

# Microscopic Noise Simulation in Nanoscale nMOSFETs

---

Von der Fakultät für Elektrotechnik und Informationstechnik der  
Rheinisch-Westfälischen Technischen Hochschule Aachen

zur Erlangung des akademischen Grades

**Dr. rer. nat.**

genehmigte Dissertation

vorgelegt von

DIPLOM-PHYSIKER

**Dino Ruić**

aus Köln

---

Berichter

**Univ.-Prof. Dr.-Ing. Christoph Jungemann**

**Prof. Dr.-Ing. Bernd Meinerzhagen**

Tag der mündlichen Prüfung

6. Juli 2017

---

# CURRICULUM VITAE

---

DINO RUIĆ

born 12 NOVEMBER 1987  
in COLOGNE, GERMANY

2017– POSTDOC  
imec, Leuven

2011–2017 DOCTORATE  
RWTH Aachen University,  
Institute of Electromagnetic Theory

2006–2011 DIPLOMA IN PHYSICS  
University of Bonn

1994–2006 UNIVERSITY ENTRANCE QUALIFICATION  
Albertus-Magnus-Gymnasium in Cologne

## ABSTRACT

The purpose of this work is the computation of the noise in a nanoscale nMOSFET whose charged carrier transport is described by the Boltzmann equation and where the confinement of carriers is determined by the Schrödinger equation.

The modeling of carrier transport in semiconductor devices has a long tradition of simulation by the Boltzmann equation or its moments but only in more recent years have deterministic solvers become feasible. In particular, solvers based on an expansion of the angular dependence of  $\mathbf{k}$ -space into spherical or Fourier harmonics have proven to be fruitful. Furthermore, it has been shown exhaustively that nanoscale transistors exhibit confining properties on the carriers that necessitate the solution of the Schrödinger equation for accurate characterization.

We build upon prior research concerning such models in order to formulate the first ever Newton-Raphson method including the Poisson, Schrödinger, and Boltzmann equations for a quadratically converging system describing a nanoscale nMOSFET. We proceed to formulate the problem for the self-consistent small signal response of the aforementioned fully coupled system. It is here that we encounter several unforeseen difficulties which are of both conceptual and numerical nature and whose investigation and proper resolution is original to this work.

The calculation of noise via the Langevin-source approach – the main purpose of this work – strongly relies on a correct understanding of how the system responds to small signal perturbations and many of our original results are necessary to formulate the problem numerically. Although the Langevin-source approach has been applied to the Boltzmann equation before, we are the first to demonstrate a fully self-consistent solution including the Schrödinger equation. Moreover, we investigate the correct treatment of degeneracy in the context of the Langevin-source approach and derive general equations to show where multiplicities need to appear. We show several optimizations to reduce the computation time, one of which is that we demonstrate that for certain symmetries in the degeneracy, the problem for the admittance parameters and for the noise is completely equivalent and can be solved in one step.

To put our approach to the test, we characterize the stationary, the small signal, and the noise properties of a nanoscale nMOSFET. We show that our full Newton-Raphson approach converges quadratically and therefore supremely faster than the usual Gummel type iteration scheme. We establish the numerical limits of the approach and validate that it fulfills essential conservation laws like current conservation, reciprocity in equilibrium, and the Nyquist theorem. Wherever possible we compare to existing results in literature which corroborate our findings.

The main advantage of a Boltzmann equation based solver for noise is that the noise is completely determined by the scattering processes. If the included scattering rate contains all essential processes that contribute to stationary solutions and to the small signal response of the system, the noise is necessarily correct. However, this argument can only hold as long as the impact of fluctuations in the distribution function on the electric potential – and vice versa – is included. Our approach contains this interdependency by solving the system of equations self-consistently which is only possible due to the deterministic formulation of the Boltzmann equation. We verify that self-consistency is indeed vital in the ballistic transport regime and show with unprecedented insight into the origins of noise in nanoscale devices how fluctuations elicit responses in the terminal currents.

This work presents the first ever fully self-consistent solver for the stationary, small signal, and noise characterization of a nanoscale nMOSFET with a confined electron gas.

## ZUSAMMENFASSUNG

Ziel dieser Arbeit ist die Berechnung des Rauschens in einem nanometergroßen nMOS-FET, dessen Ladungsträgertransport durch die Boltzmann-Gleichung beschrieben ist und bei dem das Confinement der Ladungsträger durch die Schrödingergleichung bestimmt ist.

Die Modellierung des Ladungsträgertransports in Halbleiterbauelementen mit Hilfe der Boltzmann-Gleichung oder ihrer Momente hat bereits eine längere Tradition, es ist allerdings erst seit Kurzem möglich, diese mit deterministischen Methoden sinnvoll zu lösen. Insbesondere Lösungsmethoden, die auf einer Entwicklung der Winkelabhängigkeit des  $\vec{k}$ -Raums in Kugelflächenfunktionen oder Fourierharmonischen beruhen, zeigen sich als wirkungsvoll. Außerdem wurde schon häufig gezeigt, dass Ladungsträger in nanometergroßen Transistoren Quanteneffekten ausgesetzt sind, die zur Beschreibung die Lösung der Schrödingergleichung benötigen.

Wir bauen auf vorhergegangene Arbeiten mit ähnlichen Ansätzen auf und formulieren somit die zuvor nie dagewesene vollständige Newtonmethode für das System von Poisson-, Schrödinger- und Boltzmann-Gleichung für einen nanometergroßen nMOSFET. Daraufhin formulieren wir ebenfalls das selbstkonsistente Gleichungssystem zur Bestimmung des Kleinsignalverhaltens. Hierbei treten mehrere unvorhersehbare Schwierigkeiten auf, deren Ursprung sowohl konzeptionell als auch numerisch ist und deren Behebung einen wesentlichen Teil dieser Arbeit ausmacht.

Das Hauptthema ist die Berechnung des Rauschens durch einen Langevinquellenansatz, das stark auf dem richtigen Verständnis des Verhaltens des Gleichungssystems unter Kleinsignalstörungen beruht. Viele unserer Erkenntnisse über die Berechnung des Kleinsignalverhaltens sind erforderlich, um den Langevinquellenansatz numerisch korrekt umzusetzen. Obwohl der Langevinquellenansatz bereits zuvor auf die Boltzmann-Gleichung angewandt wurde, ist es erstmals in dieser Arbeit gelungen eine vollständig selbstkonsistente Lösung, inklusive der Schrödingergleichung, zu demonstrieren. Weiterhin untersuchen wir auch die Entartung von Zuständen im Rahmen des Langevinquellenansatzes und leiten allgemeine Gleichungen her, die zeigen, wie sich Multiplizitäten von entarteten Zuständen auswirken. Ferner zeigen wir mehrere Möglichkeiten zur Optimierung der Berechnungslaufzeit. Bei einer davon zeigen wir etwa, dass unter bestimmten Symmetriebedingungen für die Entartung von Zuständen, die Rausch- und Kleinsignalverhalten mathematisch äquivalent werden und somit in bloß einem Lösungsschritt bestimmt werden können.

Um unsere Methode zu testen, charakterisieren wir das stationäre, das Kleinsignal- und das Rauschverhalten eines nanometergroßen nMOSFETs. Wir zeigen, dass die Newtonmethode mit ihrer quadratischen Konvergenz dem üblichen iterativen Gummelfahren deutlich überlegen ist. Wir ergründen die numerischen Grenzen und verifizieren Erhaltungsgesetze wie die Stromerhaltung, die Reziprozität im Gleichgewicht und das Nyquisttheorem. Wo auch immer möglich, vergleichen wir mit existierenden Resultaten aus der Literatur, die unsere Ergebnisse bekräftigen.

Ein wesentlicher Vorteil unserer Methode ist, dass das Rauschen vollständig durch die Streuprozesse in der Boltzmann-Gleichung definiert ist. Es folgt daher, dass das Rauschen notwendigerweise richtig sein muss, wenn die Streuraten alle essentiellen Streuprozesse enthält. Dies trifft allerdings nur dann zu, wenn der Einfluss von Fluktuationen in der Ladungsträgerverteilungsfunktion auf das elektrische Potential – und umgekehrt – beinhaltet ist. Unsere Methode beinhaltet diese wechselseitige Kopplung durch die selbstkonsistente Lösung, die nur mit einem deterministischen Ansatz möglich ist. Wir zeigen mit nie zuvor dagewesener Einsicht, dass diese Selbstkonsistenz eine wesentliche Rolle für ballistischen Transport spielt und wie diese zum Rauschen im Kontaktstrom führt.

Diese Arbeit stellt den ersten selbstkonsistenten Löser für sowohl das stationäre Verhalten als auch das Kleinsignal- und Rauschverhalten eines nanometergroßen nMOSFET mit Elektronenconfinement dar.

# Microscopic Noise Simulation in Nanoscale nMOSFETs

Dino Ruić



# Contents

<b>1</b>	<b>Introduction</b>	<b>7</b>
1.1	State of Device Simulation . . . . .	8
1.2	Current Work . . . . .	11
1.3	Structure of the Manuscript . . . . .	12
<b>2</b>	<b>Stationary Equations</b>	<b>15</b>
2.1	Overview . . . . .	15
2.2	Poisson Equation . . . . .	17
2.2.1	Equation . . . . .	18
2.2.2	Boundary Conditions . . . . .	18
2.2.3	Discretization . . . . .	19
2.2.4	Solution . . . . .	20
2.3	Schrödinger Equation . . . . .	22
2.3.1	Equation . . . . .	22
2.3.2	Boundary conditions . . . . .	23
2.3.3	Discretization . . . . .	24
2.3.4	Solution and Perturbation . . . . .	25
2.4	Boltzmann Equation . . . . .	25
2.4.1	Equation . . . . .	26
2.4.2	Scattering Terms . . . . .	30
2.4.3	Boundary Conditions . . . . .	44
2.4.4	Transformations . . . . .	49
2.4.5	Discretization . . . . .	56
2.4.6	Solution . . . . .	66
2.5	Iteration Schemes . . . . .	73
2.5.1	Gummel Type Iteration . . . . .	74
2.5.2	Newton-Raphson Approach . . . . .	77
<b>3</b>	<b>Small Signal Analysis</b>	<b>81</b>
3.1	Linearization . . . . .	81
3.2	Time-Derivative of the Boltzmann Equation . . . . .	85

3.3	Boundary Conditions . . . . .	88
3.4	Ramo-Shockley Theorem . . . . .	89
3.5	Reciprocity . . . . .	95
3.5.1	Proof up to the First Harmonic . . . . .	95
3.5.2	Restoring Reciprocity in Discretized $H$ -Space . . . . .	100
3.5.3	General Case . . . . .	102
3.6	Solving for Small Signal Parameters . . . . .	102
3.6.1	Direct Method . . . . .	104
3.6.2	Adjoint Method . . . . .	104
3.6.3	Common-Source Configuration . . . . .	106
<b>4</b>	<b>Noise</b>	<b>109</b>
4.1	Introduction . . . . .	110
4.2	Langevin-Source Approach . . . . .	113
4.3	Boundary Conditions . . . . .	119
4.4	Ramo-Shockley Theorem . . . . .	120
4.5	Adjoint Method . . . . .	124
4.6	Power Spectral Density . . . . .	125
4.6.1	Scattering . . . . .	126
4.6.2	Generation and Recombination . . . . .	128
4.6.3	Discretization . . . . .	130
4.6.4	Common-Source Configuration . . . . .	132
4.7	Degeneracy . . . . .	133
4.7.1	One Degenerate Group . . . . .	133
4.7.2	Multiple Degenerate Groups . . . . .	136
4.7.3	Power Spectral Density . . . . .	141
4.7.4	Application . . . . .	142
4.8	Simultaneous Small Signal Analysis . . . . .	145
<b>5</b>	<b>Results</b>	<b>149</b>
5.1	Device . . . . .	149
5.1.1	Scattering Parameters . . . . .	150
5.1.2	Contacts . . . . .	153
5.1.3	Discretization . . . . .	153
5.2	Stationary Solutions . . . . .	155
5.2.1	Numerics and Solver Requirements . . . . .	155
5.2.2	Distribution Function . . . . .	157
5.2.3	Current and Subband Energies . . . . .	160
5.2.4	Electron Density, Velocity, and Mean Energy . . . . .	162
5.3	Small Signal Analysis . . . . .	165
5.3.1	Conservation Laws and Symmetries . . . . .	167
5.3.2	Admittance Parameters . . . . .	169

5.3.3	Stability . . . . .	173
5.3.4	Cutoff and Maximum Oscillation Frequency . . . . .	174
5.4	Noise Analysis . . . . .	176
5.4.1	Power Spectral Density . . . . .	177
5.4.2	Nyquist Theorem . . . . .	178
5.4.3	Fano Factor . . . . .	180
5.4.4	Excess Noise . . . . .	181
5.4.5	Cross Correlation . . . . .	185
5.4.6	Origin of Noise . . . . .	188
5.4.7	Suppression of Noise . . . . .	191
<b>6</b>	<b>Conclusion</b>	<b>199</b>
<b>A</b>	<b>Observables</b>	<b>201</b>
A.1	Stationary . . . . .	201
A.2	Small Signal . . . . .	204
A.3	Noise . . . . .	207
<b>B</b>	<b>Normalization</b>	<b>209</b>
<b>C</b>	<b>Discretized Power Spectral Density</b>	<b>211</b>
	<b>Nomenclature</b>	<b>215</b>
	<b>Bibliography</b>	<b>224</b>



# Chapter 1

## Introduction

Since their inception in the 1940s, transistors have become essential to virtually all electronics. Their ubiquitous use in applications from logic gates to amplifiers is a testament to their indispensability to modern society. Massive orchestrated efforts spanning the whole semiconductor industry have been directed into the manufacturing of ever smaller and thus more efficient and faster transistors [1]. These efforts culminate in today's nanoscale MOSFETs used in digital and analog applications to enable the massive adoption of small, energy efficient, and inexpensive portable computers with high-bandwidth wireless communication.

For analog applications, the ongoing miniaturization has led to integrated circuits for RF applications in the GHz range. Due to the large-scale integration of CMOS transistors, they offer cost-efficient manufacturing which make them particularly suited for handling the increased demands for analog circuits [2]. Simulations significantly support the development of chips from the computer-aided design of the circuits all the way to the virtual reproduction of the manufacturing process. Somewhere in the design process also lies the simulation of individual MOSFETs with which we will occupy ourselves. As the size of these MOSFETs decreases, it becomes apparent that previously reliable models fail to accurately predict their behavior [3].

Many of the more detailed simulation approaches focus on the stationary operation of MOSFETs, yet modeling of analog circuits requires additional knowledge about the small signal response and the noise characteristics of devices. In this work, we will take an existing approach to the simulation of nanoscale MOSFETs and extend it to incorporate accurate predictions about small signal and noise characteristics. But before we elaborate on the ideas behind this, we need to understand where this work is placed in the broader context of device simulations.

## 1.1 State of Device Simulation

There are multiple approaches to the simulation of semiconductor devices, each with its own domain of applications. Since most electronic design happens at the circuit level, compact models for transistors are widely used to approximate the transistor's behavior as a circuit component [4, 5]. They consist of physics-based or fitted analytic functions, or even interpolated table lookup models. While quick to evaluate, the physics involved, if any, is a crude approximation to reality and thus they are only useful to capture the most important aspects of the transistor's operation. Free parameters of compact models are usually informed by experiments or more detailed device simulations.

On the device level, models derived as moments of the Boltzmann equation (BE) [6, 7, 8] have a history of successful simulation of transistors. The simplest one, using only the first two moments, is the drift-diffusion (DD) model [9, 10] which to this date is the most widely used and the *de facto* standard in the semiconductor industry. Its popularity is due to its simplicity yet surprising effectiveness [11].

To increase the accuracy of the DD model, it is possible to extend it with the next two moments of the BE which is called the hydrodynamic model (HD) [12]. The HD model includes carrier heating in the channel and is therefore capable of modeling velocity overshoot in the channel region. The predictive capabilities of the moments-based models are naturally linked to how well the inherent assumptions in their derivation hold. In MOSFETs, we find that devices with channel lengths of less than 100 nm cannot be reasonably simulated with these models anymore [3, 13, 14]. Nevertheless, there are applications in which the DD model is used as a fitting model for short-channel devices where most free parameters are adjusted in ways that are incompatible with their original physical meaning. This type of application would not require any predictive capabilities. Including even more moments than in the HD model is usually inadvisable [15] and it is generally a good idea to solve the BE directly when the HD model does not suffice anymore.

Solving the BE is significantly more challenging than any of the DD or HD models and usually involves a factor of around a thousand times more computational effort. The most popular way to solve the BE is to use a Monte Carlo (MC) based approach [16] due to its relatively straightforward implementation and its frugality in memory utilization. Here, an ensemble of carriers is tracked throughout the device and their respective scattering rates are stochastically applied at set time intervals. In more recent years, deterministic solvers for the BE gained some traction [17, 18, 19, 20, 21]. Deterministic solvers generally yield numerically higher quality results and they are able to solve systems which are computationally infeasible to solve with MC methods [22]. These are typically cases in which the MC based methods cannot collect statistically significant

amounts of data on a process, e.g. when considering scattering processes with two very different time scales or when resolving small currents [23]. Conversely, the downside of deterministic solvers is that they involve more sophisticated mathematics, numerical instabilities, and massive memory requirements. All of which make the implementation of a deterministic solver for the BE a formidable task.

We want to particularly emphasize advances in deterministic solvers based on an expansion of  $\mathbf{k}$ -space in spherical harmonics [24], since they are the most widely employed deterministic solvers and they are relevant to the present work (for a comprehensive review on spherical harmonics based solvers, see Ref. [25]). The basic idea is that in equilibrium the distribution function is spherically symmetric in  $\mathbf{k}$ -space and therefore it can be mapped exactly to the leading order in a spherical harmonics expansion of  $\mathbf{k}$ -space. Non-equilibria can be modeled as corrections using higher orders of the expansion, where it has been found that the series converges quickly enough for the approach to reduce the total number of unknowns significantly. Deterministic solvers based on the spherical harmonics expansion have been applied to carrier transport in semiconductor devices [26, 27, 28, 29]. Moreover, they have been used for magnetotransport simulation [30], electrothermal simulations and degradation [31, 32], coupled hot carrier and phonon systems [33], and avalanche breakdown of pn-junctions and devices [34, 35].

Despite all of its complexities, the BE still models transport classically where carriers are given definite positions and momenta and where quantum mechanical effects such as interference or tunneling are disregarded. Approaches comprising quantum transport include the solution of the Wigner equation, see e.g. Ref. [36] and references therein, or the non-equilibrium Green's function approach, see e.g. Ref. [37]. Both of these approaches are orders of magnitude more computationally intense than the BE and bring their own share of problems. We will not concern ourselves with quantum transport in this work and therefore we will not delve any further into this topic.

So far we have only discussed transport models but in nanoscale transistors a discrepancy between classical models and reality occurs that is of significant importance even in equilibrium and which needs to be addressed properly. As the oxide thickness in MOSFETs decreases, the quantum mechanical splitting of energy states into discrete levels becomes apparent where each energy level is associated with a distinct probability amplitude to find carriers. The predominant visible effect is that the carrier density close to the oxide becomes small, i.e. carriers feel the high oxide potential and they are repelled by it. While there are approaches to approximate the effect [38], the most reliable method is to solve the Schrödinger equation (SE) directly and assume transport along – as well as transitions in between – the eigenstates of the SE [39], which are often referred to as *subbands*. In such an approach, the SE is solved perpendicular

to the transport direction in order to compute the states and distribution of carriers but transport itself is still governed by the classical BE.

When solving the SE perpendicular to the transport direction – which we will refer to as the *confinement direction* –, we effectively obtain standing wave solutions with discrete energies and thus discrete wave number components in confinement direction. This means that  $\mathbf{k}$ -space is partitioned into two subspaces where the kinematically available components for the classical transport – modeled through the BE – are reduced by the dimensions where the SE is solved. Hence, when the SE is solved in one dimension, the BE will only have a two-dimensional  $\mathbf{k}$ -space, which means that the spherical harmonics expansion of  $\mathbf{k}$ -space is in fact reduced to a Fourier harmonics expansion [29, 40]. Solving the two-dimensional SE amounts to a one-dimensional  $\mathbf{k}$ -space of the BE which can be feasibly solved deterministically by a straightforward discretization. Solutions of such a system are useful to compute the properties of nanowires [41, 42].

Until now we have only discussed solvers for stationary solutions of a MOSFET's operation. To be sure, they are important to know but for analog applications, we also need to know the MOSFET's response to time-dependent biases. In particular, its properties during small signal operation which reveal parasitics that otherwise do not impact the stationary operation (see e.g. Ref. [5]). What is more, all devices generate noise, limiting the capabilities of amplifier circuits [43] which include – most prominently – wireless modems whose omnipresence is compelled by the booming smartphone market. As the smartphone market pushes to ever thinner and energy efficient devices, so do the requirements push for wireless modems with ever more advanced manufacturing nodes. However, the small signal and noise behavior of advanced manufacturing nodes is not well understood. With models that are only crude approximations to reality [44], analog circuit design becomes more difficult, increasing the necessary margins for error that need to be considered due to the discrepancies between model and reality.

For small signal and noise analyses we can – in principle – use the same simulation approaches as for the stationary case: There are compact models for the small signal and noise behavior (see e.g. Refs. [5, 45]). The DD and HD models can be linearized around some stationary operating point in order to compute small signal effects [46] and a Langevin-source approach can be used in order to compute the noise [47, 48]. Just like the DD and HD models, the BE can be linearized for a small signal simulation [49] and the Langevin-source approach can be used for the computation of the noise [50, 51].

Although at first glance it seems straightforward to compute the small signal and noise behavior, the simulation of the BE and beyond is often limited to the stationary case. In particular, since the BE is usually solved with the MC method, small signal analysis requires additional care [52] but it is still limited by the stochastic nature of the MC method. Moreover, simulation of

RF transistor properties in the GHz range are computationally expensive when plasma oscillations in the femtosecond time-scale need to be resolved with the MC method [22, 53]. On the other hand, due to the way the MC method works, the correlation function of electronic noise in a device is straightforwardly computed for every solution of a stationary operating point [54, 55]. Nevertheless, publications on the noise in devices computed using the MC method are few and far between since the MC results for electronic noise are themselves subject to statistical noise as well as the usual limitations of MC simulations such as the difficulty to compute relevant frequency ranges of noise power spectra [56, 48].

A problem occurring solely in device simulations is that self-consistency of all equations needs to be guaranteed in order to find meaningful results. Self-consistency is usually attained by a Gummel type iteration [57], as is e.g. shown in Ref. [39] for a nanoscale device. In this approach each equation is solved in isolation, one after the other, until convergence is achieved. The implicit notion being that while one equation is solved, quantities derived from the other equations are considered fixed. But in the case of a small signal or noise computation, we would need to keep track of how small time-dependent perturbations and fluctuations would impact other quantities which quickly becomes computationally infeasible. Therefore, small signal and noise analyses often ignore all equations except for the BE, which is used to compute the response in isolation. However, this can only be faithful to reality when any difference in carrier densities arising from the BE after freezing the fields cannot meaningfully impact them anymore. This assumption is known to be violated in nanoscale devices [58] and thus not only stationary results but also small signal and noise analyses must contain full self-consistency.

Deterministic solvers are uniquely positioned to compute the small signal parameters and the noise self-consistently in devices, since all equations can be written down in a single large system of equations – including all interdependencies – and then solved simultaneously [29, 59, 60, 61, 62].

## 1.2 Current Work

The work compiled in this thesis follows in the footsteps of previous deterministic solvers and attempts to solve the system consisting of Poisson equation (PE), SE, and BE for a nanoscale nMOSFET. Due to the quantization in one dimension by the SE, the BE is only solved for a two-dimensional  $\mathbf{k}$ -space and therefore we expand it in Fourier harmonics.

Previously existing work already solved the stationary problem for the self-consistent system of PE, SE, and BE [29]. The current work expands on this by implementing the first working Newton-Raphson solver over all equations for the stationary problem [63], which has the advantage of rapid quadratic convergence

compared to the usual linear convergence of the Gummel type iteration.

From there, we derive a way to self-consistently solve the small signal problem. This is all but straightforward since the naive approach leads to the violation of conservation laws that should be fulfilled by real devices. However, with a careful analysis of the equations, we are able to formulate the system of equations such that it becomes physically sound [58].

Using the Langevin-source approach, we formulate the system of PE, SE, and BE in such a way that we can compute the response of terminal currents to fluctuations within the device self-consistently. Subsequently, we can compute the noise power of terminal current fluctuations directly [58]. The advantage of a BE based approach is that the noise is fully determined by the scattering rate in the BE and hence there are no additional parameters necessary.

The deterministic approach enables us to self-consistently determine all small signal and noise related quantities for a wide range of operating points and frequencies, unencumbered by the stochastic nature of the usual MC approach. Among these are the admittance parameters, cutoff frequencies, maximum oscillation frequencies, excess noise factors and the cross-correlation of noise [64]. All of which are key figures of merit for circuit designers.

Thus, we present the first ever fully self-consistent solver for small signal and noise analyses of a system of equations consisting of PE, SE, and BE in a nanoscale nMOSFET.

### 1.3 Structure of the Manuscript

This manuscript has a strong emphasis on completeness, showing all necessary calculations and explaining all of the details necessary in order to reproduce the results. Often problematic is the discretization where ambiguities may arise and a lot of errors appear in the implementation. Hence, discretizations of all equations are provided and difficult or problematic aspects are pointed out and shown in which manner to resolve. Furthermore, significant reductions in the size of the system of equations by eliminating equations and by exploiting degeneracy are possible and are addressed in detail.

The contents are mostly arranged according to the structure of the problem. Chapter 2 introduces the Poisson, Schrödinger, and Boltzmann equations in their stationary form and gives instructions on how to transform and discretize them in order to solve them deterministically. Most of the procedure has already been shown elsewhere and references are included where appropriate. However, the full Newton-Raphson approach of Sect. 2.5.2 has first been implemented in the course of this work.

Once the stationary solution can be computed, the linearization of the equations yields a system of equations for the small signal response which is described

in Chap. 3 along with a suitable terminal current definition. It is also addressed how the symmetries of real devices put strong constraints on the symmetries of the discretized system of equations. These constraints are not fulfilled for arbitrary discretizations and therefore Chap. 3 also deals with the resolution of these issues.

In Chap. 4, the Langevin-source approach is applied to the system where we derive a computationally feasible form to solve the system directly for the Green's functions of the terminal currents. We also address the computation of the power spectral density of the terminal current noise.

To support the legibility of this manuscript, the reader is encouraged to refer to the appendix if they are in doubt about the meaning of a symbol or the definition of a quantity. A nomenclature has been included in the end which contains all symbols and their descriptions as well as references to their definition in this manuscript. Furthermore, App. A contains definitions of all quantities which are generally needed during the work on such a simulator.



## Chapter 2

# Stationary Equations

In this chapter the three main equations are introduced. The Poisson equation (PE) which yields the electric potential, the Schrödinger equation (SE) which returns the wave functions and eigenvalues of the confined electron gas, and the Boltzmann equation (BE) which describes the transport of carriers through the device.

We begin with a quick tour through the whole procedure which uses an exemplary device identical to the one used later in the results in Chap. 5. Then we will dive into the details of each equation and how they can be fitted together. Each equation has its own respective section which contains all information to successfully set up a numerical solver. This includes the discretization scheme used in this work where possible sources for errors and ambiguities are addressed and resolved.

We end this chapter by showing how the PE, SE, and BE can be solved self-consistently. This can either be achieved by an iteration over all three equations or by a Newton-Raphson approach to solve the equations simultaneously.

The contents of this chapter are well known and documented in literature. References will be given where appropriate.

### 2.1 Overview

Our goal is the simulation of the noise in a nanoscale nMOSFET. In order to compute the noise, we first need to understand how the device behaves under stationary operating conditions. This is the subject of the present chapter.

In general, the BE is so demanding that we cannot use some general solver, apply it to an arbitrary problem, and see it solved straightforwardly. As is usual in the domain of partial differential equations, methods and techniques need to be delicately tailored to a specific problem, requiring the carefulness and mental acuity of an experienced scientist to validate the success or understand

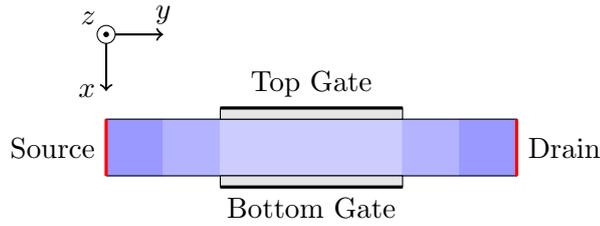


Figure 2.1: The fully depleted nanoscale double gate nMOSFET with silicon channel under consideration. Electron transport is in  $y$ -direction from source to drain. The shades in the channel region indicate the donor doping density where the contact regions are assumed to be doped heavily. Moreover, the channel width ( $x$ -direction) is assumed to be very small such that the use of the SE is adequate. The device is homogeneous in  $z$ -direction.

the failure modes of simulations.

