, and $\Gamma_{N,O} = \partial\Omega_O \backslash (\Gamma_G \cup \Gamma_{S/O})$ are the remaining boundary parts

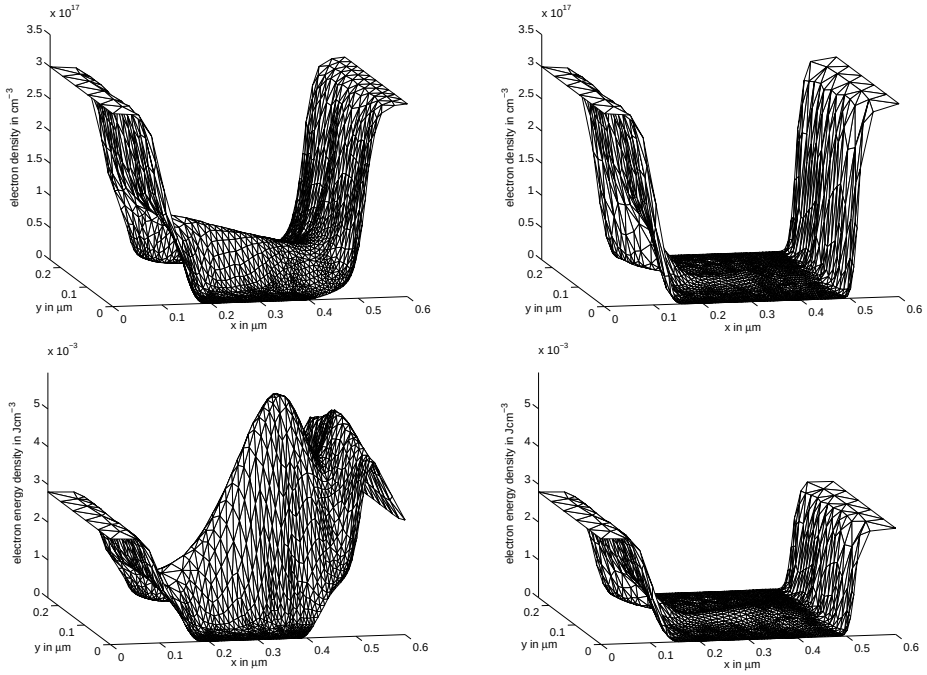

Figure 3.10: Electron density n (upper row) and energy density $\frac{3}{2}nT$ (lower row) in a double-gate MESFET (left column: open state; right column: closed state).

of Ω_O. We choose Dirichlet boundary conditions at the source, drain and bulk, i.e., the densities are set to their equilibrium values (see [**91**]). Since we assume that the particles do not penetrate the oxide, the particle and energy densities do not need to be computed in Ω_O. We impose homogeneous Neumann boundary conditions $J_\alpha \cdot \nu = 0$ for $\alpha = 1, 2, p$ on $\Gamma_{S/O}$ and on the remaining parts of the boundary of Ω_S. We solve the Poisson equation in the domain $\Omega_S \cup \Omega_O$ with a space-dependent permittivity which is constant in each Ω_i, $i = S, O$, and we specify homogeneous Neumann boundary conditions on $\Gamma_{N,O}$ and a Dirichlet condition on Γ_G.

Before we present the numerical results we show the meshes generated by our refinement strategy in Figure 3.15. The initial mesh is constructed without any knowledge of the location of the junctions. After the adaptive procedure, the final grid is refined near the junctions and near the gate oxide. The importance of the adaptive scheme becomes apparent when

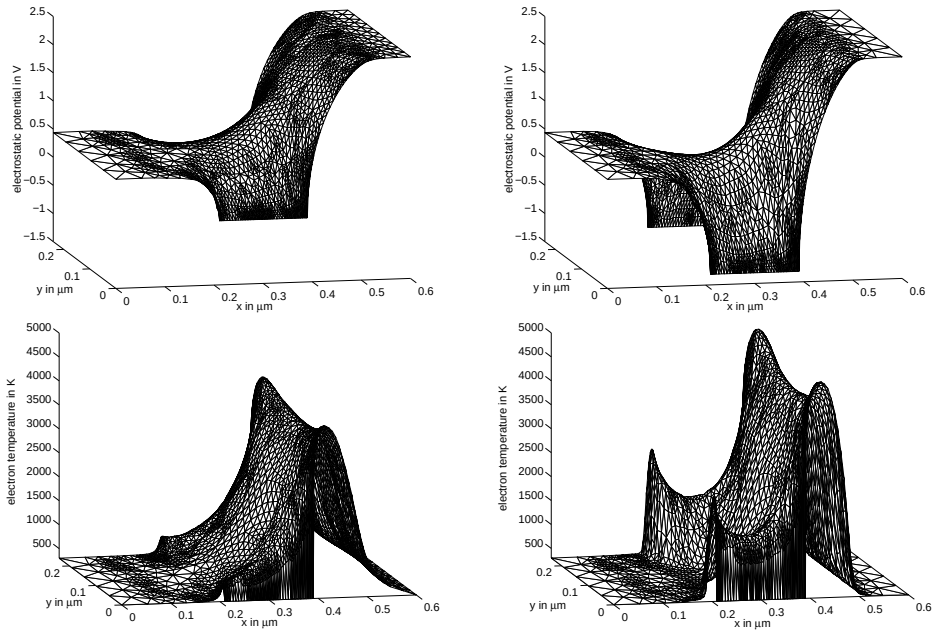

Figure 3.11: Electrostatic potential (upper row) and electron temperature (lower row) in a double-gate MESFET (left column: open state; right column: closed state).

