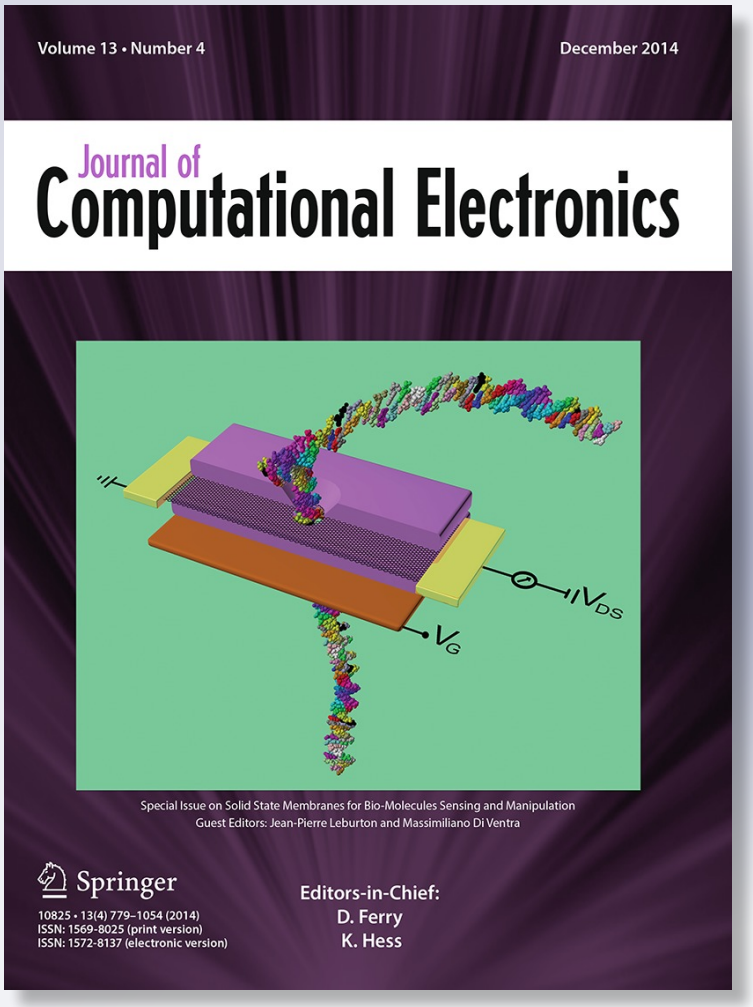
Hierarchies of transport equations for nanopores

Clemens Heitzinger & Christian Ringhofer

Journal of Computational Electronics

ISSN 1569-8025
Volume 13
Number 4

J Comput Electron (2014) 13:801-817
DOI 10.1007/s10825-014-0586-8



Springer

Your article is protected by copyright and all rights are held exclusively by Springer Science +Business Media New York. This e-offprint is for personal use only and shall not be self-archived in electronic repositories. If you wish to self-archive your article, please use the accepted manuscript version for posting on your own website. You may further deposit the accepted manuscript version in any repository, provided it is only made publicly available 12 months after official publication or later and provided acknowledgement is given to the original source of publication and a link is inserted to the published article on Springer's website. The link must be accompanied by the following text: "The final publication is available at link.springer.com".

Author's personal copy

J Comput Electron (2014) 13:801–817
DOI 10.1007/s10825-014-0586-8

Hierarchies of transport equations for nanopores

Equations derived from the Boltzmann equation and the modeling of confined structures

Clemens Heitzinger · Christian Ringhofer

Published online: 5 June 2014
© Springer Science+Business Media New York 2014

Abstract We review transport equations and their usage for the modeling and simulation of nanopores. First, the significance of nanopores and the experimental progress in this area are summarized. Then the starting point of all classical and semiclassical considerations is the Boltzmann transport equation as the most general transport equation. The derivation of the drift-diffusion equations from the Boltzmann equation is reviewed as well as the derivation of the Navier–Stokes equations. Nanopores can also be viewed as a special case of a confined structure and hence as giving rise to a multiscale problem, and therefore we review the derivation of a transport equation from the Boltzmann equation for such confined structures. Finally, the state of the art in the simulation of nanopores is summarized.

Keywords Boltzmann equation · Model hierarchy · Drift-diffusion-Poisson system · Navier–Stokes equation · Confined structure · Nanopore

1 Introduction

Artificial nanopores have received lots of attention in recent years. They are an important building block for the control of single molecules and DNA oligomers. Following the principle of the Coulter counter for counting and sizing particles suspended in electrolytes, translocation of molecules through a nanopore or binding events inside a functionalized nanopore block the ionic currents through the nanopore (see Fig. 1). This principle makes it possible to detect single molecules. Furthermore, it has been applied to DNA sequencing, which has attracted lots of attention; an overview of the history so far can be found in [1].

The appeal of these ideas is that they yield a direct detection method, i.e., the molecules are detected directly and not indirectly after being marked, which is a great advantage compared to currently employed technology. Of course, despite the great experimental progress in recent years, technical hurdles still have to be overcome. Section 2 gives an overview of some of the inspiring experiments performed recently and discusses the problems that arise when modeling this young technology.

Regarding the theoretic aspects of understanding nanopores, the starting point is the Boltzmann equation. A short introduction to the Boltzmann equation as (one of) the most general transport equations is given in Sect. 3. The drift-diffusion equations and the Navier-Stokes equations are derived from the Boltzmann equation in Sect. 4. In Sect. 5, the derivation of a transport equation for confined structures such as nanopores and ion channels is summarized.

After these modeling considerations, we summarize simulations based on these basic model equations in Sect. 6. The emphasis here is on the recent works based on advanced models. Finally, the conclusions in Sect. 7 contain an outlook of what can be expected in the future regarding the mathematical modeling and the numerical simulation of nanopores. As nanopores are a young technology, theoretic understanding is lagging behind experimental progress. Quantitative understanding, however, is important for the rational design of nanopore sensors.

C. Heitzinger (✉) · C. Ringhofer
School of Mathematical and Statistical Sciences, Arizona State University (ASU), Phoenix, AZ, USA
e-mail: Clemens.heitzinger@asu.edu

C. Ringhofer
e-mail: Ringhofer@asu.edu

C. Heitzinger
Institute for Analysis and Scientific Computing, Vienna University of Technology (TU Vienna), Wien, Austria

Author's personal copy

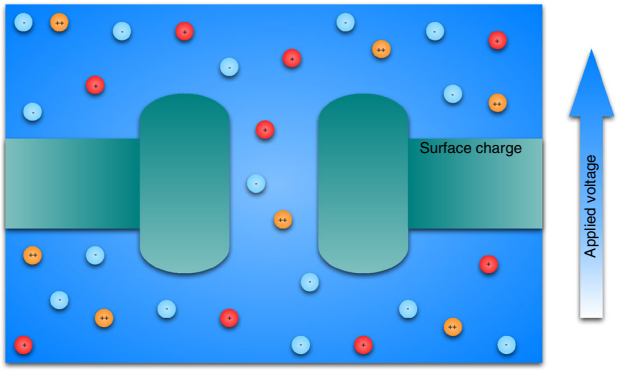


Fig. 1 Schematic diagram (not to scale) of a nanopore. Three different kinds of charge carriers (carrying $+2$, $+1$, or -1 elementary charges) are shown here. The nanopore may generally have a more irregular shape

2 Overview of experiments and modeling issues

2.1 Experimental progress

In recent years, it has become possible to manufacture single pores with diameters in the nanometer range in a reproducible manner. Although it was possible to manufacture such small pores before by various methods, the earlier experiments were not reproducible. However, well-defined diameters and sizes and reproducible fabrication are a prerequisite for well-defined functional devices that can act as sensors. Artificial nanopores are now being used in prototypes for next-generation DNA sequencing and for single-molecule detection.

A schematic diagram of a nanopore is shown in Fig. 1. In general, there are multiple species of molecules (ions or larger molecules) that translocate the nanopore due to the applied potential. A defining feature of a nanopore is that its small diameter gives rise to a multiscale problem, as particle transport in the longitudinal direction has characteristics different from transport in the transversal direction because of the confinement (see Sect. 5).

Overviews of the development of experiments in this field can be found in [1–3]. A major recent trend is the focus on the fabrication of identical and well-defined nanopores.

Generally speaking, there are four combinations: One can fabricate artificial nanopores or natural nanopores (ion channels) in artificial membranes (plastic, silicon) or in natural membranes (lipid bilayers). Solid-state nanopores are very stable, but cannot be fabricated repeatedly having the exact same geometric shape. On the other hand, all α-hemolysin pores—a protein—have the exact same geometric shape, but the lipid bilayer they are embedded in is unstable. In order to overcome the problem of lacking stability, a hybrid nanopore has been presented in [4] and reproducible DNA-origami structures are also been developed [5,6]. Because of this steady experimental progress, nanopores are taking stage as one of a few main experimental tools for controlling and investigating single molecules and DNA strands.

2.2 Complications in modeling nanopores

Due to the importance of nanopores for controlling the translocation of molecules and thus acting as a technological building block, the question of their quantitative understanding arises. As this is a new technology, its theoretical understanding is still in its infancy and simulations are complicated by several facts and considerations:

- In measurements, ionic currents are recorded. The molecular-dynamics approach, i.e., calculating the movement of each atom, is still—despite continuous advances—much too computationally expensive to calculate currents. The advantage of continuum models is that they yield the currents, i.e., the measured quantities, immediately.
- In all experiments, there are multiple ionic species that translocate the nanopore driven by an applied voltage and/or by the background medium.
- The particle species interact with one another. They also interact with the background medium that generally translocates the nanopore as well.
- Therefore a self-consistent model of the system leads to systems of equations and hence to a computationally expensive numerical problem.
- The inside of a nanopore is a highly confined structure, giving rising to a multiscale problem. In most nanopores, the length (i.e., the longitudinal direction) is much larger than the width (i.e., the transversal direction).
- The correct physical description in the form of appropriate systems of equations is non-trivial, as the equations must be coupled correctly. Many transport equations can be derived from the Boltzmann transport equation as the most general transport equation for particles, but at this point it is unknown how to derive the correct systems of equations from the Boltzmann transport equation for multiple species of particles.

For these reasons, only continuum models, i.e., (systems of) partial differential equations, are discussed here.

3 The Boltzmann equation

We start by considering an ensemble of M particles. We define the probability density $f(\mathbf{x}, \mathbf{p}, t)$ as the probability that at time t particle j is at position $x_j(t)$ and has momentum $p_j(t)$. Here $\mathbf{x} := (x_1, \ldots, x_M)$ and $\mathbf{p} := (p_1, \ldots, p_M)$. Along the particle trajectories the density f does not change

Author's personal copy

with time and therefore the total derivative vanishes, i.e.,

$$\frac{\mathrm{d}}{\mathrm{d}t} f(\mathbf{x}, \mathbf{p}, t) = 0,$$

which yields the *Liouville equation*

$$\partial_t f + \frac{1}{m}\mathbf{p} \cdot \nabla_{\mathbf{x}} f + q\mathbf{E} \cdot \nabla_{\mathbf{p}} f = 0.$$

Here $\mathbf{E} := (E_1, \ldots, E_M)$ is the field acting on each particle and we assume that $\mathbf{E}$ is a gradient field, i.e., $\mathbf{E} = -\nabla_{\mathbf{x}} V$.

Using the Hamiltonian

$$H(\mathbf{x}, \mathbf{p}) := \frac{1}{2m}|\mathbf{p}|^2 + qV(\mathbf{x}, t)$$

and the Poisson bracket

$$\{g, h\} := \nabla_{\mathbf{x}} g \cdot \nabla_{\mathbf{p}} h - \nabla_{\mathbf{p}} g \cdot \nabla_{\mathbf{x}} h,$$

we can write the Liouville equation simply as

$$\partial_t f + \{f, H\} = 0.$$

The probability density f in the Liouville equation is a function of $6M + 1$ variables and therefore not amenable to computations. In order to reduce the number of variables and thus to arrive at a feasible model, we make further assumptions that lead to the *BBGKY hierarchy*, named after Bogoliubov, Born, Green, Kirkwood, and Yvon. The idea is to replace the Liouville equation by an effective one-particle equation via a mean-field approximation. We assume that the field $\mathbf{E}$ is given by an external field E_{ext} and the sum of two-particle interactions, i.e., the field exerted on the jth particle is

$$E_j(\mathbf{x}, t) = E_{\text{ext}}(x_j, t) + \sum_{k \neq j} E_{\text{int}}(x_j, x_k), \tag{1}$$

where $E_{\text{int}}(x_j, x_k)$ gives the interaction of two particles via, e.g., the Coulomb force.