Since this work is aimed at the deterministic simulation of noise, which in itself is rather involved, we cannot hope to solve this problem in general and for arbitrary devices. Instead we follow along the approach of Ref. [29] which considers a fully depleted double gate nMOSFET with a silicon channel as shown in Fig. 2.1. The carriers are confined in  $x$ -direction, transport happens in  $y$ -direction from source to drain, and the  $z$ -direction is considered homogeneous. We assume that the width of the channel in  $x$ -direction is quite small – somewhere in the nanometer range.

In order to understand how biases applied to the contacts and the doping profile as well as carrier densities influence the electric potential, we need to solve the 2D PE. However, a qualification that comes with nanoscale devices is that carriers close to a very thin oxide do not behave classically anymore. Instead we observe that carriers are pushed away from the oxide and their previously continuous range of energy states fans out into a discrete set of energy levels [65, 39]. While there are approximations to this effect which work to some extent [38], we can only accurately describe what happens if we solve the SE itself.

We want to avoid bearing the brunt of a full quantum transport approach but rather simulate transport classically with the BE while still including the effects of the quantum mechanical carrier confinement. To this end, we will use the dimensional splitting technique employed in Ref. [65, 29, 39], where we use the electric potential to evaluate the 1D SE in  $x$ -direction, for each position in  $y$ -direction. This yields a set of eigenstates – which we will henceforth call *subbands* –, each with a discrete energy and a probability amplitude in  $x$ -direction. For this reason, the  $x$ -direction is also referred to as the *confinement direction*.

Once the energy states of the confined carriers are known, we can use their gradients in  $y$ -direction in order to find the electric force acting upon them within their respective subbands. Then we can model the transport separately for each

subband using the BE evaluated in  $y$ -direction. The hopping of carriers between these subbands is described using scattering rates that are proportional to the overlap of the respective probability amplitudes of the initial and final subbands of the scattering process. In the remainder, we will refer to the  $y$ -direction as the *transport direction*.

The whole procedure can be imagined as an interdependent set of equations consisting of the PE, SE, and BE. The PE requires a carrier density and yields the quasistatic potential. The electric potential is used to solve the SE on slices in confinement direction to obtain subband energies and probability amplitudes. The subband energies and probability amplitudes are used as input for the BE in order to compute the carrier distribution function. The distribution function is used to compute the density in order to solve the PE which completes the cyclic dependence of the equations on each other.

Although not explicitly written down, the implementation of the solver described throughout this work is handled in *simulator units*. That means any physical quantity is normalized before it is used in the implementation and the normalization is removed if the results need to be returned. The normalization is such that quantities in simulator units usually values around unity from which better numerical properties follow. Since this procedure is merely an implementation detail, it is described in App. B.

Many concepts of the present chapter are discussed – to varying degrees of depth – in Ref. [29], however, we will go into more detail and be more explicit in order to avoid any ambiguities in the numerical implementation.

## 2.2 Poisson Equation

The PE is one of the pillars of modern device simulation. It can be directly derived from the stationary Maxwell equations and it describes the electric potential in a charged medium with some permittivity. The electric potential within a semiconductor device is a fundamental quantity in many respects. First, its negative gradient yields the electric field, which is responsible for the acceleration of charged carriers. Second, through its dependence on the space charge density, it describes all built-in fields in the vicinity of interfaces of differently doped regions. Third, through Dirichlet boundary conditions, the biases applied to the contacts are contained in the electric potential. In this section, we will discuss how the PE is set up for a two-dimensional device and how it is discretized in our work.

### 2.2.1 Equation

The PE reads

$$F^{\text{PE}}(\mathbf{r}) := \nabla_{\mathbf{r}} \cdot (\kappa(\mathbf{r}) \nabla_{\mathbf{r}} V(\mathbf{r})) + \rho(\mathbf{r}) = 0, \quad (2.1)$$

where  $\nabla_{\mathbf{r}}$  is the nabla operator,  $\kappa$  is the dielectric constant,  $V$  is the electric potential, and  $\rho$  is the space charge density. The vector  $\mathbf{r} = (x, y)^t$  is considered two-dimensional since we assume that the  $z$ -direction of our device is homogeneous (cf. Fig. 2.1) and therefore any  $z$ -dependence can be integrated out of the PE. Note that for later convenience we put the PE in Eq. (2.1) in to a form, where we can easily reference it completely using  $F^{\text{PE}}$ .

Throughout this work, it will be assumed that the electric potential is *quasi-stationary*. While we will later assume time-dependence, the associated wave lengths are always assumed to be longer than the device such that Eq. (2.1) holds.

The space charge density  $\rho$  in a semiconductor is determined by the ionized donor density  $N_D$ , the ionized acceptor density  $N_A$ , as well as the electron density in the conduction band  $n_{3\text{D}}$  and the hole density in the valence band  $p_{3\text{D}}$ . Since we will only consider a fully depleted MOSFET with a doping density exclusively consisting of donors, we can set the acceptor density to zero and ignore the hole density:

$$\rho(\mathbf{r}) = q(N_D(\mathbf{r}) - n_{3\text{D}}(\mathbf{r}) - N_A(\mathbf{r}) + p_{3\text{D}}(\mathbf{r})) = q(N_D(\mathbf{r}) - n_{3\text{D}}(\mathbf{r})), \quad (2.2)$$

where  $q$  is the positive elementary charge.

To stabilize the PE, we use the usual non-linear transformation of the carrier density given by [66]

$$n_{3\text{D}}(\mathbf{r}) = n_i \exp\left(\frac{V(\mathbf{r}) - \varphi_n(\mathbf{r})}{V_T}\right), \quad (2.3)$$

where  $n_i$  is the intrinsic carrier density,  $\varphi_n$  is the quasi Fermi potential of electrons, and  $V_T = k_B T / q$  is the thermal voltage with the Boltzmann constant  $k_B$  and the temperature  $T$ . Note that, technically, Eq. (2.3) is only valid if we assume a Maxwell-Boltzmann distribution of electrons, i.e. if we omit the Pauli principle. However, in the PE, we use Eq. (2.3) solely as a mathematical procedure for stabilization. Thus, when we consider Fermi-Dirac statistics, the quasi Fermi potential  $\varphi_n$  will implicitly include the effects of the Pauli principle.

### 2.2.2 Boundary Conditions

The PE is a second order differential equation which requires us to have two boundary conditions per dimension. We will restrict our discussion to the device

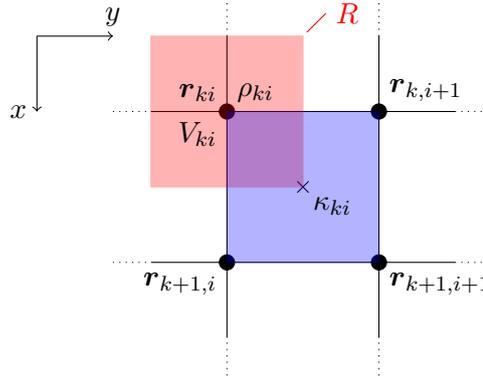


Figure 2.2: The black circles indicate direct grid points and the cross indicates an adjoint grid point. The space charge density  $\rho$  and the electric potential are defined on direct grid points, while the dielectric constant  $\kappa$  is defined on the adjoint grid points. Each quantity is considered constant in the rectangular region closest to the grid point of its definition, indicated by the red and blue rectangles.

shown in Fig. 2.1. On the gates, we assume Dirichlet boundary conditions as

$$V(\mathbf{r}) \Big|_{\mathbf{r} \in \partial D_{\text{TG/BG}}} = V_{\text{appl}}^{\text{TG/BG}} + \varphi_{\text{MS}}, \quad (2.4)$$

where  $V_{\text{appl}}^{\text{TG/BG}}$  is the applied bias at the top gate (TG) or bottom gate (BG),  $\varphi_{\text{MS}}$  is the metal-semiconductor work function difference, and  $\partial D_{\text{TG}}$  and  $\partial D_{\text{BG}}$  are the sets comprising the TG and BG contacts, respectively (cf. Fig. 2.1).

Everywhere else on the boundary of the device, we apply Neumann boundary conditions. In particular, the potential at the source and drain contacts will be adjusted through the boundary conditions of the BE, which will be discussed later on in Sect. 2.4.3.

### 2.2.3 Discretization

We use the finite volume method [67] to discretize the PE. We assume a two-dimensional rectangular tensor grid with grid points at positions denoted by  $\mathbf{r}_{ki} = (x_k, y_i)^t$ , where  $k$  and  $i$  are the indices of the grid points in  $x$ - and  $y$ -direction, respectively. The space charge density and the electric potential are defined on direct grid points while the dielectric constant is defined on adjoint grid points. By direct grid points, we refer to the actual grid points of our spatially discretized system, and by adjoint grid points, we mean points lying in between those points as in Fig. 2.2. Each function is discretized either on the direct or on the adjoint grid and is considered constant in the rectangular region surrounding the respective grid point.

Integrating Eq. (2.1) over a box  $R$  surrounding a direct grid point (cf. Fig. 2.2) and using Gauss's theorem yields

$$\int_{\partial R} d\mathbf{F} \cdot (\kappa(\mathbf{r}) \nabla_{\mathbf{r}} V(\mathbf{r})) + \int_R dV \rho(\mathbf{r}) = 0$$

which can be readily discretized on our tensor grid considering the domains of definition of each quantity:

$$\begin{aligned} F_{ki}^{\text{PE}} := & \underbrace{\kappa_{k,i-1} \frac{x_{k+1} - x_k}{2} \frac{V_{k,i-1} - V_{k,i}}{y_i - y_{i-1}}}_{A} + \underbrace{\kappa_{k,i-1} \frac{y_i - y_{i-1}}{2} \frac{V_{k+1,i} - V_{k,i}}{x_{k+1} - x_k}}_{B} \\ & + \underbrace{\kappa_{k,i} \frac{y_{i+1} - y_i}{2} \frac{V_{k+1,i} - V_{k,i}}{x_{k+1} - x_k}}_{C} + \underbrace{\kappa_{k,i} \frac{x_{k+1} - x_k}{2} \frac{V_{k,i+1} - V_{k,i}}{y_{i+1} - y_i}}_{D} \\ & + \underbrace{\kappa_{k-1,i} \frac{x_k - x_{k-1}}{2} \frac{V_{k,i+1} - V_{k,i}}{y_{i+1} - y_i}}_{E} + \underbrace{\kappa_{k-1,i} \frac{y_{i+1} - y_i}{2} \frac{V_{k-1,i} - V_{k,i}}{x_k - x_{k-1}}}_{F} \\ & + \underbrace{\kappa_{k-1,i-1} \frac{y_i - y_{i-1}}{2} \frac{V_{k-1,i} - V_{k,i}}{x_k - x_{k-1}}}_{G} + \underbrace{\kappa_{k-1,i-1} \frac{x_k - x_{k-1}}{2} \frac{V_{k,i-1} - V_{k,i}}{y_i - y_{i-1}}}_{H} \\ & + \rho_{k,i} \frac{x_{k+1} - x_{k-1}}{2} \frac{y_{i+1} - y_{i-1}}{2} \\ = & 0. \end{aligned} \tag{2.5}$$

Figure 2.3 illustrates the approach and the five-point stencil of our discretization scheme. Note that for grid points on the device boundaries, the line segments outside of the device do not contribute.

Due to the way the finite volume method works, for any grid point on a boundary, Eq. (2.5) automatically ensures that no flux past the boundary occurs and therefore Neumann boundary conditions are applied by construction.

For the gates, where we want to use Dirichlet boundary conditions, we can replace the equations referring to the respective boundary points by

$$F_{ki}^{\text{PE}} = V_{ki} - \left( V_{\text{appl}}^{\text{TG/BG}} + \varphi_{\text{MS}} \right) = 0, \quad \text{for } \mathbf{r}_{ki} \in \partial D_{\text{TG/BG}}.$$

## 2.2.4 Solution

There are two different ways in which we need to solve the PE in this work. First, we need to solve it in isolation and, second, we need to solve it as part of a larger system of equations with the BE and SE. If we solve the PE as a standalone equation, we express the charge  $\rho$  using Eq. (2.3) and employ the

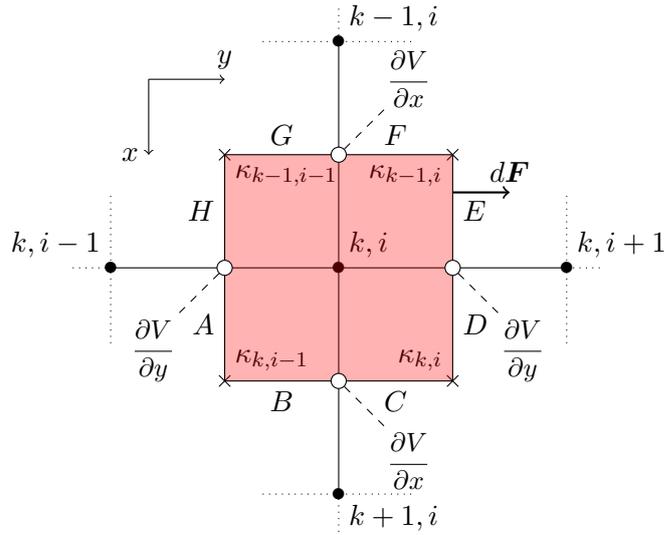


Figure 2.3: The box-integration of the PE. The integration measure  $d\mathbf{F}$  points outwards of the box. The dielectric constants  $\kappa$  are defined in the regions in between grid points. The derivatives evaluated on the surface of the box are defined as indicated by the white circles, aligned with the measure  $d\mathbf{F}$ . The eight line segments denoted by A to H correspond to the eight underbraced terms of Eq. (2.5).

Newton-Raphson method. The linear system to solve for each iteration is given by

$$\sum_{\ell, m} \frac{\partial F_{ki}^{\text{PE}}}{\partial V_{\ell m}} \delta V_{\ell m} = -F_{ki}^{\text{PE}}, \quad (2.6)$$

where the space charge derivative is given by

$$\frac{\partial \rho_{ki}}{\partial V_{\ell m}} = -\frac{q n_i}{V_T} \exp\left(\frac{V_{ki} - (\varphi_n)_{ki}}{V_T}\right) \delta_{k\ell} \delta_{im}.$$

The PE is an elliptic differential equation and therefore it is guaranteed to be solved uniquely with any initial guess. Thus, with any initial  $V$ , we can solve Eq. (2.6) and then update the potential as

$$V_{ki} \longrightarrow V_{ki} + \delta V_{ki}.$$

Then solve Eq. (2.6) again and update the potential. This process is repeated until the difference  $\delta V$  after each iteration becomes negligible which means we can assume that we are close enough to the solution. During this work, we use the sparse linear solver ILUPACK [68] to solve Eq. (2.6).

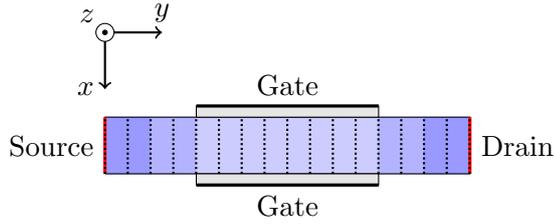


Figure 2.4: Slices (dotted lines) on which the one-dimensional SE is solved in  $x$ -direction.

If we solve the PE in conjunction with the BE, we need to express the space charge density as a function of the distribution function. This will be discussed in Sect. 2.5.2.

## 2.3 Schrödinger Equation

The SE is important for devices with very thin oxides where the details of the inversion density close to the oxide contribute significantly. It describes how carriers are pushed away from the oxide interface and how their states are quantized into a discrete set of energy levels. Since we are solely interested in the properties of electrons, we will restrict ourselves to this case.

The present discussion of the SE and its implementation is one of the simplest possible cases which can also be found in, e.g., Ref. [69].

### 2.3.1 Equation

A discussion about electrons in periodic crystal lattice potentials usually involves the simplification of the SE by splitting up the wave function into a known Bloch function and an unknown envelope function and subsequently reducing the full SE into an eigenvalue equation for the envelope function. The whole procedure is well known and widely documented in literature. Therefore we will not concern ourselves with the rationale leading to the application of the SE to this case and refer the reader to, e.g., Refs. [6, 65, 69, 70].

We will solve the one-dimensional SE in confinement direction along slices as indicated in Fig. 2.4. The stationary SE for the envelope function in one dimension given by

$$\varepsilon^s(y_0)\Psi^s(x, y_0) = \hat{H}\Psi^s(x, y_0) = \left( -\frac{\hbar^2}{2m_x} \frac{\partial^2}{\partial x^2} + \mathcal{V}(x, y_0) \right) \Psi^s(x, y_0), \quad (2.7)$$

where  $\varepsilon$  is the energy eigenvalue which is also referred to as the subband energy,  $\Psi$  is the envelope function and it will also be referred to as the wave function

or eigenfunction,  $\hat{H}$  is the Hamilton operator and  $m_x$  is the effective electron mass in  $x$ -direction. The superscript  $s$  indicates that we obtain an infinite set of eigenvalues and eigenfunctions from the SE. All quantities have an additional dependence on a position  $y_0$  in  $y$ -direction, since we cut the device into slices and solve the 1D SE along each of these slices. Therefore we obtain separate eigenvalues and eigenfunctions for each position  $y_0$  as the potential energy  $\mathcal{V}$  varies in this direction. The potential energy is simply given by the electric potential acting on electrons:

$$\mathcal{V}(x, y) = -qV(x, y). \quad (2.8)$$

Note that we assume that the mass  $m_x$  is constant, which will hold true throughout this work.

An important underlying assumption of the Hamiltonian of Eq. (2.7) is that the kinetic energy of an electron in the crystal lattice is given by  $p^2/2m_x$ , where  $p$  is its momentum. This is not true in general but in the case of silicon we can approximate the dispersion relation as a parabola – as is the case for free particles – and use the *effective* mass to parameterize the inertia resulting from the compound interaction of the underlying crystal lattice. The discussion of the band structure is given in more detail in Sect. 2.4.1.

A straightforward way to improve the accuracy of the band structure is to include a non-parabolicity factor as in Ref. [16, 71], however, this increases the difficulty of solving the SE significantly. In this work we will restrict ourselves to the case of a parabolic band structure where the effective mass  $m_x$  is constant throughout the semiconductor region, i.e. the Hamiltonian can be expressed as in Eq. (2.7).

We will only ever refer to the stationary SE since we assume – just as for the PE – that the subband energies  $\varepsilon^s$  and the wave functions  $\Psi^s$  are quasistationary, i.e. they have sufficiently short relaxation times that even in the presence of a time variation of the applied bias, they can follow along the perturbations in the steady state.

### 2.3.2 Boundary conditions

The potential barriers at each end of the potential well in confinement direction are formed by insulating materials. These potential barriers are in reality not infinitely high and therefore the wave function actually penetrates the barriers resulting in a non-zero chance to encounter carriers within the oxides. However, for practical purposes we make the assumption that they are indeed infinitely high and the wave functions vanish at the boundaries [69]. Thus, if we solve the SE on a domain  $x \in [0, L]$ , the boundary conditions read

$$\Psi^s(0, y_0) = \Psi^s(L, y_0) = 0. \quad (2.9)$$

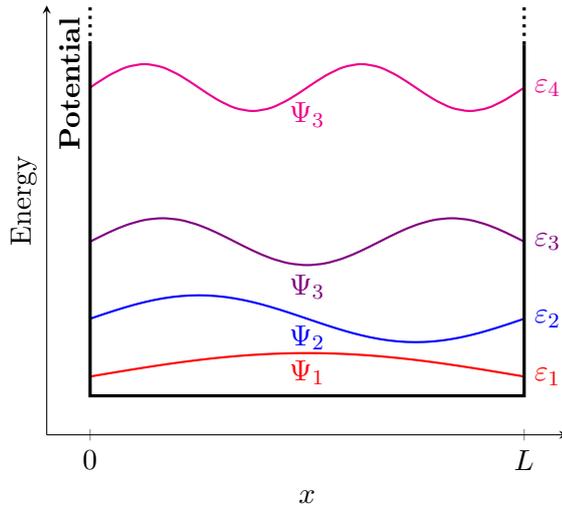


Figure 2.5: Illustration of the infinite square well potential with the solutions of the SE associated with the four lowest eigenvalues.

Thus, the basic problem we are dealing with is essentially similar to the infinite potential well problem whose solutions are illustrated in Fig. 2.5. Such a potential well gives rise to a set of bound states which are standing wave solutions associated with a set of increasing energy levels.

### 2.3.3 Discretization

The discretization of the SE is straightforward. For a set of grid points  $x_k = x_1, \dots, x_{N_x}$  and associated wave functions  $\Psi_k^s = \Psi_1^s, \dots, \Psi_{N_x}^s$ , we express the second derivative on a grid point  $x_k$  as a three-point stencil:

$$\left( \frac{\partial^2}{\partial x^2} \Psi(x) \right)_k = \frac{2}{x_{k+1} - x_{k-1}} \left( \frac{\Psi_{k+1} - \Psi_k}{x_{k+1} - x_k} - \frac{\Psi_k - \Psi_{k-1}}{x_k - x_{k-1}} \right),$$

where we omitted the superscript  $s$  and the dependence on  $y_0$  for brevity. Then the SE reads

$$\varepsilon \Psi_k = -\frac{\hbar^2}{m_x} \frac{1}{x_{k+1} - x_{k-1}} \left( \frac{\Psi_{k+1} - \Psi_k}{x_{k+1} - x_k} - \frac{\Psi_k - \Psi_{k-1}}{x_k - x_{k-1}} \right) - qV_k \Psi_k.$$

The boundary conditions, given by Eq. (2.9), translate to the discretized system as

$$\Psi_1 = \Psi_{N_x} = 0,$$

which can be applied by simply inserting them into the equations for the grid points 2 and  $N_x - 1$ .

### 2.3.4 Solution and Perturbation

The SE is an eigenvalue problem and in the discretized form, its Hamiltonian is given by a real, symmetric, and tridiagonal matrix which we solve with the routine `DSTEVX` contained in the `LAPACK` implementation of the Intel Math Kernel Library [72].

The one dimensional SE is almost trivial to solve. In practice, the elapsed time to solve the SE is negligible compared to what the BE needs.

Since the SE is an eigenvalue equation, we cannot straightforwardly include it in a self-consistent system of equations like the PE and BE – the details of which will be discussed in Sect. 2.5.2. For now, let us note that we need to know how the subband energies and wave functions change if the potential changes slightly. This is known as perturbation theory and to leading order – which is all we will need in this work – the expressions are fairly simple [73].

Consider a Hamiltonian  $\hat{H}$  of the stationary SE that has known solutions  $\Psi^s$  associated with known energy eigenvalues  $\varepsilon^s$ . Suppose we change the potential energy by a very small perturbation  $\delta\mathcal{V} = -q\delta V$ :

$$\hat{H} \longrightarrow \hat{H} + \delta\mathcal{V}.$$

Then, the eigenvalues and eigenfunctions will change in leading order as

$$\delta\varepsilon^s(y_0) = -q \int dx |\Psi^s(x, y_0)|^2 \delta V(x, y_0), \quad (2.10)$$

$$\delta\Psi^s(x, y_0) = -q \sum_{s' \neq s} \frac{\int dx' \Psi^{s'}(x', y_0) \delta V(x', y_0) \Psi^s(x', y_0)}{\varepsilon^s(y_0) - \varepsilon^{s'}(y_0)} \Psi^{s'}(x, y_0). \quad (2.11)$$

In order to speed up the evaluation of the SE, it is tempting to truncate the series of subband energies and wave functions at some subband index, however, this is not possible when we use Eq. (2.11) due to the sum over all subbands. Nevertheless, this is not a problem in practice as the one dimensional SE can be solved nearly instantly on modern computers.

## 2.4 Boltzmann Equation

The BE is the central equation of our model. The BE is far more difficult to understand than the PE or SE and it requires a more detailed explanation as well as a plethora of transformations to be tamed into a numerically feasible form.

This section is going to extend a bit further as quite a few important concepts have to be introduced in order to make sense of the rest of this manuscript. A certain amount of familiarity with the BE and the notation is required for the later chapters which we hope the reader will be able to pick up in this section.

However, the contents of this section are by no means new but have been extensively studied in Refs. [26, 28, 29, 40] for various cases. Here, we are going to consider the BE only for the case of a two-dimensional electron gas in order to remove any unnecessary distractions that come from a treatment aspiring to be general. Nevertheless, the concepts introduced here are applicable to other cases as well.

### 2.4.1 Equation

In the most general case, the stationary BE is given by

$$\left[ F^{\text{BE}} \right]^\nu(\mathbf{r}, \mathbf{k}) := L^\nu(\mathbf{r}, \mathbf{k}) - S^\nu(\mathbf{r}, \mathbf{k}) - \Gamma^\nu(\mathbf{r}, \mathbf{k}) = 0, \quad (2.12)$$

where

$$L^\nu(\mathbf{r}, \mathbf{k}) := \frac{1}{\hbar} \mathbf{F}^\nu(\mathbf{r}) \cdot \nabla_{\mathbf{k}} f^\nu(\mathbf{r}, \mathbf{k}) + \mathbf{v}^\nu(\mathbf{k}) \cdot \nabla_{\mathbf{r}} f^\nu(\mathbf{r}, \mathbf{k})$$

is the free streaming term,  $S$  is the scattering term discussed later in this section,  $\Gamma$  is a generation and recombination term discussed in Sect. 2.4.3,  $f$  is the distribution function of carriers,  $\mathbf{F}$  is the force acting on carriers,  $\mathbf{v}$  is the group velocity  $\nabla_{\mathbf{r}}$  is the Nabla operator w.r.t. to the spatial coordinates, and  $\nabla_{\mathbf{k}}$  is the Nabla operator w.r.t.  $\mathbf{k}$ -space. Note that we cast the BE into a form where we can easily refer to it as  $F^{\text{BE}}$  which will prove convenient later on.

The superscript  $\nu = (v, s)$  indicates that transport can happen along different channels which, in our case, involve the valleys of the band structure of silicon  $v$  and the subbands  $s$  emerging from the electron confinement. Unlike the free streaming term  $L$ , the scattering term  $S$  can couple these channels by establishing a mutual dependence by a scattering process.

In principle we would also need to carry an index denoting the electron spin, but none of the scattering transitions we incorporate will have the ability to flip the spin and therefore spin remains unchanged and can be ignored. It follows that we may regard the BE of Eq. (2.12) as describing only, say, spin up electrons and then multiply the resulting observables, like electron densities and currents, with a factor of two since for every spin up electron, there is also a spin down electron. We will discuss the multiplicities arising from degeneracies of the BE in more detail in Sect. 2.4.6.

Now let us simplify the BE for the case we are interested in. Recall that the device in Fig. 2.1 was assumed to be homogeneous in  $z$ -direction. Moreover, we solve the SE in  $x$ -direction. Therefore, no quantity can depend on either  $x$  or  $z$  and it follows that the force can only be exerted in  $y$ -direction from source to drain, i.e.  $\mathbf{F}(\mathbf{r}) \equiv F(y)\mathbf{e}_y$ , and  $\nabla_{\mathbf{r}} f^\nu(\mathbf{r}, \mathbf{k}) = \frac{\partial}{\partial y} f^\nu(y, \mathbf{k})\mathbf{e}_y$ . Then the free

streaming term of the BE reduces to

$$L^\nu(y, \mathbf{k}) = \frac{1}{\hbar} F^\nu(y) \frac{\partial}{\partial k_y} f^\nu(y, \mathbf{k}) + v_y^\nu(\mathbf{k}) \frac{\partial}{\partial y} f^\nu(y, \mathbf{k}), \quad (2.13)$$

where we defined the  $y$ -direction of the group velocity as  $v_y$ , i.e.  $\mathbf{v} \cdot \mathbf{e}_y =: v_y$ .

Furthermore, the standing wave solutions of the SE in  $x$ -direction are equivalent to superpositions of waves propagating parallel and antiparallel to the  $x$ -direction which are associated with wave numbers  $\pm k_x \mathbf{e}_x$ . Thus, by solving the SE, we already fixed the wave number in  $x$ -direction, leaving only the other two dimensions as variables for the BE which is equivalent to a two-dimensional electron gas:

$$\mathbf{k} := \begin{pmatrix} k_y \\ k_z \end{pmatrix}.$$

The Lorentz force  $F$  is classically given by the electric field which is the negative gradient of the electric potential. But in our case, we use the SE to obtain a spectrum of discrete eigenvalues that are occupied by electrons to various degrees. These so called subband energies may vary along the transport direction of the device representing the varying potential energy similar to the potential energy a charge feels in an electric field. Therefore our force in the channels of transports, i.e. in the various subbands, is given by the gradient of the subband energy:

$$F^\nu(y) = - \frac{\partial}{\partial y} \varepsilon^\nu(y) \quad (2.14)$$

and therefore the force is both valley and subband dependent.