comparing the current-voltage curves for different meshes (Figure 3.16). The characteristics seem to stabilize for meshes with about 5000 triangles for the estimator η_{CR} on the left. On the right of Figure 3.16 the estimator η_{ZZ} was used for the electron current density in the system indicator to control the refinement process. We clearly see that the IV-curves stabilize already for a smaller number of elements, and the maximum current is slightly larger for the curves computed under refinement control of η_{ZZ}. The strongly localized current flow in the MOSFET device is the main reason for this advantageous behaviour. Keeping the applied voltage fixed we see for both estimators, that the current is slowly increasing with refining the mesh.

First we present current-voltage curves of the drain current $I_D(V_D)$ depending on the drain voltage V_D for different applied gate voltages (Figure 3.17). For larger gate voltages the drain current reaches a higher level before saturation effects diminish the slope of the curves [**112**].

The influence of the electron temperature is shown in Figures 3.18 and 3.19

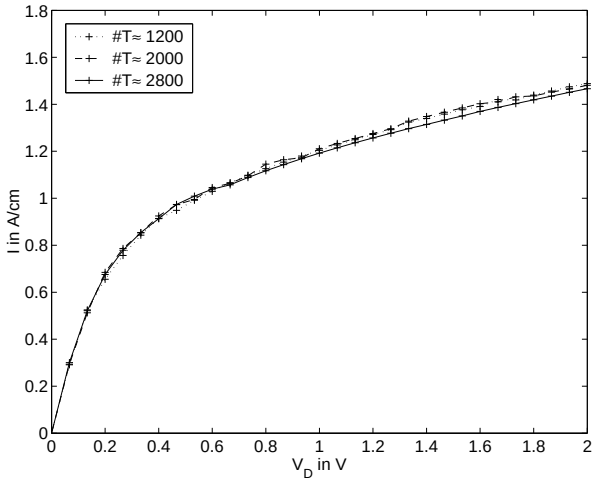

Figure 3.12: MESFET drain current as a function of the drain voltage for various meshes ($V_G = 0$V).

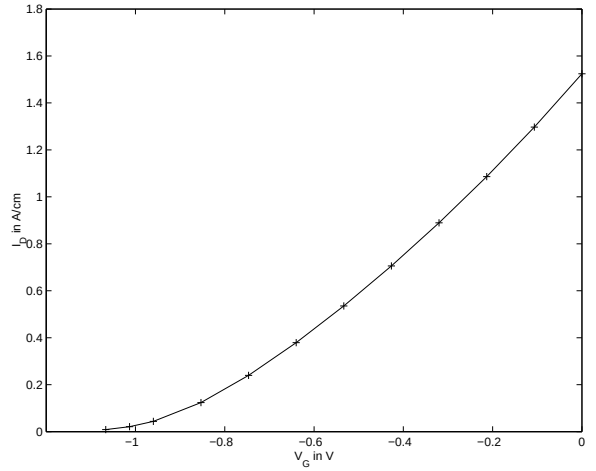

Figure 3.13: MESFET drain current as a function of the gate voltage ($V_D = 2$V).

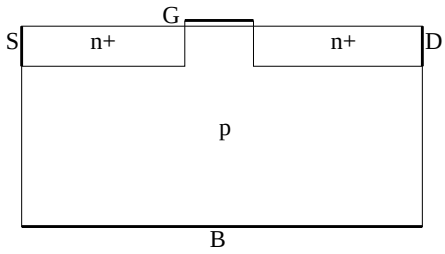

Figure 3.14: Geometry of the MOSFET with source S, drain D, gate G and bulk B contacts.

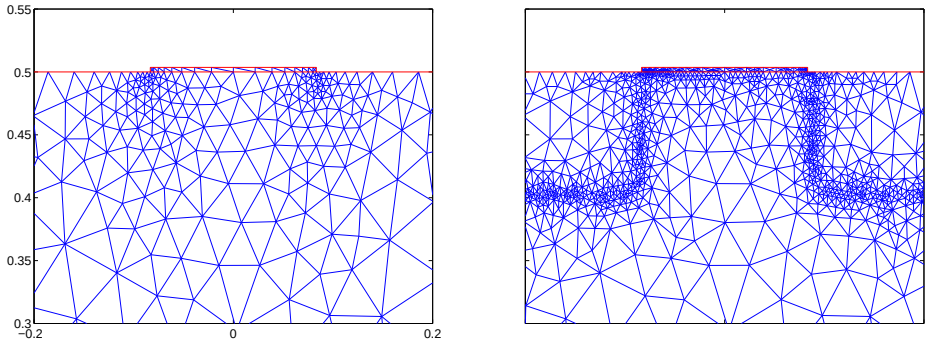

Figure 3.15: Adaptively refined triangulations for the MOSFET (zoom). Left: initial mesh (about 1000 triangles); right: final mesh (about 4000 triangles).