This effective one-particle equation is the *Vlasov equation*

$$\partial_t F + v \cdot \nabla_x F + \frac{q}{m} E_{\text{eff}} \cdot \nabla_v F = 0 \tag{2}$$

and can be derived from the Liouville equation. Here the effective field is

$$E_{\text{eff}}(x, t) := E_{\text{ext}}(x, t) + \int_{\mathbb{R}^3} n(x^*, t) E_{\text{int}}(x, x^*) \mathrm{d}x^*, \tag{3}$$

and $F(x, v, t)$ is the particle number density in phase space, i.e., the number of particles per unit volume in an infinitesimal neighborhood of (x, v) at time t. The number density n in position space is defined as

$$n(x, t) := \int_{\mathbb{R}^3} F(x, v, t) \mathrm{d}v$$

and it is interpreted as the number of particles per unit volume in an infinitesimal neighborhood of x at time t.

Importantly, the Vlasov Eq. (2) has the form of a single-particle Liouville equation. Many-body physics enter the model only via the effective field E_{eff} after the mean-field approximation. Equation (2) has the characteristics

$$\dot{x} = v, \tag{4a}$$

$$\dot{v} = \frac{q}{m} E_{\text{eff}}(x, t) \tag{4b}$$

and they are the trajectories of the particles moving in the field E_{eff}.

The Vlasov equation is a macroscopic description of the motion of many-particle systems under the assumption of weak interactions caused by a long-range force. This implies that it does not account for scattering of particles due to strong short-range forces. For systems where scattering is important, it is only a useful model on a time scale much shorter than the mean time between consecutive scattering or collision events.

It is important to note that the Vlasov equation only accounts for long-range interactions. Short-range interactions such as collisions between particles and collisions of a particle with the background environment are neglected. Collisions are events where particles are instantaneously scattered from one state into another so that their velocity vector (and consequently their momentum and wave vectors) change extremely fast, while the position changes only slowly.

We now derive a generalization of the Vlasov equation that includes long-range interactions and a statistical account of scattering events. The following derivation was first made by Ludwig Boltzmann in 1872 where he considered equilibrium phenomena in dilute gases. We start by noting that the rate of change of the number density $F(x, v, t)$ of the ensemble due to the convection caused by the effective field E_{eff} vanishes along the characteristics (4) when collisions are neglected, i.e.,

$$\left(\frac{\partial F}{\partial t}\right)_{\text{conv}} = 0.$$

When collisions are included, we postulate that the rate of change of the number density F due to convection and the rate of change of F due to collisions are equal, i.e.,

$$\left(\frac{\partial F}{\partial t}\right)_{\text{conv}} = \left(\frac{\partial F}{\partial t}\right)_{\text{coll}}.$$

Author's personal copy

This yields

$$\partial_t F + v \cdot \nabla_x F + \frac{q}{m} E_{\text{eff}} \cdot \nabla_v F = \left(\frac{\partial F}{\partial t}\right)_{\text{coll}},$$

where E_{eff} is given by (3).

We now model and define the *scattering rate* s. The rate

$$P(x, v' \mapsto v, t)$$

is the rate that a particle with position x changes its velocity v' to v at time t due to a scattering event. It is assumed that it is proportional to the occupation probability $F(x, v', t)$ of the state (x, v') at time t. Furthermore, also to account for the Pauli principle, it is assumed to be proportional to $1 - F(x, v, t)$, which is the probability that the target state (x, v) is unoccupied at time t. This yields the model

$$P(x, v' \mapsto v, t) = s(x, v', v) F(x, v', t)\big(1 - F(x, v, t)\big), \tag{5}$$

where s is the scattering rate. More precisely, $s(x, v', v)\mathrm{d}v'$ is the transition rate that a particle with position x changes its velocity v' in the volume element $\mathrm{d}v'$, i.e., around v', to the velocity v. The model of the scattering rate obviously depends on the physical situation under consideration and an abundance of scattering mechanisms are known.

Using this model for the rate P, we can now give an expression for the collisions $\left(\frac{\partial F}{\partial t}\right)_{\text{coll}}$. The rate of change of the number density $F(x, v, t)$ due to collisions is given by the sum of the rates of particles being scattering from all possible states (x, v') into the target state (x, v) at time t minus the sum of the rates of the particles being scattering from the state (x, v) into any possible state (x, v') at the same time. This yields

$$\left(\frac{\partial F}{\partial t}\right)_{\text{coll}} (x, v, t) = \int_{\mathbb{R}^3} P(x, v' \mapsto v, t) - P(x, v \mapsto v', t)\mathrm{d}v'.$$

Inserting (5) into this expression and defining

$$Q(F) := \left(\frac{\partial F}{\partial t}\right)_{\text{coll}}$$

yields

$$Q(F)(x, v, t) = \int_{\mathbb{R}^3} s(x, v', v) F'(1 - F) - s(x, v, v') F(1 - F')\mathrm{d}v',$$

where $F := F(x, v, t)$ and $F' := F'(x, v', t)$. The operator Q is called the *collision operator* and $Q(F)$ the *collision integral*.

In summary, we find the *Boltzmann equation*

$$\partial_t F + v \cdot \nabla_x F + \frac{q}{m} E_{\text{eff}} \cdot \nabla_v F = Q(F) \tag{6}$$

for all $x \in \mathbb{R}^3$, for all $v \in \mathbb{R}^3$, and for all times $t > 0$, where E_{eff} is given by (3). When the Coulomb force is used as the model for the long-range interactions via the Poisson equation, the resulting system of equations is called the *Boltzmann–Poisson system*. The self-consistent field causes a nonlinearity and the collision integral $Q(F)$ introduces another quadratic nonlinearity which is nonlocal in the velocity direction. A good reference for the Boltzmann equation is [7].

4 Macroscopic approximations to the Boltzmann equation

Macroscopic (fluid dynamic type) equations can be derived from the Boltzmann Eq. (6) in regimes where the collision operator Q is dominant. We write the Boltzmann equation for such regimes as

$$\partial_t F(x, v, t) + v \cdot \nabla_x F + \frac{q}{m} E_{\text{eff}} \cdot \nabla_v F = \frac{1}{\varepsilon} Q(F), \qquad 0 < \varepsilon \ll 1. \tag{7}$$

Here the small dimensionless parameter ε denotes the mean free path between collisions relative to the length scale under consideration, i.e., collision events are very frequent for $\varepsilon \ll 1$. The key to finding equations for macroscopic densities are quantities which are conserved in the collision process, so called *collision invariants*.

We have to distinguish between elastic collisions between particles of the same species conserving mass, momentum, and energy on the one hand and inelastic collisions, where momentum and energy are exchanged with a background, on the other hand. Thus we write the collision operator $Q(F)$ as

$$Q = Q^{\text{el}} + Q^{\text{inel}},$$

where the elastic collision operator Q^{el} conserves mass, momentum and energy, i.e.,

$$\int \left(1, mv, m\frac{|v|^2}{2}\right) Q^{\text{el}}(F)\mathrm{d}v = (0, 0, 0)$$

holds for all density functions F, and the inelastic operator Q^{inel} only satisfies

Author's personal copy

$$\int Q^{\text{inel}}(F)\mathrm{d}v = 0 \qquad \forall F.$$

Accordingly, we consider the macroscopic (fluid) densities (ρ, u, θ) for mass, velocity, and temperature given by

$$\begin{aligned}\rho(x,t) &= \int F(x,v,t)\mathrm{d}v,\\ \rho u(x,t) &= \int vF(x,v,t)\mathrm{d}v,\\ \left(\frac{3}{2}k_B\theta + \frac{m}{2}|u|^2\right)\rho(x,t) &= \int \frac{m}{2}|v|^2F(x,v,t)\mathrm{d}v.\end{aligned}$$

Integrating the Boltzmann Eq. (7) against the corresponding powers of the velocity vector v gives the balance laws for particle density, momentum, and energy of the form

$$\begin{aligned}&\partial_t\begin{pmatrix}\rho\\ m\rho u\\ \frac{\rho}{2}(m|u|^2+3k_B\theta)\end{pmatrix} + \begin{pmatrix}\nabla_x\cdot(\rho u)\\ \nabla_x\cdot\Phi_p\\ \nabla_x\cdot\varphi_e\end{pmatrix} - \begin{pmatrix}0\\ q\rho E\\ 2q\rho E\cdot u\end{pmatrix}\\ &= \frac{1}{\varepsilon}\begin{pmatrix}0\\ \int mvQ^{\text{inel}}\mathrm{d}v\\ \int \frac{m|v|^2}{2}Q^{\text{inel}}\mathrm{d}v\end{pmatrix}.\end{aligned} \tag{8}$$

Equation (8) is not closed, since it still contains the higher moments of the transport operator given by the 3×3 matrix $\Phi_p = \int mvv^\top F\mathrm{d}v$ and the vector $\varphi_e = \int \frac{m}{2}|v|^2vF\mathrm{d}v$.

These closure terms have to be expressed in terms of the primary variables (ρ, u, θ) via asymptotic expansions of the solution of the kinetic Eq. (7) in the small parameter ε. The procedure consists of a functional expansion of the form

$$F(x,v,t) = \psi_\varepsilon(v,\rho(x,t),u(x,t),\theta(x,t)),$$

where the shape function ψ_ε is expanded into

$$\psi_\varepsilon = \psi_0 + \varepsilon\psi_1 + \cdots.$$

This leads in zeroth order to the so called *hydrodynamic equations* and in first order to the *Navier–Stokes equations*.

In the absence of inelastic collisions, this has been carried out originally by Maxwell [8]. More modern derivations for different collision operators can be found in the books [7,9]. The results are closure terms of the form

$$\begin{aligned}\Phi_p &= m\rho uu^\top - \rho P,\\ \varphi_e &= \rho ue - \rho Pu + \varepsilon\rho S,\end{aligned}$$

with the energy e, the energy stress tensor P, and the heat flux S given by

$$e := \frac{3}{2}k_B\theta + \frac{m}{2}|u|^2, \tag{9a}$$

$$P := -\left(k_B\theta + \frac{2\varepsilon m\kappa}{3}\nabla_x\cdot u\right)I + \varepsilon m\kappa(\nabla_x u^\top + (\nabla_x u^\top)^\top), \tag{9b}$$

$$S := -\nu\nabla_x(k_B\theta). \tag{9c}$$

Here I is the 3×3 identity matrix and $^\top$ denotes the transpose of a vector or matrix. κ is the viscosity and ν is the heat conduction in the fluid. The terms on the right side of (8), due to inelastic collisions with the background, depend on the relative size of the elastic and inelastic collision mechanisms.

It has been shown in [10,11] that they can be approximated by relaxation terms of the form

$$\left(\frac{1}{\varepsilon}\int mvQ^{\text{inel}}\mathrm{d}v, \frac{1}{\varepsilon}\int\frac{m|v|^2}{2}Q^{\text{inel}}\right) = \left(-\frac{m}{\tau_p}\rho u, -\frac{m}{\tau_e}\rho e\right)$$

in an appropriate regime. This gives the Navier-Stokes type system

$$\begin{aligned}&\partial_t\begin{pmatrix}\rho\\ m\rho u\\ \rho e\end{pmatrix} + \begin{pmatrix}\nabla_x\cdot(\rho u)\\ \nabla_x\cdot(m\rho uu^\top - \rho P)\\ \nabla_x\cdot(\rho ue - \rho Pu + \varepsilon\rho S)\end{pmatrix} - \begin{pmatrix}0\\ q\rho E\\ 2q\rho E\cdot u\end{pmatrix}\\ &= -\begin{pmatrix}0\\ \frac{m}{\tau_p}\rho u\\ \frac{1}{\tau_e}\rho e\end{pmatrix},\end{aligned} \tag{10}$$

with the energy e, the stress tensor P, and the heat flux S given by (9).