The scattering term  $S$  of Eq. (2.12) in 2D  $\mathbf{k}$ -space and for one spatial dimension is given by

$$S^\nu(y, \mathbf{k}) = \Omega \sum_{\eta, \nu'} \int \frac{d^2 k'}{(2\pi)^2} \left[ (1 - f^\nu(y, \mathbf{k})) S_{\eta}^{\nu, \nu'}(y; \mathbf{k}, \mathbf{k}') f^{\nu'}(y, \mathbf{k}') \right. \\ \left. - (1 - f^{\nu'}(y, \mathbf{k}')) S_{\eta}^{\nu', \nu}(y; \mathbf{k}', \mathbf{k}) f^\nu(y, \mathbf{k}) \right], \quad (2.15)$$

where  $\Omega$  is the system area and  $S_{\eta}^{\nu, \nu'}(y; \mathbf{k}, \mathbf{k}')$  is the transition rate at position  $y$  from the state  $(\nu', \mathbf{k}')$  to the state  $(\nu, \mathbf{k})$  by the scattering process  $\eta$ . The specific transition rates used in this work will be discussed in Sect. 2.4.2. Note that Eq. (2.15) includes the Pauli exclusion principle via the factors  $(1 - f)$  which reduce the scattering rate if the distribution function  $f$  of the final state is close to unity.

In the remainder of this work, we will usually include the Pauli principle but occasionally we will compare to the case without the Pauli principle to acknowledge its importance. In that case we use the scattering term

$$\begin{aligned} \left[ S^{\text{no Pauli}} \right]^\nu(y, \mathbf{k}) = \Omega \sum_{\eta, \nu'} \int \frac{d^2 k'}{(2\pi)^2} & \left[ S_\eta^{\nu, \nu'}(y; \mathbf{k}, \mathbf{k}') f^{\nu'}(y, \mathbf{k}') \right. \\ & \left. - S_\eta^{\nu', \nu}(y; \mathbf{k}', \mathbf{k}) f^\nu(y, \mathbf{k}) \right]. \end{aligned} \quad (2.16)$$

Any derivations shown in the remainder of this work are equally applicable to the scattering term without the Pauli principle.

### Band Structure

The band structure of silicon is well known and described in many text books on semiconductor devices [6, 69, 74]. In this work, we will only consider the six most important valleys – named  $X$ -valleys – which represent the six absolute minima of the silicon band structure. In addition, we assume that the crystal coordinate system is aligned with the device coordinate system, i.e. two  $X$ -valleys are located on each coordinate axis in  $\mathbf{k}$ -space. Figure 2.6 shows equienergy surfaces for the six  $X$ -valleys of the silicon band structure. The equienergy surfaces can be approximated as prolate ellipsoids of revolution where the large semi-axes are aligned with the axes of  $\mathbf{k}$ -space. An electron located in one of these valleys will appear to have a large mass when it is accelerated in the direction of the large semi-axis, while its mass appears small when it travels in the direction of one of the small semi-axes.

Close to the minimum of an  $X$ -valley, we can assume that the band structure is approximately parabolic where the inertia of electrons is quantified through the longitudinal and transversal *effective masses* given by [75]

$$m_l = 0.919 m_0, \quad m_t = 0.19 m_0, \quad (2.17)$$

respectively. These effective masses are the cumulative effect of the underlying silicon lattice on an electron. Therefore the dispersion relation – or kinetic energy – in a valley  $v$  can be expressed as [69]

$$\varepsilon_{\text{tot}}^v(k_x, k_y, k_z) = \frac{\hbar^2}{2} \left( \frac{(k_x - K_x^v)^2}{m_x^v} + \frac{(k_y - K_y^v)^2}{m_y^v} + \frac{(k_z - K_z^v)^2}{m_z^v} \right),$$

where  $\mathbf{K}^v = (K_x^v \ K_y^v \ K_z^v)^t$  is the displacement of the valley minimum  $v$  in 3D  $\mathbf{k}$ -space. Note that the effective mass of electrons is non-isotropic and valley-dependent due to the orientation of the ellipsoidal shape of the equienergy surfaces of the valleys (cf. Fig. 2.6).

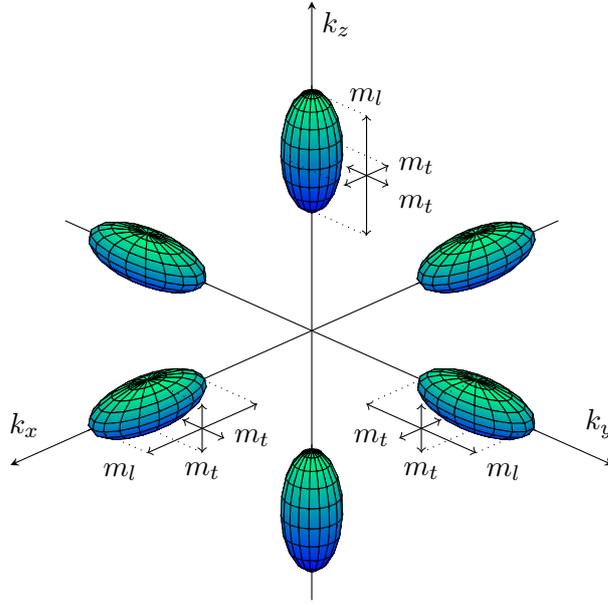


Figure 2.6: Illustration of the equienergy surfaces of the six  $X$ -valleys of the silicon band structure. The equienergy surfaces can be approximated as prolate ellipsoids of revolution where the large semi-axes align with the axes of  $\mathbf{k}$ -space. Due to the non-isotropy of the valleys, electrons have different effective masses –  $m_l$ ,  $m_t$ , or a combination thereof – depending on their direction of travel as indicated in the figure.

To simplify matters, we can treat the kinematics of each valley separately and use a valley dependent coordinate translation given by

$$(k_x \ k_y \ k_z)^t \longrightarrow (k_x \ k_y \ k_z)^t + (\mathbf{K}^v)^t.$$

Then, the dispersion relation in the valley  $v$  simplifies to

$$\varepsilon_{\text{tot}}^v(\mathbf{k}) = \frac{\hbar^2}{2} \left( \frac{k_x^2}{m_x^v} + \frac{k_y^2}{m_y^v} + \frac{k_z^2}{m_z^v} \right), \quad (2.18)$$

where from here on out it is implicitly understood that vectors in  $\mathbf{k}$ -space are always given relative to the minimum  $\mathbf{K}^v$  of the valley  $v$  under consideration.

Due to the dimensional splitting into the SE and BE, the  $x$ -component  $k_x$  is no longer a kinematic variable but is fixed by the standing wave solution of the SE. Therefore  $\mathbf{k}$ -space is reduced to the 2D case and the equienergy surfaces of the band structure are projected on the  $yz$ -plane of  $\mathbf{k}$ -space as in Fig. 2.7. The total energy of an electron is the sum of its eigenvalue in confinement direction and the kinetic energy in the remaining kinematic  $yz$ -plane:

$$\varepsilon_{\text{tot}}^{v,s}(\mathbf{k}) = \varepsilon^{v,s} + E^v(\mathbf{k})$$

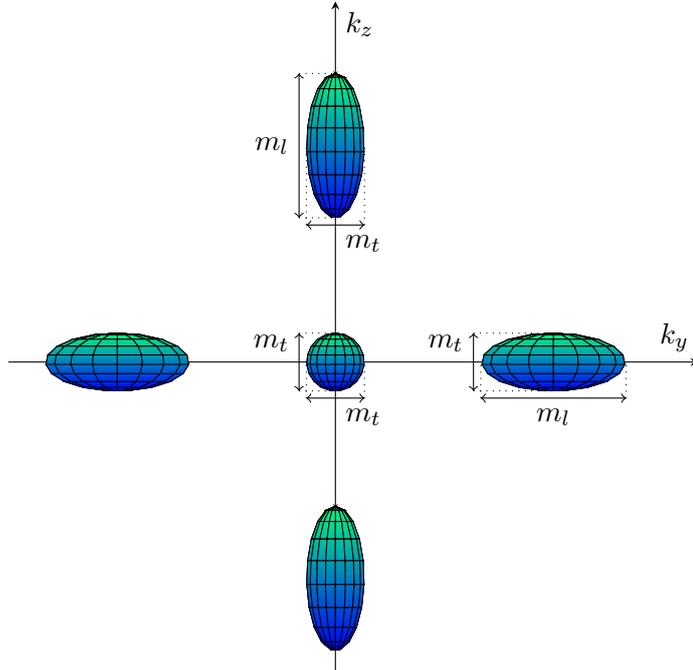


Figure 2.7: Illustration of the equienergy surfaces of the six X-valleys of the silicon band structure in the 2D  $\mathbf{k}$ -space of the BE. The effective masses of electrons in the two directions of travel are as indicated.

where  $\varepsilon^{v,s}$  is the energy eigenvalue of the SE of Eq. (2.7) and the dispersion relation in the remaining kinematic  $yz$ -plane reads

$$E^v(\mathbf{k}) = \frac{\hbar^2}{2} \left( \frac{k_y^2}{m_y^v} + \frac{k_z^2}{m_z^v} \right). \quad (2.19)$$

Note how the continuum of values of the  $k_x$ -component is reduced to the set of discrete subbands  $s$  associated with a fixed energy and therefore a fixed value of  $k_x$ .

### 2.4.2 Scattering Terms

When moving through matter, charged carriers experience a resistance<sup>1</sup> which is a consequence of microscopic scattering processes. In the following, we will describe the scattering of carriers which is well understood for silicon and can be looked up in, e.g., Ref. [69]. One of the dominant effects in semiconductors is the scattering with phonons, which are the quasiparticles of lattice vibrations. Consider a semiconductor like silicon with a finite temperature. Then its crystal

<sup>1</sup>Neglecting extreme phenomena such as superconductivity.

lattice of silicon atoms vibrates and these vibrations come in different forms which can be characterized as acoustical and optical as well as transversal and longitudinal modes. The vibrational modes – or phonons – warp the underlying crystal lattice and carriers traversing the lattice will experience the electromagnetic effects of the warped lattice. The exposure of carriers to the presence of an electromagnetic force, like when a phonon and an electron experience each other's presence, can be modeled by a one-time scattering event and using the statistical distributions of electrons and phonons, we can find the rate at which these scattering events occur. The rate of scattering in the BE is tantamount to the resistance carriers experience in the crystal.

Another important scattering process that we are interested in is specific to MOSFETs. The fabrication of the gate oxide on the silicon channel always leads to an imperfect oxide-silicon interface. Usually this is referred to as the surface roughness which is quantified statistically using a mean height of the imperfections and a correlation length. The surface roughness warps the silicon crystal on a microscopic scale which the carriers feel in the vicinity of the interface. Similar to the phonon scattering process, we can model the surface roughness scattering as instantaneous one-time scattering events and subsequently obtain a rate of scattering with the carrier distribution.

Finally, there is ionized impurity scattering. Transistors are doped with foreign atoms to ensure adequate amounts of carriers in the conduction band. But these doping atoms are ionized defects in the crystal lattice which perturb the uniformity of the crystal lattice and which carriers experience as Coulomb scattering targets. The difficulty in the impurity scattering is that the computed scattering terms diverge unless the screening of the ionized impurities by carriers is included. This presents a problem in the approach employed in this work. For now, it suffices to say that in the channel region of our devices, where the interesting transport phenomena happen, the ionized impurity scattering is negligible due to the low density of dopants. In the contact regions, where the doping is high, it acts as a resistance and can be modeled via other means.

All other scattering processes can be neglected for the purposes of our present work.

### General Approach

The scattering processes discussed in this work have been studied extensively in the quantum mechanical framework and they rely on a deep knowledge of the underlying interactions. One of the main assumptions of scattering theory is that the incoming and outgoing particles can be considered *free* when they are still far apart. Usually we know the exact solution of free particles and therefore we model them as approaching each other from infinity as free particles and

treat the interaction as a perturbation of the free state.<sup>2</sup>

Carriers in condensed matter are never free in the sense that there is nothing to interact with in their environment. Instead, they are subject to the sum of the potentials of the underlying lattice. However, the sum of all interactions with the underlying crystal lattice gives rise to the band structure, i.e. a dispersion relation that describes how particles experience inertia when they are accelerated. Moreover, considering that particles prefer occupying the lowest accessible energies in equilibrium, we can often approximate the band structure as a dispersion relation similar to the dispersion relation of free particles by, e.g., defining effective masses as in the parabolic band structure model (see Sect. 2.4.1). Once we know a dispersion relation for a band, a scattering process can be modeled as a perturbation of that state.

Perturbation theory results in an infinite series of terms that – in its totality – represents the integral over the interaction of the particles' trajectories. Individually each of these terms represents a certain amount of scattering vertices. If the interaction is very weak, we can safely ignore it at large distances, however, when the two particles are at a very close range, even the faintest interaction will quickly change both particles' trajectories. Omitting all but the leading terms of the perturbation expansion is called the Born approximation [73] which is equivalent to assuming that the interaction is instantaneous.

Now, if we assume that we have a carrier in the state  $\alpha$  and a scattering target that creates a potential  $\mathcal{V}$  felt by the carrier, we may formulate the rate of such a scattering process with Fermi's Golden Rule [73]

$$S_{\alpha \rightarrow \beta} = \frac{2\pi}{\hbar} |M_{\alpha \rightarrow \beta}|^2 \delta(E_{\alpha} - E_{\beta}), \quad (2.20)$$

where  $S_{\alpha \rightarrow \beta}$  is the transition rate of a scattering process from state  $\alpha$  to state  $\beta$ ,  $M_{\alpha \rightarrow \beta} = \langle \beta | \mathcal{V} | \alpha \rangle$  is the matrix element of the incoming particle in state  $\alpha$  scattering via the potential  $\mathcal{V}$  to an outgoing state  $\beta$ , and  $E_{\alpha}$ ,  $E_{\beta}$  are the energies of the respective states.

In order to streamline our notation, we are going to introduce a factor we call the transition rate coefficient  $c_{\alpha \rightarrow \beta}$  as

$$S_{\alpha \rightarrow \beta} =: \frac{1}{\Omega} c_{\alpha \rightarrow \beta} \delta(E_{\alpha} - E_{\beta}) \quad (2.21)$$

which will be useful later on. Here,  $\Omega$  is the system area present in all transition rates which cancels with the system area found in the scattering integral of Eq. (2.15).

We will not concern ourselves with the details of the derivation of the matrix elements for each scattering process as they can be found in literature. Hence,

---

<sup>2</sup>Perturbation theory is only valid if the interaction is weak enough so that the perturbation series converges.

we are going to present the matrix elements and reference appropriate text books for the reader to study if need be.

### Phonon Scattering

Phonon scattering is of great importance to our work as it is the mechanism by which the temperature of the crystal lattice affects carriers and vice versa. The derivation of the matrix element takes some effort and in order to get correct results it has to be approached quantum mechanically by means of the second quantization [69]. This results in two distinct matrix elements for absorption and emission of a phonon by an electron which are given by

$$|M^{(\text{ab});\nu,\nu'}(y; \mathbf{k}, \mathbf{k}')|^2 = \delta_{\mathbf{k}',(\mathbf{k}+\mathbf{q})} \frac{\pi\hbar}{\rho\Omega} \int_{q_x} \frac{|D(\mathbf{Q})|^2}{\omega_\lambda(\mathbf{Q})} n_\lambda(\mathbf{Q}) |G^{\nu,\nu'}(y; q_x)|^2 dq_x, \quad (2.22)$$

$$|M^{(\text{em});\nu,\nu'}(y; \mathbf{k}, \mathbf{k}')|^2 = \delta_{\mathbf{k}',(\mathbf{k}-\mathbf{q})} \frac{\pi\hbar}{\rho\Omega} \int_{q_x} \frac{|D(\mathbf{Q})|^2}{\omega_\lambda(\mathbf{Q})} (n_\lambda(\mathbf{Q}) + 1) |G^{\nu,\nu'}(y; -q_x)|^2 dq_x, \quad (2.23)$$

where  $\nu, \nu'$  are the compound valley and subband indices;  $\mathbf{k}, \mathbf{k}'$  are the initial and final electron two-momenta, respectively;  $\lambda$  indicates the polarization of the phonon;  $\mathbf{Q} = (q_x, \mathbf{q})$  is the phonon three-momentum;  $\rho$  is the mass density of the silicon crystal;  $\omega_\lambda$  is the dispersion relation of the phonon with polarization  $\lambda$ ;  $n_\lambda(\mathbf{Q})$  is the number of phonons with polarization  $\lambda$  and three-momentum  $\mathbf{Q}$ ; and

$$G^{\nu,\nu'}(y; q_x) = \frac{1}{2\pi} \int dx (\Psi^\nu(x, y))^* \Psi^{\nu'}(x, y) e^{iq_x x}$$

is the overlap integral of the wave functions determined by the SE. The deformation potential of the phonons is given by

$$D(\mathbf{Q}) = \begin{cases} \frac{i}{2} \sum_{i,j=1}^3 \Xi_{ij} (e_{\lambda,i} Q_j + e_{\lambda,j} Q_i), & \text{for acoustical phonons,} \\ D_{\text{op}}, & \text{for optical phonons,} \end{cases}$$

where  $\Xi_{ij}$  are the components of a  $3 \times 3$ -matrix collecting the deformation potentials of acoustical phonons and  $e_{\lambda,i}$  and  $Q_i$  are the components of the polarization vector  $\mathbf{e}_\lambda$  and the phonon three-momentum  $\mathbf{Q}$ , respectively.

With the matrix elements, we can write down the transition rate according to Fermi's Golden Rule of Eq. (2.20) as

$$S_{\text{phonon}}^{\lambda;\nu,\nu'}(y; \mathbf{k}, \mathbf{k}') = \frac{2\pi}{\hbar} |M^{(\text{ab});\nu,\nu'}(y; \mathbf{k}, \mathbf{k}')|^2 \delta(\varepsilon_{\text{tot}}^\nu(y, \mathbf{k}) - \varepsilon_{\text{tot}}^{\nu'}(y, \mathbf{k}') + \hbar\omega_\lambda(\mathbf{Q})) \\ + \frac{2\pi}{\hbar} |M^{(\text{em});\nu,\nu'}(y; \mathbf{k}, \mathbf{k}')|^2 \delta(\varepsilon_{\text{tot}}^\nu(y, \mathbf{k}) - \varepsilon_{\text{tot}}^{\nu'}(y, \mathbf{k}') - \hbar\omega_\lambda(\mathbf{Q})), \quad (2.24)$$

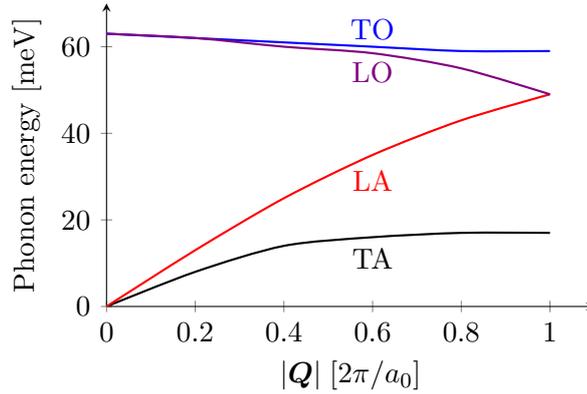


Figure 2.8: Illustration of phonon dispersion relations in silicon along the [100] direction of transverse optical (TO), longitudinal optical (LO), longitudinal acoustic (LA), transversal acoustic (TA) phonons (see Refs. [69, 76, 77]). The lattice spacing of silicon is given by  $a_0 \approx 0.54$  nm.

where the total energy is defined as the sum of the subband energy and the kinetic energy, i.e.

$$\varepsilon_{\text{tot}}^{\nu}(y, \mathbf{k}) = \varepsilon^{\nu}(y) + E^{\nu}(\mathbf{k}).$$

In practice the phonon transition rate of Eq. (2.24) is often simplified further. One observation about phonons is that they come in two flavors, acoustical and optical, which have very different dispersion relations. The dispersion relations of acoustic modes are – at low energies – approximately proportional to the magnitude of the momentum, while the optical modes have approximately constant energy, irrespective of their momenta, see Fig. 2.8.

The distinction between acoustical and optical modes leads to the approximation of Eq. (2.24) into two separate transition rates. The first one is the scattering by low-energy acoustical phonons which, in leading order, can be regarded as elastic and due to the insufficient energy transfer electrons remain in the valley they originated in. The second one is the scattering by optical modes which can be approximated to have fixed energy transfers. These phonons have sufficient energy to scatter electrons into another valley. In the following two sections we will find adequate representations for these transition rates. The content of these sections is already documented in literature in, e.g., Ref. [69].

### Intra-Valley Acoustic Phonon Scattering

To simplify the transition rate of Eq. (2.24) to something computationally feasible, we will use multiple approximations. First, as Fig. 2.8 suggests, the dis-

persion relation of acoustical phonons can be approximated as

$$\hbar\omega_\lambda(\mathbf{Q}) \approx \hbar v_\lambda |\mathbf{Q}|,$$

where  $v_\lambda$  is the sound velocity of acoustical phonons with polarization  $\lambda$ . Second, we use the energy equipartition approximation which states that the average number of phonons in equilibrium, which follow the Bose-Einstein statistics, can be approximated as

$$n_\lambda(\mathbf{Q}) = \frac{1}{\exp\left(\frac{\hbar\omega_\lambda(\mathbf{Q})}{k_B T}\right) - 1} = \frac{k_B T}{\hbar\omega_\lambda(\mathbf{Q})} - \frac{1}{2} + \mathcal{O}(\mathbf{Q}). \quad (2.25)$$

Third, we assume the transition rate to be isotropic which means that the deformation potential for acoustic phonons takes the form

$$D(\mathbf{Q}) \approx D_{\text{ac}} |\mathbf{Q}|, \quad (2.26)$$

where  $D_{\text{ac}}$  is the effective deformation potential.

Inserting these approximations into Eqs. (2.22), (2.23) and using Fermi's Golden Rule of Eq. (2.24), we find for the transition rate of acoustic phonons

$$\begin{aligned} S_{\text{ac}}^{\nu,\nu'}(y; \mathbf{k}, \mathbf{k}') &= \frac{2\pi}{\hbar} \delta(\varepsilon_{\text{tot}}^\nu(y, \mathbf{k}) - \varepsilon_{\text{tot}}^{\nu'}(y, \mathbf{k}')) \\ &\quad \times \left( |M^{(\text{ab});\nu,\nu'}(y; \mathbf{k}, \mathbf{k}')|^2 + |M^{(\text{em});\nu,\nu'}(y; \mathbf{k}, \mathbf{k}')|^2 \right) \\ &= \frac{2\pi}{\hbar} \frac{\pi\hbar}{\rho\Omega} \delta_{v,v'} \delta(\varepsilon_{\text{tot}}^\nu(y, \mathbf{k}) - \varepsilon_{\text{tot}}^{\nu'}(y, \mathbf{k}')) \\ &\quad \times \left[ \delta_{\mathbf{k}',(\mathbf{k}+\mathbf{q})} \int \frac{D_{\text{ac}}^2 |\mathbf{Q}|^2}{v_\lambda |\mathbf{Q}|} \left( \frac{k_B T}{\hbar v_\lambda |\mathbf{Q}|} - \frac{1}{2} \right) |G^{\nu,\nu'}(y; q_x)|^2 dq_x \right. \\ &\quad \left. + \delta_{\mathbf{k}',(\mathbf{k}-\mathbf{q})} \int \frac{D_{\text{ac}}^2 |\mathbf{Q}|^2}{v_\lambda |\mathbf{Q}|} \left( \frac{k_B T}{\hbar v_\lambda |\mathbf{Q}|} + \frac{1}{2} \right) |G^{\nu,\nu'}(y; -q_x)|^2 dq_x \right] \end{aligned}$$

Note that we added a Kronecker delta  $\delta_{v,v'}$  since electrons cannot scatter to other valleys. Since the only dependencies on  $q_y$  and  $q_z$  are through the magnitude of the phonon three-momentum  $|\mathbf{Q}|$ , the Kronecker deltas  $\delta_{\mathbf{k}',(\mathbf{k}+\mathbf{q})}$  and  $\delta_{\mathbf{k}',(\mathbf{k}-\mathbf{q})}$  are interchangeable. Moreover, the identity  $|G^{\nu,\nu'}(y; q_x)|^2 = |G^{\nu,\nu'}(y; -q_x)|^2$  can be trivially derived since the wave functions  $\Psi^\nu$  are real. Lastly, we can simplify the integral over  $q_x$  as

$$\int |G^{\nu,\nu'}(y; q_x)|^2 dq_x = \frac{1}{2\pi} \int dx |\Psi^{\nu'}(x, y)|^2 |\Psi^\nu(x, y)|^2 =: \frac{1}{2\pi} R^{\nu,\nu'}(y). \quad (2.27)$$

Thus, the final expression for intra-valley elastic acoustic phonon scattering reads

$$S_{\text{ac}}^{\nu,\nu'}(y; \mathbf{k}, \mathbf{k}') \approx \frac{1}{\Omega} \delta_{v,v'} \frac{2\pi k_B T D_{\text{ac}}^2}{\hbar \rho v_\lambda^2} R^{\nu,\nu'}(y) \delta(\varepsilon_{\text{tot}}^\nu(y, \mathbf{k}) - \varepsilon_{\text{tot}}^{\nu'}(y, \mathbf{k}'))$$

We summarize the scattering process in the transition rate coefficient for later reference:

**Intra-Valley Elastic Acoustic Phonon Scattering**

$$c_{\eta}^{\nu,\nu'}(y) \Big|_{\eta=\text{ac}} = \delta_{v,v'} \frac{2\pi k_B T D_{\text{ac}}^2}{\hbar \rho v_{\lambda}^2} R^{\nu,\nu'}(y), \quad (2.28)$$

quantity	value	description
$k_B$	$8.617\,330 \times 10^{-5} \text{ eV K}^{-1}$	Boltzmann constant
$T$	300 K	temperature
$\delta_{v,v'}$	—	intra-valley scattering
$D_{\text{ac}}$	20 eV	effective deformation potential
$\rho$	$2.33 \times 10^3 \text{ kg m}^{-3}$	mass density of silicon crystal
$v_{\lambda}$	$9.05 \text{ m s}^{-1}$	sound velocity of acoustic phonons
$R^{\nu,\nu'}$	Eq. (2.27)	overlap integral

where  $\eta$  indicates the type of scattering which we will refer to as  $\eta = \text{ac}$  in the case at hand.

The value for the deformation potential  $D_{\text{ac}}$  depends on the device in question since it has been experimentally confirmed that it is actually position dependent: closer to the oxide interface carriers feel larger deformation potentials than further away [78, 79]. Averaging over such a position dependence to find a constant effective deformation potential  $D_{\text{ac}}$  leads to a device dependence where devices with thin channels will require a larger deformation potentials than bulk devices in order to obtain simulated mobilities approximating experiments. Hence, simulations of bulk silicon typically use values around  $D_{\text{ac}} \approx 10 \text{ eV}$  [80], while bulk MOSFETs where carriers accumulate close to the oxide use values of around  $D_{\text{ac}} = 12 \text{ eV}$  [81] to  $D_{\text{ac}} = 14.8 \text{ eV}$  [82]. Silicon-on-Insulator MOSFETs with ultrathin channels require even higher deformation potentials.

Since we will be simulating a double gate MOSFET with an ultrathin channel (cf. Fig. 5.1), the deformation potential in this work is rather high at  $D_{\text{ac}} = 20 \text{ eV}$ . This is necessary in order to approximate experimentally measured mobilities for similar devices (cf. Sect. 5.1.1).

**Inter-Valley Phonon Scattering**

Inter-valley phonon scattering has been studied thoroughly and the approximations presented in this section can be found in Refs. [54, 69, 83]. We are going to repeat the most important aspects for completeness and for later reference.

The band structure of silicon has six  $X$ -valleys lying at  $\pm 0.85 \frac{2\pi}{a_0}$  on each axis of 3D  $\mathbf{K}$ -space<sup>3</sup> in the crystal coordinate system (see Fig. 2.6), where  $a_0 \approx 0.54$  nm is the lattice spacing of the silicon crystal. These  $X$ -valleys represent the six absolute minima of the silicon band structure and they are all we will concern ourselves with during this work. A scattering process of an electron in the conduction band with a phonon may exchange sufficient momentum such that the electron changes the valley it occupies. The phonon momentum of such a process is given by

$$\mathbf{Q} = \pm(\mathbf{K}' - \mathbf{K}) + \mathbf{G}, \quad (2.29)$$

where  $\mathbf{K}$  and  $\mathbf{K}'$  are the electron three-momenta in the crystal coordinate system before and after scattering, respectively, and  $\mathbf{G}$  is the reciprocal lattice vector. The upper and lower sign correspond to absorption and emission, respectively.

In the following it will be assumed that the deviations of  $\mathbf{K}$  and  $\mathbf{K}'$  from the positions of the valley minima in  $\mathbf{K}$ -space are negligible compared to the distance between the valleys. Thus, each transition between valleys can be identified with a constant momentum transfer  $\mathbf{Q}$ . Moreover, if we take the perspective from an electron in one of the six  $X$ -valleys, we find that we only have to distinguish two different transitions as shown in Fig. 2.9. First, the transition to the opposing valley, which is referred to as a  $g$ -type transition. Second, the  $f$ -type transitions to one of the neighboring valleys.