(using $V_D = 1\text{V}$ and $V_G = 1\text{V}$). Due to the high temperature region at the drain junction the effective electron mobility $\mu(T) = \mu_{n,0}/T$ decreases and a region with a higher electron concentration is formed in the channel. The temperature near the right end of the drain junction is larger than at the source junction since the electrons gain more energy from the electric field during their flow through the device.

Finally we want to mention two fields for further investigations, where we exclude already possible extensions to quantum fluid models. Considering for instance the simulation results for the MESFET device, the boundary layer of the temperature near the gate contact indicates that different boundary conditions may be physically more appropriate. One possibility is to employ Robin boundary conditions at the contacts, since they are second-

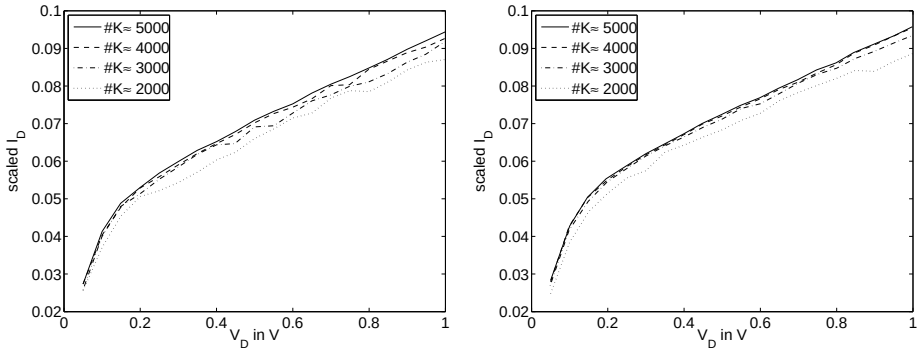

Figure 3.16: Drain current depending on the drain voltage for different meshes ($V_G = 0.95$V). For each curve the maximum element count is fixed. The refinement starts always from the same coarse mesh. On the left the refinement is derived from η_{CR} and on the right from η_{ZZ}.

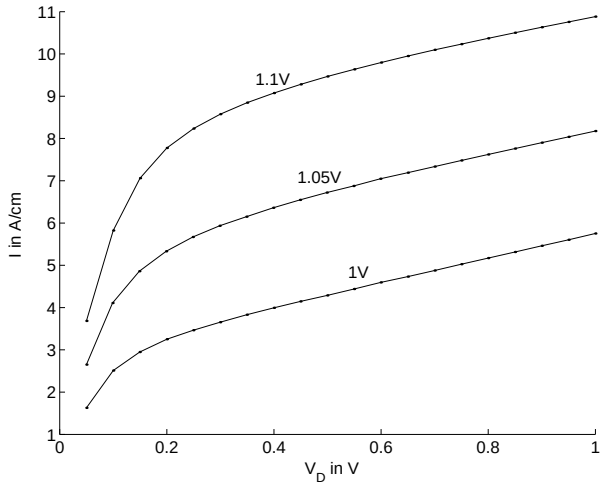

Figure 3.17: Current-voltage curves for different applied gate voltages.

order approximations of boundary conditions for the Boltzmann equation [**51**] and the energy-transport equations are itself derived from the Boltzmann equation [**17**]. Robin boundary conditions have already been used in the drift-diffusion model [**119**]. A second very interesting point for future extensions is to derive a DWR-estimator for the convection-diffusion problem and to apply it to a linearized version of the full nonlinear system. Summarizing this section we can say that numerical method is in good

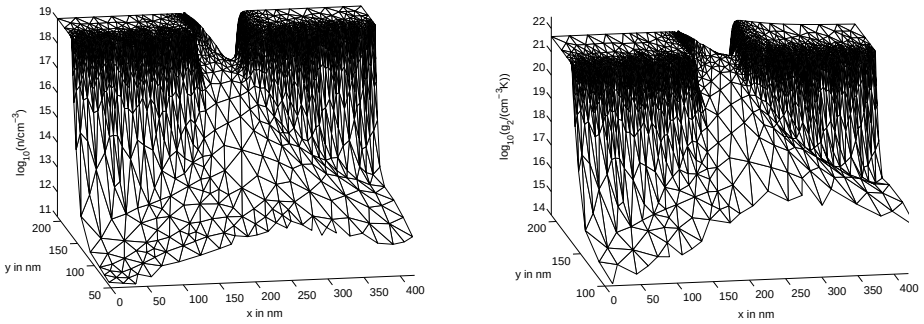

Figure 3.18: Electron density (left) and electron energy density (right) in a MOSFET with 70nm channel length (logarithmic plot). Notice that a part of the bulk region is not shown.