Various approximations to the system (10) are important and are discussed in the following.

The hydrodynamic equations. Disregarding the $O(\varepsilon)$ terms in (10) and (9) gives the (hyperbolic) hydrodynamic system with $P = -k_B\theta I$, $\nabla_x\cdot(\rho P) = -k_B\nabla_x(\rho\theta)$, and $\rho eu - \rho Pu = \rho(\frac{5}{2}k_B\theta + \frac{m}{2}|u|^2)u$, which results in the hydrodynamic system

$$\begin{aligned}&\partial_t\begin{pmatrix}\rho\\ m\rho u\\ \rho e\end{pmatrix} + \begin{pmatrix}\nabla_x\cdot(\rho u)\\ \nabla_x\cdot(m\rho uu^\top) + k_B\nabla_x(\rho\theta)\\ \nabla_x\cdot(\rho(\frac{5}{2}k_B\theta + \frac{m}{2}|u|^2)u)\end{pmatrix}\\ &- \begin{pmatrix}0\\ q\rho E\\ 2q\rho E\cdot u\end{pmatrix} = -\begin{pmatrix}0\\ \frac{m}{\tau_p}\rho u\\ \frac{1}{\tau_e}\rho e\end{pmatrix}.\end{aligned} \tag{11}$$

The energy transport system. Assuming small velocities, i.e., $u = O(\varepsilon)$, considering much larger time scales, i.e., $t_1 = t/\varepsilon$ and $\partial_t = \varepsilon\partial_{t_1}$, replaces the momentum balance in (11) in the limit $\varepsilon \to 0$ by the constitutive relation

Author's personal copy

$k_B\nabla_x(\rho\theta) - q\rho E = -\frac{m}{\tau_p}\rho u$ and yields the energy transport system

$$\partial_{t_1}\begin{pmatrix}\rho\\ \frac{3}{2}\rho k_B\theta\end{pmatrix} + \begin{pmatrix}\nabla_x\cdot(\rho u)\\ \nabla_x\cdot(\frac{5}{2}k_B\theta\rho u)\end{pmatrix} - \begin{pmatrix}0\\ 2q\rho E\cdot u\end{pmatrix} = -\begin{pmatrix}0\\ \frac{3}{2\tau_e}\rho k_B\theta\end{pmatrix}, \tag{12a}$$

$$\rho u = \frac{\tau_p}{m}\big(q\rho E - k_B\nabla_x(\rho\theta)\big). \tag{12b}$$

The drift-diffusion equations. Finally, the drift-diffusion equations, which are widely used for conventional semiconductor device simulation, are obtained by neglecting fluctuations in the temperature θ, i.e., by setting θ to a (given) constant and solving only the particle conservation equation

$$\partial_{t_1}\rho + \nabla_x\cdot(\mu q\rho E - D\nabla_x\rho) = 0 \tag{13}$$

with

$$\mu = \frac{\tau_p}{m}, \qquad D = \frac{\tau_p}{m}k_B\theta.$$

From a practical point of view, the mobilities μ and the temperature θ can usually be measured, and the relaxation times τ_p and τ_e have to be inferred from measurements.

5 A transport equation for confined structures derived from the Boltzmann equation

In this section, a transport equation for confined structures is derived. Throughout [12], the calculations were performed using dimensionless variables and the theoretical feasibility of this approach was demonstrated. Here we extend the derivation of the transport equation to variables with units. These calculations yield the expressions for the fluxes in the resulting diffusion-type equation that can be applied to real-world problems. At the same time, the present calculations provide a verification of the correctness of the derivation by showing that all the physical units are consistent.

The starting point is the 3D Boltzmann equation with a suitable scattering operator and a suitable scaling of the confining potential. The scaling is chosen so that transport and scattering occur in the longitudinal direction, while the particles are confined in the two transversal directions. The result is a diffusion-type equation for the concentration and equations for the fluxes as functions of position in the longitudinal direction and local energy.

The transport coefficients depend on the geometry that is given by an arbitrary harmonic confinement potential. An important feature of the resulting transport equation is that its coefficients are given explicitly. This has the important computational consequence that the six position and momentum dimensions of the original 3D Boltzmann equation are reduced to a 2D problem.

The applications are any geometrically complex structures, where transport occurs in a longitudinal direction and confinement in two transversal directions. Of course, this includes nanopores and ion channels.

5.1 The Boltzmann equation as the starting point

The starting point is the Boltzmann transport equation in the form

$$\partial_t f + \{\mathcal{E}, f\}_{XP} + \mathcal{Q}[f] = 0, \tag{14}$$

where the Poisson bracket, or commutator, is defined as

$$\{g, f\}_{XP} := \nabla_P g\cdot\nabla_X f - \nabla_X g\cdot\nabla_P f. \tag{15}$$

Here $f(X, P, t)$ is the kinetic particle density, $X \in \mathbb{R}^3$ is position, $P \in \mathbb{R}^3$ is momentum, t is time, $\mathcal{E}(X, P)$ is the energy, and $\mathcal{Q}$ is the scattering operator. The energy is given by

$$\mathcal{E}(X, P) := V(X) + \frac{|P|^2}{2m}.$$

Three-dimensional structures confined in two dimensions with one-dimensional transport are modeled here. Therefore we split position X and momentum P into

$$X = (x, y) = (x, y_1, y_2),$$
$$P = (p, q) = (p, q_1, q_2),$$

where x is the longitudinal direction of charge transport and y_1 and y_2 are the two transverse directions of confinement. Accordingly, p is the momentum in the longitudinal direction and q_1 and q_2 are the momenta in the transverse directions. The potential V determines the geometry of the structure and it has the form

$$V(x, y) = V_0(x) + V_1(x, y).$$

We also split the energy $\mathcal{E}$ into two contributions from the longitudinal and transverse directions, i.e.,

$$\mathcal{E}(X, P) = \mathcal{E}_x(x, p) + \mathcal{E}_y(x, y, q),$$
$$\mathcal{E}_x(x, p) := V_0(x) + \frac{|p|^2}{2m},$$
$$\mathcal{E}_y(x, y, q) := V_1(x, y) + \frac{|q|^2}{2m}.$$

Author's personal copy

The units of these variables are

$$
\begin{aligned}
&[t] := \mathrm{s},\\
&[X_i] := \mathrm{m} \quad \forall i \in \{1,2,3\},\\
&[x] = \mathrm{m},\\
&[y_i] = \mathrm{m} \quad \forall i \in \{1,2\},\\
&[P_i] := \frac{\mathrm{kg}\cdot\mathrm{m}}{\mathrm{s}} \quad \forall i \in \{1,2,3\},\\
&[p] = [v] = \frac{\mathrm{kg}\cdot\mathrm{m}}{\mathrm{s}},\\
&[q_i] = [w_i] := \frac{\mathrm{kg}\cdot\mathrm{m}}{\mathrm{s}} \quad \forall i \in \{1,2\},\\
&[\mathcal{E}] = \mathrm{J} = \frac{\mathrm{kg}\cdot\mathrm{m}^2}{\mathrm{s}^2},\\
&[f] := \frac{1}{[X_1][X_2][X_3]}\frac{1}{[P_1][P_2][P_3]} = \frac{\mathrm{s}^3}{\mathrm{kg}^3\cdot\mathrm{m}^6},\\
&[\partial_t f] = \frac{\mathrm{s}^2}{\mathrm{kg}^3\cdot\mathrm{m}^6}.
\end{aligned}
$$

Therefore the Poisson bracket has the units

$$
[\{\mathcal{E}, f\}_{xp}] = \frac{[\mathcal{E}][f]}{[X_1][P_1]} = \frac{\mathrm{s}^2}{\mathrm{kg}^3\cdot\mathrm{m}^6}.
$$

The scattering operator $\mathcal{Q}$ is a relaxation operator defined as

$$
\mathcal{Q}[f](x,y,p,q,t) := \frac{1}{\tau}\left(f - M(p)\frac{u_f(x,\mathcal{E}_y(x,y,q),t)}{N(x,\mathcal{E}_y(x,y,q))}\right), \tag{16}
$$

where $M(p)$ is a Maxwellian distribution of the form

$$
M(p) := c\exp\left(-\frac{|p|^2}{2mkT}\right)
$$

such that

$$
\int_{\mathbb{R}^3} M(p)\mathrm{d}p = 1
$$

holds. Here τ is the relaxation time, c is a constant, m is the particle mass, k is the Boltzmann constant, and T is the temperature. Furthermore, N is the density of states

$$
N(x,\eta) := \int \delta(\mathcal{E}_y(x,y,q) - \eta)\mathrm{d}yq \tag{17}
$$

and u_f is the function

$$
u_f(x,\eta,t) := \int \delta(\mathcal{E}_y(x,y,q) - \eta) f(x,y,p,q,t)\mathrm{d}ypq
$$

defined such that the scattering operator $\mathcal{Q}$ conserves the transverse energy.

The units of these variables and constants are

$$
\begin{aligned}
&[\tau] = \mathrm{s},\\
&[M] = \frac{1}{[p]} = \frac{\mathrm{s}}{\mathrm{kg}\cdot\mathrm{m}},\\
&[c] = \frac{1}{[p]} = \frac{\mathrm{s}}{\mathrm{kg}\cdot\mathrm{m}},\\
&[N] = \frac{[y_1][y_2][q_1][q_2]}{[\mathcal{E}]} = \mathrm{kg}\cdot\mathrm{m}^2,\\
&[u_f] = \frac{[f][y_1][y_2][p][q_1][q_2]}{[\mathcal{E}]} = \frac{\mathrm{s}^2}{\mathrm{kg}\cdot\mathrm{m}^3},\\
&[Q(f)] = \frac{[f]}{[\tau]} = \frac{\mathrm{s}^2}{\mathrm{kg}^3\cdot\mathrm{m}^6}
\end{aligned}
$$

by noting that $[\delta(f)] = 1/[f]$. The units in (16) are consistent iff $[f] = [Mu_f/N]$, which is easily verified.

In [12, Sect. 2.2], all variables were scaled and transformed into a dimensionless formulation. Here, however, we only scale the confinement direction y and the time t by setting

$$
y_s := \frac{y}{\epsilon}, \qquad t_s := \epsilon t, \qquad [\epsilon] := 1.
$$

As usual, we simplify notation by using the same variable names as before the scaling. In order to be consistent with the notation in [12], we set $v := p$ and $w := q$, but note that v and w denote momenta.

Using this notation, the Boltzmann equation becomes

$$
\partial_t f + \{\mathcal{E}_x + \mathcal{E}_y, f\}_{xp} + \{\mathcal{E}_y, f\}_{yq} + \mathcal{Q}[f] = 0. \tag{18}
$$

The units in this equation are consistent, since

$$
[\{\mathcal{E}_x + \mathcal{E}_y, f\}_{xp}] = [\{\mathcal{E}_y, f\}_{yq}] = \frac{[\mathcal{E}][f]}{[x][p]} = \frac{\mathrm{s}^2}{\mathrm{kg}^3\cdot\mathrm{m}^6}.
$$

5.2 The Chapman–Enskog expansion

The linear relaxation operator $\mathcal{Q}$ in (16) is a projection. We define a projection operator $\mathcal{P}$ by

$$
\mathcal{P}[f](x,y,v,w,t) := \frac{\rho_f(x,\mathcal{E}_y(x,y,w),t)}{N(x,\mathcal{E}_y(x,y,w))}M(v), \tag{19a}
$$

$$
\rho_f(x,\eta,t) := \int \delta(\mathcal{E}_y(x,y,w) - \eta) \cdot f(x,y,v,w,t)\mathrm{d}yvw. \tag{19b}
$$

A straightforward calculation checks that $\mathcal{P}$ is indeed a projection operator, i.e., that $\mathcal{P}^2 = \mathcal{P}$ holds. Furthermore, $[\mathcal{P}(f)] = [f]$.