As the phonon momentum resides within the first Brillouin zone, there is only one possible value for  $\mathbf{Q}$  for each type of transition. A vector  $\mathbf{Q}$  is defined to be in the first Brillouin zone if it fulfills

$$|Q_x|, |Q_y|, |Q_z| \leq \frac{2\pi}{a_0} \quad \text{and} \quad |Q_x| + |Q_y| + |Q_z| \leq \frac{3}{2} \frac{2\pi}{a_0}.$$

In the case of a  $g$ -type transition we may, without loss of generality, assume the initial and final momenta of the electron as

$$\mathbf{K} = (0.85, 0, 0) \frac{2\pi}{a_0}, \quad \mathbf{K}' = (-0.85, 0, 0) \frac{2\pi}{a_0}.$$

For absorption, Eq. (2.29) with the upper sign suggests

$$\mathbf{Q}_g = (0.3, 0, 0) \frac{2\pi}{a_0} \quad \text{with} \quad \mathbf{G} = (2, 0, 0) \frac{2\pi}{a_0},$$

where  $\mathbf{G}$  was chosen such that  $\mathbf{Q}_g$  resides in the first Brillouin zone. In fact,  $\mathbf{Q}_g$  is the only solution of Eq. (2.29) in the first Brillouin zone.

---

<sup>3</sup>We use an upper case  $\mathbf{K}$  to denote the 3D case.

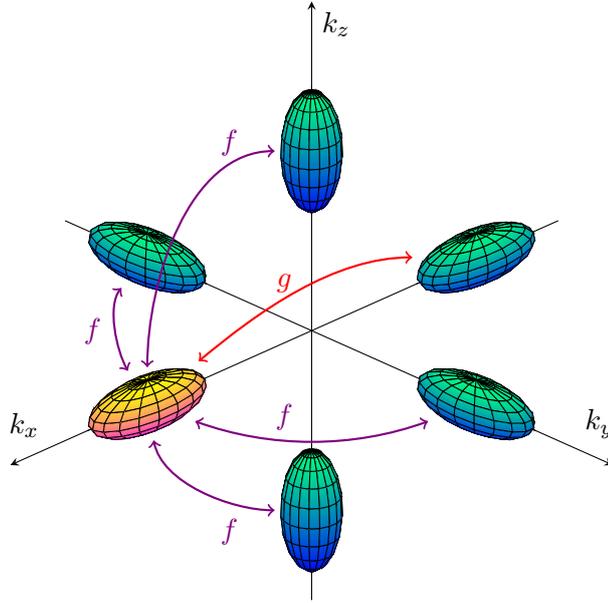


Figure 2.9: Illustration of  $f$ - and  $g$ -type inter-valley phonon scattering processes from the perspective of the  $X$ -valley on the positive  $k_x$ -axis (yellow valley). A  $g$ -type process scatters the electron to the opposing valley, while an  $f$ -type process scatters it to an adjacent valley.

Likewise, for an  $f$ -type transition, we may choose two neighboring  $X$ -valleys as

$$\mathbf{K} = (0, 0.85, 0) \frac{2\pi}{a_0}, \quad \mathbf{K}' = (0, 0, 0.85) \frac{2\pi}{a_0},$$

which also has a unique solution in the first Brillouin zone given by

$$\mathbf{Q}_f = (1, 0.15, -0.15) \frac{2\pi}{a_0} \quad \text{with } \mathbf{G} = (1, 1, -1) \frac{2\pi}{a_0}.$$

Although we looked at only one  $g$ -type transition and  $f$ -type transition, it is obvious that transitions from and to other valleys are completely analogous resulting in

$$|\mathbf{Q}_g| = 0.3 \frac{2\pi}{a_0}, \quad |\mathbf{Q}_f| = 1.02225 \frac{2\pi}{a_0}.$$

Since we know the magnitudes of the momentum transfers, we can use the associated deformation potentials of the various kinds of phonons available to read off the associated energy transfer from their respective dispersion relations. Table 2.1 summarizes the deformation potentials and transfer energies for the six available processes that need to be considered in our work [54, 71, 84, 85].

Phonon process	$\eta$	$\hbar\omega_\eta[\text{meV}]$	$D_\eta[10^{10} \text{ eV m}^{-1}]$
$g$ -type, transversal acoustic	$g_1$	12.1	0.47
$g$ -type, longitudinal acoustic	$g_2$	18.5	0.74
$g$ -type, longitudinal optical	$g_3$	62.0	10.23
$f$ -type, transversal acoustic	$f_1$	19.0	0.280
$f$ -type, longitudinal acoustic	$f_2$	47.4	1.860
$f$ -type, longitudinal optical	$f_3$	58.6	1.860

Table 2.1: Summary of phonon energies and deformation potentials involved in inter-valley phonon scattering processes in silicon [54, 71, 84, 85].

Using Fermi's Golden Rule of Eq. (2.24) and the phonon scattering matrix elements of Eqs. (2.22) and (2.23), we obtain for the inter-valley phonon scattering transition rate of type  $\eta$ :

$$\begin{aligned}
S_\eta^{\nu,\nu'}(y; \mathbf{k}, \mathbf{k}') &= \frac{\pi}{\rho\Omega} \frac{D_\eta^2}{\omega_\eta} R^{\nu,\nu'}(y) \left[ n_\lambda(\hbar\omega_\eta) \delta(\varepsilon_{\text{tot}}^\nu(y, \mathbf{k}) - \varepsilon_{\text{tot}}^{\nu'}(y, \mathbf{k}') + \hbar\omega_\eta) \right. \\
&\quad \left. + (n_\lambda(\hbar\omega_\eta) + 1) \delta(\varepsilon_{\text{tot}}^\nu(y, \mathbf{k}) - \varepsilon_{\text{tot}}^{\nu'}(y, \mathbf{k}') - \hbar\omega_\eta) \right].
\end{aligned} \tag{2.30}$$

Note that inter-valley transitions obviously have to occur between *different* valleys which was left implicit in this equation. Furthermore, the phonon number  $n_\lambda$  of Eq. (2.25) needs to be evaluated without the equipartition approximation due to the high phonon energies.

We can write down the transition rate using the transition rate coefficient yielding

$$\begin{aligned}
S_{\eta\sigma}^{\nu,\nu'}(y; \mathbf{k}, \mathbf{k}') \Big|_{\substack{\eta=f_i, g_i \\ \sigma=\pm 1}} &= \frac{1}{\Omega} c_\eta^{\nu,\nu'}(y; \mathbf{k}, \mathbf{k}') \\
&\quad \times \delta(\varepsilon_{\text{tot}}^\nu(y, \mathbf{k}) - \varepsilon_{\text{tot}}^{\nu'}(y, \mathbf{k}') + \sigma\hbar\omega_\eta) \Big|_{\substack{\eta=f_i, g_i \\ \sigma=\pm 1}},
\end{aligned}$$

where  $\sigma = \pm 1$  determines the sign of the energy transfer. Thus, the transition rate coefficient reads:

**Inter-Valley Phonon Scattering**

$$c_{\eta}^{\nu,\nu'}(y; \mathbf{k}, \mathbf{k}') \Big|_{\eta=f_i, g_i} = \frac{\pi D_{\eta}^2}{\rho \omega_{\eta}} R^{\nu,\nu'}(y) \left[ n_{\lambda}(\hbar\omega_{\eta}) + \Theta(\varepsilon_{\text{tot}}^{\nu'}(y, \mathbf{k}') - \varepsilon_{\text{tot}}^{\nu}(y, \mathbf{k})) \right] \quad (2.31)$$

quantity	value	description
$\rho$	$2.33 \times 10^3 \text{ kg m}^{-3}$	mass density of silicon crystal
$f_i, g_i$	Tab. 2.1	type of transition
$D_{\eta}$	Tab. 2.1	effective deformation potential for phonon of transition type $\eta$
$\omega_{\eta}$	Tab. 2.1	angular frequency of phonon type $\eta$
$R^{\nu,\nu'}$	Eq. (2.27)	overlap integral
$n_{\lambda}(\hbar\omega_{\eta})$	Eq. (2.25)	phonon number (without equipartition approximation)
$\Theta$	—	Heaviside step function

Note that the transition rate coefficient of Eq. (2.31) was written down somewhat peculiarly with a Heaviside step function  $\Theta$ . To understand the reasoning for this, let us describe the scattering process phenomenologically. The scattering integral of Eq. (2.15) describes the rate of scattering in a state  $(\nu, \mathbf{k})$  at some position  $y$ . The first term on the r.h.s. of Eq. (2.15) is the rate of in-scattering, i.e. how many electrons transition from any other state  $(\nu', \mathbf{k}')$  to the state  $(\nu, \mathbf{k})$ . The second term on the r.h.s. is the out-scattering term which is the rate at which electrons in the state  $(\nu, \mathbf{k})$  transition to any other state.

For inelastic scattering like the inter-valley phonon scattering, the distinction between in-scattering  $S_{\eta}^{\nu,\nu'}(y; \mathbf{k}, \mathbf{k}')$  and out-scattering  $S_{\eta}^{\nu',\nu}(y; \mathbf{k}', \mathbf{k})$  is crucial due to the difference in emission and absorption in Eq. (2.30).

If we assume two different energy levels where, say,  $\varepsilon_{\text{tot}}^{\nu}(y, \mathbf{k}) > \varepsilon_{\text{tot}}^{\nu'}(y, \mathbf{k}')$ , then the phonon is *absorbed* during in-scattering because the electron transitions from  $(\nu', \mathbf{k}')$  to  $(\nu, \mathbf{k})$ , gaining energy in the process. However, if we describe the out-scattering, we find that the electron transitions from  $(\nu, \mathbf{k})$  to  $(\nu', \mathbf{k}')$ , i.e. the electron loses energy in the process which means a phonon is *emitted*.

Therefore, the crux of writing the transition rate down compactly with the transition rate coefficient of Eq. (2.31) is to differentiate the cases of absorption and emission adequately. The deciding property that tells us whether we are dealing with absorption or emission is whether the electron gains or loses energy during scattering, which is why we used the Heaviside step function in Eq. (2.31).

### Surface Roughness Scattering

The idea behind surface roughness scattering is that the interface between semiconductor and oxide is not perfectly smooth but exhibits miniscule variations. These variations, which are usually called roughness, alter the potential perceived by carriers and therefore lead to scattering. At the time of this writing, the exact computation of the surface roughness scattering is still not fully understood as there exist multiple descriptions and the details of the approximations are still somewhat under debate [86].

However, all of the approaches depend on a statistical modeling of the roughness and involve the details of how the density of carriers, i.e. the wave functions, behave close to the interface.

For the purposes of this work, we will be satisfied with the simplest description that goes back to Prange and Nee [87]. Thus, we can write down the matrix element of surface roughness scattering as

$$|M_{\text{SR}}^{\nu,\nu'}(y; \mathbf{k}, \mathbf{k}')|^2 \approx \frac{1}{\Omega} \langle |\Delta_{\mathbf{q}}|^2 \rangle |\Gamma_{\text{PN}}^{\nu,\nu'}(y)|^2,$$

where  $\mathbf{q} = \mathbf{k}' - \mathbf{k}$  is the momentum transfer. We assume an exponential model for the power spectrum of the surface roughness given by

$$\langle |\Delta_{\mathbf{q}}|^2 \rangle = \pi \Delta^2 L^2 \left( 1 + \frac{|\mathbf{q}|^2 L^2}{2} \right)^{-\gamma},$$

where  $\Delta$  and  $L$  are the root mean square height and correlation length of the roughness, respectively. Note that sometimes the power spectrum is also chosen to have a Gaussian form [88], though, Ref. [89] shows that the exponential form with an exponent of  $\gamma = 1.5$  is a better fit to experimental measurements. Both  $\Delta$  and  $L$  are fitting parameters that have to be adjusted to obtain the correct low-field mobility. Furthermore, the Prange-Nee term is given by

$$\Gamma_{\text{PN}}^{\nu,\nu'}(y) = \frac{\hbar^2}{2m_x^v} \frac{\partial \Psi^\nu(x, y)}{\partial x} \frac{\partial \Psi^{\nu'}(x, y)}{\partial x} \Big|_{x=x_0},$$

where  $x_0$  denotes the position of the interface. Since surface roughness scattering is considered to be elastic, electrons cannot change valleys. Thus, with Fermi's Golden Rule of Eq. (2.20) the transition rate reads

$$\begin{aligned} S_{\eta,\sigma}^{\nu,\nu'}(y; \mathbf{k}, \mathbf{k}') \Big|_{\substack{\eta=\text{SR} \\ \sigma=0}} &= \delta_{v,v'} \frac{1}{\Omega} \frac{\pi \hbar^3}{2(m_x^v)^2} \langle |\Delta_{\mathbf{q}}|^2 \rangle \left| \frac{\partial \Psi^\nu(x, y)}{\partial x} \frac{\partial \Psi^{\nu'}(x, y)}{\partial x} \right|^2 \\ &\times \delta(\varepsilon_{\text{tot}}^{\nu'}(y, \mathbf{k}') - \varepsilon_{\text{tot}}^\nu(y, \mathbf{k})) \Big|_{x=x_0} \end{aligned} \quad (2.32)$$

We can express the magnitude of  $\mathbf{q}$  as

$$|\mathbf{q}|^2 = |\mathbf{k}'|^2 + |\mathbf{k}|^2 - 2|\mathbf{k}'||\mathbf{k}| \cos(\phi' - \phi),$$

where it becomes evident that not only does the transition rate depend on the initial and final energy of the electron but it is also angular dependent, which increases the computational load significantly (cf. the projection onto Fourier harmonics in Sect. 2.4.4).

Since we have to fit the parameters  $\Delta$  and  $L$  of the surface roughness scattering to the low-field mobility anyway, we can justify another approximation. If we assume that surface roughness scattering is an intra-subband process, the elasticity of the scattering suggests  $|\mathbf{k}'| = |\mathbf{k}|$ . If we assume additionally that the equienergy lines are circles (cf. Sect. 2.4.4 on the Herring-Vogt transformation) represented by the Herring-Vogt mass  $m_d^v = \sqrt{m_y^v m_z^v}$ , we find

$$|\mathbf{q}|^2 = 2|\mathbf{k}|^2(1 - \cos(\alpha)) = \frac{4m_d^v E}{\hbar^2}(1 - \cos(\alpha)),$$

where  $\alpha = \phi' - \phi$  is the opening angle between the initial  $\mathbf{k}$  and final  $\mathbf{k}'$  of the electron. Note that usually  $|\mathbf{k}(E)|$  is angular dependent due to the ellipsoidal shape of the valleys but we anticipate that we will later, in Sect. 2.4.4, use the Herring-Vogt transformation accompanied by the approximation that we can evaluate the scattering processes as if  $\mathbf{k}$ -space were Herring-Vogt transformed (cf. Eq. (2.52)).

By using the microscopic relaxation time approximation [90, 91, 92, 93], we can construct a transition rate that is angular independent but whose total rate coincides with the total rate obtained from the transition rate of Eq. (2.32). The microscopic relaxation time for intra-subband processes can be expressed as

$$\begin{aligned} \frac{1}{\tau^\nu(E)} &= \Omega \int dE' d\alpha Z^v S^{\nu,\nu}(\mathbf{k}'(E'), \mathbf{k}(E)) (1 - \cos(\alpha)) \\ &= Z^v \frac{\pi^2 \hbar^3}{2(m_x^v)^2} \left| \frac{\partial \Psi^\nu(x, y)}{\partial x} \right|_{x=x_0}^4 \\ &\quad \times \Delta^2 L^2 \int_0^{2\pi} d\alpha \frac{1 - \cos(\alpha)}{\left(1 + \frac{2m_d^v E}{\hbar^2} L^2 (1 - \cos(\alpha))\right)^\gamma}, \end{aligned} \quad (2.33)$$

where – in general – the integral over  $\alpha$  has to be solved numerically due to the free parameter  $\gamma$ .

Thus the angular independent transition rate that yields the same total scattering rate is given by

$$\bar{S}_{\eta,\sigma}^{\nu,\nu'}(y; \mathbf{k}, \mathbf{k}') \Big|_{\sigma=0}^{\eta=\text{SR}} = \frac{1}{\Omega} \frac{\delta_{\nu,\nu'}}{2\pi} \frac{1}{Z^v \tau^\nu(E(\mathbf{k}))} \delta(\varepsilon_{\text{tot}}^{\nu'}(y, \mathbf{k}') - \varepsilon_{\text{tot}}^\nu(y, \mathbf{k})), \quad (2.34)$$

which can be easily verified by inserting Eq. (2.34) into Eq. (2.33). Subsequently, the transition rate coefficient reads

**Surface Roughness Scattering**

$$c_{\eta}^{\nu,\nu'}(y) \Big|_{\eta=\text{SR},x_0} = \frac{\delta_{\nu,\nu'}}{2\pi} \frac{\pi^2 \hbar^3}{2(m_x^v)^2} \left| \frac{\partial \Psi^\nu(x, y)}{\partial x} \right|^4 \Big|_{x=x_0} \Delta^2 L^2 \quad (2.35)$$

$$\times \int_0^{2\pi} d\alpha \frac{1 - \cos(\alpha)}{\left(1 + \frac{2m_d^v E}{\hbar^2} L^2 (1 - \cos(\alpha))\right)^\gamma},$$

quantity	value	description
$\Delta$	0.4 nm	root mean square height of interface roughness
$L$	1 nm	correlation length of roughness
$\gamma$	1.5	exponent of power spectrum
$x_0$	—	position of interface
$\Psi$	—	wave function
$E$	—	kinetic energy
$m_x^v$	Eq. 2.17	mass of electrons in direction perpendicular to interface
$m_d^v$	Eq. (2.48)	Herring-Vogt mass

Since we will consider a double gate MOSFET for most of this work, we actually have to deal with two interfaces. We assume that the roughness of one interface is not correlated with the roughness of the other interface and therefore we may simply add the scattering rates of both interfaces [94].

The root mean square height and correlation length of the roughness are chosen such that experimentally measured mobilities for an SOI MOSFET with an ultrathin channel can be reproduced (cf. Sect. 5.1.1). The final values of  $\Delta = 0.3$  nm and  $L = 1$  nm are in the vicinity of the values found in literature [86].

We omit the effect of screening in the surface roughness scattering because of the reason explained in the following section.

### Ionized Impurity Scattering

In MOSFETs, we usually have to deal with ionized impurity scattering due to the doping. When we dope the semiconductor, we essentially insert foreign atoms which are usually ionized and populate the conduction band with electrons or the valence band with holes, depending on the kind of doping. These ionized

atoms represent Coulomb scattering targets which have to be considered in the scattering rate of the BE (see e.g. [69]).

Computing the scattering rate with Coulomb targets straightforwardly leads to divergences, unless the screening by the surrounding electron gas is included. However, screening represents a hurdle in self-consistent solvers as it couples each scattering term to the density of carriers and therefore fills up the matrix of the system of equations a lot more than other scattering terms (cf. Sect. 2.4.6 for the setup of the solver). Since this work is more concerned with a proper self-consistent description of noise, we will approximate the effect of ionized impurity scattering using the deformation potential of acoustic elastic phonon scattering.

Let us first note that impurity scattering mainly contributes in the highly doped contact regions as a reduction of mobility since the channel region is only lightly doped. However, the highly doped contact regions also have high numbers of carriers and therefore the distribution function is close to equilibrium. And as long as the distribution function is close to equilibrium, we can interpret the mobility reduction of the ionized impurity scattering as a mere global resistance. Thus, we can effectively increase the deformation potential of the elastic acoustic phonon scattering of Eq. (2.28) in order to obtain the desired mobility in the contact regions.

Meanwhile, in the channel, where we will find more complex dynamics, we can safely ignore ionized impurity scattering due to the low doping. In Sect. 5.2.2 we are going to verify that our assumptions hold.

### 2.4.3 Boundary Conditions

We want to describe the boundary conditions of the BE as generation and recombination (GR) terms relative to a thermal bath of electrons in the contacts. To this end, we could use Robin boundary conditions with some flux crossing the devices boundaries at the source and drain contacts. However, this is equivalent to assuming that no flux crosses the boundaries but a singular GR rate  $\Gamma$  (cf. Eq. (2.12)) is located on the ohmic contacts, i.e. the source and drain contacts in Fig. 2.1. This GR rate acts as a source and sink of electrons.

For the sake of the argument, let us assume that there exists an additional subband  $C$  for each subband  $\nu$ , denoting the thermal electrons in the source or drain,  $C = S/D$ , with equilibrium distribution  $f_{\text{eq}}^{\nu}(y_C, \mathbf{k})$  as depicted in Fig. 2.10. Then, we can express the rate of transfer from this thermal bath to the inside of the device as a scattering rate analogous to Eq. (2.15) only that the rate does not transition between two subbands  $\nu$  and  $\nu'$  but between a subband  $\nu$  and the

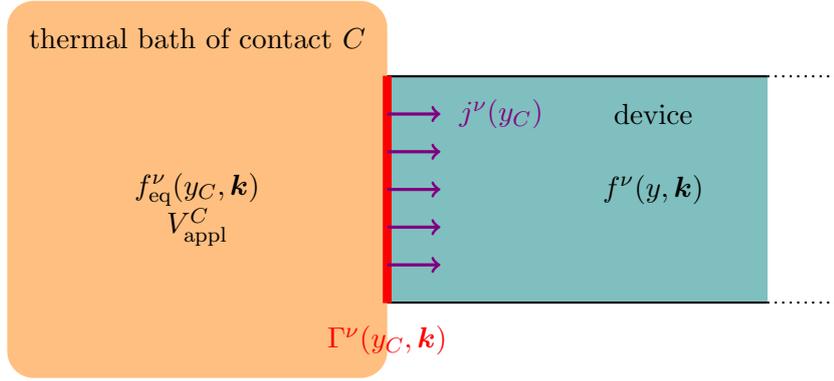


Figure 2.10: Illustration of the hypothetical thermal bath representing a contact  $C$  with an applied bias  $V_{\text{appl}}^C$  and an equilibrium distribution function  $f_{\text{eq}}^\nu(y_C, \mathbf{k})$ . On the contact resides the GR rate  $\Gamma^\nu(y_C, \mathbf{k})$  connecting the thermal bath with the distribution function  $f^\nu(y, \mathbf{k})$  inside the device and creating a current density  $j^\nu(y_C)$ .

thermal bath band  $C$ :

$$\Gamma^\nu(y, \mathbf{k}) = \Omega \sum_{C=S,D} \int \frac{d^2 k'}{(2\pi)^2} \left[ (1 - f^\nu(y, \mathbf{k}, t)) \gamma^{\nu,C}(y; \mathbf{k}, \mathbf{k}') f_{\text{eq}}^\nu(y, \mathbf{k}') - (1 - f_{\text{eq}}^\nu(y, \mathbf{k}')) \gamma^{C,\nu}(y; \mathbf{k}', \mathbf{k}) f^\nu(y, \mathbf{k}, t) \right]. \quad (2.36)$$

Assuming the transition rate  $\gamma^{\nu,C}$  is constant for all  $\mathbf{k}$ , we can write it down as

$$\gamma^{\nu,C}(y; \mathbf{k}, \mathbf{k}') = \gamma^{C,\nu}(y; \mathbf{k}', \mathbf{k}) = \frac{1}{\Omega} v_{\text{GR}} (2\pi)^2 \delta(\mathbf{k} - \mathbf{k}') \delta(y - y_C), \quad (2.37)$$

where  $y_C = y_S$  and  $y_C = y_D$  are the  $y$ -coordinates of the source and drain, respectively. The constant  $v_{\text{GR}}$  is called the *recombination velocity*. It is the velocity of injection and extraction of carriers at the source and the drain which can be obtained by approximating the corresponding  $\mathbf{k}$ -dependent rate [29]. The higher  $v_{\text{GR}}$ , the lower the resistance of the contact, and vice versa.

Inserting Eq. (2.37) into Eq. (2.36), the GR rate simplifies to

$$\Gamma^\nu(y, \mathbf{k}) = -v_{\text{GR}} (f^\nu(y, \mathbf{k}) - f_{\text{eq}}^\nu(y, \mathbf{k})) (\delta(y - y_S) + \delta(y - y_D)), \quad (2.38)$$

which is equivalent to the expression given in Ref. [24].

### Contact Current

As is evident from Eq. (2.38),  $\Gamma$  is singular on the contacts and vanishes everywhere else. The distribution function outside the device  $D$  vanishes,

$$f^\nu(y, \mathbf{k}, t) \Big|_{y \notin D} = 0,$$

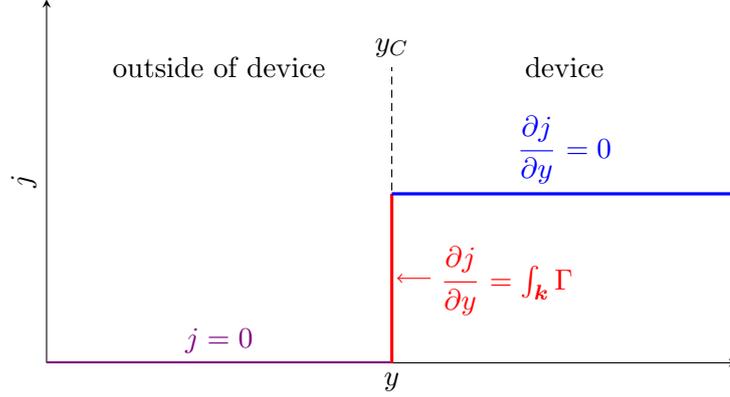


Figure 2.11: Illustration of the current density at the contact. Outside the device it vanishes and inside of the device it is constant. Whatever current flows inside the device needs to be created by the singular GR rate on the contact at  $y_C$ .

such that all flux terminates or originates in the singular  $\Gamma$  as depicted in Fig. 2.11. Naturally, we do not want to exclude the possibility of a current within the device and therefore the current density needs to be discontinuous – as a Heaviside step function – where the GR rate  $\Gamma$  is located. Thus the derivative of the distribution function w.r.t.  $y$  is necessarily singular – as a delta-distribution function – on the contacts.

Integration of the whole BE over an infinitely small interval containing, say, the source contact position  $y_S$ , will leave only the singular terms as

$$\begin{aligned}
0 &= \lim_{\ell \rightarrow 0} \int_{y_S - \ell/2}^{y_S + \ell/2} dy F^{\text{BE}} = \lim_{\ell \rightarrow 0} \int_{y_S - \ell/2}^{y_S + \ell/2} dy \left[ v_y^\nu(\mathbf{k}) \frac{\partial}{\partial y} f^\nu(y, \mathbf{k}, t) - \Gamma^\nu(y, \mathbf{k}) \right] \\
&= \lim_{\ell \rightarrow 0} v_y^\nu(\mathbf{k}) \left[ \underbrace{f^\nu(y_S - \ell/2, \mathbf{k}, t) - f^\nu(y_S + \ell/2, \mathbf{k}, t)}_{=0} \right] \\
&\quad + v_{\text{GR}} (f^\nu(y_S, \mathbf{k}, t) - f_{\text{eq}}^\nu(y_S, \mathbf{k})) \\
&= -v_y^\nu(\mathbf{k}) f^\nu(y_S, \mathbf{k}, t) + v_{\text{GR}} (f^\nu(y_S, \mathbf{k}, t) - f_{\text{eq}}^\nu(y_S, \mathbf{k})),
\end{aligned}$$

where the distribution function outside the device vanishes.

Integrating this over  $\mathbf{k}$ -space, we see that the GR rate  $\Gamma$  gives rise to an electron current

$$j^\nu(y_S, t) = - \lim_{\ell \rightarrow 0} \int_{y_S - \ell/2}^{y_S + \ell/2} dy \int \frac{d^2 k}{(2\pi)^2} \Gamma^\nu(y, \mathbf{k}) = v_{\text{GR}} (n^\nu(y_S, t) - n_{\text{eq}}) \quad (2.39)$$

and analogously if the BE were integrated around the drain contact position  $y_D$

$$j^\nu(y_D, t) = \lim_{\ell \rightarrow 0} \int_{y_D - \ell/2}^{y_D + \ell/2} dy \int \frac{d^2 k}{(2\pi)^2} \Gamma^\nu(y, \mathbf{k}) = -v_{\text{GR}} (n^\nu(y_D, t) - n_{\text{eq}}). \quad (2.40)$$

The opposite sign in the equation for the drain contact current stems from the fact that the generation of charge elicits a current flowing from the contact inwards of the device, which is the negative  $y$ -direction as seen from the drain contact.