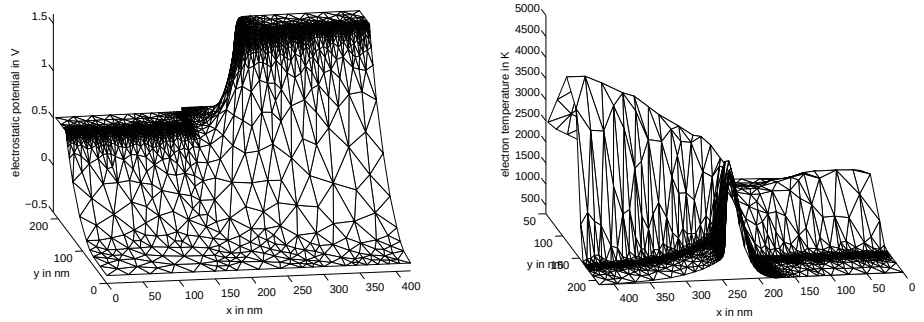

Figure 3.19: Electrostatic potential (left) and electron temperature (right) in a MOSFET with 70nm channel length.

agreement with the results available in the literature. It is very flexible in the choice of the underlying physical model as well as regarding the adaptive refinement strategy. Especially the error estimator may be configured easily by adding weighting factors in the definitions (3.25) and (3.26) to address specific simulation interests.

Bibliography

[1] D.A. Adams. *Sobolev Spaces*. Academic Press, New York, 1975.

[2] M. Ainsworth and J.T. Oden. *A Posteriori Error Estimation in Finite Element Analysis*. Wiley, 2000.

[3] W. Allegretto and H. Xie. Nonisothermal semiconductor systems. In X. Liu and D. Siegel, editors, *Comparison methods and stability theory*, volume 162 of *Lecture Notes in Pure and Applied Mathematics*, New York, 1994. Marcel Dekker.

[4] L. Angermann and S. Wang. Three-dimensional exponentially fitted conforming tetrahedral finite elements for the semiconductor continuity equations. *Appl. Numer. Math.* 46:19–43, 2003.

[5] Y. Apanovich, P. Blakey, R. Cottle, E. Lyumkis, B. Polsky, A. Shur, and A. Tcherniaev. Numerical simulations of submicrometer devices including coupled nonlocal transport and nonisothermal effects. *IEEE Trans. El. Dev.* 42:890–897, 1995.

[6] D. Arnold and F. Brezzi. Mixed and nonconforming finite element methods: implementation, postprocessing and error estimates. *RAIRO Modél. Math. Anal. Numér.* 19:7–32, 1977.

[7] G. Baccarani and M. Wordemann. An investigation on steady-state velocity overshoot in silicon. *Solid-State Electr.* 29:970–977, 1982.

[8] C. Baiocchi and A. Capelo. Variational and Quasivariational Inequalities. *Wiley*, 1984.

[9] R.E. Bank, D.J. Rose, and W. Fichtner. Numerical methods for semiconductor device simulation. *IEEE Trans. Electr. Dev.* ED-30: 1031–1041, 1983.

[10] R.E. Bank, J.F. Burgler, W. Fichtner, and R.K. Smith.Some upwinding techniques for finite element approximation of convection–diffusion equations. *Numer. Math.* 58:185–202, 1990.

[11] R.E. Bank, A.H. Sherman, and A. Weiser. Refinement algorithms and data structures for regular local mesh refinement. In R. Steplemen et al., editors, *Scientific Computing*, pages 3–17, Amsterdam, 1983. IMACS/North Holland.

[12] S. Bartels and C. Carstensen. Each averaging technique yields reliable a posteriori error control in FEM on unstructured grids. II. Higher order FEM. *Math. Comp.* 71, 239:971–994, 2002.

[13] R. Becker and R. Rannacher. A feed-back approach to error control in finite element methods: Basic analysis and examples. *East-West J. Numer. Math.* 4,4: 237–264, 1996.

[14] R. Becker and R. Rannacher. Weighted a posteriori error control in FE methods. In H. Bock et al., editors, *Proc. of ENUMATH 97*, pages 621–637, Singapore, 1998, World Scientific.

[15] R. Becker, R. Rannacher. An optimal control approach to a posteriori error estimation in finite element methods. In: A. Iserles (Ed.), *Acta Numerica 2001*, pages 1–102, Cambridge. 2001, Cambridge University Press.

[16] N. Ben Abdallah, P. Degond, and S. Génieys. An energy transport model for semiconductors derived from the Boltzmann equation. *J. Stat. Phys.* 84,1-2:205–231, 1996.

[17] N. Ben Abdallah and P. Degond. On a hierarchy of macroscopic models for semiconductors. *J. Math. Phys.* 37:3308–3333, 1996.

[18] K. Bløtekjær. Transport equations for electrons in two-valley semiconductors. *IEEE Trans. El. Dev.* 17:38–47, 1970.

[19] F. Bornemann, D. Lauri, S. Wagon, J. Waldvogel. The SIAM 100-Digit Challenge (A Study in High-Accuracy Numerical Computing). *SIAM*, 2004.