Author's personal copy

Hence the relaxation operator $\mathcal{Q}$ is a projection operator, since $\mathcal{Q} = \mathcal{I} - \mathcal{P}$ holds, where $\mathcal{I}$ is the identity. The projection $\mathcal{P}$ projects onto the linear manifold of functions which are multiples of the Maxwellian $M(v)$ and depend on y and w only through the energy $\mathcal{E}_x(x, y, w)$.

The Chapman–Enskog expansion consists of deriving the asymptotic form of the Boltzmann equation in local coordinates on this manifold. To this end, we split the density function $f(x, y, v, w, t)$ into

$$f = f_0 + \epsilon f_1,$$

where

$$f_0(x, y, v, w, t) := \mathcal{P}[f](x, y, v, w, t),$$
$$f_1(x, y, v, w, t) := \frac{1}{\epsilon}(\mathcal{I} - \mathcal{P})[f](x, y, v, w, t).$$

Of course, $[f] = [f_0] = [f_1]$. Then we split the evaluation Eq. (18) by applying the projections $\mathcal{P}$ and $\mathcal{I}-\mathcal{P}$. After some calculations, this yields the system

$$\epsilon \partial_t f_0 + \mathcal{P}[\{\mathcal{E}_x + \mathcal{E}_y, f_0 + \epsilon f_1\}_{xv}] = 0, \tag{20a}$$
$$\epsilon^2 \partial_t f_1 + (\mathcal{I} - \mathcal{P})[\{\mathcal{E}_x + \mathcal{E}_y, f_0 + \epsilon f_1\}_{xv}] + \frac{1}{\epsilon}\{\mathcal{E}_y, f_0 + \epsilon f_1\}_{yw} + \frac{1}{\tau} f_1 = 0. \tag{20b}$$

It is simplified by using various properties of the projection $\mathcal{P}$ and the Poisson bracket, resulting in

$$\epsilon \partial_t f_0 + \mathcal{P}[\{\mathcal{E}_x + \mathcal{E}_y, \epsilon f_1\}_{xv}] = 0, \tag{21a}$$
$$\epsilon^2 \partial_t f_1 + \{\mathcal{E}_x + \mathcal{E}_y, f_0 + \epsilon f_1\}_{xv} - \mathcal{P}[\{\mathcal{E}_x + \mathcal{E}_y, \epsilon f_1\}_{xv}] + \frac{1}{\epsilon}\{\mathcal{E}_y, \epsilon f_1\}_{yw} + \frac{1}{\tau} f_1 = 0. \tag{21b}$$

Equation (21a) describes the evaluation on the kernel manifold of the operator $\mathcal{Q}$, and Eq. (21b) describes the evolution on the orthogonal complement. The macroscopic approximation is obtained by dropping the $O(\epsilon)$ terms in (21b). This means that the term ϵf_1 stays small for all times assuming that we start on the kernel manifold, i.e., $\epsilon f_1 = 0$ holds at $t = 0$. This gives the system

$$\partial_t f_0 + \mathcal{P}[\{\mathcal{E}_x + \mathcal{E}_y, f_1\}_{xv}] = 0, \tag{22a}$$
$$\{\mathcal{E}_x + \mathcal{E}_y, f_0\}_{xv} + \{\mathcal{E}_y, f_1\}_{yw} + \frac{1}{\tau} f_1 = 0. \tag{22b}$$

The consistency of the units is seen from comparing with (18) and noting that $[\mathcal{P}(f)] = [f]$.

5.3 The conservation law

The first equation in the system, (22a), can be written as a conservation law for the unknown density $\rho_f(x, \eta, t)$. To verify this, Eq. (22a) is integrated against the test function $\psi = \delta(\mathcal{E}_y(x, y, w) - \eta)$ with respect to y, v, and w yielding

$$\partial_t \rho_{f_0}(x, \eta, t) + \Phi(x, \eta, t) = 0, \tag{23}$$

where

$$\Phi(x, \eta, t) := \int \delta(\mathcal{E}_y(x, y, w) - \eta)\{\mathcal{E}_x + \mathcal{E}_y, f_1\}_{xv} \mathrm{d}y v w.$$

This is indeed a conservation law, since

$$\int \Phi(x, \eta, t) \mathrm{d}x \eta = \int \{\mathcal{E}_x + \mathcal{E}_y, f_1\}_{xv} \mathrm{d}x y v w = 0.$$

Hence it should be possible to write Φ in divergence form as $\Phi = \nabla_x \cdot F^x + \partial_\eta F^\eta$.

In order to find the fluxes F^x and F^η, we integrate $\Phi(x, \eta, t)$ against a test function $\psi(x)$ and calculate

$$\begin{aligned}\int \psi(x)\Phi(x, \eta, t)\mathrm{d}x &= \int \psi(x) \underbrace{\nabla_x \cdot (\delta(\mathcal{E}_y - \eta)\frac{v}{m} f_1)}_{=F^x(x,\eta,t)} \mathrm{d}x y v w \\ &\quad + \int \psi(x) \underbrace{\partial_\eta\big(\delta(\mathcal{E}_y - \eta)(\nabla_x V_1 \cdot \frac{v}{m}) f_1\big)}_{=F^\eta(x,\eta,t)} \mathrm{d}x y v w. \end{aligned} \tag{24}$$

This equation is the weak definition of the fluxes F^x and F^η.

Therefore the conservation law for the mesoscopic density ρ_{f_0} is

$$\partial_t \rho_{f_0}(x, \eta, t) + \nabla_x \cdot F^x(x, \eta, t) + \partial_\eta F^\eta(x, \eta, t) = 0 \tag{25}$$

with the fluxes

$$F^x(x, \eta, t) := \int \delta(\mathcal{E}_y - \eta)\frac{v}{m} f_1 \mathrm{d}y v w, \tag{26a}$$
$$F^\eta(x, \eta, t) := \int \delta(\mathcal{E}_y - \eta)(\nabla_x V_1 \cdot \frac{v}{m}) f_1 \mathrm{d}y v w. \tag{26b}$$

The density ρ_{f_0} in (25) depends on the free energy $\eta = \mathcal{E}_y$ and hence the resulting model is similar to a SHE (spherical-harmonics expansion) model [13] for the Boltzmann equation with the difference that only part of the energy appears as an independent variable.

Of course, the challenge is to compute the fluxes F^x and F^η. In other words, the density f_1 from (22b) must be computed for a given f_0 of the form $f_0(x, y, v, w, t) = M(v)\rho_{f_0}(x, \mathcal{E}_y, t)/N(x, \mathcal{E}_y)$.

Author's personal copy

Equations (25) and (26) imply that the mesoscopic equation for ρ_{f_0} will contain second-order derivatives with respect to x and η. This raises the question of well-posedness of equation (25), which is solved by an entropy estimate for the system (see Proposition 1 below).

The units in (25) and (26) are consistent. By definition (19b), the unit of ρ_{f_0} is

$$[\rho_{f_0}] = \frac{[f][y][v][w]}{[\mathcal{E}]} = \frac{\mathrm{s}^2}{\mathrm{kg}\cdot\mathrm{m}^3}.$$

Since $\rho_{f_0}(x,\eta,t)$ is a concentration per position and energy, we also expect

$$[\rho_{f_0}] = \frac{1}{[x][\mathcal{E}]} = \frac{\mathrm{s}^2}{\mathrm{kg}\cdot\mathrm{m}^3}.$$

The units of the fluxes are

$$[F^x] = \frac{[v][f_1][y][v][w]}{[\mathcal{E}][m]} = \frac{\mathrm{s}}{\mathrm{kg}\cdot\mathrm{m}^2},$$
$$[F^\eta] = \frac{[V_1][v][f_1][y][v][w]}{[\mathcal{E}][x][m]} = \frac{\mathrm{J}\cdot\mathrm{s}}{\mathrm{kg}\cdot\mathrm{m}^3} = \frac{1}{\mathrm{m}\cdot\mathrm{s}}.$$

In summary, we have

$$[\partial_t \rho_{f_0}] = [\Phi] = [\nabla_x \cdot F^x] = [\partial_\eta F^\eta] = \frac{\mathrm{s}}{\mathrm{kg}\cdot\mathrm{m}^3}$$

for the units in (23) and (25).

5.4 The entropy estimate

In the next step, an entropy estimate, Proposition 1, is shown. It was shown in [12] that there is a convex functional of the density ρ, an entropy, which decays in time. The entropy estimate implies that the system (22) is well-posed.

We start from the system (22) and we drop the subscript f_0 of the mesoscopic density ρ_{f_0} from here on for notational simplicity.

After defining the two linear operators $\mathcal{L}_1$ and $\mathcal{L}_2$

$$\mathcal{L}_1[f_1](x,\eta,t) := \int \delta(\mathcal{E}_y(x,y,w)-\eta)\cdot \cdot \{\mathcal{E}_x+\mathcal{E}_y, f_1\}_{xv}\mathrm{d}y v w, \tag{27a}$$

$$\mathcal{L}_2[\rho](x,y,v,w,t) := \left\{\mathcal{E}_x+\mathcal{E}_y, M(v)\frac{\rho(x,\mathcal{E}_y,t)}{N(x,\mathcal{E}_y)}\right\}_{xv} \tag{27b}$$

with the units

$$[\mathcal{L}_1(f_1)] = \frac{[\mathcal{E}][f_1][y][v][w]}{[\mathcal{E}][x][v]} = \frac{\mathrm{s}}{\mathrm{kg}\cdot\mathrm{m}^3},$$
$$[\mathcal{L}_2(\rho)] = \frac{[\mathcal{E}][M][\rho]}{[N][x][v]} = \frac{\mathrm{s}^2}{\mathrm{kg}^3\cdot\mathrm{m}^6},$$

the system (22) becomes

$$\partial_t \rho(x,\eta,t) + \mathcal{L}_1[f_1](x,\eta,t) = 0, \tag{28a}$$
$$\mathcal{L}_2[\rho](x,y,v,w,t) + \{\mathcal{E}_y, f_1\}_{yw} + \frac{1}{\tau} f_1 = 0. \tag{28b}$$

The operator $\mathcal{L}_1$ maps functions of the form $f_1(x,y,v,w,t)$ to functions of the form $\rho(x,\eta,t)$, while $\mathcal{L}_2$ operates in the opposite direction. The operator $\mathcal{L}_2$ is related to $\mathcal{L}_1$ by the following adjoint property.

Lemma 1 (adjoint property) *Let $\mathcal{L}_1^{\mathrm{adj}}$ denote the adjoint of $\mathcal{L}_1$ with respect to the L^2 inner product. Then the operators $\mathcal{L}_1$ and $\mathcal{L}_2$ are related by the equation*

$$\mathcal{L}_2[\rho](x,y,v,w,t) = -c\mathrm{e}^{-(\mathcal{E}_x+\mathcal{E}_y)/(kT)} \cdot \mathcal{L}_1^{\mathrm{adj}}\left[\frac{\mathrm{e}^{(V_0(x)+\eta)/(kT)}\rho(x,\eta,t)}{N(x,\eta)}\right](x,y,v,w,t).$$

Furthermore the identity

$$\Re\left(\int \mathrm{e}^{(\mathcal{E}_x+\mathcal{E}_y)/(kT)} f^*\{\mathcal{E}_y, f\}_{yw}\mathrm{d}y w\right) = 0 \quad \forall x$$

holds for all complex functions $f(y,w)$, where $\Re$ denotes the real part.