### Contact Fermi Level

In order to determine the GR rate  $\Gamma$  of Eq. (2.38), we need to know the Fermi level of the equilibrium electrons in the thermal bath representing the ohmic source and drain contacts and we need to understand how this Fermi level impacts the electron distribution function within the device. Assuming that the contacts are made from the same metal, the shift in their Fermi levels due to an applied bias is given by

$$E_F^{S/D} - E_F^{\text{ref}} = -qV_{\text{appl}}^{S/D},$$

where  $E_F^{S/D}$  is the Fermi level of the source or drain contact;  $E_F^{\text{ref}}$  is the Fermi level of some reference metal contact, usually chosen as the source contact itself; and  $V_{\text{appl}}^{S/D}$  is the applied bias between the source or drain and the reference contact.

We want to understand how the Fermi levels of the source and drain contacts relate to the adjacent highly doped silicon regions within the device. To describe the behavior of electrons within a semiconductor device, the quasi Fermi level  $-q\varphi_n$  is a useful quantity. It is a chemical potential which gives rise to the total electron current, i.e. it contains both diffusion and conduction phenomena. In the equilibrium case, when there is no conduction current flowing through the source or drain contacts, the quasi Fermi level in the highly doped silicon adjacent to the contacts will adjust itself to the Fermi level in the contacts. Once a bias is applied, the quasi Fermi potential and the metal contact can be different, however, since our silicon regions next to the contacts are highly doped, we can assume that the carriers are close to equilibrium. Furthermore, the source and drain contacts are defined to be ohmic which means that the exchange rate between the metal and the semiconductor is high which in turn means that metal and semiconductor are in thermodynamic equilibrium. It follows that the Fermi levels of the contacts are the same as the quasi Fermi levels in the adjacent highly doped silicon region, even under non-equilibrium conditions:

$$E_F^S = -q\varphi_n \Big|_{y=y_S}, \quad E_F^D = -q\varphi_n \Big|_{y=y_D}.$$

And therefore the applied bias can be expressed in terms of the quasi Fermi potentials as

$$\varphi_n \Big|_{y=y_S} - \varphi_n^{\text{ref}} = V_{\text{appl}}^S, \quad \varphi_n \Big|_{y=y_D} - \varphi_n^{\text{ref}} = V_{\text{appl}}^D, \quad (2.41)$$

where  $\varphi_n^{\text{ref}} = -E_F^{\text{ref}}/q$  is the quasi Fermi potential corresponding to the Fermi level of the reference contact.

With the understanding of how the Fermi level impacts the semiconductor region, we can find an expression for the Fermi level by substituting the metal region with an infinitely large three-dimensional thermal bath of electrons with density

$$n_{3D}^{\text{S/D}} = N_{3D}^{\text{eff}} \exp\left(\frac{E_F^{\text{S/D}} - E_C}{k_B T}\right), \quad (2.42)$$

where  $V$  is the electric potential,

$$E_C = -qV - E_F^{\text{ref}} \quad (2.43)$$

is the conduction band energy measured relative to the reference Fermi level, and

$$N_{3D}^{\text{eff}} = 3\sqrt{2} \left( \frac{\sqrt[3]{m_x^v m_y^v m_z^v}}{\hbar^2 \pi} k_B T \right)^{3/2} \quad (2.44)$$

is the effective density of states of silicon in three dimensions. Choosing silicon for the material of the thermal bath certainly simplifies things because the contact just becomes an extension of the device and therefore we do not have to consider interface conditions.

Close to the contacts but on the inside of the device, we can express the electron density with the quasi Fermi potential as

$$n_{3D} = n_i \exp\left(\frac{V - \varphi_n}{V_T}\right) \quad (2.45)$$

and since we treat the contact just as an extension of the silicon region, we require that the density is continuous and therefore we find with Eqs. (2.42) and (2.45):

$$E_F^{\text{S/D}} = k_B T \log\left(\frac{n_i}{N_{3D}^{\text{eff}}}\right) + E_C \Big|_{y=y_{\text{S/D}}} + qV \Big|_{y=y_{\text{S/D}}} - q\varphi_n \Big|_{y=y_{\text{S/D}}}.$$

Note that the conduction band energy  $E_C$  of Eq. (2.43) is constant within the thermal bath and therefore we chose to evaluate it on the contacts. With Eq. (2.41) we finally find

$$E_F^{\text{S/D}} = k_B T \log\left(\frac{n_i}{N_{3D}^{\text{eff}}}\right) - qV_{\text{appl}}^{\text{S/D}}. \quad (2.46)$$

In the above derivation we neglected the Pauli principle but the same reasoning also applies in that case as well. With the Fermi level of Eq. (2.46), we can finally formulate the equilibrium Fermi-Dirac distribution functions of the source and drain thermal baths as

$$f_{\text{eq}}^\nu(y_{\text{S/D}}, \mathbf{k}) = \frac{1}{\exp\left(\frac{E^\nu(\mathbf{k}) + \varepsilon^\nu(y_{\text{S/D}}) + qV_{\text{appl}}^{S/D}}{k_B T} - \log\left(\frac{n_i}{N_{3\text{D}}^{\text{eff}}}\right)\right) + 1}, \quad (2.47)$$

which completes the definition of the GR rate  $\Gamma$  of Eq. (2.38).

#### 2.4.4 Transformations

We cannot solve the BE of Eq. (2.13) efficiently and numerically stable if we use a straightforward approach. Instead, we are going to perform a series of transformations described in the following.

##### Herring-Vogt Transformation

The dispersion relation of Eq. (2.19) is inconvenient to work with due to the potentially differing effective masses in  $y$ - and  $z$ -direction which stem from the ellipsoidal  $X$ -valleys. However, we can transform  $\mathbf{k}$ -space in order to distort the ellipsoidal valleys into spherical ones using the Herring-Vogt transformation [95]. In our case, the transformation in a valley  $v$  is given by

$$\mathbf{k}_{\text{HV}} := T_{\text{HV}}^v \mathbf{k}, \quad \text{with } T_{\text{HV}}^v = \text{diag}\left(\sqrt{\frac{m_d^v}{m_y^v}}, \sqrt{\frac{m_d^v}{m_z^v}}\right), \quad m_d^v = \sqrt{m_y^v m_z^v}, \quad (2.48)$$

where  $\mathbf{k}_{\text{HV}}$  is a 2D vector in the Herring-Vogt transformed  $\mathbf{k}$ -space.

The kinetic energy can be readily transformed as

$$E^v(\mathbf{k}) = E^v\left((T_{\text{HV}}^v)^{-1} \mathbf{k}_{\text{HV}}\right) = \frac{\hbar^2}{2m_d^v} |\mathbf{k}_{\text{HV}}|^2 =: E_{\text{HV}}^v(\mathbf{k}_{\text{HV}}). \quad (2.49)$$

Note how  $E_{\text{HV}}^v$  has the same mass  $m_d^v$  irrespective of the direction. Likewise the group velocity can be transformed as

$$\mathbf{v}^v(\mathbf{k}) = \frac{1}{\hbar} \nabla_{\mathbf{k}} E(\mathbf{k}) = \hbar \begin{pmatrix} 1/m_y^v & 0 \\ 0 & 1/m_z^v \end{pmatrix} \mathbf{k} = T_{\text{HV}}^v \frac{\hbar}{m_d^v} \mathbf{k}_{\text{HV}} = T_{\text{HV}}^v \mathbf{v}^v(\mathbf{k}_{\text{HV}}) \quad (2.50)$$

and the force remains the same, i.e.  $F^\nu \equiv F_{\text{HV}}^\nu$ , since it doesn't depend on  $\mathbf{k}$ -space.

The stationary free streaming operator in the Herring-Vogt transformed  $\mathbf{k}$ -space is thus given by

$$L^v(y, \mathbf{k}) = (T_{\text{HV}}^v)_{yy} \left[ \frac{1}{\hbar} F_{\text{HV}}^\nu(y) \frac{\partial}{\partial k_{\text{HV},y}} f_{\text{HV}}^\nu(y, \mathbf{k}_{\text{HV}}) + v_y^v(\mathbf{k}_{\text{HV}}) \frac{\partial}{\partial y} f_{\text{HV}}^\nu(y, \mathbf{k}_{\text{HV}}) \right], \quad (2.51)$$

where  $(T_{\text{HV}}^v)_{yy} = \sqrt{m_d^v/m_y^v}$  and

$$f^\nu(y, \mathbf{k}) = f^\nu\left(y, (T_{\text{HV}}^v)^{-1} \mathbf{k}_{\text{HV}}\right) = f_{\text{HV}}^\nu(y, \mathbf{k}_{\text{HV}}).$$

Note that structurally nothing changed in the free streaming term except for the appearance of the diagonal  $y$ -component of the Herring-Vogt transformation matrix.

The scattering term of Eq. (2.15) also needs to be Herring-Vogt transformed. The measure  $d^2k/(2\pi)^2$  stays invariant, however, the transition rate – in general – does depend on the shape of the band structure. Nevertheless, since we already employ significant simplifications to the scattering process, we will further approximate that we can compute the transition rates as if  $\mathbf{k}$ -space is Herring-Vogt transformed:

$$S^\nu(y, \mathbf{k}) \approx S^\nu(y, \mathbf{k}_{\text{HV}}) \quad (2.52)$$

Finally, the GR rate  $\Gamma$  of Eq. (2.38) remains invariant under the Herring-Vogt transformation:

$$\Gamma^\nu(y, \mathbf{k}) = \Gamma^\nu(y, \mathbf{k}_{\text{HV}}).$$

In the remainder of this work, all quantities will be given in the Herring-Vogt transformed  $\mathbf{k}$ -space, however, we will omit the HV subscript to avoid cluttering the notation.

### Fourier Harmonics Expansion

The BE in its general form of Eq. (2.12) is impervious to numerical brute force approaches due to its high dimensionality of three spatial dimensions and three wave vector dimensions. Even after the simplifications leading up to Eq. (2.13) it is highly inefficient to solve due to the two-dimensional  $\mathbf{k}$ -space.

To ease the computational burden, we can exploit the fact that in equilibrium, the distribution function is radially symmetric in energy space and non-equilibrium distribution functions can be formulated as corrections to a radially symmetric function.

To this end, we are going to express the BE in terms of the kinetic energy  $E$  and the angle  $\phi$  describing the orientation of  $\mathbf{k}$  in our two-dimensional  $\mathbf{k}$ -space. Thereafter we are going to expand the angle-dependence in a series of Fourier harmonics, where the zeroth order – which is independent of the angle – is sufficient to express the exact equilibrium solution. Higher orders are corrections that become important in non-equilibrium. As it turns out, such a series converges quickly enough – even far from equilibrium – that we can truncate the series to make computations feasible.

In the previous section we saw that the dispersion relation in the Herring-Vogt transformed  $\mathbf{k}$ -space takes on the simple form of Eq. (2.49). It is thus strictly monotonically increasing and can be inverted as

$$\mathbf{k}^\nu(E, \phi) = \sqrt{\frac{2m_d^v E}{\hbar^2}} \begin{pmatrix} \cos(\phi) \\ \sin(\phi) \end{pmatrix}, \quad (2.53)$$

where  $\phi$  gives the orientation of  $\mathbf{k}$ . We can use this inversion to define all quantities in terms of polar coordinates  $E$  and  $\phi$ , rather than cartesian coordinates  $k_y$  and  $k_z$ . Then each quantity, say  $X(E, \phi)$ , can be expanded in Fourier harmonics as

$$X(E, \phi) = \sum_{m=-\infty}^{\infty} X_m(E) Y_m(\phi), \quad (2.54)$$

where  $Y_m$  is the  $m$ -th Fourier harmonic given by

$$Y_m(\phi) = c_m \cos(m\phi + \phi_m), \quad \text{with } c_m = \sqrt{\frac{1}{(1 + \delta_{m,0})\pi}}, \quad \phi_m = \begin{cases} 0, & m \geq 0 \\ \frac{\pi}{2}, & m < 0 \end{cases} \quad (2.55)$$

and fulfilling the orthonormality relation

$$\int_0^{2\pi} d\phi Y_m(\phi) Y_n(\phi) = \delta_{m,n} \quad (2.56)$$

with the Kronecker delta  $\delta_{m,n}$ . Equation (2.54) can be inverted using the orthonormality of Eq. (2.56) which yields

$$X_m(E) = \int_0^{2\pi} d\phi X(E, \phi) Y_m(\phi).$$

In order to switch from a cartesian  $\mathbf{k}$ -space to the above polar representation, we also need to transform the measure of the  $\mathbf{k}$ -space integration as

$$\frac{d^2 k}{(2\pi)^2} = \frac{1}{(2\pi)^2} d|\mathbf{k}| d\phi |\mathbf{k}| = dE d\phi Z^v \quad (2.57)$$

where  $Z^v = m_d^v/(2\pi)^2\hbar$  is the two-dimensional valley-dependent density of states for a parabolic band structure.

Using all of the above, we can project any quantity  $X(\mathbf{k})$  in  $\mathbf{k}$ -space<sup>4</sup> onto an equienergy circle  $E$  and onto a Fourier harmonic  $m$  as

$$\begin{aligned} \int \frac{d^2k}{(2\pi)^2} X(\mathbf{k}) \delta(E - E^v(\mathbf{k})) Y_m(\phi(\mathbf{k})) \\ &= \int dE' d\phi Z^v X(\mathbf{k}(E', \phi)) \delta(E - E') Y_m(\phi) \\ &= \int d\phi Z^v X(E, \phi) Y_m(\phi) \\ &= Z^v X_m(E), \end{aligned}$$

where  $\delta$  is the Dirac-distribution.

The projection of the Herring-Vogt transformed free streaming term of Eq. (2.51) is more involved. The full treatment is exercised in Ref. [29] and will be omitted in this work. Nevertheless, we will record the result for the projected free-streaming term which is given by

$$\begin{aligned} L^v(y, E) &:= \int \frac{d^2k}{(2\pi)^2} L^v(y, \mathbf{k}) \delta(E - E^v(\mathbf{k})) Y_m(\phi) \\ &= Z^v \sum_{m'=m-1, m+1} \left[ \left( F^v(y) \frac{\partial}{\partial E} + \frac{\partial}{\partial y} \right) a_{m, m'}^v(y, E) f_{m'}^v(y, E) \right. \\ &\quad \left. - b_{m, m'}^v(y, E) f_{m'}^v(y, E) \right], \end{aligned} \quad (2.58)$$

where

$$a_{m, m'}^v(y, E) := (T_{\text{HV}}^v)_{yy} (v_y^v)_1(E) \int d\phi Y_{m'}(\phi) Y_m(\phi) Y_1(\phi), \quad (2.59)$$

$$b_{m, m'}^v(y, E) := (T_{\text{HV}}^v)_{yy} F^v(y) \frac{m\sqrt{\pi}}{\hbar k^v(E)} \int d\phi Y_{m'}(\phi) Y_{-m}(\phi) Y_{-1}(\phi), \quad (2.60)$$

with the first Fourier harmonic coefficient of the group velocity  $(v_y^v)_1(E)$  defined by

$$v_y^v(E, \phi) = \frac{\hbar}{m_d^v} k_y(E, \phi) = \sqrt{\frac{2\pi E}{m_d^v}} Y_1(\phi) =: (v_y^v)_1(E) Y_1(\phi). \quad (2.61)$$

---

<sup>4</sup>Note that  $X(\mathbf{k})$  is technically another function than  $X(E, \phi)$  but we nevertheless use the same symbol as the reader should easily be able to discern these based upon the function arguments.

The integral over the three Fourier harmonics can be evaluated using [29]

$$\int_0^{2\pi} d\phi Y_m(\phi)Y_n(\phi)Y_\ell(\phi) = \frac{\pi}{2}c_m c_n c_\ell \sum_{\sigma, \tau \in \{1, -1\}} \delta_{0, m + \sigma n + \tau \ell} \cos(\phi_m + \sigma \phi_n + \tau \phi_\ell). \quad (2.62)$$

Note that this means that the only non-zero contributions to the sum over all the harmonics in the free streaming term comes from the terms  $m' \in \{m-1, m+1\}$  since in all other cases the integrals over the Fourier harmonics of Eqs. (2.59) and (2.60) vanish.

Likewise, we can project the scattering term of Eq. (2.15). Usually we would have to expand the transition rates in Fourier harmonics as well but the transition rates discussed in Sect. 2.4.2 neither depend on the angle of the initial state nor on the angle of the final state, i.e.

$$S_\eta^{\nu, \nu'}(y; \mathbf{k}, \mathbf{k}') \equiv S_\eta^{\nu, \nu'}(y; k, k'),$$

where  $|\mathbf{k}| = k$  and  $|\mathbf{k}'| = k'$ . Therefore, the projection reads

$$\begin{aligned} S_m^\nu(y, E) &:= \int \frac{d^2 k}{(2\pi)^2} S^\nu(y, \mathbf{k}) \delta(E - E^\nu(\mathbf{k})) Y_m(\phi) \\ &= \Omega \sum_{\eta, \nu'} \int \frac{d^2 k}{(2\pi)^2} \frac{d^2 k'}{(2\pi)^2} \left[ (1 - f^\nu(y, \mathbf{k})) S_\eta^{\nu, \nu'}(y; k, k') f^{\nu'}(y, \mathbf{k}') \right. \\ &\quad \left. - (1 - f^{\nu'}(y, \mathbf{k}')) S_\eta^{\nu', \nu}(y; k', k) f^\nu(y, \mathbf{k}) \right] \delta(E - E^\nu(\mathbf{k})) Y_m(\phi) \\ &= \frac{\Omega}{Y_0} \sum_{\eta, \nu'} Z^\nu Z^{\nu'} \int dE' \left[ \left( \frac{\delta_{m,0}}{Y_0} - f_m^\nu(y, E) \right) S_\eta^{\nu, \nu'}(y; E, E') f_0^{\nu'}(y, E') \right. \\ &\quad \left. - \left( \frac{1}{Y_0} - f_0^{\nu'}(y, E') \right) S_\eta^{\nu', \nu}(y; E', E) f_m^\nu(y, E) \right], \end{aligned} \quad (2.63)$$

where we used the orthonormality of Fourier harmonics of Eq. (2.56) extensively. Note how the angular independence of the transition rate immediately leads to a scattering term where the initial state in the  $m$ -th harmonic can only connect to final states in the zeroth harmonic, i.e. states that are radially symmetric.

The projection of the contact GR rate of Eq. (2.38) is straightforward and yields

$$\begin{aligned} \Gamma_m^\nu(y, E) &:= \int \frac{d^2 k}{(2\pi)^2} \Gamma^\nu(y, \mathbf{k}) \delta(E - E^\nu(\mathbf{k})) Y_m(\phi) \\ &= -v_{\text{GR}} Z^\nu \left( f_m^\nu(y, E) - \frac{\delta_{m,0}}{Y_0} f_{\text{eq}}^\nu(y, E) \right) (\delta(y - y_{\text{S}}) + \delta(y - y_{\text{D}})). \end{aligned} \quad (2.64)$$

With Eqs. (2.58), (2.63), and (2.64), the total BE projected onto Fourier harmonics and equienergy circles is given by

$$0 = \left[ F^{\text{BE}} \right]_m^\nu(y, E) = L_m^\nu(y, E) - S_m^\nu(y, E) - \Gamma_m^\nu(y, E).$$

### ***H*-Transformation**

The *H*-transformation [18] has been widely employed to stabilize the BE in deterministic solvers. Fundamentally, it is a simple transformation of the BE from the kinetic energy *E* to the total energy *H*, which – in our case – is defined as the sum of the subband energy and the kinetic energy

$$H^\nu(y, E) = \varepsilon^\nu(y) + E. \quad (2.65)$$

Therefore the transformation is position and subband dependent and reads

$$E \longrightarrow \varepsilon^\nu(y) + E = H$$

The advantage of this representation reveals itself in the free streaming term. Consider a non-interacting carrier whose transport is described by the free streaming term. While the potential energy and therefore the kinetic energy of the carrier might change, the total energy is conserved. Thus, in the ballistic regime, we can describe the transport of the carrier with a single value of *H*. Effectively, we decouple the free streaming term in energy space, making it numerically stable. Even in the case where scattering is involved, the numerical properties of the BE will be better with the *H*-transformation. Particularly in our case, i.e. in nanoscale devices, where transport is close to ballistic.

To see this, we are going to perform the *H*-transformation on the BE as was done in Ref. [29]. The free streaming term of Eq. (2.58) contains two derivatives which behave under *H*-transformation as

$$\frac{\partial}{\partial E} \longrightarrow \frac{\partial}{\partial H}, \quad \frac{\partial}{\partial y} \longrightarrow \frac{\partial}{\partial y} + \frac{\partial \varepsilon^\nu(y)}{\partial y} \frac{\partial}{\partial H}.$$

If we use Eq. (2.14) for the force and apply the *H*-transformation to the projected free streaming term of Eq. (2.58), we find

$$L_m^\nu(y, E) \longrightarrow L_m^\nu(y, H) = Z^v \sum_{m'=m-1, m+1} \left[ \frac{\partial}{\partial y} (a_{m, m'}^\nu(y, H) f_{m'}^\nu(y, H)) - b_{m, m'}^\nu(y, H) f_{m'}^\nu(y, H) \right], \quad (2.66)$$

where the term with the energy derivative has conveniently cancelled. Note that – technically – functions of *H* are different from functions of *E* but we reused

the symbols to avoid cluttering the notation, since the reader should be able to discern the functions by their arguments.

The  $H$ -transformation of the scattering term of Eq. (2.63) is trivial and yields

$$S_m^\nu(y, E) \longrightarrow S_m^\nu(y, H) = \frac{\Omega}{Y_0} \sum_{\eta, \nu'} Z^\nu Z^{\nu'} \int_{\varepsilon^\nu(y)}^{\infty} dH' \\ \times \left[ \left( \frac{\delta_{m,0}}{Y_0} - f_m^\nu(y, H) \right) S_{\eta}^{\nu, \nu'}(y; H, H') f_0^{\nu'}(y, H') \right. \\ \left. - \left( \frac{1}{Y_0} - f_0^{\nu'}(y, H') \right) S_{\eta}^{\nu', \nu}(y; H', H) f_m^\nu(y, H) \right].$$

However, one aspect to keep in mind about the  $H$ -transformation – and which will be of significance later in this work – is that the support of integrals in  $H$ -space in a subband  $\nu$  at position  $y$  is defined as

$$H \in [\varepsilon^\nu(y), \infty). \quad (2.67)$$

Obviously the kinetic energy must be zero or positive and therefore carriers can only occupy states on or above the subband edge.

Using the common representation of transition rates defined by Eq. (2.21), we can express the scattering term as

$$S_m^\nu(y, H) = \frac{1}{Y_0} \sum_{\eta, \sigma, \nu'} Z^\nu Z^{\nu'} \\ \times \left[ \left( \frac{\delta_{m,0}}{Y_0} - f_m^\nu(y, H) \right) c_{\eta}^{\nu, \nu'}(y; H, H + \sigma \hbar \omega_\eta) f_0^{\nu'}(y, H + \sigma \hbar \omega_\eta) \right. \\ \left. - \left( \frac{1}{Y_0} - f_0^{\nu'}(y, H - \sigma \hbar \omega_\eta) \right) c_{\eta}^{\nu', \nu}(y; H - \sigma \hbar \omega_\eta, H) f_m^\nu(y, H) \right], \quad (2.68)$$

where the transition rate coefficients  $c$  are given by Eqs. (2.28), (2.31), and (2.35). Note that for elastic processes we assume  $\hbar \omega_\eta = 0$  and  $\sigma = 0$ . For the inelastic inter-valley phonon scattering, we have  $\hbar \omega_\eta$  given by Tab. 2.1 and  $\sigma = \pm 1$ .

Finally, the  $H$ -transformation of the contact GR rate of Eq. (2.64) is straightforward and reads

$$\Gamma_m^\nu(y, E) \longrightarrow \Gamma_m^\nu(y, H) = -v_{\text{GR}} Z^\nu \left( f_m^\nu(y, H) - \frac{\delta_{m,0}}{Y_0} f_{\text{eq}}(y, H) \right) \\ \times \left( \delta(y - y_{\text{S}}) + \delta(y - y_{\text{D}}) \right) \quad (2.69)$$

with the equilibrium distribution function of Eq. (2.47) transformed into  $H$ -space as

$$f_{\text{eq}}(y_{\text{S/D}}, H) = \frac{1}{\exp \left( \frac{H + qV_{\text{appl}}^{\text{S/D}}}{k_B T} - \log \left( \frac{n_i}{N_{\text{3D}}^{\text{eff}}} \right) \right) + 1}. \quad (2.70)$$

Thus, with Eqs. (2.66), (2.68), and (2.68), the total  $H$ -transformed BE reads

$$0 = \left[ F^{\text{BE}} \right]_m^\nu(y, H) = L_m^\nu(y, H) - S_m^\nu(y, H) - \Gamma_m^\nu(y, H). \quad (2.71)$$

### 2.4.5 Discretization

To complete the discussion of the BE, we will discuss how to properly discretize the BE. The whole treatment has been performed in Ref. [29], however, not in as much detail as here where we will attempt to remove any ambiguities and sources for errors.

The discretization of the BE is based on the finite volume method – or box-integration method. We start out with the Herring-Vogt transformed,  $H$ -transformed, and projected BE of Eq. (2.71). We define a *direct* spatial grid and an *adjoint* spatial grid in transport direction. Quantities on the direct grid points represent densities while quantities on the adjoint grid represent fluxes. Furthermore, we use a *constant interpolation* scheme where each quantity is considered constant in the box surrounding the grid point it is defined on.

#### Grid

Let us define the set of grid points in transport direction

$$y_i \in \{y_1, y_2, \dots, y_{N_y}\}, \quad (2.72)$$

which may be non-equidistant, and a set of grid points in  $H$ -space

$$H_j \in \{H_0, H_1, H_2, \dots, H_{N_H}, H_{N_H+1}\}, \quad (2.73)$$

which we will require to be equidistant and which contains dummy grid points  $H_0$  and  $H_{N_H+1}$  for later convenience. Recall that  $H$  is the total energy as defined by Eq. (2.65). The grid looks as in Fig. 2.12, where the subband energy  $\varepsilon^\nu(y)$  of some subband  $\nu$  has been schematically drawn. Below the subband energy  $\varepsilon^\nu(y)$  no electron states can exist and therefore all transport happens at or above the subband energy.

It is important to point out that the  $H$ -grid needs to be constant throughout the whole computation. In fact, the whole idea of the  $H$ -grid is to express all energies w.r.t. a constant grid and therefore it needs to be globally defined and cannot change between iterations.

Figure 2.12 also shows the grid points of the adjoint spatial grid as small crosses in between the direct grid points. We will use the following notation to reference adjoint grid points:

$$y_{i+} = \frac{y_{i+1} + y_i}{2}, \quad i \in \{1, \dots, N_y - 1\},$$

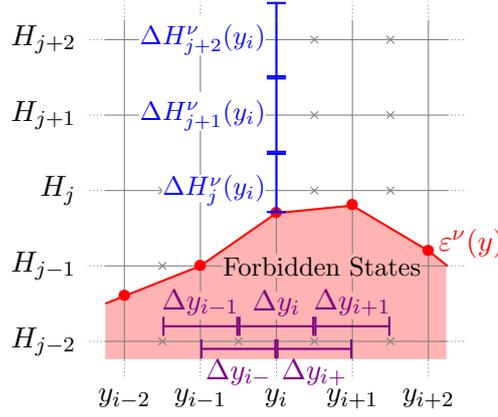


Figure 2.12: Illustration of the tensor grid of the BE comprising the transport direction and  $H$ -space. Adjoint grid points in the transport direction are marked with crosses. Boxes in  $y$ -space for direct and adjoint grids are shown surrounding the direct and adjoint grid points, respectively. Since there is no adjoint grid in  $H$ -space, there are only boxes surrounding the direct grid points. Note that there are no boxes in  $H$ -space below the subband energy and that the lowest box is truncated by the subband energy.

$$y_{i-} = \frac{y_i + y_{i-1}}{2}, \quad i \in \{2, \dots, N_y\},$$

where the boundary points are defined as

$$y_{1-} = y_1, \quad y_{N_y+} = y_{N_y}.$$

Obviously, we have duplicate definitions since, e.g.,  $y_{i+} = y_{(i+1)-}$ . Nevertheless, this notation will prove convenient later on.