[20] F. Bosisio, R. Sacco, F. Saleri, and E. Gatti. Exponentially fitted mixed finite volumes for energy balance models in semiconductor device simulation. In H. Bock et al., editors, *Proc. of ENUMATH 97*, pages 188–197, Singapore, 1998. World Scientific.

[21] D. Braess, R. Verfürth. A posteriori error estimators for the Raviart-Thomas element. *SIAM J. Numer. Anal.* 33, 6:2431-2444, 1996.

[22] S. Brenner and R.L. Scott. *The Mathematical Theory of Finite Element Methods.* Springer, 1994.

[23] F. Brezzi and M. Fortin. *Mixed and Hybrid Finite Element Methods* Springer, 1991.

[24] F. Brezzi, L. Marini, S. Micheletti, P. Pietra, R. Sacco, and S. Wang. Discretization of Semiconductor Device Problems (I). *Handbook of Numerical Analysis,Vol. XIII* (Numerical Methods for Electrodynamical Problems), W.H.A. Schilders and E.J.W. Maten, eds., Elsevier Science, 2005.

[25] F. Brezzi, L.D. Marini, P. Pietra. Two-dimensional exponential fitting and applications to drift-diffusion models. *SIAM J. Numer. Anal.* 26, 6:1342–1355, 1989.

[26] F. Brezzi, L.D. Marini, and P. Pietra. Numerical simulation of semiconductor devices. *Comp. Meth. Appl. Mech Eng.* 75,1-3:493–514, 1989.

[27] F. Brezzi and A. Russo. Choosing bubbles for advection diffusion problems. *Math. Models Meth. Appl. Sci.* 4:571–587, 1994.

[28] C. Carstensen. A posteriori error estimate for the mixed finite element method. *Math. Comput.* 66,218: 465–476, 1997.

[29] C. Carstensen. All first-order averaging techniques for a posteriori finite element error control on unstructured grids are efficient and reliable. *Math. Comp* 73, 247: 1153-1165, 2004.

[30] C. Carstensen. A unifying theory of a posteriori finite element error control. Electronic publication in *Numer. Math.* (DOI) 10.1007/s00211-004-0577-y, 2005.

[31] C. Carstensen and S. Bartels. Each averaging technique yields reliable a posteriori error control in FEM on unstructured grids. I. Low order conforming, nonconforming, and mixed FEM. *Math. Comp.* 71, 239:945–969, 2002.

[32] C. Carstensen and S. Funken: Averaging technique for FE - a posteriori error control in elasticity. Part I: Conforming FEM. *Comput. Methods Appl. Mech. Engrg.* 190:2483–2498, 2001. Part II: -independent estimates. *Comput. Methods Appl. Mech. Engrg.* 190:4663–4675, 2001. Part III: Locking-free nonconforming FEM. *Comput. Methods Appl. Mech. Engrg.* 191, 8-10:861–877, 2001.

[33] C. Carstensen and S. Funken. A posteriori error control in low-order finite element discretisations of incompressible stationary flow problems. *Math. Comp.* 70:1353–1381, 2001.

[34] C. Carstensen and S. Jansche. An a posteriori estimate for nonconforming finite element methods. *Z. Angew. Math. Mech.* 78:S871–S872, 1998.

[35] E. Cassan, S. Galdin, P. Dollfus, and P. Hesto. Comparison between device simulators for gate current calculation in ultra-thin gate oxide n-MOSFETs. *IEEE Trans. Elect. Dev.* 83:1194–1202, 2000.

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

[37] D. Chen, E. Sangiorgi, M. Pinto, E. Kan, U. Ravaioli, and R. Dutton. Analysis of spurious velocity overshoot in hydrodynamic simulations. *NUPAD IV*, pages 109–114, 1992.

[38] Z. Chen, B. Cockburn, J. Jerome, and C. Shu. Finite element computation of the hydrodynamic model of semiconductor devices. *VLSI Design*, 3:145–158, 1995.

[39] L. Hsiao and L. Chen. The Solution of Lyumkis Energy Transport Model in Semiconductor Science. *Math. Method Appl. Sci.* 26:1421–1433, 2003.

[40] L. Chen and Y. Li. Global Existence and Asymptotic Behavior to the solution of 1-D Energy Transport Model for Semiconductors. *J. Partial Diff. Eqs.* 15:81-95, 2002

[41] I. Choquet, P. Degond, and C. Schmeiser. Energy–transport models for carrier involving impact ionization in semiconductors. *Transport Theory Statist. Phys.* 32,2:99–132, 2003.

[42] P.G. Ciarlet. *The Finite Element Method for Elliptic Problems.* North-Holland, Amsterdam (1978).

[43] B. Cockburn, J. Gopalakrishnan. A characterization of hybridized mixed methods for second order elliptic problems. *SIAM J. Numer. Anal.* , 42,1:283–301, 2004.

[44] B. Cockburn, J. Gopalakrishnan. Error analysis of variable degree mixed methods for elliptic problems via hybridization. *Math. Comp.* 74:1653–1677, 2005.