Using this lemma, the system (28) takes the form

$$\partial_t \rho(x,\eta,t) + \mathcal{L}_1[f_1](x,\eta,t) = 0, \tag{29a}$$
$$-c\mathrm{e}^{-(\mathcal{E}_x+\mathcal{E}_y)/(kT)} \cdot \mathcal{L}_1^{\mathrm{adj}}\left[\frac{\mathrm{e}^{(V_0(x)+\eta)/(kT)}\rho(x,\eta,t)}{N(x,\eta)}\right](x,y,v,w,t) + \{\mathcal{E}_y, f_1\}_{yw} + \frac{1}{\tau} f_1 = 0. \tag{29b}$$

This form of the system is important for two reasons. First, it makes it possible to prove the entropy estimate, Proposition 1 below, and therefore ensures that the resulting diffusion system is well-posed. Second, solving for f_1 in terms of ρ in the second equation (29) is nontrivial and must to be done numerically in general. We will make sure that this numerical approximation obeys the same entropy estimate by using the form (29).

Author's personal copy

Proposition 1 (entropy estimate) *Solutions* (ρ, f_1) *of the system (29) satisfy the entropy estimate*

$$\frac{1}{2}\partial_t \int \frac{\mathrm{e}^{(V_0(x)+\eta)/(kT)}}{N(x,\eta)}|\rho(x,\eta,t)|^2 \mathrm{d}x\eta = -\frac{1}{c\tau}\int \mathrm{e}^{(\mathcal{E}_x+\mathcal{E}_y)/(kT)}|f_1|^2 \mathrm{d}xyvw \le 0.$$

5.5 Eliminating the lateral velocity

The Poisson bracket in the definition (27b) of $\mathcal{L}_2[\rho]$ simplifies to

$$\mathcal{L}_2[\rho](x,y,v,w,t) = M(v)\frac{v}{m}\cdot\left(\nabla_x\left(\frac{\rho(x,\mathcal{E}_y,t)}{N(x,\mathcal{E}_y)}\right) + \frac{\rho(x,\mathcal{E}_y,t)}{N(x,\mathcal{E}_y)}\nabla_x\frac{V_0+V_1}{kT}\right).$$

Furthermore, the Poisson bracket $\{\mathcal{E}_y, f_1\}_{yw}$ does not operate on the velocity component v, and hence Eq. (28b) allows a solution of the form

$$f_1(x,y,v,w,t) = M(v)\frac{v}{m}\cdot g(x,y,w,t),$$

where the function $g(x,y,w,t) \in \mathbb{R}^d$ is vector valued. Correspondingly, we define the operators Λ_1 and Λ_2 as

$$\Lambda_1[g](x,\eta,t) := \mathcal{L}_1[M(v)\frac{v}{m}\cdot g](x,\eta,t), \tag{30a}$$

$$\mathcal{L}_2[\rho](x,y,v,w,t) = M(v)\frac{v}{m}\cdot\Lambda_2[\rho](x,y,w,t), \tag{30b}$$

$$\Lambda_2[\rho](x,y,w,t) := \nabla_x\left(\frac{\rho(x,\mathcal{E}_y,t)}{N(x,\mathcal{E}_y)}\right) + \frac{\rho(x,\mathcal{E}_y,t)}{N(x,\mathcal{E}_y)}\nabla_x\frac{V_0+V_1}{kT}. \tag{30c}$$

With these definitions, the system (28) becomes

$$\partial_t\rho(x,\eta,t) + \Lambda_1[g](x,\eta,t) = 0, \tag{31a}$$

$$\Lambda_2[\rho](x,y,w,t) + \{\mathcal{E}_y, g\}_{yw} + \frac{1}{\tau}g = 0. \tag{31b}$$

The units of the new variables are

$$[g] = \frac{[m][f_1]}{[M][v]} = \frac{\mathrm{s}^3}{\mathrm{kg}^2\cdot\mathrm{m}^6},$$

$$[\Lambda_1(g)] = [\mathcal{L}_1(f_1)] = \frac{\mathrm{s}}{\mathrm{kg}\cdot\mathrm{m}^3},$$

$$[\Lambda_2(\rho)] = [\mathcal{L}_2(\rho)]\frac{[m]}{[M][v]} = \frac{\mathrm{s}^2}{\mathrm{kg}^2\cdot\mathrm{m}^6}$$

so that all terms in (31b) have the unit $\mathrm{s}^2\cdot\mathrm{kg}^{-2}\cdot\mathrm{m}^{-6}$.

The adjoint property still holds.

Lemma 2 (adjoint property) *The operator* Λ_1 *is of the form*

$$\Lambda_1[g](x,\eta,t) = \frac{kT}{m}\int \nabla_x\cdot\big(\delta(\mathcal{E}_y-\eta)g(x,y,w,t)\big) + \partial_\eta\big(\delta(\mathcal{E}_y-\eta)\nabla_x V_1(x,y)\cdot g\big)\mathrm{d}yw$$

and the operator Λ_2 *is given in terms of the adjoint of* Λ_1 *by*

$$\Lambda_2[\rho](x,y,w,t) = -\frac{m}{kT}\mathrm{e}^{-(V_0+\mathcal{E}_y)/(kT)}\cdot\Lambda_1^{\mathrm{adj}}\left[\mathrm{e}^{(V_0+\eta)/(kT)}\frac{\rho(x,\eta,t)}{N(x,\eta)}\right](x,y,w,t).$$

Furthermore the identity

$$\Re\Big(\int \mathrm{e}^{(\mathcal{E}_x+\mathcal{E}_y)/(kT)}g^H\{\mathcal{E}_y,g\}_{yw}\mathrm{d}yw\Big) = 0 \quad \forall x \tag{32}$$

holds for all complex functions $g(y,w)$.

5.6 Energy and angle variables

To obtain a closed equation for the charge concentration $\rho(x,\eta,t)$, the second Eq. in the system must be solved for g in terms of ρ and substituted into the first equation. For a general confinement potential $V_1(x,y)$, it is only possible to do this approximatively. The approximation takes the form of a series expansion, i.e., a Galerkin solution, of the second equation in [12].

To make the calculations tractable, we use a bijective variable transformation mapping the 4-dimensional vector (y,w) to the scalar energy $u = \mathcal{E}_y(x,y,w)$ and a 3-dimensional angular variable θ. In other words, we will find an x-dependent variable transformation $(u,\theta) = \Gamma(x,y,w)$ and its functional inverse $(y,w) = \Omega(x,u,\theta)$ satisfying

$$\mathcal{E}_y(x,\Omega(x,u,\theta)) = u,$$
$$\Gamma(x,\Omega(x,u,\theta)) = (u,\theta),$$
$$\Omega(x,\Gamma(x,y,w)) = (y,w)$$

for all $x\in\mathbb{R}$, $y,w\in\mathbb{R}^2$, $u\in\mathbb{R}$, and $\theta\in\mathbb{R}^3$. Employing the transformation $(y,w) = \Omega(x,u,\theta)$ still to be determined, we define the operator $\mathcal{A}$ as

$$g_1(x,u,\theta,t) := g(x,\Omega(x,u,\theta),t), \tag{33a}$$

$$\mathcal{A}[g_1](x,\eta,t) := \Lambda_1[g](x,\eta,t), \tag{33b}$$

where

$$[g_1] = [g] = \frac{\mathrm{s}^3}{\mathrm{kg}^2\cdot\mathrm{m}^6},$$

$$[\mathcal{A}(g_1)] = [\Lambda_1(g)] = \frac{\mathrm{s}}{\mathrm{kg}\cdot\mathrm{m}^3}.$$

Author's personal copy

To find the system (31) after the transformation $(y, w) = \Omega(x, u, \theta)$, we have to transform the operators $\Lambda_1, \Lambda_1^{\text{adj}}$ (after using Lemma 2 to replace Λ_2 by Λ_1^{adj}), and $\{\mathcal{E}_y, g\}_{yw}$.

For the Jacobian matrices $\partial\Omega$ and $\partial\Gamma$, we introduce the notation

$$\partial\Omega(x,u,\theta) = \frac{\partial(y,w)}{\partial(u,\theta)} = \begin{pmatrix} \frac{\partial y}{\partial u} & \frac{\partial y}{\partial \theta} \\ \frac{\partial w}{\partial u} & \frac{\partial w}{\partial \theta} \end{pmatrix} =: \begin{pmatrix} \omega_{11} & \omega_{12} \\ \omega_{21} & \omega_{22} \end{pmatrix},$$
$$\partial\Gamma(x,y,w) = \frac{\partial(u,\theta)}{\partial(y,w)} = \begin{pmatrix} \frac{\partial u}{\partial y} & \frac{\partial u}{\partial w} \\ \frac{\partial \theta}{\partial y} & \frac{\partial \theta}{\partial w} \end{pmatrix} =: \begin{pmatrix} \gamma_{11} & \gamma_{12} \\ \gamma_{21} & \gamma_{22} \end{pmatrix}.$$

After several calculations (see [12, Sect. 4.2]), the system (31) becomes

$$\partial_t \rho(x,\eta,t) + \mathcal{A}[g_1](x,\eta,t) = 0, \quad (34a)$$
$$-\frac{m}{kT} e^{-\frac{V_0+u}{kT}} \mathcal{A}^{\text{adj}}\left[e^{\frac{V_0+\eta}{kT}} \frac{\rho(x,\eta,t)}{N(x,\eta)} \right](x,u,\theta,t) + \sigma(S\cdot\nabla_\theta)g_1 + \frac{\sigma}{\tau} g_1(x,u,\theta,t) = 0, \quad (34b)$$

where

$$\mathcal{A}[g_1](x,\eta,t) = \frac{kT}{m}\int \nabla_x \cdot \big(\sigma g_1(x,\eta,\theta,t)\big) + \partial_\eta\big(\sigma(x,\eta,\theta)\nabla_1 V_1(x,\Omega_y(x,\eta,\theta))\cdot g_1\big)\mathrm{d}\theta,$$
$$S(x,u,\theta) = (\gamma_{21}\gamma_{12}^\top - \gamma_{22}\gamma_{11}^\top)(x,\Omega(x,u,\theta)),$$
$$\sigma(x,u,\theta) = |\det(\partial\Omega(x,u,\theta))|.$$

The units are

$$[u] = [\mathcal{E}] = \frac{\mathrm{kg}\cdot\mathrm{m}^2}{\mathrm{s}^2},$$
$$[\theta] = 1,$$
$$[S] = [\gamma_{21}\gamma_{12}^\top] = [\gamma_{22}\gamma_{11}^\top] = \frac{[\theta][u]}{[y_1][w_1]} = \frac{1}{\mathrm{s}},$$
$$[\sigma] = \frac{[y][w]}{[u][\theta]} = \mathrm{kg}\cdot\mathrm{m}^2,$$

and hence all three terms in (34b) consistently have the unit $\mathrm{s}^2\cdot\mathrm{kg}^{-1}\cdot\mathrm{m}^{-4}$.

5.7 Series expansions

The basis functions $\kappa_K(\theta)$, $K \in \mathcal{K}$, are chosen as an orthonormal system satisfying

$$\int \kappa_K(\theta)^* \kappa_{K'}(\theta)\mathrm{d}\theta = \delta_{KK'},$$

where K is a multiindex varying in a 3-dimensional index set $\mathcal{K}$. We expand g_1 into the basis functions by writing

$$g_1(x,\eta,\theta,t) = \sum_{K\in\mathcal{K}} \kappa_K(\theta) G_K(x,\eta,t).$$

We choose the units as

$$[G_K] := [g_1] = \frac{\mathrm{s}^3}{\mathrm{kg}^2\cdot\mathrm{m}^6},$$
$$[\kappa_K] := 1.$$

After the series expansion and with all units, the second equation (34b) in the system becomes

$$\frac{m}{kT} a_K(x,\eta)^* e^{-V_0/(kT)} \nabla_x\big(e^{V_0/(kT)}\frac{\rho}{N}\big) + \frac{m}{kT} A_K(x,\eta)^* e^{-\eta/(kT)}\partial_\eta\big(e^{\eta/(kT)}\frac{\rho}{N}\big) + \sum_{K'\in\mathcal{K}} C_{KK'}(x,\eta) G_{K'}(x,\eta,t) = 0 \quad \forall K\in\mathcal{K}, \quad (35)$$

where

$$a_K(x,\eta) := \frac{kT}{m}\int \kappa_K(\theta)\sigma(x,\eta,\theta)\mathrm{d}\theta, \quad (36a)$$
$$A_K(x,\eta) := \frac{kT}{m}\int \kappa_K(\theta)\sigma(x,\eta,\theta)\nabla_1 V_1(x,\Omega_y(x,\eta,\theta))\mathrm{d}\theta, \quad (36b)$$
$$C_{KK'}(x,\eta) := \int \kappa_K(\theta)^*\sigma(x,\eta,\theta)\big((S\cdot\nabla_\theta)\kappa_{K'}(\theta) + \frac{1}{\tau}\kappa_{K'}(\theta)\big)\mathrm{d}\theta, \quad (36c)$$
$$S(x,\eta,\theta) = (\gamma_{21}\gamma_{12}^T - \gamma_{22}\gamma_{11}^T)(x,\Omega(x,\eta,\theta)), \quad (36d)$$
$$\sigma(x,\eta,\theta) = |\det(\partial\Omega(x,\eta,\theta))|. \quad (36e)$$

The units are

$$[a_K] = \frac{\mathrm{kg}\cdot\mathrm{m}^4}{\mathrm{s}^2},$$
$$[A_K] = \frac{\mathrm{kg}^2\cdot\mathrm{m}^5}{\mathrm{s}^4},$$
$$[C_{KK'}] = \frac{\mathrm{kg}\cdot\mathrm{m}^2}{\mathrm{s}}$$

and all the three terms in (35) have the unit $\mathrm{s}^2\cdot\mathrm{kg}^{-1}\cdot\mathrm{m}^{-4}$.