For the box-integration, we define boxes in transport direction as illustrated in Fig. 2.12. Around direct grid points, the boxes are given by

$$\Delta y_i = y_{i+} - y_{i-}, \quad i \in \{1, \dots, N_y\} \quad (2.74)$$

and around adjoint grid points by

$$\begin{aligned} \Delta y_{i+} &= y_{i+1} - y_i, & i &\in \{1, \dots, N_y - 1\}, \\ \Delta y_{i-} &= y_i - y_{i-1}, & i &\in \{2, \dots, N_y\}. \end{aligned}$$

The boxes of the  $H$ -grid are more complicated to express due to the subband energy as shown in Fig. 2.12. As implied by Eq. (2.67), the support of  $H$ -space integrations is limited by the subband energy which must be reflected in the boxes. Hence, the  $H$ -box is either zero if it is completely below the subband energy, or it is truncated by the subband energy if the upper limit is above

the subband energy and the lower limit is below, or the box is complete if the subband energy is below the lower limit. To quantify, the  $H$ -box is given by

$$\Delta H^\nu(y, H_j) = \begin{cases} 0, & \text{if } \varepsilon^\nu(y) \geq H_{j+}, \\ H_{j+} - \varepsilon^\nu(y), & \text{if } H_{j+} > \varepsilon^\nu(y) > H_{j-}, \\ H_{j+} - H_{j-}, & \text{else.} \end{cases} \quad (2.75)$$

Here, the intermediary  $H$ -grid points are defined analogously to the adjoint spatial points as

$$\begin{aligned} H_{j+} &= \frac{H_{j+1} + H_j}{2}, & j \in \{1, \dots, N_H - 1\} \\ H_{j-} &= \frac{H_j + H_{j-1}}{2}, & j \in \{2, \dots, N_H\}, \\ H_{1-} &= \frac{H_1 + H_0}{2}, & H_{N_H+} = \frac{H_{N_H+1} + H_{N_H}}{2}. \end{aligned}$$

Note that the definition of the  $H$ -box of Eq. (2.75) does not suggest whether the  $H$ -box is on a direct or adjoint grid point in transport direction. This is because we need to be able to evaluate both kinds of boxes. If the box is on a direct grid point  $y_i$  the reference subband energy is given by  $\varepsilon^\nu(y_i)$ . However, if we need to evaluate the box on an adjoint point  $y_{i+}$  the reference energy must be given by the average,

$$\varepsilon^\nu(y_{i+}) = \frac{\varepsilon^\nu(y_{i+1}) + \varepsilon^\nu(y_i)}{2}, \quad (2.76)$$

since we only solved the SE on direct grid points.

Another idea important to the  $H$ -space discretization is that we can truncate the  $H$ -grid at high energy. Technically energy states stretch all the way to infinity, however, it becomes increasingly unlikely to find carriers occupying states of higher energies. In fact, the probability to find carriers at high energies decreases exponentially (cf. the Fermi-Dirac distribution function of Eq. (2.47)). Therefore we can argue that at some energy the occupation is so low that its contribution to observables drops below the numerical precision of our computation and can be safely neglected. Bear in mind that this means that the distribution function vanishes and not that the energy grid is limited. Hence, the box in  $H$ -space surrounding the uppermost grid point  $H_{N_H}$  is a complete box extending up to  $H_{N_H+}$ , which is different from the way the boxes on the boundaries of  $y$ -space work.

Now let us take a look at which quantities constitute densities and which ones are fluxes and how they are associated with the direct and adjoint grids in  $y$ -space. First off, any quantity depending on  $\mathbf{k}$ -space should be expanded into Fourier harmonics as in Eq. (2.54). Our choice of the Fourier harmonics

of Eq. (2.55) yields that under inversion, i.e.  $\mathbf{k} \rightarrow -\mathbf{k}$ , the even harmonics are invariant – like densities – while the odd harmonics will switch their signs – like fluxes. A flux describes a change in density from one grid point to another and therefore, we are going to define fluxes on the adjoint grid points. Consequently any quantity with an odd Fourier harmonic index is defined on the adjoint grid points, while the even ones are defined on direct grid points.

As a last remark, it might seem surprising that adjoint grid points do not take into consideration energy space, i.e. adjoint grid points are simply in between direct  $y$ -grid points but at the same energy. The reason for this is that the  $H$ -transformation eliminates the derivative in energy space as shown in Sect. 2.4.4. Therefore, it is not necessary to describe fluxes in  $H$ -space and there's no need for an adjoint grid in  $H$ -space.

### Free Streaming Term

The structure of the discretized free streaming term needs to be chosen carefully in order to find a numerically sound representation. We will use the maximum entropy dissipation scheme, which requires us to use a certain form for the free streaming term. The full treatment of this is beyond the scope of this work and therefore we refer the reader to Refs. [24, 29, 96, 97, 98, 99] for more details. Here, we will only present the results thereof.

We start out from the free streaming term of Eq. (2.66) to find a suitable expression on the grid points  $y_i$  and  $H_j$  in the subband  $\nu$  and for the *even* Fourier harmonic  $m$ . To this end, let us write down the expression for the box-integration

$$\bar{L}_m^\nu(y_i, H_j) := \int_{\Delta H^\nu(y_i, H_j)} dH \int_{y_{i-}}^{y_{i+}} dy L_m^\nu(y, H) \Big|_{m \text{ even}},$$

where  $\int_{\Delta H^\nu(y_i, H_j)}$  denotes the integral over the appropriate box surrounding the grid point  $H_j$  as given by Eq. (2.75). As previously discussed, quantities on direct grid points can only be coefficients of even Fourier harmonics. Thus it is understood that  $\bar{L}_m^\nu(y_i, H_j)$  is only defined for even harmonics  $m$  since it is evaluated on the direct grid point  $y_i$ .

The resulting expression has been derived in [29] and is given by

$$\begin{aligned} \bar{L}_m^\nu(y_i, H_j) = Z^\nu \sum_{m'=m\pm 1} & \left[ f_{m'}^\nu(y_{i+}, H_j) \int_{\Delta H^\nu(y_{i+}, H_j)} dH a_{m,m'}(y_{i+}, H) \right. \\ & - f_{m'}^\nu(y_{i-}, H_j) \int_{\Delta H^\nu(y_{i-}, H_j)} dH a_{m,m'}(y_{i-}, H) \\ & - \frac{1}{2} f_{m'}^\nu(y_{i+}, H_j) \Delta y_{i+} \int_{\Delta H^\nu(y_{i+}, H_j)} dH b_{m,m'}(y_{i+}, H) \\ & \left. - \frac{1}{2} f_{m'}^\nu(y_{i-}, H_j) \Delta y_{i-} \int_{\Delta H^\nu(y_{i-}, H_j)} dH b_{m,m'}(y_{i-}, H) \right] \Big|_{m \text{ even}}, \end{aligned} \quad (2.77)$$

where it becomes obvious that the free streaming term on the direct grid point  $y_i$  with even harmonic  $m$ , only couples to distribution functions on the surrounding adjoint grid points  $y_{i\pm}$  with odd harmonics  $m' = m \pm 1$ .

There is a difficulty with the remaining integrals over the coefficients  $a$  and  $b$  of Eq. (2.77) that needs to be addressed before we proceed. Note that we cannot simply assume that the integral over  $H$  is the center value of the integrand times the width of the  $H$ -box since it would lead to problems in the stability and quality of the solution. Instead, we need to perform these integrals analytically and with special care for the integration boundaries.

The free streaming term describes how free carriers propagate from one grid point to another. Apart from constants and the distribution function, the coefficients  $a$  and  $b$  govern the rate of transfer from one grid point in transport direction to the next one. In particular, if certain energy ranges are forbidden due to the subband energy, it will be reflected in the coefficients  $a$  and  $b$ . Therefore, it is of great importance to determine the discretization of these coefficients correctly.

To understand how the box-integration works for the coefficients, let us take a look at Fig. 2.13. It depicts how a transition from the grid point  $y_i$  to the grid point  $y_{i+1}$  according to the free streaming term works. Remember, that the free streaming term only describes free carriers which move along constant total energy lines, i.e. at constant  $H$  or horizontally in Fig. 2.13. In a device, we are confronted with the case that the subband energy changes from one grid point to the next one. Therefore states at certain energies might exist at position  $y_i$  but might be below the subband energy at the grid point  $y_{i+1}$ .

Obviously, such transitions cannot happen and therefore their rate is zero which means they have to be excluded from the box-integration in  $H$ -space. To see how this can be accommodated in our discretization, let us set up the

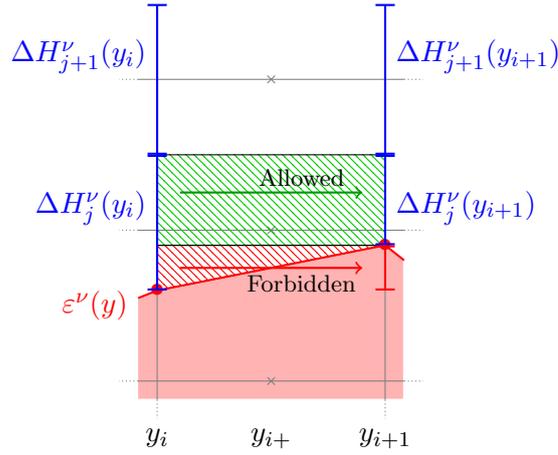


Figure 2.13: Allowed and forbidden transitions of the coefficients  $a_{m,m'}(y_{i+}, H)$  and  $b_{m,m'}(y_{i+}, H)$ .

integral over the coefficient  $a$  of Eq. (2.59) which can be expressed as follows

$$\begin{aligned}
 A_{m,m'}^\nu(y_{i+}, H_j) &:= \int_{\Delta H^\nu(y_i, H_j)} dH a_{m,m'}(y_{i+}, H) \\
 &= (T_{\text{HV}}^v)_{yy} \int d\phi Y_{m'}(\phi) Y_m(\phi) Y_1(\phi) \\
 &\quad \times \int_{\Delta H^\nu(y_{i+}, H_j)} dH (v_y^\nu)_1(y_{i+}, H),
 \end{aligned}$$

where the group velocity has been transformed to  $H$ -space, thereby making it position dependent. Here the quantity of interest is the integral over the group velocity of Eq. (2.61). In  $H$ -space the group velocity reads

$$(v_y^\nu)_1(y_{i+}, H) = \sqrt{\frac{2\pi(H - \varepsilon^\nu(y_{i+}))}{m_d^\nu}}$$

and the subband energy on an adjoint grid point is simply given by the arithmetic mean of the surrounding direct grid points as given in Eq. (2.76).

Taking into account that carriers at energies below the subband energy of either surrounding grid point cannot propagate, we find that the integral over the group velocity reads

$$\int_{\Delta H^\nu(y_{i+}, H_j)} dH (v_y^\nu)_1(y_{i+}, H)$$

$$= \begin{cases} \sqrt{\frac{2\pi}{m_d^v} \frac{2}{3}} (H - \varepsilon^\nu(y_{i+}))^{\frac{3}{2}} \Big|_{H=H_{j-}}^{H_{j+}}, & \text{if } H_{j-} \geq \max(\varepsilon^\nu)_{i+}, \\ \sqrt{\frac{2\pi}{m_d^v} \frac{2}{3}} (H - \varepsilon^\nu(y_{i+}))^{\frac{3}{2}} \Big|_{H=\max(\varepsilon^\nu)_{i+}}^{H_{j+}}, & \text{if } H_{j+} > \max(\varepsilon^\nu)_{i+} > H_{j-}, \\ 0, & \text{else,} \end{cases} \quad (2.78)$$

with

$$\max(\varepsilon^\nu)_{i+} := \max(\varepsilon^\nu(y_i), \varepsilon^\nu(y_{i+1})).$$

Note that even if the  $H$ -boxes are measured w.r.t. the availability of states on the surrounding grid points, the velocity itself is measured w.r.t. the distance to the subband energy  $\varepsilon^\nu(y_{i+})$  on the adjoint grid point.

Completely analogous, we can obtain the  $H$ -box integral over the coefficient  $b$  of Eq. (2.60) given by

$$\begin{aligned} B_{m,m'}^\nu(y_{i+}, H_j) &:= \int_{\Delta H^\nu(y_{i+}, H_j)} dH b_{m,m'}(y_{i+}, H) \\ &= (T_{\text{HV}}^v)_{yy} F^\nu(y_{i+}) \frac{m\sqrt{\pi}}{\hbar} \int d\phi Y_{m'}(\phi) Y_{-m}(\phi) Y_{-1}(\phi) \\ &\quad \times \int_{\Delta H^\nu(y_{i+}, H_j)} \frac{dH}{k^\nu(y_{i+}, H)} \end{aligned}$$

where the discretized force of Eq. (2.14) is simply given by the finite difference

$$F^\nu(y_{i+}) = -\frac{\varepsilon^\nu(y_{i+1}) - \varepsilon^\nu(y_i)}{\Delta y_{i+}}$$

and the magnitude of the wave vector of Eq. (2.53) transformed into  $H$ -space is given by

$$k^\nu(y_{i+}, H) = \sqrt{\frac{2m_d^v(H - \varepsilon^\nu(y_{i+}))}{\hbar^2}}$$

and therefore the  $H$ -box integration reads

$$\begin{aligned} &\int_{\Delta H^\nu(y_{i+}, H_j)} dH \frac{dH}{k^\nu(y_{i+}, H)} \\ &= \begin{cases} \sqrt{\frac{2\hbar^2}{m_d^v}} (H - \varepsilon^\nu(y_{i+}))^{\frac{1}{2}} \Big|_{H=H_{j-}}^{H_{j+}}, & \text{if } H_{j-} \geq \max(\varepsilon^\nu)_{i+}, \\ \sqrt{\frac{2\hbar^2}{m_d^v}} (H - \varepsilon^\nu(y_{i+}))^{\frac{1}{2}} \Big|_{H=\max(\varepsilon^\nu)_{i+}}^{H_{j+}}, & \text{if } H_{j+} > \max(\varepsilon^\nu)_{i+} > H_{j-}, \\ 0, & \text{else.} \end{cases} \end{aligned}$$

Let us now turn to the free streaming term on the adjoint grid point. It might be tempting to simply reuse the free streaming term of Eq. (2.66) and evaluate it on the adjoint grid point but this approach brings along inconsistencies. Reference [29] points out that we need to use an equivalent adjoint form of the free streaming term projected onto odd Fourier harmonics. It can be derived with the continuum relation

$$\frac{\partial}{\partial y} a_{m,m'}^\nu(y, H) = b_{m,m'}(y, H) + b_{m',m}(y, H),$$

which can be applied to Eq. (2.66) to obtain

$$(L^{\text{ad}})_m^\nu(y, H) = Z^\nu \sum_{m'=m-1, m+1} \left[ a_{m,m'}^\nu(y, H) \frac{\partial}{\partial y} f_{m'}^\nu(y, H) + b_{m',m}^\nu(y, H) f_{m'}^\nu(y, H) \right]. \quad (2.79)$$

In the continuum Eqs. (2.66) and (2.79) are equivalent, however, only the adjoint form of Eq. (2.79) will yield a discretization in accordance with the maximum entropy dissipation scheme. This is important as free – or non-interacting – carriers as described by the free streaming term cannot change the entropy and hence this should also be reflected in the discretized system.

Applying the box-integration yields the expression [29]

$$\begin{aligned} \bar{L}_m^\nu(y_{i+}, H_j) &:= \int_{\Delta H^\nu(y_i, H_j)} dH \int_{y_i}^{y_{i+1}} dy (L^{\text{ad}})_m^\nu(y, H) \Big|_{m \text{ odd}} \\ &= Z^\nu \sum_{m'=m\pm 1} \left[ A_{m,m'}(y_{i+}, H_j) (f_{m'}^\nu(y_{i+1}, H_j) - f_{m'}^\nu(y_i, H_j)) \right. \\ &\quad \left. + \frac{1}{2} B_{m',m}(y_{i+}, H_j) (f_{m'}^\nu(y_{i+1}, H_j) + f_{m'}^\nu(y_i, H_j)) \Delta y_{i+} \right] \Big|_{m \text{ odd}}, \end{aligned} \quad (2.80)$$

where it is understood that  $m$  is odd on the adjoint grid point.

### Scattering Term

The discretization of the scattering term is mostly straightforward but there are still some pitfalls of which to be wary of. First of all, we are going to consider only energy transfers that are multiples of our  $H$ -grid spacing. This is not a necessary simplification but one that will simplify the discretization of the scattering term while only minorly impacting scattering rates when the  $H$ -grid is chosen fine enough. Hence, typical transitions between two energy states appear as depicted by Fig. 2.14.

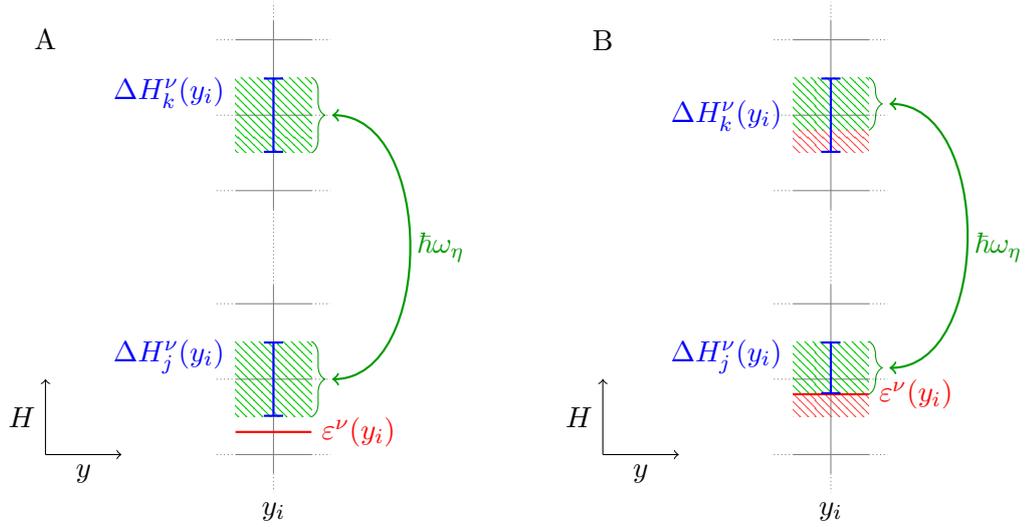


Figure 2.14: Scattering transition between the  $H$ -grid points  $H_j$  and  $H_k$  with an energy transfer  $\hbar\omega_\eta$  of a scattering process  $\eta$  which is mapped to a multiple of the grid spacing. A: Subband energy is below the lower energy state. B: Subband energy truncates the lower end of the box leading to a reduced box size.

You can see that the mapped phonon energies in Fig. 2.14A naturally transfer one  $H$ -grid box into another. In Fig. 2.14B we visualized the case when the lower box is truncated by the subband energy. Remember that states below the subband energy are not accessible to carriers, therefore we cannot integrate over these energy states. In that case, only a subset of the box can be considered during the box-integration. Thus, the size of the box for a scattering process with energy transfer  $\hbar\omega$  from the energy  $H_j$  to  $H_j + \hbar\omega$  is given by

$$\min(\Delta H^\nu(y, H_j), \Delta H^{\nu'}(y, H_j + \hbar\omega)),$$

with the  $H$ -boxes defined by Eq. (2.75).

It follows immediately that the box-integrated scattering integral of Eq. (2.68) on a direct grid point with even harmonic  $m$  is given by

$$\begin{aligned} \bar{S}_m^\nu(y_i, H_j) &= \int_{H_{j-}}^{H_{j+}} dH \int_{y_{i-}}^{y_{i+}} dy S_m^\nu(y, H) \Big|_{m \text{ even}} \\ &= \frac{1}{Y_0} \sum_{\sigma, \eta, \nu'} Z^\nu Z^{\nu'} \\ &\quad \times \left[ \left( \frac{\delta_{m,0}}{Y_0} - f_m^\nu(y_i, H_j) \right) c_{\eta}^{\nu, \nu'}(y_i; H_j, H_j + \sigma \hbar\omega_\eta) f_0^{\nu'}(y_i, H_j + \sigma \hbar\omega_\eta) \right. \\ &\quad \left. \times \min(\Delta H^\nu(y_i, H_j), \Delta H^{\nu'}(y_i, H_j + \sigma \hbar\omega_\eta)) \right] \end{aligned}$$

$$\begin{aligned}
& - \left( \frac{1}{Y_0} - f_0^{\nu'}(y_i, H_j - \sigma \hbar \omega_\eta) \right) c_\eta^{\nu', \nu}(y_i; H_j - \sigma \hbar \omega_\eta, H_j) f_m^\nu(y_i, H_j) \\
& \quad \times \min(\Delta H^\nu(y_i, H_j), \Delta H^{\nu'}(y_i, H_j - \sigma \hbar \omega_\eta)) \Big] \Delta y_i \Big|_{m \text{ even}}. \quad (2.81)
\end{aligned}$$

Likewise, on the adjoint grid and with an odd harmonic  $m$ , we obtain

$$\begin{aligned}
\bar{S}_m^\nu(y_{i+}, H_j) &= \int_{H_{j-}}^{H_{j+}} dH \int_{y_i}^{y_{i+1}} dy S_m^\nu(y, H) \Big|_{m \text{ odd}} \\
&= \frac{1}{Y_0} \sum_{\sigma, \eta, \nu'} Z^\nu Z^{\nu'} \\
& \quad \times \left[ -\frac{1}{2} f_m^\nu(y_{i+}, H_j) \left( c_\eta^{\nu', \nu}(y_i; H_j, H_j + \sigma \hbar \omega_\eta) f_0^{\nu'}(y_i, H_j + \sigma \hbar \omega_\eta) \right. \right. \\
& \quad \quad \left. \left. + c_\eta^{\nu', \nu}(y_{i+1}; H_j, H_j + \sigma \hbar \omega_\eta) f_0^{\nu'}(y_{i+1}, H_j + \sigma \hbar \omega_\eta) \right) \right. \\
& \quad \quad \times \min(\Delta H^\nu(y_{i+}, H_j), \Delta H^{\nu'}(y_{i+}, H_j + \sigma \hbar \omega_\eta)) \\
& \quad - \frac{1}{Y_0} \frac{1}{2} \left( c_\eta^{\nu', \nu}(y_i; H_j - \sigma \hbar \omega_\eta, H_j) \right. \\
& \quad \quad \left. + c_\eta^{\nu', \nu}(y_{i+1}; H_j - \sigma \hbar \omega_\eta, H_j) \right) f_m^\nu(y_{i+}, H_j) \\
& \quad \quad \times \min(\Delta H^\nu(y_{i+}, H_j), \Delta H^{\nu'}(y_{i+}, H_j - \sigma \hbar \omega_\eta)) \\
& \quad + \frac{1}{2} \left( f_0^{\nu'}(y_i, H_j - \sigma \hbar \omega_\eta) c_\eta^{\nu', \nu}(y_i; H_j - \sigma \hbar \omega_\eta, H_j) \right. \\
& \quad \quad \left. + f_0^{\nu'}(y_{i+1}, H_j - \sigma \hbar \omega_\eta) c_\eta^{\nu', \nu}(y_{i+1}; H_j - \sigma \hbar \omega_\eta, H_j) \right) f_m^\nu(y_{i+}, H_j) \\
& \quad \quad \left. \times \min(\Delta H^\nu(y_{i+}, H_j), \Delta H^{\nu'}(y_{i+}, H_j - \sigma \hbar \omega_\eta)) \right] \Delta y_{i+} \Big|_{m \text{ odd}}. \quad (2.82)
\end{aligned}$$

When trying to discretize the scattering term, we run into a possible ambiguity in the implementation concerning the domains of definition of quantities relating to even and odd harmonics. Fourier components with even harmonics are defined on direct grid points, while Fourier components with odd harmonics are defined on the adjoint grid. Keeping this in mind, the integration over the spatial grid behaves as follows: Consider a product of some functions  $d_n$  on the direct grid and some functions  $a_m$  on the adjoint grid. Then the integral over a box surrounding a direct grid point  $y_i$  needs to be computed as

$$\begin{aligned}
& \int_{i-}^{i+} dy \left( \prod_n d_n(y) \right) \left( \prod_m a_m(y) \right) \\
& \quad = \Delta y_i \left( \prod_n d_n(y_i) \right) \frac{\prod_m a_m(y_{i+}) + \prod_m a_m(y_{i-})}{2}.
\end{aligned}$$

If we were not interested in the noise in the device, we might also average all quantities individually and then multiply them together. However, anticipating the calculation of the power spectral density (cf. Sect. 4.6.3) in the device and the verification through the Nyquist theorem (cf. Sect. 5.4.2), we need to keep in mind that the averaging scheme needs to be consistent, which cannot be guaranteed unless we use proper box-integration.

### Boundary Conditions

The boundary GR term of Eq. (2.69) can only exist on the direct grid points on the contacts due to the delta-distributions. Therefore we find

$$\begin{aligned}\bar{\Gamma}_m^\nu(y_i, H_j) &= \int_{H_{j-}}^{H_{j+}} dH \int_{y_{i-}}^{y_{i+}} dy \Gamma_m^\nu(y, H) \Big|_{m \text{ even}} \\ &= -v_{\text{GR}} (f^\nu(y_i, H_j) - f_{\text{eq}}^\nu(y_i, H_j)) (\delta_{i,1} + \delta_{i,N_y}) \Delta H^\nu(y_i, H_j),\end{aligned}\quad (2.83)$$

where the Kronecker-deltas assure that the grid point  $i$  is either the source contact grid point  $y_1 = y_S$  or the drain contact grid point  $y_{N_y} = y_D$ . Note that the box-integration effectively transforms  $\Gamma$  from a singular GR rate on the contacts – equivalent to a boundary condition – to a volume GR rate  $\bar{\Gamma}$  defined in the volumes  $\Delta y_1$  and  $\Delta y_{N_y}$ .

The box-integrated GR rate vanishes on adjoint grid points:

$$\bar{\Gamma}_m^\nu(y_{i+}, H_j) = \int_{H_{j-}}^{H_{j+}} dH \int_{y_i}^{y_{i+1}} dy \Gamma_m^\nu(y, H) \Big|_{m \text{ odd}} = 0.$$

### 2.4.6 Solution

We applied the Herring-Vogt transformation to the BE, then we projected it onto Fourier harmonics and equienergy surfaces, and subsequently used the box-integration to discretize it. With the resulting free streaming term of Eqs. (2.77) and (2.80), the scattering term of Eqs. (2.81) and (2.82), and the boundary term of Eq. (2.83), we can express the even harmonics of the BE as

$$0 = [\bar{F}^{\text{BE}}]_m^\nu(y_i, H_j) \Big|_{m \text{ even}} = \bar{L}_m^\nu(y_i, H_j) - \bar{S}_m^\nu(y_i, H_j) - \bar{\Gamma}_m^\nu(y_i, H_j) \Big|_{m \text{ even}} \quad (2.84)$$

and the odd harmonics as

$$0 = [\bar{F}^{\text{BE}}]_m^\nu(y_{i+}, H_j) \Big|_{m \text{ odd}} = \bar{L}_m^\nu(y_{i+}, H_j) - \bar{S}_m^\nu(y_{i+}, H_j) \Big|_{m \text{ odd}}. \quad (2.85)$$

These equations constitute a set of equations which can be solved to obtain the distribution functions  $f_m^\nu(y_i, H_j) \Big|_{m \text{ even}}$  and  $f_m^\nu(y_{i+}, H_j) \Big|_{m \text{ odd}}$  for a set of

subbands and valleys  $\nu = (v, s)$ , a set of harmonics  $m$ , a set of direct spatial grid points  $y_i$  and their associated adjoint grid points  $y_{i+}$ , as well as a set of total energies  $H_j$ .

In order to streamline the notation, let us define an aggregate index

$$\alpha := \begin{cases} (\nu, y_i, H_j, m), & m \text{ even,} \\ (\nu, y_{i+}, H_j, m), & m \text{ odd,} \end{cases} \quad (2.86)$$

which runs over all subbands in all valleys  $\nu$ , all energy grid points  $H_j$ , all harmonics  $m$ , and all direct grid points  $y_i$  in the case of an even harmonic or all adjoint grid points  $y_{i+}$  in the case of an odd harmonic. Then, the discretized BE can be expressed as

$$F_\alpha^{\text{BE}} := \begin{cases} [\bar{F}^{\text{BE}}]_m^\nu(y_i, H_j), & m \text{ even,} \\ [\bar{F}^{\text{BE}}]_m^\nu(y_{i+}, H_j), & m \text{ odd.} \end{cases} \quad (2.87)$$

The non-linear system of equations  $F_\alpha^{\text{BE}} = 0$  is directly solvable with the iterative Newton-Raphson approach, i.e. we set some kind of reasonable starting condition for  $f_\alpha$  – e.g. the equilibrium distribution function – and then solve the system of equations of the Newton-Raphson approach given by

$$\sum_\beta \frac{\partial F_\alpha^{\text{BE}}}{\partial f_\beta} \delta f_\beta = -F_\alpha^{\text{BE}}, \quad (2.88)$$

where  $\beta$  is an aggregate index like  $\alpha$ . Thereafter we update the distribution function as

$$f_\alpha \longrightarrow f_\alpha + \delta f_\alpha \quad (2.89)$$

and return to Eq. (2.88) for the next iteration. Although solving Eq. (2.88) directly is possible, it is inefficient since we can still achieve a significant reduction in the size of the linear system by considering degeneracies and the coupling of even and odd harmonics of the BE.

### Degeneracy

Intuitively, it is obvious that states which cannot be distinguished need not be accounted for separately. In the case of the BE of this work, we cannot distinguish between spin up and spin down electrons. A spin up electron will remain a spin up electron and likewise a spin down electron will remain a spin down electron, since neither the free streaming term nor the scattering term are able to flip it. Furthermore, the GR term generates and recombines electrons of both spins equally. Therefore we may compute the distribution function for,

say, spin up electrons and multiply the resulting observables, like densities and currents, with a factor of two.