[45] M. Crouzeix and P.A. Raviart. Conforming and nonconforming finite element methods for solving the stationary Stokes equation. *RAIRO*, 7:33–76, 1973.

[46] E. Dari, R. Duran, C. Padra, and V. Vampa. A posteriori error estimation for nonconforming finite element methods. *Math. Mod. Num. Anal.* 30:385–400, 1996.

[47] P. Degond, S. Génieys, and A. Jüngel. A system of parabolic equations in nonequilibrium thermodynamics including thermal and electrical effects. *J. Math. Pures Appl.* 76:991–1015, 1997.

[48] P. Degond, S. Génieys, and A. Jüngel. A steady-state system in nonequilibrium thermodynamics including thermal and electrical effects. *Math. Meth. Appl. Sci.* 21:1399–1413, 1998.

[49] P. Degond, A. Jüngel, and P. Pietra. Numerical discretization of energy-transport model for semiconductors with non-parabolic band structure. *SIAM J. Sci. Comp.* 22:986–1007, 2000.

[50] P. Degond, C.D. Levermore, and C. Schmeiser. A note on the energy-transport limit of the semiconductor Boltzmann equation. *Proc. of Transport in transition regimes (Minneapolis, MN, 2000)*, IMA Vol. Math. Appl., 135, Springer, 2004.

[51] P. Degond and C. Schmeiser. Kinetic boundary layers and fluid-kinetic coupling in semiconductors. *Transp. Theory Stat. Phys.* 28:31–55, 1999.

[52] A. Ern and J.-L. Guermond. Theory and Practice of Finite Elements. *Springer*, 2004.

[53] K. Eriksson, D. Estep, P. Hansbo, and C. Johnson. Introduction to adaptive methods for differential equations. *Acta Numerica*, 4:105-158, 1995

[54] W. Fang and K. Ito. Existence of stationary solutions to an energy drift-diffusion model for semiconductor devices. *Math. Models Meth. Appl. Sci.* 11:827–840, 2001.

[55] H. Federer. Geometric Measure Theory. *Springer*, 1969.

[56] M.V. Fischetti and S.E. Laux. DAMOCLES Theoretical Manual. IBM, Yorktown Heights, 1994.

[57] A. Forghieri, R. Guerrieri, P. Ciampolini, A. Gnudi, M. Rudan, and G. Baccarani. A new discretization strategy of the semiconductor equations comprising momentum and energy balance. *IEEE Trans. Comp. Aided Design Integr. Circuits Sys.* 7:231–242, 1988.

[58] M. Fournié. *Construction et analyse de schémas compacts d'ordre élevé pour des problèmes fortement convectifs. Application à la simulation de semiconducteurs.* PhD thesis, Université Paul Sabatier, Toulouse, France, 1999.

[59] M. Fournié. Numerical discretization of energy-transport model for semiconductors using high-order compact schemes. *Appl. Math. Lett.* 15:727–734, 2002.

[60] C. Gardner, J. Jerome, and D. Rose. Numerical methods for the hydrodynamic device model: subsonic flow. *IEEE Trans. CAD*, 8:501–507, 1989.

[61] C. Gardner, P. Lanzkorn, and D. Rose. A parallel block iterative method for the hydrodynamic device model. *IEEE Trans. Comp. Aided Design Integr. Circuits Sys.* 10:1187–1192, 1991.

[62] I. Gasser, and R. Natalini. The energy-transport and the drift-diffusion equations as relaxation limits of the hydrodynamic model for semiconductors. *Quart. Appl. Math.* 57:269–282, 1999.

[63] P. Gérard, P. Markowich, N. Mauser, F. Poupaud. Homogenization Limits and Wigner Transforms. *Comm. Pure Appl. Math.* 50:323–379, 1997.

[64] D. Gilbarg, N.S. Trudinger. Elliptic partial differential equations of second order. *Springer*, 2001.

[65] J. Gopalakrishnan. A Schwartz Preconditioner for a hybridized mixed method. *Comp. Meth. Appl. Math.* 3,1:116–134, 2003.

[66] P. Grisvard. Elliptic Problems in Nonsmooth Domains *Pitman*, 1985.

[67] P. Grisvard. Singularities in boundary value problems. *Springer*, 1992.

[68] J. Griepentrog. An application of the implicit function theorem to an energy model of the semiconductor theory. *Z. Angew. Math. Mech.* 79:43–51, 1999.

[69] H.K. Gummel. A Self-Contained Iterative Scheme for One-Dimensional Steady-State Transistor Calculations. *IEEE Trans. El. Dev.,* ED-11:455–465, 1964.

[70] P. Heres and W. Schilders. Reduced order modelling of RLC-networks using an SVD-Laguerre based method. In W. Schilders et al., eds, *Proc. of the SCEE 2002 Conf.* Eindhoven, Springer, 2002.

[71] M. Hinze, R. Pinnau. An Optimal Control Approach to Semiconductor Design. *Math. Mod. Meth. Appl. Sc.* 12,1:89–107, 2002.