The 2-dimensional flux vector F^x and the scalar flux F^η are

$$F^x = \frac{kT}{m}\int \sigma g_1(x,\eta,\theta,t)\mathrm{d}\theta,$$
$$F^\eta = \frac{kT}{m}\int \sigma\nabla_1 V_1(x,\Omega_y(x,\eta,\theta))\cdot g_1\mathrm{d}\theta.$$

Author's personal copy

After the series expansion, they become

$$F^x(x,\eta,t) = -\frac{m}{kT}\sum_{K,K'\in\mathcal{K}} a_K(x,\eta)C^{-1}_{KK'}(x,\eta) \cdot\Big(a_{K'}(x,\eta)^* \mathrm{e}^{-\frac{V_0}{kT}}\nabla_x\big(\mathrm{e}^{\frac{V_0}{kT}}\frac{\rho}{N}\big) + A_{K'}(x,\eta)^*\mathrm{e}^{-\frac{\eta}{kT}}\partial_\eta\big(\mathrm{e}^{\frac{\eta}{kT}}\frac{\rho}{N}\big)\Big)$$

and

$$F^\eta(x,\eta,t) = -\frac{m}{kT}\sum_{K,K'\in\mathcal{K}} A_K(x,\eta)^T C^{-1}_{KK'}(x,\eta) \cdot\Big(a_{K'}(x,\eta)^* \mathrm{e}^{-\frac{V_0}{kT}}\nabla_x\big(\mathrm{e}^{\frac{V_0}{kT}}\frac{\rho}{N}\big) + A_{K'}(x,\eta)^*\mathrm{e}^{-\frac{\eta}{kT}}\partial_\eta\big(\mathrm{e}^{\frac{\eta}{kT}}\frac{\rho}{N}\big)\Big).$$

5.8 Harmonic confinement potentials and the variable transformation

The formulas for the fluxes F^x and F^η can be simplified dramatically whenever the confinement potential $V_1(x,y)$ is quadratic in y, i.e., when it is a harmonic potential at each x. This means it is of the form

$$V_1(x,y) = \frac{1}{2}\big(y-b(x)\big)^\top B(x)\big(y-b(x)\big), \tag{37}$$

where $y, b \in \mathbb{R}^2$ and the diagonal matrix $B(x)$ is

$$B(x) = \begin{pmatrix} B_1(x) & 0 \\ 0 & B_2(x)\end{pmatrix}.$$

Of course, we require that $B_1(x) > 0$ and $B_2(x) > 0$ for all x so that the particles are indeed confined. Recalling $[y_i] = \mathrm{m}$, the units of the B_i are defined as

$$[B_i] := \frac{\mathrm{kg}}{\mathrm{s}^2}$$

such that $[V_1] = \mathrm{J}$.

We now define the variable transformation $(y,w) = \Omega(x,\eta,\theta)$ in two steps. In the first step, we transform (y_1,y_2,w_1,w_2) to $(r_1,r_2,\theta_1,\theta_2)$ using polar coordinates by setting

$$y_j =: b_j + \sqrt{\frac{2r_j}{B_j}}\cos\theta_j, \qquad w_j =: \sqrt{2mr_j}\sin\theta_j$$

where $r_j \in [0,\infty)$, $\theta_j \in [-\pi,\pi)$, $[r_j] = \mathrm{J}$, and $[\theta_j] = 1$ for all $j \in \{1,2\}$. The energy $\mathcal{E}_y$ becomes $\mathcal{E}_y(x,y,w) = r_1 + r_2$.

In the second step, we define the transformation

$$\eta := r_1 + r_2, \quad \theta_3 := \frac{r_2 - r_1}{r_2 + r_1},$$
$$r_1 = \eta\frac{1-\theta_3}{2}, \quad r_2 = \eta\frac{1+\theta_3}{2},$$

where $\eta \in [0,\infty)$ and $\theta_3 \in [-1,1]$. Thus the energy $\mathcal{E}_y$ becomes $\mathcal{E}_y(x,y,w) = \eta$. Both transformations result in

$$\begin{pmatrix} y \\ w\end{pmatrix} = \Omega(x,\eta,\theta) = \begin{pmatrix} b_1 + \sqrt{\frac{\eta(1-\theta_3)}{B_1}}\cos\theta_1 \\ b_2 + \sqrt{\frac{\eta(1+\theta_3)}{B_2}}\cos\theta_2 \\ \sqrt{m\eta(1-\theta_3)}\sin\theta_1 \\ \sqrt{m\eta(1+\theta_3)}\sin\theta_2\end{pmatrix}.$$

After applying the transformation Ω, the volume elements become

$$\mathrm{d}yw = \frac{m}{\sqrt{B_1B_2}}\mathrm{d}r_1r_2\theta_1\theta_2 = \frac{m\eta}{2\sqrt{B_1B_2}}\mathrm{d}\eta\theta_1\theta_2\theta_3,$$

and hence

$$\sigma(x,\eta,\theta) = |\det(\partial\Omega(x,\eta,\theta))| = \frac{m\eta}{2\sqrt{B_1(x)B_2(x)}},$$
$$N(x,\eta) = \int\sigma(x,\eta,\theta)\mathrm{d}\theta = \frac{4\pi^2 m\eta}{\sqrt{B_1(x)B_2(x)}}.$$

With this definition of the transformation, the units are consistent.

We choose the basis functions κ_K as

$$\kappa_K(\theta) := \frac{1}{2\pi}\mathrm{e}^{\mathrm{i}k_1\theta_1 + \mathrm{i}k_2\theta_2}L_{k_3}(\theta_3), \quad K = (k_1,k_2,k_3),$$

where the $L_{k_3}(\theta_3)$ are the Legendre polynomials of degree k_3 normalized in the L^2-norm on the interval $[-1,1]$ and the units are $[\kappa_K] = 1$ and $[L_{k_3}] = 1$.

With these definitions, the coefficients a_K become

$$a_K(x,\eta) = \frac{kT}{m}\int\kappa_K\sigma\mathrm{d}\theta = \frac{\sqrt{2}\pi kT\eta}{\sqrt{B_1B_2}}\int\kappa_0^*\kappa_K\mathrm{d}\theta = \frac{\sqrt{2}\pi kT\eta}{\sqrt{B_1B_2}}\delta_{0,K}.$$

The coefficients A_K become

$$\begin{aligned} A_K(x,\eta) = \frac{\pi kT\eta}{2\sqrt{B_1B_2}}\cdot &\Big(\frac{1}{4}\eta\partial_x(\ln B_1)\delta_{k_2,0}(\delta_{k_1,-2} \\ &+ 2\delta_{k_1,0} + \delta_{k_1,2})(\sqrt{2}\delta_{k_3,0} - \sqrt{2/3}\delta_{k_3,1}) \\ &+ \frac{1}{4}\eta\partial_x(\ln B_2)\delta_{k_1,0}(\delta_{k_2,-2} \\ &+ 2\delta_{k_2,0} + \delta_{k_2,2})(\sqrt{2}\delta_{k_3,0} + \sqrt{2/3}\delta_{k_3,1}) \\ &- \sqrt{\eta B_1}\partial_x b_1\delta_{k_2,0}(\delta_{k_1,-1} + \delta_{k_1,1})L^-_{k_3} \\ &- \sqrt{\eta B_2}\partial_x b_2\delta_{k_1,0}(\delta_{k_2,-1} + \delta_{k_2,1})L^+_{k_3}\Big). \end{aligned}$$

Author's personal copy

To compute the coefficients $C_{KK'}$, we note that

$$S(x,\eta,\theta)=\begin{pmatrix}-\sqrt{\frac{B_1(x)}{m}}\\ -\sqrt{\frac{B_2(x)}{m}}\\ 0\end{pmatrix}.$$

After integration, the coefficients become

$$\begin{aligned}C_{KK'}(x,\eta)=\frac{m}{8\pi^2}\frac{\eta}{\sqrt{B_1B_2}}\int \mathrm{e}^{-\mathrm{i}k_1\theta_1-\mathrm{i}k_2\theta_2}L_{k_3}(\theta_3)\\ \cdot\left(-\mathrm{i}k_1'\sqrt{\frac{B_1}{m}}-\mathrm{i}k_2'\sqrt{\frac{B_2}{m}}\right.\\ \left.+\frac{1}{\tau}\right)\mathrm{e}^{\mathrm{i}k_1'\theta_1+\mathrm{i}k_2'\theta_2}L_{k_3'}(\theta_3)\mathrm{d}\theta,\end{aligned}$$

meaning that this particular choice of Galerkin basis functions makes the matrix $C(x,\eta)$ diagonal. This is of great computational importance, as otherwise the matrix $C(x,\eta)$ would have to be inverted numerically at each grid point (x,η). The elements of the inverse $C^{-1}_{KK'}(x,\eta)$ are calculated as

$$C^{-1}_{KK'}(x,\eta)=\frac{2\tau\sqrt{B_1B_2}}{m\eta}\frac{1+\tau\left(k_1\sqrt{\frac{B_1}{m}}+k_2\sqrt{\frac{B_2}{m}}\right)\mathrm{i}}{1+\tau^2\left(k_1\sqrt{\frac{B_1}{m}}+k_2\sqrt{\frac{B_2}{m}}\right)^2}\delta_{KK'},$$

whose unit is

$$[C^{-1}_{KK'}]=\frac{\mathrm{s}}{\mathrm{kg}\cdot\mathrm{m}^2}.$$

The expressions for the fluxes simplify to

$$\begin{aligned}F^x(x,\eta,t)=-\frac{m}{kT}a_0(x,\eta)C^{-1}_{00}(x,\eta)\\ \cdot\left(a_0(x,\eta)^*\mathrm{e}^{-\frac{V_0}{kT}}\nabla_x\left(\mathrm{e}^{\frac{V_0}{kT}}\frac{\rho}{N}\right)\right.\\ \left.+A_0(x,\eta)^*\mathrm{e}^{-\frac{\eta}{kT}}\partial_\eta\left(\mathrm{e}^{\frac{\eta}{kT}}\frac{\rho}{N}\right)\right)\end{aligned}$$

and

$$\begin{aligned}F^\eta(x,\eta,t)=-\frac{m}{kT}\sum_{k_1=-2}^{2}\sum_{k_2=-2}^{2}\sum_{k_3=0}^{\infty}A_K(x,\eta)C^{-1}_{KK}(x,\eta)\\ \cdot\left(\delta_{0,K}a_0(x,\eta)^*\mathrm{e}^{-\frac{V_0}{kT}}\nabla_x\left(\mathrm{e}^{\frac{V_0}{kT}}\frac{\rho}{N}\right)\right.\\ \left.+A_K(x,\eta)^*\mathrm{e}^{-\frac{\eta}{kT}}\partial_\eta\left(\mathrm{e}^{\frac{\eta}{kT}}\frac{\rho}{N}\right)\right).\end{aligned}$$

Further simplifications yield the fluxes summarized in the next section.