By omitting a spin index, we have already implicitly assumed that the spin is a degenerate quantum number but in principle we could have exercised all derivations of the present chapter for, say, spin up electrons and would have found that the very same equations also hold for spin down electrons.

A case where it is a bit more difficult to follow how degeneracy works is when we consider the valley degeneracy of silicon. Recall that the band structure of silicon in  $\mathbf{k}$ -space, depicted in Fig. 2.6, has two valleys along each axis where it is not distinguishable whether an electron is in the valley on the positive side of the axis or on the negative side. In both cases the dispersion relations are identical and in both cases there are identical intra- and inter-valley scattering processes with identical rates. Therefore we may conclude that each pair of valleys on one of the axes of  $\mathbf{k}$ -space is degenerate. However, care must be taken with how to reduce the BE in that case. We cannot simply remove the equations w.r.t. one of the degenerate valleys from the system of equations since they are coupled to each other by the inter-valley scattering.

To see how we can reduce the size of the system of equations in that case, let us consider the general case with the following terminology: We will refer to an individual state as a *true state*. If a set of true states are indistinguishable in terms of their description by the BE, we call these states a *degenerate group*. A single true state, which is part of a degenerate group, will be called a *degenerate state*.

Assume we have a BE with  $m$  degenerate groups, where the  $i$ -th group consists of a set of  $\mu_i$  true states. Then we can express Eq. (2.88) as

$$\begin{pmatrix} L^{S,1} & S^{12} & \dots & \dots & S^{1m} \\ S^{21} & L^{S,2} & \dots & \dots & S^{2m} \\ \vdots & & \ddots & & \vdots \\ \vdots & & & \ddots & S^{m-1,m} \\ S^{m1} & \dots & \dots & S^{m,m-1} & L^{S,m} \end{pmatrix} (\delta f) = -(F^{\text{BE}}), \quad (2.90)$$

where the submatrices  $L^{S,i}$  contain  $\mu_i$  degenerate states and possibly balanced scattering process in between the degenerate states of the group. They can be further split up into the subspaces of the true states as

$$L^{S,i} = \begin{pmatrix} L^i & s^i & \dots & s^i \\ s^i & \ddots & & \vdots \\ \vdots & & \ddots & s^i \\ s^i & \dots & s^i & L^i \end{pmatrix},$$

where the  $\mu_i$  free streaming terms  $L^i$  on the main diagonal and the scattering rates  $s^i$  in between the true states are necessarily identical to each other since the states are part of a degenerate group. The submatrices  $S^{ij}$  are scattering processes in between the true states of different degenerate groups. However, since the true states of a degenerate group are indistinguishable, the scattering submatrices can be expressed with the same scattering rates in each of the true state subspaces as

$$S^{ij} = \begin{pmatrix} s^{ij} & \dots & s^{ij} \\ \vdots & \ddots & \vdots \\ s^{ij} & \dots & s^{ij} \end{pmatrix},$$

where  $S^{ij}$  is a  $\mu_i \times \mu_j$ -matrix and  $s^{ij}$  is the scattering submatrix in between a true state of the degenerate group  $i$  and a true state of the degenerate group  $j$ .

Now we can also split up the distribution function and the r.h.s. of Eq. (2.90) into the  $m$  subspaces as

$$\begin{aligned} (\delta f) &= ((\delta f)_1 \quad (\delta f)_2 \quad \dots \quad (\delta f)_m)^t, \\ (F^{\text{BE}}) &= ((F^{\text{BE}})_1 \quad (F^{\text{BE}})_2 \quad \dots \quad (F^{\text{BE}})_m)^t. \end{aligned}$$

Since the states contained in a degenerate group are indistinguishable, the distribution functions and the r.h.s. of the BE must be indistinguishable for each of the true states of a degenerate group, i.e.

$$\begin{aligned} (\delta f)_i &= (\delta f_i \quad \delta f_i \quad \dots \quad \delta f_i)^t, \\ (F^{\text{BE}})_i &= (F_i^{\text{BE}} \quad F_i^{\text{BE}} \quad \dots \quad F_i^{\text{BE}})^t. \end{aligned}$$

Note that both  $\delta f_i$  and  $F_i^{\text{BE}}$  are still vectors in the subspace of a true state.

Then, we can express the BE for the true states of the  $i$ -th degenerate group of Eq. (2.90) as

$$\begin{aligned} -(F_\alpha^{\text{BE}})_i &= L^{S,i}(\delta f)_i + \sum_{\substack{j=1 \\ j \neq i}}^m S^{ij}(\delta f)_j \\ &= \begin{pmatrix} L^i \delta f_i + (\mu_i - 1) s^i \delta f_i \\ \vdots \\ L^i \delta f_i + (\mu_i - 1) s^i \delta f_i \end{pmatrix} + \sum_{\substack{j=1 \\ j \neq i}}^m \begin{pmatrix} \mu_j s^{ij} \delta f_j \\ \vdots \\ \mu_j s^{ij} \delta f_j \end{pmatrix}, \end{aligned}$$

where we obtain  $\mu_i$  subspaces containing identical equations. Therefore, we may simply use one of each of the equations of a degenerate group and reduce the

size of the degenerate BE of Eq. (2.90) by the amount of degeneracy contained in the equation:

$$\begin{pmatrix} L^1 + (\mu_1 - 1) s^1 & \mu_2 s^{12} & \cdots & \cdots & \mu_m s^{1m} \\ \mu_1 s^{21} & L^2 + (\mu_2 - 1) s^2 & \cdots & \cdots & \mu_m s^{2m} \\ \vdots & & \ddots & & \vdots \\ \vdots & & & \ddots & \mu_m s^{m-1,m} \\ \mu_1 s^{m1} & \cdots & \cdots & \mu_{m-1} s^{m,m-1} & L^m + (\mu_m - 1) s^m \end{pmatrix} \times \begin{pmatrix} \delta f_1 \\ \delta f_2 \\ \vdots \\ \delta f_m \end{pmatrix} = - \begin{pmatrix} F_1^{\text{BE}} \\ F_2^{\text{BE}} \\ \vdots \\ F_m^{\text{BE}} \end{pmatrix}. \quad (2.91)$$

Equation (2.91) gives a very simple prescription to treat the spin degeneracy of the BE. We have one degenerate group of spins,  $m = 1$ , comprising the two true states of spin up and spin down,  $\mu_1 = 2$ . We do not have any scattering rates which can flip the spin, therefore  $s^1 = 0$ . From Eq. (2.91) it follows immediately that it is sufficient to calculate the BE of a single spin state, completely ignoring the other spin state since its BE is identical and therefore its solution is identical. Furthermore, any observable that we derive from the distribution function which depends on both spin states, can be computed by replacing one of the spin states with the other one. Usually this means that the observable can be computed with one spin direction and then multiplied by a factor of two.

Treatment of the valley degeneracy is not as straightforward due to the inter-valley scattering involved. Instead of the six  $X$ -valleys, we only want to consider three  $X$ -valleys – one per axis in 3D  $\mathbf{k}$ -space. That means we have three degenerate groups,  $m = 3$ , each containing two valleys,  $\mu_1 = \mu_2 = \mu_3 = 2$ .

Within a degenerate group, the  $g$ -type inter-valley phonon scattering can occur, i.e.  $s^i$  is the discretized scattering term of the  $g$ -type scattering for the  $i$ -th degenerate group. Now, Eq. (2.91) tells us that we can simply consider the  $g$ -type scattering term as if it would scatter from one of the true states into the same true state with an additional factor of  $(\mu_i - 1) = 1$ .

Electrons can scatter in between the degenerate groups through the  $f$ -type inter-valley scattering. Thus we can identify the scattering term  $s^{ij}$  with the discretized  $f$ -type scattering from a single true state in the original degenerate group  $i$ , to a single true state in the final degenerate group  $j$ . Then, Eq. (2.91) tells us that we can compute the BE using only one true state per degenerate group, if we multiply the scattering term  $s^{ij}$  with a factor of  $\mu_j = 2$ . Figure 2.15 illustrates the scattering processes and their associated multiplicities if only one true state per degenerate group is considered.

In the remainder of this work, we will refer to the number of true states in

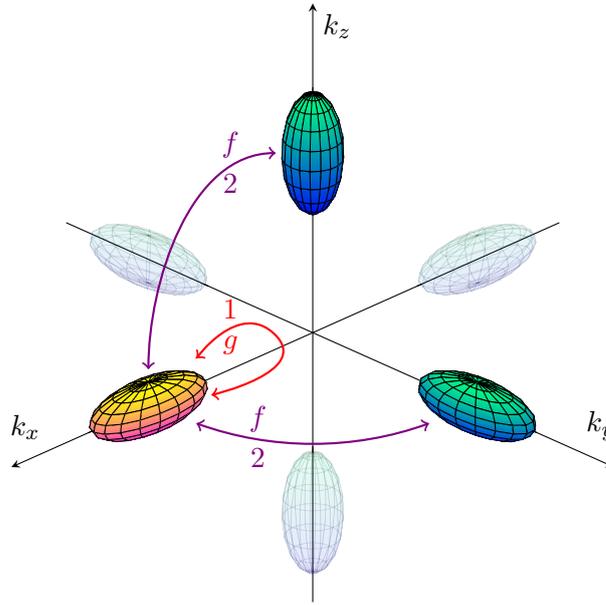


Figure 2.15: Illustration of  $f$ - and  $g$ -type inter-valley phonon scattering processes from the perspective of the X-valley on the positive  $k_x$ -axis (yellow valley) when degeneracy is considered, i.e. only states in the opaque valleys are considered. The states in the transparent valleys are degenerate with the states in the respective opaque valley on the same axis and need not be tracked. Thus, a  $g$ -type inter-valley scattering process scatters into the same valley with a multiplicity of ‘1’, while the  $f$ -type scattering processes scatter into the adjacent two adjacent valleys, each with a multiplicity of ‘2’.

a degenerate group as the *multiplicity* of a state. The spin multiplicity will be referenced as

$$\mu_{\text{spin}} = 2 \quad (2.92)$$

and the valley multiplicity as

$$\mu_{\text{val}} = 2. \quad (2.93)$$

### Elimination of Odd Equations

We can reduce the size of the BE system for the Newton-Raphson approach of Eq. (2.88) by using the inherent structure of the BE when only isotropic scattering rates are considered. For this approach to work, we need to truncate the series of Fourier harmonics in such a way that there is an equal number of even and odd harmonics, which we will assume we are doing from now on.

To this end, consider the discretized BE of Eq. (2.85) for the index  $\alpha = (\nu, y_{i+}, H_j, m)$ , where  $m$  is an *odd* Fourier harmonic. Then the free streaming

term of Eq. (2.80) will only contain distribution functions on the adjacent *direct* grid points  $y_i$  and  $y_{i+1}$  with *even* Fourier harmonics  $m \pm 1$ . The associated isotropic scattering shown in Eq. (2.82) contains even distribution functions of the zeroth Fourier harmonic as well as the only *odd* distribution function  $f_\alpha$  in the whole equation.

Since there is only one odd distribution function in the whole equation, we can use Eq. (2.88) to express it in terms of the even distribution functions as

$$\delta f_\alpha \Big|_{m \text{ odd}} = \frac{1}{A_{\alpha\alpha}^{\text{BE}}} \left( -F_\alpha^{\text{BE}} - \sum_{\substack{\beta=1 \\ \beta \neq \alpha}}^{N_{\text{BE}}} A_{\alpha\beta}^{\text{BE}} \delta f_\beta \right), \quad (2.94)$$

where the elements of the Jacobian are defined by

$$A_{\alpha\beta}^{\text{BE}} := \frac{\partial F_\alpha^{\text{BE}}}{\partial f_\beta}. \quad (2.95)$$

Now let us consider the discretized BE of Eq. (2.84) for the index  $\alpha = (\nu, y_i, H_j, m)$ , where  $m$  is an *even* Fourier harmonic. While the scattering term of Eq. (2.81) and the boundary term of Eq. (2.83) only contains even distribution functions, the free streaming term of Eq. (2.77) contains the *odd* distribution functions  $f_{m\pm 1}^\nu(y_{i\pm}, H_j)$ .

Using Eq. (2.94), we can eliminate all odd distribution functions from equations projected onto even harmonics of Eq. (2.88). For later convenience, let us denote this linear transformation as  $S^{\text{BE}}$  given by

$$A_{\alpha=(\nu, y_i, H_j, m), \beta}^{\text{BE}} \Big|_{m \text{ even}} \xrightarrow{S^{\text{BE}}} A_{\alpha, \beta}^{\text{BE}} - \sum_{\substack{m'=m\pm 1 \\ y_{i'}=y_{i\pm}}} \frac{A_{\alpha, (\nu, y_{i'}, H_j, m')}^{\text{BE}}}{A_{(\nu, y_{i'}, H_j, m'), (\nu, y_{i'}, H_j, m')}^{\text{BE}}} A_{(\nu, y_{i'}, H_j, m'), \beta}^{\text{BE}}, \quad (2.96)$$

such that the rows of  $S^{\text{BE}} A^{\text{BE}}$  associated with indices of even harmonics, do not contain any elements in the columns associated with odd harmonics. The transformation  $S^{\text{BE}}$  acts analogously on the r.h.s. of Eq. (2.88) as

$$F_{\alpha=(\nu, y_i, H_j, m)}^{\text{BE}} \Big|_{m \text{ even}} \xrightarrow{S^{\text{BE}}} F_\alpha^{\text{BE}} - \sum_{\substack{m'=m\pm 1 \\ y_{i'}=y_{i\pm}}} \frac{A_{\alpha, (\nu, y_{i'}, H_j, m')}^{\text{BE}}}{A_{(\nu, y_{i'}, H_j, m'), (\nu, y_{i'}, H_j, m')}^{\text{BE}}} F_{(\nu, y_{i'}, H_j, m')}^{\text{BE}}. \quad (2.97)$$

Since the equations with even harmonics are completely decoupled from the equations with odd harmonics, we can determine the distribution functions using only half the equations.

Later on we are going to need a formal description of how the system of equations is reduced, therefore we will introduce some of the necessary notation here. Let  $N_{\text{BE}}$  be the number of indices of the BE. Then the Jacobian  $A^{\text{BE}}$  is an  $N_{\text{BE}} \times N_{\text{BE}}$ -matrix and Eq. (2.88) can be written down in matrix form as

$$A^{\text{BE}} \delta \mathbf{f} = -\mathbf{F}^{\text{BE}} \quad (2.98)$$

with  $\mathbf{f}, \mathbf{F}^{\text{BE}} \in \mathbb{R}^{N_{\text{BE}}}$ .

Now let us define a compression matrix  $C_e^{\text{BE}} \in \mathbb{R}^{N_{\text{BE}}/2 \times N_{\text{BE}}}$  that removes all odd rows, i.e. a unit matrix where the row index only runs over even equations such that – ignoring subband and energy indices for a moment – we can write down the compression matrix  $C_e^{\text{BE}}$  in terms of the harmonic index  $m$  in a subspace of a subband  $\nu$  at some position  $y_i$  or  $y_{i+}$  and energy  $H_j$  as

$$C_e^{\text{BE}} = \begin{pmatrix} m=0 & m=1 & m=2 & m=3 & \dots \\ 1 & 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & 1 & 0 & \dots & 0 \\ \vdots & & \ddots & & \vdots & \end{pmatrix} \begin{matrix} m=0 \\ m=2 \\ \vdots \end{matrix}. \quad (2.99)$$

Thus, we can reduce Eq. (2.98) of the Newton-Raphson approach to only contain even equations with even distribution functions as

$$\underbrace{C_e^{\text{BE}} S^{\text{BE}} A^{\text{BE}} (C_e^{\text{BE}})^t}_{=: A_{e/e}^{\text{BE}}} \underbrace{C_e^{\text{BE}} \delta \mathbf{f}}_{=: \delta \mathbf{f}_e} = - \underbrace{C_e^{\text{BE}} S^{\text{BE}} \mathbf{F}^{\text{BE}}}_{=: \mathbf{F}_e^{\text{BE}}}, \quad (2.100)$$

where  $A_{e/e}^{\text{BE}} \in \mathbb{R}^{N_{\text{BE}}/2 \times N_{\text{BE}}/2}$  only contains rows and columns of even harmonics,  $\delta \mathbf{f}_e \in \mathbb{R}^{N_{\text{BE}}/2}$  only contains distribution functions of even harmonics, and  $\mathbf{F}_e^{\text{BE}} \in \mathbb{R}^{N_{\text{BE}}/2}$  only contains the BE projected onto even harmonics.

Once Eq. (2.100) is solved using a sparse linear solver such as ILUPACK [68], we can calculate the odd distribution functions with Eq. (2.94). Thereafter, we update the distribution function as in Eq. (2.89) and return to Eq. (2.100) for the next iteration.

## 2.5 Iteration Schemes

Up until this section, we have shown how to discretize and solve the stationary PE, SE, and BE in Sects. 2.2, 2.3, and 2.4, respectively. But solving each of the equations only once cannot yield the correct solution since the equations are interdependent. The PE requires an electron density and yields a potential, the SE requires a potential and yields subband energies and wave functions, and the BE requires subband energies and wave functions and yields a distribution

function, which in turn can be used to compute the density in the PE. Only if all variables, i.e. the potential, the subband energies, the wave functions, and the distribution function, simultaneously solve the PE, SE, and BE, respectively, do we actually have the correct solution. In this case, we call the solution *self-consistent*.

Due to the cyclic dependence, it is not possible to achieve self-consistency by solving each equation only once, but we must solve them iteratively in order to approach the self-consistent solution. We will discuss two possible ways of iteration: the popular Gummel type iteration and a rapidly converging Newton-Raphson approach. The latter is an essential feature in achieving self-consistency for small signal and noise related computations discussed in the following chapters.

### 2.5.1 Gummel Type Iteration

The Gummel type iteration is a simple scheme that was originally introduced to solve the PE and the drift-diffusion model self-consistently [57]. Nonetheless, its basic premise is applicable also in our case [29].

In Gummel's method, we take an initial guess for the density in the non-linear PE, solve it and proceed to use the potential in the SE. Thereafter we take the subbands and wave functions and solve the BE in order to obtain the distribution function. The distribution functions are used to compute the density for the PE, which initiates the next iteration step. The Gummel type iteration is remarkably stable and – absent of inappropriate initial guesses for the distribution function – solidly converges to a self-consistent solution.

However, before we attempt to solve the BE, we usually do a pre-iteration including only the PE and SE as shown in Fig. 2.16. This is useful since the BE can diverge, therefore we want to be as close to the actual solution as possible when the BE is attempted to solve. Thus, we start by solving the non-linear PE of Eq. (2.1) with an initial quasi Fermi potential  $\varphi_n^{\text{init}}$  fulfilling the boundary conditions.<sup>5</sup> We obtain a potential  $V$  and insert it into the SE of Eq. (2.7), which yields the wave functions and the subband energies. For the next iteration, we need to compute the density as

$$n(x, y) = \frac{\mu_{\text{spin}} \mu_{\text{val}}}{Y_0^2} \sum_{\nu} \int_{\varepsilon^{\nu}(y)}^{\infty} dH Z^{\nu} f_{\text{eq}}^{\nu}(y, H) |\Psi^{\nu}(x, y)|^2,$$

which can be discretized as

$$n(x_k, y_i) = \frac{\mu_{\text{spin}} \mu_{\text{val}}}{Y_0^2} \sum_{\nu} \sum_j \Delta H^{\nu}(y_i, H_j) Z^{\nu} (f_{\text{eq}}^{2\text{D}})^{\nu}(x_k, y_i, H_j) |\Psi^{\nu}(x_k, y_i)|^2 \quad (2.101)$$

---

<sup>5</sup>We obtain the initial quasi Fermi potential  $\varphi_n^{\text{init}}$  from a drift-diffusion simulation by the GALENE III device simulation suite [100, 101] which also exports our grid and material data.

with the equilibrium distribution function

$$(f_{\text{eq}}^{2\text{D}})^\nu(x, y, H) = \frac{1}{\exp\left(\frac{H + q\varphi_n^{\text{init}}(x, y)}{k_B T} - \log\left(\frac{n_i}{N_{3\text{D}}^{\text{eff}}}\right)\right) + 1},$$

where the Fermi energy is given as in Eq. (2.46) except that we insert our initial quasi Fermi potential. With the density of Eq. (2.101), we can use the non-linear transformation of Eq. (2.3) in order to compute the quasi Fermi potential used for the PE in the next iteration as

$$\varphi_n(x_k, y_i) = V(x_k, y_i) - V_T \log\left(\frac{n(x_k, y_i)}{n_i}\right). \quad (2.102)$$

Note that the equilibrium distribution function strictly uses the quasi Fermi potential  $\varphi_n^{\text{init}}$  from the initialization. Using the quasi Fermi potential  $\varphi_n$  which is updated every iteration, would result in a situation where both  $\varphi_n$  and the potential  $V$  through the subband energies and wave functions influence the density which results in an unstable iteration. Therefore, we must measure the equilibrium distribution function against a baseline quasi Fermi potential  $\varphi_n^{\text{init}}$ . The closer  $\varphi_n^{\text{init}}$  is to the actual solution, the better the potential estimate of the PE and SE iteration.

Once the pre-Gummel iteration of Fig. 2.16 has converged, we will have a good estimate as to how the electron density – and therefore the potential – is modified in the vicinity of the gate oxides. The next step is to start the Gummel type iteration where the BE is included in the iteration loop as shown in Fig. 2.17. Once again, we start out by using the quasi Fermi potential to solve the non-linear PE of Eq. (2.1), then we use the potential to solve the SE of Eq. (2.7). With the subband energies  $\varepsilon^\nu$  and wave functions  $\Psi^\nu$ , we can set up the BE of Eq. (2.12). The BE can be solved using the Newton-Raphson approach detailed in Sect. 2.4.6 with an initial guess for the distribution function. Once again, we use the equilibrium distribution function but shifted by the applied biases at the contacts and linearly interpolated in between

$$(f^{\text{init}})_m^\nu(y, H) = \frac{\delta_{m,0}}{Y_0} \frac{1}{\exp\left(\frac{H + q\varphi_{\text{bias}}(y)}{k_B T} - \log\left(\frac{n_i}{N_{3\text{D}}^{\text{eff}}}\right)\right) + 1},$$

with

$$\varphi_{\text{bias}}(y) := \frac{1}{y_D - y_S} \left( (y_D - y)V_{\text{appl}}^S + (y - y_S)V_{\text{appl}}^D \right).$$

Note that this is not the optimal way to start the iteration but it has worked well for our case.

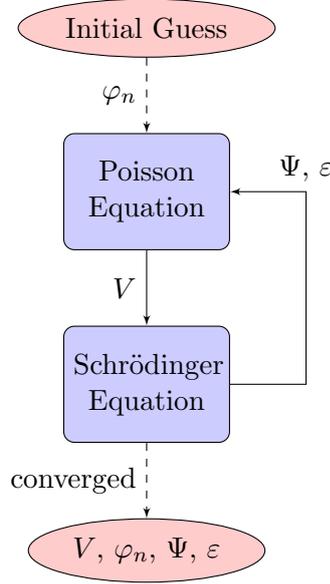


Figure 2.16: Pre-Gummel iteration loop consisting of PE and SE to approach the correct solution for the subband energies  $\varepsilon^\nu$  and wave functions  $\Psi^\nu$  before the solution of the BE is attempted.

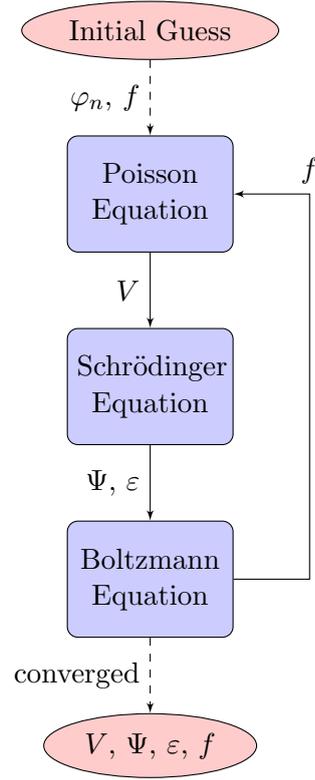


Figure 2.17: Gummel type iteration procedure for the PE, SE, and BE. We use an initial guess for the quasi Fermi potential  $\varphi_n$  for the PE. The initial guess for the distribution function is needed since the BE is non-linear and it is solved using the Newton-Raphson approach.

When the BE is solved, we can use the resulting distribution function to compute the density given by

$$\begin{aligned} n(x, y) &= \mu_{\text{spin}} \mu_{\text{val}} \sum_{\nu} \int \frac{d^2k}{(2\pi)^2} f^{\nu}(y, \mathbf{k}) |\Psi^{\nu}(x, y)|^2 \\ &= \mu_{\text{spin}} \mu_{\text{val}} \sum_{\nu} \frac{1}{Y_0} \int_{\varepsilon^{\nu}(y)}^{\infty} dH Z^{\nu} f_0^{\nu}(y, H) |\Psi^{\nu}(x, y)|^2, \end{aligned}$$

which can be discretized as

$$n(x_k, y_i) = \mu_{\text{spin}} \mu_{\text{val}} \sum_{\nu} \frac{1}{Y_0} \sum_j \Delta H^{\nu}(y_i, H_j) Z^{\nu} f_0^{\nu}(y_i, H_j) |\Psi^{\nu}(x_k, y_i)|^2. \quad (2.103)$$

Inserting this density into Eq. (2.102), we can determine the quasi Fermi potential  $\varphi_n$  for the PE in the next iteration of the Gummel loop and repeat the whole procedure until the potential, subband energies, wave functions, and distribution functions have converged towards the self-consistent solution.

Note that far away from the solution, during the iteration of the Newton-Raphson approach for the BE, unphysical distribution functions may occur. It is important to chop these values, i.e. ensure that the distribution function always satisfies  $0 \leq f^\nu(y, \mathbf{k}) \leq 1$ , or otherwise the Newton-Raphson approach is prone to diverge.

### 2.5.2 Newton-Raphson Approach

Far from equilibrium, the Gummel type iteration has the disadvantage that it takes an excessive amount of iterations to converge. It is also inadequate to determine the response of the system of equations to perturbations or fluctuations since the interdependencies of the potential, the wave functions and subband energies, and the distribution functions are not explicit. To address our needs for the eventual formulation of the small signal and noise problems, we are going to introduce a Newton-Raphson approach including the PE, SE, and BE, which also provides a rapidly converging solver for the stationary problem.

The idea behind the application of the Newton-Raphson approach is simple. We treat the PE together with the BE as one set of equations which needs to be solved simultaneously, i.e.

$$\begin{aligned} \sum_{\beta} \frac{\partial F_{\alpha}^{\text{BE}}}{\partial f_{\beta}} \delta f_{\beta} + \sum_b \frac{\partial F_{\alpha}^{\text{BE}}}{\partial V_b} \delta V_b &= -F_{\alpha}^{\text{BE}}, \\ \sum_{\beta} \frac{\partial F_a^{\text{PE}}}{\partial f_{\beta}} \delta f_{\beta} + \sum_b \frac{\partial F_a^{\text{PE}}}{\partial V_b} \delta V_b &= -F_a^{\text{PE}}, \end{aligned} \quad (2.104)$$

where  $\alpha$  and  $\beta$  are aggregate indices running over all coordinates of the BE as in Eq. (2.86) and  $a$  and  $b$  run over all coordinates of the PE as

$$a, b \in \{(x_k, y_j) \mid k \in \{1, \dots, N_x\}, j \in \{1, \dots, N_y\}\}. \quad (2.105)$$

Note that the PE depends directly on the distribution  $f$  through the density. However, until now we expressed the density through the non-linear transformation of Eq. (2.3) via the quasi Fermi potential. But in the Newton-Raphson approach of Eq. (2.104), we need an explicit dependency on the distribution function and therefore we use Eq. (2.103) directly to compute the density. For this to work, we already need to be close enough to the solution that stabilization via the exponential form of the density is not necessary.

The SE cannot be included in Eq. (2.104) like the PE and BE since it is an eigenvalue equation. Moreover, the BE does not explicitly depend on the

potential but only on the wave functions and subband energies resulting from the SE. Thus, in order to set up Eq. (2.104), we need to understand how the wave functions and subband energies change if the potential changes. But this is given by the first order time-independent perturbation theory of Eqs. (2.10) and (2.11). Note that since the Newton-Raphson approach given by Eq. (2.104) is simply the Taylor expansion up to first order, the first order perturbation theory is sufficient to capture the influence of the potential up to this order completely.