[72] R. Hoppe and B. Wohlmuth. Element-oriented and edge-oriented local error estimators for nonconforming finite element methods. *Mod. Math. Anal. Num.* 30:237–263, 1996.

[73] R. Hoppe and B. Wohlmuth. A comparison of a posteriori error estimators for mixed finite element discretization by Raviart-Thomas elements. *Math. Comp.* 68:1347–1378, 1999.

[74] P. Houston, R. Rannacher, and E. Süli. A posteriori error analysis for stabilised finite element approximations of transport problems. *Comp. Meth. Appl. Mech. Eng.* 190:1483–1508, 2000.

[75] P. Houston and E. Süli. A note on the design of *hp*-adaptive finite element methods for elliptic partial differential equations. *Comp. Meth. Appl. Mech. Eng.* 190:229–243, 2005.

[76] J. Jerome. *Analysis of charge transport. A mathematical study of semiconductor devices*. Springer, Berlin, 1996.

[77] J. Jerome and C.-W. Shu. Energy models for one-carrier transport in semiconductor devices. In W. Coughran, J. Colde, P. Lloyd, and J. White, eds, *Semiconductors, Part II*, volume 59 of *IMA Volumes in Mathematics and its Applications*, pages 185–207, New York, 1994. Springer.

[78] J. Jerome and C.-W. Shu. Energy transport systems for semiconductors: Analysis and simulation. In V. Lakshmikantham, editor, *Proc. of the First World Congress of Nonlinear Analysts*, pages 3835–3846, Berlin, 1996. Walter de Gruyter.

[79] A. Jüngel. *Quasi-hydrodynamic Semiconductor Equations*. Progress in Nonlinear Differential Equations. Birkhäuser, Basel, 2001.

[80] A. Jüngel and P. Pietra. A discretization scheme of a quasi-hydrodynamic semiconductor model. *Math. Models Meth. Appl. Sci.* 7:935–955, 1997.

[81] A. Jüngel and S. Tang. A relaxation scheme for the hydrodynamic equations for semiconductors. *Appl. Num. Math.* 43:229–252, 2002.

[82] E. Kane. Band structure of indium-antimonide. *J. Phys. Chem. Solids*, 1:249–261, 1957.

[83] T. Kerkhoven. Mathematical modelling of quantum wires in periodic hetero junction structures. In W. Coughran, J. Colde, P. Lloyd, and J. White,eds. *Semiconductors Part II*, volume 59 of *IMA Volumes in Mathematics and its Applications*, pages 237-253. , New York, 1994, Springer.

[84] T. Kerkhoven and Y. Saad. On acceleration methods for coupled nonlinear elliptic systems. *Numer. Math.* 60:525–548, 1992.

[85] C. Lab and P. Caussignac. An energy-transport model for semiconductor heterostructure devices: application to AlGaAs/GaAs MODFETs. *Compel*, 18:61–76, 1999.

[86] M. Lundstrom. Fundamentals of Carrier Transport. 2nd edition, *Cambridge University Press*, 2000.

[87] E. Lyumkis, B. Polsky, A. Shur, and P. Visocky. Transient semiconductor device simulation including energy balance equation. *Compel*, 11:311–325, 1992.

[88] J. Malý, D. Swanson, W. Ziemer. The Coarea formula for Sobolev mappings. *Trans. Amer. Math. Soc.* 355:477-492, 2003.

[89] L. D. Marini and P. Pietra. An abstract theory for mixed approximations of second order elliptic equations. *Mat. Aplic. Comp.* 8:219–239, 1989.

[90] L. D. Marini and P. Pietra. New mixed finite element schemes for current continuity equations. *COMPEL*, 9:257–268, 1990.

[91] P. A. Markowich, C. A. Ringhofer, and C. Schmeiser. *Semiconductor Equations*. Springer, 1990.

[92] A. Marrocco, P. Montarnal, and B. Perthame. Simulation of the energy-transport and simplified hydrodynamic models for semiconductor devices using mixed finite elements. In *Proc. of ECCOMAS 96*, Wiley, 1996.

[93] J. J. Miller and S. Wang. A new non-conforming Petrov-Galerkin finite-element method with triangular elements for a singularly perturbated advection-diffusion problem. *IMA J. Numer. Anal.* 14:257–276, 1996.

[94] K. W. Morton. *Numerical solution of convection-diffusion problems.* Applied Mathematics and Mathematical Computation 12, Chapman & Hall, 1996.

[95] S. Nicaise. *Polynomial Interface Problems.* Peter Lang, Frankfurt am Main, 1993.

[96] R.R.P. van Nooyen. A Petrov-Galerkin mixed finite element method with exponential fitting. *Numer. Meth. Part. Differ. Eq.* 11,5:501–524, 1995.

[97] F. Poupaud. Diffusion approximation of the linear semiconductor Boltzmann equation. *J. on Asympt. Anal.* 4:293–317, 1991.