5.9 The diffusion-type equation for confined structures

It is quite surprising that explicit expressions for the fluxes F^x and F^η can be found in the case of harmonic confinement potentials using this approach.

In summary, we have derived the conservation law

$$\partial_t\rho(x,\eta,t)+\partial_xF^x(x,\eta,t)+\partial_\eta F^\eta(x,\eta,t)=0 \tag{38}$$

with the fluxes F^x and F^η given by

$$\begin{aligned}F^x(x,\eta,t)=-\frac{4\pi^2kT\tau\eta}{\sqrt{B_1B_2}}T_1\\ -\frac{\pi^2kT\tau\eta^2}{\sqrt{B_1B_2}}\big(\partial_x(\ln B_1)+\partial_x(\ln B_2)\big)T_2\end{aligned} \tag{39}$$

and

$$\begin{aligned}F^\eta(x,\eta,t)=-\frac{\pi^2kT\tau\eta^2}{\sqrt{B_1B_2}}\big(\partial_x(\ln B_1)+\partial_x(\ln B_2)\big)T_1\\ -\frac{\pi^2kT\tau\eta^2}{6\sqrt{B_1B_2}}\left(\frac{12mB_1(\partial_xb_1)^2}{m+\tau^2B_1}+\frac{12mB_2(\partial_xb_2)^2}{m+\tau^2B_2}\right.\\ +2\eta\partial_x(\ln B_1)\partial_x(\ln B_2)\\ +\frac{\eta(3m+8\tau^2B_1)(\partial_x(\ln B_1))^2}{m+4\tau^2B_1}\\ \left.+\frac{\eta(3m+8\tau^2B_2)(\partial_x(\ln B_2))^2}{m+4\tau^2B_2}\right)T_2\end{aligned} \tag{40}$$

after defining

$$\begin{aligned}T_1&:=\mathrm{e}^{-V_0/(kT)}\nabla_x\left(\mathrm{e}^{V_0/(kT)}\frac{\rho}{N}\right),\\ T_2&:=\mathrm{e}^{-\eta/(kT)}\partial_\eta\left(\mathrm{e}^{\eta/(kT)}\frac{\rho}{N}\right).\end{aligned}$$

The confinement potential V is given by

$$V(x,y)=V_0(x)+V_1(x,y),$$

where V_0 is the applied potential and $V_1(x,y)$ has the form (37).

The units in the transport Eq. (38) are

$$\begin{aligned}[\rho]&=\frac{1}{[x][\eta]}=\frac{1}{\mathrm{m}\cdot\mathrm{J}}=\frac{\mathrm{s}^2}{\mathrm{kg}\cdot\mathrm{m}^3},\\ [F^x]&=\frac{\mathrm{s}}{\mathrm{kg}\cdot\mathrm{m}^2},\\ [F^\eta]&=\frac{1}{\mathrm{m}\cdot\mathrm{s}}\end{aligned}$$

so that all three terms have the unit $\mathrm{s}/(\mathrm{kg}\cdot\mathrm{m}^3)$.

This resulting equation is a diffusion equation because of the first-order partial derivatives with respect to x and η in the expressions for the fluxes (39) and (40). The resulting diffusion matrix has automatically the right eigenvalues, since the entropy estimate Proposition 1 holds.

Author's personal copy

Importantly, numerical solutions of the diffusion-type Eq. (38) can be calculated as fast as solutions of a two-dimensional diffusion equation.

6 Simulation studies

After the overview of the modeling approaches in the previous Sects. 3, 4, and 5, simulation studies are summarized here.

6.1 Using the drift-diffusion–Poisson system

The drift-diffusion equations have a long history under various names. They are also known as the convection-diffusion equations, the Fokker–Planck equations, the Nernst–Planck equations, the scalar transport equation, and the Smoluchowski equations. They always describe—with minor differences in the particular area—the transport of species of charged particles governed by two effects: drift (due to an electric field) and diffusion (due to a concentration gradient). Then coupling the drift-diffusion equations with the Poisson equation and solving iteratively ensures self-consistency. As we have seen in Sect. 4, the drift-diffusion equations can be derived from the Boltzmann equation.

The drift-diffusion–Poisson system [14,15] is the system

$$-\nabla\cdot(A\nabla V) = q(C+p-n), \tag{41a}$$
$$\frac{\partial p}{\partial t} = -\nabla\cdot J_p/q, \tag{41b}$$
$$\frac{\partial n}{\partial t} = -\nabla\cdot J_n/(-q), \tag{41c}$$
$$J_p/q = -D_p\nabla p - \mu_p p\nabla V, \tag{41d}$$
$$J_n/(-q) = -D_n\nabla n + \mu_n n\nabla V, \tag{41e}$$

where V denotes the electric potential (and hence the electric field is $-\nabla V$), A is the permittivity, n and p are the concentrations of the free carriers, namely electrons n and holes p, C is the concentration of fixed charges, $q > 0$ is the elementary charge, and J_n and J_p are the densities of the electron and hole electric currents, D_n and D_p are the diffusion coefficients of electrons and holes, and μ_n and μ_p are the mobilities of electrons and holes. Hence $J_n/(-q)$ and J_p/q are the particle currents of electrons and holes, for which the continuity Eq. (41b) and (41c) hold.

The diffusion coefficients and the mobilities are related by the Einstein relations

$$D_n = \mu_n\frac{k_BT}{q},$$
$$D_p = \mu_p\frac{k_BT}{q},$$

where k_B is the Boltzmann constant and T is the absolute temperature.

The first Eq. (41a) is the Poisson equation. The second and third Eq. (41b) and (41c) are continuity equations meaning that electrons and holes are conserved. The last two Eq. (41d) and (41e) mean that the particle currents J_p/q and $J_n/(-q)$ are proportional to the gradients of the particle concentrations (the diffusion part) and to the electric field $-\nabla V$ (the drift or advection part of the model), i.e.,

$$J_p/q = J_p^{\text{diff}} + J_p^{\text{drift}},$$
$$J_n/(-q) = J_n^{\text{diff}} + J_n^{\text{drift}}.$$

How are these four currents modeled? The current densities J_p^{diff} and J_n^{diff} of the diffusion process are proportional to the concentration gradient and opposite in direction, since the gradient points towards higher concentrations. Therefore we have

$$J_p^{\text{diff}} = -D_p\nabla p,$$
$$J_n^{\text{diff}} = -D_n\nabla n.$$

The electric field $E = -\nabla V$ gives rise to the drift currents J_p^{drift} and J_n^{drift}. The drift currents are the products of the carrier concentrations and the average drift velocities $v_p = \mu_p E$ and $v_n = -\mu_n E$. The drift of holes has the same direction as the electric field, whereas the drift of electrons has opposite direction. Furthermore, the drift velocities are proportional to the electric field. This yields

$$J_p^{\text{drift}} = pv_p = \mu_p pE = -\mu_p p\nabla V, \tag{42a}$$
$$J_n^{\text{drift}} = nv_n = -\mu_n nE = \mu_n n\nabla V. \tag{42b}$$

In summary, these expressions for the four currents yield the two current relations (41d) and (41e) for the particle currents J_p/q and $J_n/(-q)$.

The boundary $\partial\Omega$ of the domain Ω where the problem (41) is posed commonly consists of a Dirichlet part $\partial\Omega_D$ and a Neumann part $\partial\Omega_N$ with $\partial\Omega_D \cap \partial\Omega_N = \emptyset$. Contacts give rise to Dirichlet boundary conditions, whereas zero Neumann boundary conditions are usually used everywhere else.

Existence and local uniqueness (whenever the Dirichlet boundary conditions, i.e., the applied potentials are sufficiently small) can be shown for the system (41). Existence and local uniqueness for an extended system with interface conditions whose size is determined in a self-consistent loop were shown in [16].

Is the solution globally unique? It is not. There are reasons why global uniqueness of the solution cannot be expected:

- There is numerical evidence that the solution is not unique when large voltages are applied. It is usually observed

Author's personal copy

that the numerical solution starts to oscillate between two functions when the voltages are stepped up.
– There are also good reasons to expect the drift-diffusion equations to fail to be a good model when the applied voltages are too large or when the electron or hole currents are too large. Recalling the derivation from the Boltzmann equation in Sect. 4, we see that the drift-diffusion equations implicitly assume that velocities of the particles are distributed according to a Maxwellian distribution, i.e., they are in thermal equilibrium. In other words, the concentrations $p(x,t)$ and $n(x,t)$ do not depend on velocity or momentum. However, large applied voltages (yielding large currents) result in particles with high speed and these particles are obviously modeled insufficiently by a Maxwellian distribution centered at $v = 0$.

Of course, the drift-diffusion–Poisson system has been applied to the simulation of ion channels and artificial nanopores as one of its many applications. It has become the standard model for electro-diffusion in ion channels and nanopores. References to works on the drift-diffusion–Poisson system in various contexts are [14,16–19]; references to works specific to ion channels and nanopores are [20–29] among several others.

Numerical methods for the drift-diffusion–Poisson system are well-known. The work by Scharfetter and Gummel was motivated by applications of the drift-diffusion–Poisson system to early semiconductor devices [30–33].

It is important to note that the drift-diffusion–Poisson system neglects the background medium apart from the diffusion constants and the mobilities of the particles in the background medium. In other words, the background medium is assumed to be stationary and homogeneous. In particular, the background medium is a homogeneous bulk in contrast to the real situation where particles move from bulk conditions into a highly confined geometry and then into a bulk again. These limitations lead to the next section.

6.2 Using the drift-diffusion-Stokes–Poisson system

More recently, the effect of the background medium has begun to be included in continuum models by adding the Stokes equation for the background medium, i.e., water, to the drift-diffusion–Poisson system, arriving at the drift-diffusion–Stokes–Poisson system.

In [34], the drift-diffusion–Stokes–Poisson system in cylindrical coordinates was used to simulate ionic flow through a nanopore. In [35], the authors solved the drift-diffusion–Stokes–Poisson system in two dimensions using an arbitrary Lagrangian–Eulerian (ALE) method.

In [36], the authors coupled the Poisson–Boltzmann equation with the Stokes equation. The nonlinear Poisson–Boltzmann equation was solved in one-dimensional geometries. Since the Poisson–Boltzmann equation can be obtained from the drift-diffusion equations by assuming that the system is in equilibrium, assumptions on the interaction of the ions with the background medium are made.

In [37], the drift-diffusion–Poisson equations were coupled with the Navier–Stokes equations in an axisymmetric geometry. In [38], the authors used the drift-diffusion–Stokes–Poisson system to investigate the cause for the high ionic current through a single-walled carbon nanotube.

In [39], the drift-diffusion–Stokes–Poisson system was used to investigate electrically gated nanopores and nanopores acting as sensors. The system of equations used there is the drift-diffusion–Poisson system

$$\nabla \cdot (\epsilon_w \nabla \psi) + q(C_+ - C_-) = 0, \tag{43a}$$

$$q\nabla \cdot (-D_+ \nabla C_+ - \mu_+ C_+ \nabla \psi + C_+ \mathbf{u}) = 0, \tag{43b}$$

$$-q\nabla \cdot (-D_- \nabla C_- + \mu_- C_- \nabla \psi + C_- \mathbf{u}) = 0, \tag{43c}$$

where ψ is the electrostatic potential, ϵ_w the fluid permittivity, $C_\pm$ are the ion concentrations, $\mu_\pm$ the ion mobilities, $D_\pm$ the ion diffusion coefficients, and $\mathbf{u}$ is the fluid velocity. The fluid is modeled as an incompressible, Newtonian Stokes flow governed by the Stokes divergence equations

$$-\nabla p + \gamma \Delta \mathbf{u} - q(C_+ - C_-)\nabla \psi - k_B T \nabla (C_+ + C_-) = 0, \tag{44a}$$

$$\nabla \cdot \mathbf{u} = 0, \tag{44b}$$

where p is the pressure and γ the viscosity. Eqs. (43) and (44) together constitute the drift-diffusion–Stokes–Poisson system. A descreening effect interesting for sensors applications was found in the numerical results in [39].