[98] F. Poupaud and C. Ringhofer. Semi-classical limits in a crystal with exterior potentials and effective mass theorems. *Commun. Partial Differ. Equations* 21,No. 11-12:1897-1918, 1996.

[99] P. Raviart and J. Thomas. A mixed finite element method for second order elliptic equations. In *Mathematical Aspects of the Finite Element Method*, volume 606 of *Lecture Notes in Math.* pages 292–315. Springer, 1977.

[100] L. Reggiani. Hot-Electron Transport in Semiconductors. *Springer*, 1985.

[101] C. Ringhofer. An entropy-based finite difference method for the energy transport system. *Math. Models Meth. Appl. Sci.* 11:769–796, 2001.

[102] W. Van Roosbroeck. Theory of flow of electron and holes in germanium and other semiconductors. *Bell Syst. Techn. J.* 29:560–607, 1950.

[103] M. Rudan, A. Gnudi, and W. Quade. A generalized approach to the hydrodynamic model of semiconductor equations. In G. Baccarani, editor, *Process and Device Modeling for Microelectronics*, Amsterdam, 1993. Elsevier.

[104] R. Sacco and F. Saleri. Stabilization of mixed finite elements for convection-diffusion problems. *CWI Quarterly*, Vol 10:301-315, 1997.

[105] D. Scharfetter and H.K. Gummel. Large-signal analysis of a silicon read diode oscillator. *IEEE Trans. Elec. Dev.* ED-16:64–77, 1969.

[106] C. Schmeiser and A. Zwirchmayr. Elastic and drift-diffusion limits of electron-phonon interaction in semiconductors. *Math. Models Meth. Appl. Sci.* 8:37–53, 1998.

[107] C. Schwab. *p− and hp− Finite Element Methods.* Theory and Applications to Solid and Fluid Mechanics, Oxford University Press, 1998.

[108] S. Selberherr. *Analysis and Simulation of Semiconductor Devices.* Springer, 1984.

[109] M. Selva Soto and C. Tischendorf. Numerical Analysis of DAEs from Coupled Circuit and Semiconductor Simulation. *Appl. Numer. Math.* 53,2-4:471–488, 2005.

[110] K. Souissi, F. Odeh, H. Tang, and A. Gnudi. Comparative studies of hydrodynamic and energy transport models. *COMPEL*, 13:439–453, 1994.

[111] R. Stratton. Diffusion of hot and cold electrons in semiconductor barriers. *Phys. Rev.* 126:2002–2014, 1962.

[112] S. Sze. *Physics of Semiconductor Devices.* John Wiley, New York, 1981.

[113] R. Verfürth. *A Review of a Posteriori Error Estimation and Adaptive Mesh Refinement Techniques.* Wiley, Teubner, Germany, 1996.

[114] P. Visocky. A method for transient semiconductor device simulation using hot-electron transport equations. In J. Miller, editor, *Proc. of the Nasecode X Conf.* Dublin, 1994. Boole Press.

[115] W. Walus. Computational methods for the Boltzmann equation. In N. Bellomo, editor, *Lecture Notes on the Mathematical Theory of the Boltzmann Equation*, pages 179–223. World Scientific, Singapore, 1995.

[116] S. Wang and L. Angermann. On convergence of the exponentially fitted finite volume method with an anisotropic mesh refinement for a singularly perturbed convection-diffusion equation. *Comp. Meth. Appl. Math.* 3:493–512, 2003.

[117] B. Wohlmuth. A residual based error estimator for mortar finite element discretizations. *Numer. Math.* 84:143–171, 1999.

[118] J. Xu and L. Zikatanov. A monotone finite element method for convection–diffusion equations. *Math. Comp.* 68,228:149–1446, 1999.

[119] A. Yamnahakki. Second order boundary conditions for the drift-diffusion equations of semiconductors. *Math. Models Meth. Appl. Sci.* 5:429–455, 1995.

[120] O.C. Zienkiewich and J.Z. Zhu. A simple error estimator and adaptive procedure for practical engineering analysis. *Int. J. Numer. Methods Eng.* 24:337–357, 1987.

Curriculum Vitae

Name:	Stefan Holst
Born:	9.6.1973 in Berlin - Germany
Nationality:	German

Education

1979-85	Alfred-Brehm-Primary School
1985-92	Carl-Friedrich-v.-Siemens-Grammar School
1990	Participation at the "Mathematical Seminar for pupils" at the FU Berlin
1992	Finishing grammar school with "A-Level", main subjects: Mathematics, Physics, English and Political Science
1993-99	Study of Applied Mathematics at the TU Berlin
11/5/1999	Degree with distinction in applied mathematics (Dipl. Math. techn.) at the TU Berlin

Professional Experience

07/1996-11/1999	student assistant for software-development at the computer science faculty of the TU Berlin
05/1999-10/1999	student assistant for teaching at the math faculty of the TU Berlin
11/1999-1/9/2000	research associate at the computer science faculty of the TU Berlin for finalizing a running project
1/2000-10/2002	research associate at the math faculty of the University of Konstanz
since 11/2002	research associate at the math faculty of the University of Mainz