Although the drift-diffusion–Stokes–Poisson system captures the physical properties of nanopores more realistically than the drift-diffusion–Poisson system, it still does not take the momenta of the particles into account. It makes the assumption that the momenta of all particles are distributed according to a Maxwellian distribution.

6.3 Using a transport equation for confined structures

The transport equation for confined structures discussed in Sect. 5 was solved numerically. The significance of the derivation in Sect. 5 for computations is that the $(6+1)$-dimensional problem of the original Boltzmann equation (three space dimensions and three momentum dimensions and time) is reduced to a $(2+1)$-dimensional diffusion-type problem (position and local energy and time), which can be solved quickly. The significance for nanopores and ion channels is that the current can be calculated from the solution of (38) immediately in contrast to molecular-dynamics calculations.

Author's personal copy

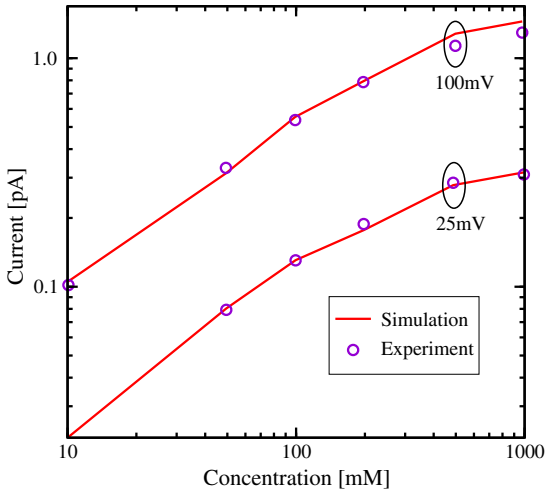


Fig. 2 The simulated current through Gramicidin A versus experiments as a function of the ionic concentration for applied potentials of 25 mV and 100 mM

It is quite surprising that such a transport equation can be derived at all and it is even more surprising that explicit formulas for the geometry dependent fluxes can be given. These fluxes play the role of geometry dependent transport coefficients. The advantage of this fact for calculating numerical solutions is that the solutions can be calculated as fast as 2D diffusion problems without any additional preprocessing or other computational steps.

To calculate the current through realistic structures, an approximate harmonic confinement potential must be constructed first. How to construct such a potential, i.e., how to calculate the functions $b_1(x)$, $b_2(x)$, $B_1(x)$, and $B_2(x)$ in Sect. 5.8, was discussed in [12]. These functions can also be estimated from the geometries and potential walls to be modeled. These coefficients then appear in the fluxes in (39) and (40).

The transport Eq. (38) was applied to two structures, namely Gramicidin A (an antibiotic) and the KcsA channel (a potassium channel). Very good agreement between simulation and measurement was found (see Fig. 2).

6.4 Other approaches

The other main approach in addition to continuum models are molecular-dynamics simulations representing an atomic approach. Nanopores have been simulated using the molecular-dynamics approach (see, e.g., [40,41]). Although molecular dynamics is indispensable for understanding biomolecules and great progress has been made in speeding it up, it calculates the movement of each atom of the background medium and the results are known to depend on the force field used.

An interesting alternative approach to continuum models is reviewed in [42]. There the authors use the variational formulation and constructed the total-energy functional of a charge-transport system including the polar and non-polar free energies of solvation and chemical-potential related energy. They derive coupled Laplace–Beltrami and Poisson–Nernst–Planck equations using the Euler–Lagrange variation. The solution of these equations leads to the minimization of total free energy.

7 Conclusions

Their quantitative understanding is still in its infancy. It can be argued that the modeling of nanopores is more complicated than, e.g., the one of semiconductor devices, since the systems of equations become more involved.

Here an overview of the mathematical modeling and the simulation of nanopores using continuum models was given. The starting point was the Boltzmann equation, to which all other transport models here can eventually be traced back.

Acknowledgments The first author acknowledges support by the FWF (Austrian Science Fund) START project No. Y660 *PDE Models for Nanotechnology*. The second author acknowledges support by NSF Grant 11-07465 (KI-Net).

References

1. Pennisi, E.: Search for pore-fection. Science **336**(6081), 534–537 (2012)
2. Branton, D., et al.: The potential and challenges of nanopore sequencing. Nat. Biotechnol. **26**(10), 1146–1153 (2008)
3. Wanunu, M.: Nanopores: a journey towards DNA sequencing. Phys. Life Rev. **9**(2), 125–158 (2012)
4. Hall, A.R., Scott, A., Rotem, D., Mehta, K.K., Bayley, H., Dekker, C.: Hybrid pore formation by directed insertion of α-haemolysin into solid-state nanopores. Nat. Nanotechnol. **5**, 874–877 (2010)
5. Bell, N.A.W., Engst, C.R., Ablay, M., Divitini, G., Ducati, C., Liedl, T., Keyser, U.F.: DNA origami nanopores. Nano Lett. **12**(1), 512–517 (2012)
6. Burns, J.R., Stulz, E., Howorka, S.: Self-assembled DNA nanopores that span lipid bilayers. Nano Lett. **13**(6), 2351–2356 (2013)
7. Cercignani, C.: The Boltzmann Equation and Its Applications. Springer-Verlag, New York (1988)
8. Maxwell, J.C.: On the dynamical theory of gases. Phil. Trans. Roy. Soc. (London) **157**, 49–88 (1867)
9. Cercignani, C., Illner, R., Pulvirenti, M.: The Mathematical Theory of Dilute Gases. Springer-Verlag, Berlin (1994)
10. Ben Abdallah, N.: On a hierarchy of macroscopic models for semiconductors. J. Math. Phys. **37**(7), 3306–3333 (1996)
11. Ben Abdallah, N., Degond, P., Génieys, S.: An energy-transport model for semiconductors derived from the Boltzmann equation. J. Statist. Phys. **84**(1–2), 205–231 (1996)
12. Heitzinger, Clemens, Ringhofer, Christian: A transport equation for confined structures derived from the Boltzmann equation. Commun. Math. Sci. **9**(3), 829–857 (2011)

Author's personal copy

13. Ben Abdallah, N., Degond, P., Markowich, P., Schmeiser, C.: High field approximations of the spherical harmonics expansion model for semiconductors. Z. Angew. Math. Phys. **52**(2), 201–230 (2001)
14. Markowich, P.A., Ringhofer, C.A., Schmeiser, C.: Semiconductor Equations. Springer-Verlag, Wien (1990)
15. Anile, A.M., Allegretto, W., Ringhofer, C.: Mathematical Problems in Semiconductor Physics. Springer-Verlag, Berlin (2003)
16. Baumgartner, Stefan, Heitzinger, Clemens: Existence and local uniqueness for 3d self-consistent multiscale models for field-effect sensors. Commun. Math. Sci. **10**(2), 693–716 (2012)
17. Markowich, P.A.: The Stationary Semiconductor Device Equations. Springer-Verlag, Wien (1986)
18. Degond, P., Méhats, F., Ringhofer, C.: Quantum energy-transport and drift-diffusion models. J. Statist. Phys. **118**, 625–667 (2005)
19. Baumgartner, Stefan, Heitzinger, Clemens: A one-level FETI method for the drift-diffusion-Poisson system with discontinuities at an interface. J. Comput. Phys. **243**, 74–86 (June 2013)
20. Rubinstein, I.: Electro-Diffusion of Lons. SIAM, Philadelpha, PA (1990)
21. Ramírez, P., Mafé, S., Aguilella, V.M., Alcaraz, A.: Synthetic nanopores with fixed charges: an electrodiffusion model for ionic transport. Phys. Rev. E (3), 68(1):011910/1-8, (July 2003)
22. van der Straaten, T.A., Tang, J.M., Ravaioli, U., Eisenberg, R.S., Aluru, N.R.: Simulating ion permeation through the ompF porin ion channel using three-dimensional drift-diffusion theory. J. Comp. Electron. **2**, 29–47 (2003)
23. Gardner, C.L., Nonner, W., Eisenberg, R.S.: Electrodiffusion model simulation of ionic channels: 1D simulations. J. Comput. Electron. **3**, 25–31 (2004)
24. Coalson, R.D., Kurnikova, M.G.: Poisson–Nernst–Planck theory approach to the calculation of current through biological ion channels. IEEE Trans. Nanobiosci. **4**(1), 81–93 (2005)
25. Cervera, J., Schiedt, B., Ramírez, P.: A Poisson/Nernst–Planck model for ionic transport through synthetic conical nanopores. Europhys. Lett. **71**(1), 35–41 (2005)
26. Liu, Y., Sauer, J., Dutton, R.W.: Effect of electro-diffusion current flow on electrostatic screening in aqueous pores. J. Appl. Phys. **103**, 084701 (2008)
27. Cervera, J., Ramírez, P., Manzanares, J.A., Mafé, S.: Incorporating ionic size in the transport equations for charged nanopores. Microfluid. Nanofluid. **9**(1), 41–53 (2010)
28. Gardner, C.L., Jones, J.R.: Electrodiffusion model simulation of the potassium channel. J. Theoret. Biol. **291**, 10–13 (2011)
29. Lu, B., Hoogerheide, D.P., Zhao, Q., Yu, D.: Effective driving force applied on DNA inside a solid-state nanopore. Phys. Rev. E **86**(1), 011921 (2012)
30. Gummel, H.K.: A self-consistent iterative scheme for one-dimensional steady state transistor calculations. IEEE Trans. Electron Devices **ED–11**, 455–465 (1964)
31. Scharfetter, D.L., Gummel, H.K.: Large-signal analysis of a silicon read diode oscillator. IEEE Trans. Electron Devices ED **16**(1), 64–77 (1969)
32. Kerkhoven, T.: A proof of convergence of Gummel's algorithm for realistic device geometries. SIAM J. Numer. Anal. **23**(6), 1121–1137 (1986)
33. Gartland Jr, E.C.: On the uniform convergence of the Scharfetter–Gummel discretization in one dimension. SIAM J. Numer. Anal. **30**(3), 749–758 (1993)
34. Mao, M., Ghosal, S., Hu, G.: Hydrodynamic flow in the vicinity of a nanopore induced by an applied voltage. Nanotechnology **24**(24), 245202 (2013)
35. Ai, Y., Qian, S.: Electrokinetic particle translocation through a nanopore. Phys. Chem. Chem. Phys. **13**(9), 4060–4071 (2011)
36. van Dorp, S., Keyser, U.F., Dekker, N.H., Dekker, C., Lemay, S.G.: Origin of the electrophoretic force on DNA in solid-state nanopores. Nat. Phys. **5**, 347–351 (2009)
37. Jubery, T.Z., Prabhu, A.S., Kim, M.J., Dutta, P.: Modeling and simulation of nanoparticle separation through a solid-state nanopore. Electrophoresis **33**(2), 325–333 (2012)
38. Bearden, S., Zhang, G.: The effects of the electrical double layer on giant ionic currents through single-walled carbon nanotubes. Nanotechnology **24**(12), 125204 (2013)
39. Liu, Y., Huber, D.E., Tabard-Cossa, V., Dutton, R.W.: Descreening of field effect in electrically gated nanopores. Appl. Phys. Lett. **97**(1–3), 143109 (2010)
40. Bhatia, S., Nicholson, D.: Molecular transport in nanopores. J. Chem. Phys. **119**(3), 1719–1730 (2013)
41. Joseph, S., Guan, W., Reed, M.A., Krstic, P.S.: A long DNA segment in a linear nanoscale Paul trap. Nanotechnology **21**(1), 015103 (2010)
42. Wei, G.-W., Zheng, Q., Chen, Z., Xia, K.: Variational multiscale models for charge transport. SIAM Rev. **54**(4), 699–754 (2012)