From the variations of Eqs. (2.10) and (2.11) it follows that the derivative w.r.t. the potential can be reformulated in terms of subband energies and wave functions in order to yield the system of equations

$$\begin{aligned} \sum_{\beta} \frac{\partial F_{\alpha}^{\text{BE}}}{\partial f_{\beta}} \delta f_{\beta} + \sum_b \left( \sum_k \frac{\partial F_{\alpha}^{\text{BE}}}{\partial \varepsilon_k} \frac{\partial \varepsilon_k}{\partial V_b} + \sum_{\ell} \frac{\partial F_{\alpha}^{\text{BE}}}{\partial \Psi_{\ell}} \frac{\partial \Psi_{\ell}}{\partial V_b} \right) \delta V_b &= -F_{\alpha}^{\text{BE}}, \\ \sum_{\beta} \frac{\partial F_a^{\text{PE}}}{\partial f_{\beta}} \delta f_{\beta} + \sum_b \frac{\partial F_a^{\text{PE}}}{\partial V_b} \delta V_b &= -F_a^{\text{PE}}, \end{aligned} \quad (2.106)$$

where  $k$  runs over all indices of  $\varepsilon$ ,  $\ell$  runs over all indices of  $\Psi$  and

$$\sum_b \frac{\partial \varepsilon_k}{\partial V_b} \delta V_b = \delta \varepsilon_k, \quad \sum_b \frac{\partial \Psi_{\ell}}{\partial V_b} \delta V_b = \delta \Psi_{\ell},$$

which means the derivatives of  $\varepsilon$  and  $\Psi$  w.r.t.  $V$  can be simply read off of Eqs. (2.10) and (2.11). Note that the setting up the system in Eq. (2.106) is quite cumbersome in practice since we find that each term of the BE is dependent on the subband energy through the boxes in  $H$ -space.

In order to express the elimination of the odd equations reasonably and for later convenience, we are going to introduce some additional notation at this point. Let us denote the linear system of Eq. (2.104) as

$$A^{\text{DC}} \delta \mathbf{x} = -\mathbf{F}, \quad (2.107)$$

where

$$\delta \mathbf{x} = (\delta \mathbf{f} \quad \delta \mathbf{V})^t, \quad \mathbf{F} = (\mathbf{F}^{\text{BE}} \quad \mathbf{F}^{\text{PE}})^t \quad (2.108)$$

are the vectors containing all discretized distribution functions and the potential as well as all discretized equations, respectively. Note that  $\mathbf{F}$  should already contain the right hand sides of the boundary conditions. Thus,  $A^{\text{DC}}$  is the matrix defined in such a way that Eq. (2.107) is equal to Eq. (2.104) except for where the boundary conditions of Eq. (2.4) are applied.

Assume that we already applied any reduction concerning the degeneracy of the BE as discussed in Sect. 2.4.6. We also want to eliminate all the odd

equations from the BE as was also discussed in Sect. 2.4.6. Since, we consider a system of equations including the PE, we need to extend the transformation matrix  $S^{\text{BE}}$ , defined by Eqs. (2.96) and (2.97), as

$$S := \begin{pmatrix} S^{\text{BE}} & 0 \\ 0 & \mathbb{I}_{\text{PE}} \end{pmatrix} \in \mathbb{R}^{N \times N}, \quad (2.109)$$

where  $\mathbb{I}_{\text{PE}} \in \mathbb{R}^{N_{\text{PE}} \times N_{\text{PE}}}$  is the unit matrix and  $N = N_{\text{BE}} + N_{\text{PE}}$ . Likewise, the compression matrix of Eq. (2.99) is given by

$$C_e := \begin{pmatrix} C_e^{\text{BE}} & 0 \\ 0 & \mathbb{I}_{\text{PE}} \end{pmatrix} \in \mathbb{R}^{n \times N}, \quad (2.110)$$

where  $n = N_{\text{BE}}/2 + N$ . Thus, analogous to Eq. (2.100), we solve the system of equations given by

$$\underbrace{C_e S A^{\text{DC}} (C_e)^t}_{=: A_{e/e}^{\text{DC}}} \underbrace{C_e \delta \mathbf{x}}_{=: \delta \mathbf{x}_e} = - \underbrace{C_e S \mathbf{F}}_{=: \mathbf{F}_e}, \quad (2.111)$$

where  $A^{\text{DC}} \in \mathbb{R}^{n \times n}$  is the system of equations consisting of only the even rows and columns of the BE as well as the PE,  $\delta \mathbf{x}_e$  contains only the even distribution functions as well as the potential, and  $\mathbf{F}_e$  contains only the even equations of the BE as well as the PE.

Once again, we use the sparse linear solver ILUPACK [68] in order to solve Eq. (2.111). Thereafter, we compute the odd distribution functions from the odd rows of the BE similar to Eq. (2.94), only now, we need to sum over the potential as well, i.e.

$$f_\alpha \Big|_{\alpha \text{ odd}} = \frac{1}{A_{\alpha,\alpha}^{\text{BE}}} \left( -F_\alpha^{\text{BE}} - \sum_{\substack{\beta=1 \\ \beta \neq \alpha}}^{N_{\text{BE}}} A_{\alpha,\beta}^{\text{BE}} f_\beta - \sum_{b=1}^{N_{\text{PE}}} A_{\alpha, N_{\text{BE}}+b}^{\text{BE}} V_b \right). \quad (2.112)$$

Then we update our variables as

$$f \longrightarrow f + \delta f, \quad V \longrightarrow V + \delta V$$

and solve the next iteration of Eq. (2.111) until convergence is achieved.

Figure 2.18 shows the flowchart of the algorithm for the Newton-Raphson approach of PE, SE, and BE. We take an initial value for the potential and the distribution function, preferably from a previous Gummel type iteration, and we use the potential to solve the SE. Thereafter we take the wave functions and subband energies, as well as the distribution function from the Gummel

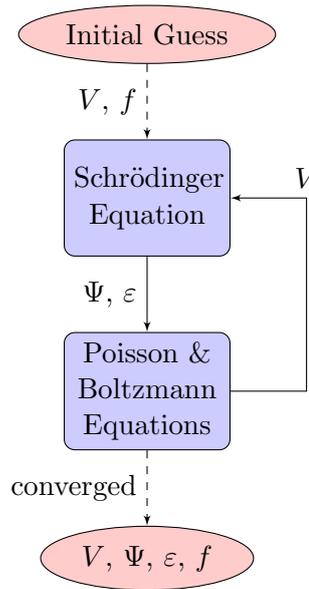


Figure 2.18: Newton-Raphson approach for solving the PE, SE, and BE. We take an initial guess for the potential  $V$  and the distribution function  $f$  and update the subband energies  $\varepsilon$  and wave functions  $\Psi$ . Thereafter we solve the combined system of PE and BE and return to the SE, unless our solution has sufficiently converged.

type iteration, to setup the system of equations of Eq. (2.111) and solve for the potential and distribution functions.

The reason we need to solve the SE after each Newton step is that, although the effects of a change in the potential on the wave functions and subband energies are correctly included in the Newton-Raphson approach, they are only valid up to linear order. Therefore, it is necessary to update the unperturbed wave functions and subband energies with the exact values at each iteration.

Since all three equations are included in the Newton-Raphson approach, it is expected to converge with the usual rapid quadratic rate in the vicinity of the solution which will be shown in Sect. 5.2.1.

## Chapter 3

# Small Signal Analysis

One of the main achievements of this work is to understand how small signal quantities can be computed in the framework of Poisson equation (PE), Schrödinger equation (SE), and Boltzmann equation (BE) with an  $H$ -transformed energy space. The  $H$ -transformation is a boon for stationary computations but it makes it difficult to derive discretized equations suitable for the small signal analysis. Not only do we obtain an additional term due to the time-dependence of the  $H$ -transformation but a naive discretization also breaks important continuum symmetries and conservation laws.

Another difficulty is the dimensional splitting where the PE is defined in 2D, the SE is defined in confinement direction, and the BE is defined in transport direction. Usually the Ramo-Shockley theorem is used to compute terminal currents in devices [102, 103, 104] but it is not obvious how to apply the theorem in the case of our split dimensions. Therefore we derive a form of the Ramo-Shockley theorem that is suitable for our case.

Lastly, we are going to introduce the adjoint method of solving the system of equations for the admittance parameters. This is going to enable us to solve both the noise and the small signal problem in one step.

### 3.1 Linearization

The idea of the small signal analysis is that we have a stationary bias applied to our contacts and on top of that we introduce a small sinusoidal perturbation on one of the contacts. Let the stationary applied bias at some contact is given by  $V_{\text{appl}}^C$ , then we can express the total applied bias as

$$\tilde{V}_{\text{appl}}^C(t) = V_{\text{appl}}^C + \text{Re}(\underline{V}_{\text{appl}}^C e^{i\omega t}), \quad (3.1)$$

where  $\underline{V}_{\text{appl}}^C$  is the phasor of the small signal bias and  $\omega$  is the angular frequency of the applied bias.

We expect the system to respond to the additional small signal bias. The response is – of course – governed by the PE, SE, and BE. For the PE and SE we assume that any small signal bias is slow enough such that the time-dependence can be assumed as a quasistationary change in the potential, wave functions, and subband energies. Thus, the PE of Eq. (2.1) and the SE of Eq. (2.7) are still valid. In contrast, we need to include the time-derivative in the BE, yielding

$$0 = \left[ \tilde{F}^{\text{BE}} \right]^\nu (y, \mathbf{k}, t) := \tilde{T}^\nu(y, \mathbf{k}, t) + \tilde{L}^\nu(y, \mathbf{k}, t) - \tilde{S}^\nu(y, \mathbf{k}, t) - \tilde{\Gamma}^\nu(y, \mathbf{k}, t) \quad (3.2)$$

where we define the time-derivative as

$$\tilde{T}^\nu(y, \mathbf{k}, t) := \frac{\partial}{\partial t} \tilde{f}^\nu(y, \mathbf{k}, t) \quad (3.3)$$

for future reference. Moreover, the time-dependent free streaming term  $\tilde{L}$ , scattering term  $\tilde{S}$ , and boundary term  $\tilde{\Gamma}$  can be obtained by replacing the stationary distribution function with a time-dependent one as

$$\begin{aligned} L^\nu(y, \mathbf{k}) &\xrightarrow{f^\nu(y, \mathbf{k}) \rightarrow \tilde{f}^\nu(y, \mathbf{k}, t)} \tilde{L}^\nu(y, \mathbf{k}, t), \\ S^\nu(y, \mathbf{k}) &\xrightarrow{f^\nu(y, \mathbf{k}) \rightarrow \tilde{f}^\nu(y, \mathbf{k}, t)} \tilde{S}^\nu(y, \mathbf{k}, t), \\ \Gamma^\nu(y, \mathbf{k}) &\xrightarrow{f^\nu(y, \mathbf{k}) \rightarrow \tilde{f}^\nu(y, \mathbf{k}, t)} \tilde{\Gamma}^\nu(y, \mathbf{k}, t), \end{aligned}$$

with  $L$ ,  $S$ , and  $\Gamma$  defined by Eqs. (2.13), (2.15), and (2.38), respectively.

Since we assume that the small signal perturbation is given by a harmonic time-dependence as in Eq. (3.1), we can express the distribution function and potential in a similar way as

$$\tilde{f}^\nu(y, \mathbf{k}, t) = f^\nu(y, \mathbf{k}) + \text{Re} \left( \underline{f}^\nu(y, \mathbf{k}) e^{i\omega t} \right), \quad (3.4)$$

$$\tilde{V}(\mathbf{r}, t) = V(\mathbf{r}) + \text{Re} \left( \underline{V}(\mathbf{r}) e^{i\omega t} \right), \quad (3.5)$$

where  $\underline{f}$  and  $\underline{V}$  are small. Naturally, the wave functions and subband energies also respond to the changed potential as

$$\tilde{\varepsilon}^\nu(y, t) = \varepsilon^\nu(y) + \text{Re} \left( \underline{\varepsilon}^\nu(y) e^{i\omega t} \right), \quad (3.6)$$

$$\tilde{\Psi}^\nu(x, y, t) = \Psi^\nu(x, y) + \text{Re} \left( \underline{\Psi}^\nu(x, y) e^{i\omega t} \right), \quad (3.7)$$

where the phasors can be directly computed since  $\underline{V}$  is a small perturbation. Completely analogous to Eqs. (2.10) and (2.11) we find

$$\underline{\varepsilon}^\nu(y) = -q \int dx |\Psi^\nu(x, y)|^2 \underline{V}(x, y) =: \int dx \Delta_{\underline{\varepsilon}}^\nu(x, y) \underline{V}(x, y), \quad (3.8)$$

$$\begin{aligned}
\underline{\Psi}^\nu(x, y) &= -q \sum_{\substack{\nu' \neq \nu \\ \nu' = v}} \frac{\int dx' \underline{\Psi}^{\nu'}(x', y) \underline{V}(x', y_0) \underline{\Psi}^\nu(x', y)}{\varepsilon^\nu(y) - \varepsilon^{\nu'}(y)} \underline{\Psi}^{\nu'}(x, y) \\
&=: \int dx' \Delta_{\underline{\Psi}}^\nu(x, x', y) \underline{V}(x', y),
\end{aligned} \tag{3.9}$$

where we also defined the abbreviations  $\Delta_\varepsilon$  and  $\Delta_{\underline{\Psi}}$  for later convenience. Bear in mind that the sum in the wave function only runs over all subbands of the *same* valley.

Since we are only interested in the small signal response up to linear order, it is not necessary to solve the SE since perturbation theory already gives us all the information we need about the quasistationary wave functions and subband energies.

Assuming the stationary solution of the BE is known, we can insert the linear expansions of Eqs. (3.4) and (3.5) into the full BE of Eq. (3.2) and drop all terms of higher than linear order, yielding

$$0 = \left[ \tilde{F}^{\text{BE}} \right]^\nu(y, \mathbf{k}, t) \approx \underbrace{\left[ F^{\text{BE}} \right]^\nu(y, \mathbf{k})}_{=0} + \text{Re} \left( \left[ \underline{F}^{\text{BE}} \right]^\nu(y, \mathbf{k}) e^{i\omega t} \right), \tag{3.10}$$

where  $\underline{F}^{\text{BE}}$  only contains terms linear in  $\underline{f}$  or  $\underline{V}$ . It follows trivially that

$$0 = \left[ \underline{F}^{\text{BE}} \right]^\nu(y, \mathbf{k}) \tag{3.11}$$

needs to be fulfilled which is the equation we will attempt to solve.

The whole linearization of  $\tilde{F}^{\text{BE}}$  is equivalent to a Taylor expansion up to first order around the stationary solution which is the same as the setup of the Newton-Raphson approach of Eq. (2.106). Thus, we can reuse all parts of the BE except for the time derivative. Then, after the Herring-Vogt transformation, the projection onto equienergy lines and Fourier harmonics, the  $H$ -transformation, and the discretization, Eq. (3.11) can be expressed as

$$\begin{aligned}
\underline{T}_\alpha + \sum_\beta \frac{\partial F_\alpha^{\text{BE}}}{\partial f_\beta} \underline{f}_\beta + \sum_b \left( \sum_k \frac{\partial F_\alpha^{\text{BE}}}{\partial \varepsilon_k} \frac{\partial \varepsilon_k}{\partial V_b} + \sum_\ell \frac{\partial F_\alpha^{\text{BE}}}{\partial \Psi_\ell} \frac{\partial \Psi_\ell}{\partial V_b} \right) \underline{V}_b \\
= \sum_{C=S,D} \frac{\partial \Gamma_\alpha}{\partial V_{\text{appl}}^C} V_{\text{appl}}^C,
\end{aligned} \tag{3.12}$$

where  $\underline{T}_\alpha$  is the transformed and discretized time-derivative discussed in Sect. 3.2,  $F_\alpha^{\text{BE}}$  is defined by Eq. (2.87), the term proportional to  $V_{\text{appl}}^C$  is the small signal response of the generation and recombination (GR) term to a perturbation in the applied bias which will be derived in Sect. 3.3,  $\alpha$  and  $\beta$  are aggregate indices

running over all indices of the BE as in Eq. (2.86), and  $a$  and  $b$  are aggregate indices running over all indices of the PE as in Eq. (2.105). Furthermore, the subband energy index is given by  $k = (\nu, y_i)$  and the wave function index is given by  $\ell = (\nu, x_k, y_i)$ . Note that the derivatives of the subband energy and wave functions are simply given by the response to perturbations as

$$\varepsilon_k = \sum_b \frac{\partial \varepsilon_k}{\partial V_b} V_b, \quad \underline{\Psi}_\ell = \sum_b \frac{\partial \Psi_\ell}{\partial V_b} V_b.$$

Completely analogously, we can linearize the stationary PE as

$$0 = \tilde{F}^{\text{PE}}(\mathbf{r}) \approx \underbrace{F^{\text{PE}}(\mathbf{r})}_{=0} + \text{Re}(\underline{F}^{\text{PE}}(\mathbf{r}) e^{i\omega t}) \quad (3.13)$$

and therefore it follows that we need to solve

$$0 = \underline{F}^{\text{PE}}(\mathbf{r}), \quad (3.14)$$

where  $\underline{F}^{\text{PE}}$  is linear in  $\underline{f}$  and  $\underline{V}$ . As in the case of the BE, the linearization is equivalent to a Taylor expansion around the stationary state and thus we can express Eq. (3.14) with the help of the Jacobian of the Newton-Raphson approach as

$$\sum_\beta \frac{\partial F_a^{\text{PE}}}{\partial f_\beta} f_\beta + \sum_b \frac{\partial F_a^{\text{PE}}}{\partial V_b} V_b = 0. \quad (3.15)$$

For later reference, we want to record that the linearized small signal PE takes the shape

$$\underline{F}^{\text{PE}}(\mathbf{r}) := \nabla_{\mathbf{r}} \cdot (\kappa(\mathbf{r}) \nabla_{\mathbf{r}} \underline{V}(\mathbf{r})) - q n_{3\text{D}}(\mathbf{r}) = 0. \quad (3.16)$$

The 3D small signal electron density is given by

$$\underline{n}_{3\text{D}}(\mathbf{r}) = \sum_\nu \left( \underline{n}^\nu(y) |\Psi^\nu(\mathbf{r})|^2 + 2n^\nu(y) \Psi^\nu(\mathbf{r}) \underline{\Psi}^\nu(\mathbf{r}) \right), \quad (3.17)$$

where the wave functions were assumed to be real and the small signal sheet density per subband reads

$$\underline{n}^\nu(y) = \left( \frac{Z^\nu}{Y_0} \int_{\varepsilon^\nu(y)}^\infty dH f_0^\nu(y, H) \right) - \frac{Z^\nu}{Y_0} f_0^\nu(y, \varepsilon^\nu(y)) \underline{\varepsilon}^\nu(y). \quad (3.18)$$

Equations (3.12) and (3.15) form a linear system of equations that can be directly solved to obtain the small signal perturbations  $\underline{f}$  and  $\underline{V}$ . The treatment of the discretized time-derivative  $T_\alpha$  is postponed to Sect. 3.2. Furthermore, boundary conditions are discussed in Sect. 3.3. Setting up the equations for small signal analysis naively as in Eqs. (3.12) and (3.15) will lead to inconsistencies with symmetries and conservation laws. The resolution of these issues is discussed in Sect. 3.5.

### 3.2 Time-Derivative of the Boltzmann Equation

In order to determine the time-derivative of the BE, we need to apply the same transformations to  $\tilde{T}^\nu(y, \mathbf{k}, t)$  of Eq. (3.3) as to the rest of the BE. These include the Herring-Vogt transformation, the projection onto equienergy lines and Fourier harmonics as well as the  $H$ -transformation (see Sect. 2.4.4) and the discretization via the box-integration method (see Sect. 2.4.5).

The Herring-Vogt transformation leaves  $\tilde{T}$  invariant. After the projection onto equienergy lines and Fourier harmonics, we find

$$\tilde{T}_m^\nu(y, E, t) := \int \frac{d^2k}{(2\pi)^2} \tilde{T}^\nu(y, \mathbf{k}, t) \delta(E - E^\nu(\mathbf{k})) Y_m(\phi) = Z^v \frac{d}{dt} \tilde{f}_m^\nu(y, E, t). \quad (3.19)$$

Regarding the  $H$ -transformation, we have to be careful since Eq. (2.65) states that  $H$  depends on the subband energy  $\varepsilon$  which in turn is time-dependent as in Eq. (3.6). That means the order of the time-derivative in Eq. (3.19) and the transformation to  $H$ -space matters. Therefore we obtain

$$\begin{aligned} \tilde{T}_m^\nu(y, E, t) &\longrightarrow \tilde{T}_m^\nu(y, H, t) = Z^v \frac{d}{dt} \tilde{f}_m^\nu(y, H^\nu(y, E, t), t) \Big|_{H^\nu(y, E, t) = H} \\ &= Z^v \frac{d}{dt} \left[ \underline{f}_m^\nu(y, H^\nu(y, E, t)) \right. \\ &\quad \left. + \operatorname{Re} \left( \underline{f}_m^\nu(y, H^\nu(y, E, t)) e^{i\omega t} \right) \right] \Big|_{H^\nu(y, E, t) = H} \\ &= Z^v \frac{\partial \underline{f}_m^\nu(y, H)}{\partial H} \frac{dH^\nu(y, E, t)}{dt} \\ &\quad + Z^v \operatorname{Re} \left( i\omega \underline{f}_m^\nu(y, H) e^{i\omega t} \right) \\ &\quad + Z^v \operatorname{Re} \left( \frac{\partial \underline{f}_m^\nu(y, H)}{\partial H} \frac{dH^\nu(y, E, t)}{dt} e^{i\omega t} \right) \\ &= \operatorname{Re} \left[ \underbrace{i\omega Z^v \left( \frac{\partial \underline{f}_m^\nu(y, H)}{\partial H} \underline{\varepsilon}^\nu(y) + \underline{f}_m^\nu(y, H) \right)}_{=: \underline{T}_m^\nu(y, H)} e^{i\omega t} \right] \\ &\quad + \mathcal{O} \left( \underline{f}_m^\nu(y, H) \underline{\varepsilon}^\nu(y) \right), \end{aligned} \quad (3.20)$$

where we only kept terms up to linear order and used that

$$\frac{dH^\nu(y, E, t)}{dt} = \operatorname{Re} \left( i\omega \underline{\varepsilon}^\nu(y) e^{i\omega t} \right),$$

which follows immediately from Eqs. (2.65) and (3.6). In the last line, we also defined the phasor of the small signal perturbation of the time derivative as  $\underline{T}$ .

Now we can box-integrate  $\underline{T}$  around a direct grid point to find

$$\begin{aligned}
\bar{\underline{T}}_m^\nu(y_i, H_j) &:= \int_{H_{j-}}^{H_{j+}} dH \int_{y_{i-}}^{y_{i+}} dy \underline{T}_m^\nu(y, H) \Big|_{m \text{ even}} \\
&= i\omega Z^v \Delta y_i \left( \int_{H_{j-}}^{H_{j+}} dH \frac{\partial f_m^\nu(y_i, H)}{\partial H} \underline{\varepsilon}^\nu(y_i) \right. \\
&\quad \left. + \Delta H^\nu(y_i, H_j) \underline{f}_m^\nu(y_i, H_j) \right) \Big|_{m \text{ even}} \\
&= i\omega Z^v \Delta y_i \left( (f_m^\nu(y_i, H_{j+}) - f_m^\nu(y_i, H_{j-})) \underline{\varepsilon}^\nu(y_i) \right. \\
&\quad \left. + \Delta H^\nu(y_i, H_j) \underline{f}_m^\nu(y_i, H_j) \right) \Big|_{m \text{ even}} \tag{3.21}
\end{aligned}$$

and likewise for the adjoint grid

$$\begin{aligned}
\bar{\underline{T}}_m^\nu(y_{i+}, H_j) &:= \int_{H_{j-}}^{H_{j+}} dH \int_{y_i}^{y_{i+1}} dy \underline{T}_m^\nu(y, H) \Big|_{m \text{ odd}} \\
&= i\omega Z^v \Delta y_{i+} \left( (f_m^\nu(y_{i+}, H_{j+}) - f_m^\nu(y_{i+}, H_{j-})) \underline{\varepsilon}^\nu(y_{i+}) \right. \\
&\quad \left. + \Delta H^\nu(y_{i+}, H_j) \underline{f}_m^\nu(y_{i+}, H_j) \right) \Big|_{m \text{ odd}}. \tag{3.22}
\end{aligned}$$

Note that the integral over the  $H$ -box cannot be trivially resolved since we do not have an adjoint  $H$ -grid. Instead, we need resort to an averaging scheme that is consistent with an integration in the continuum. To this end, let us recall that the BE of Eq. (2.13) integrated over  $\mathbf{k}$ -space and summed over all subbands, must yield the continuity equation:

$$\frac{\partial}{\partial y} \tilde{j}(y, t) + \frac{\partial}{\partial t} \tilde{n}(y, t) = 0. \tag{3.23}$$

Since our scattering terms described in Sect. 2.4.2 are charge conserving, only the term  $\tilde{T}$  can give rise to the rate in charge generation or recombination and thus we have<sup>1</sup>

$$\frac{\partial \tilde{n}(y, t)}{\partial t} = \sum_\nu \int \frac{d^2 k}{(2\pi)^2} \tilde{T}^\nu(y, \mathbf{k}, t).$$

---

<sup>1</sup>When considering degeneracies, we will also find spin and valley multiplicities in the expression for the density.

Using

$$\tilde{n}(y, t) = n(y) + \text{Re}(\underline{n}(y)e^{i\omega t}) \quad (3.24)$$

and the usual transformations and the box-integration, it follows

$$i\omega \underline{n}(y_i) \Delta y_i = \sum_{\nu} \frac{1}{Y_0} \sum_j \bar{T}_0^{\nu}(y_i, H_j). \quad (3.25)$$

On the other hand, the straightforward linearization of the electron density yields

$$\begin{aligned} \underline{n}(y_i) = \sum_{\nu} \frac{1}{Y_0} \sum_j Z^{\nu} \left( f_0^{\nu}(y_i, H_j) \frac{\partial \Delta H^{\nu}(y_i, H_j)}{\partial \varepsilon^{\nu}(y_i)} \underline{\varepsilon}^{\nu}(y_i) \right. \\ \left. + \underline{f}_0^{\nu}(y_i, H_j) \Delta H^{\nu}(y_i, H_j) \right). \end{aligned} \quad (3.26)$$

Comparing Eq. (3.25) and Eq. (3.26) we find that the following relation must hold if we want the time-derivative in the BE to be consistent with the definition of the density:

$$\begin{aligned} \sum_{j=j_{\min}}^{N_H} (f_0^{\nu}(y_i, H_{j+}) - f_0^{\nu}(y_i, H_{j-})) &\stackrel{!}{=} \sum_j f_0^{\nu}(y_i, H_j) \frac{\partial \Delta H^{\nu}(y_i, H_j)}{\partial \varepsilon^{\nu}(y_i)} \\ &= -f_0^{\nu}(y_i, H_{j_{\min}}), \end{aligned} \quad (3.27)$$

where it was used that the derivative of the  $H$ -box is only non-zero at the subband edge (cf. Eq. (2.75)) and  $j_{\min}$  is the index of the box surrounding the subband energy

$$H_{j_{\min}-} \leq \varepsilon^{\nu}(y_i) < H_{j_{\min}+}.$$

Note that Eq. (3.27) would be fulfilled in the continuum but in discretized  $H$ -space, we need to carefully define what  $f_0^{\nu}(y_i, H_{j\pm})$  is in terms of the distribution function on the direct  $H$ -grid points. Let us define it as

$$\begin{pmatrix} f_m^{\nu}(y, H_{j+}) \\ f_m^{\nu}(y, H_{j-}) \end{pmatrix} = \begin{cases} \begin{pmatrix} 0 \\ \frac{1}{2} [f_m^{\nu}(y, H_j) + f_m^{\nu}(y, H_{j-1})] \end{pmatrix}, & \text{if } H_{j+} > H_{N_H} > H_{j-}, \\ \begin{pmatrix} \frac{1}{2} [f_m^{\nu}(y, H_{j+1}) + f_m^{\nu}(y, H_j)] \\ f_m^{\nu}(y, H_j) \end{pmatrix}, & \text{if } H_{j+} > \varepsilon^{\nu}(y) \geq H_{j-}, \\ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, & \text{if } \varepsilon^{\nu}(y) \geq H_{j+}, \\ \begin{pmatrix} \frac{1}{2} [f_m^{\nu}(y, H_{j+1}) + f_m^{\nu}(y, H_j)] \\ \frac{1}{2} [f_m^{\nu}(y, H_j) + f_m^{\nu}(y, H_{j-1})] \end{pmatrix}, & \text{else,} \end{cases} \quad (3.28)$$

where  $y$  can be either a direct point, if  $m$  is even, or an adjoint point, if  $m$  is odd. Only if we choose the distribution function on the  $H$ -grid points  $H_{j\pm}$  as in Eq. (3.28), can we get a discretized time derivative as in Eqs. (3.21) and (3.22) that is consistent with a rate of change of the electron density.

In this work we will often refer to the transformed and discretized time-derivative of the BE with the aggregate index  $\alpha$  of Eq. (2.86) as

$$\underline{T}_\alpha := \bar{T}_m^\nu(y_i, H_j).$$

This form of the time-derivative has already made an appearance in the system of equations of the small signal BE of Eq. (3.12).

### 3.3 Boundary Conditions

Equations (3.12) and (3.15) together with the time-derivative shown in Eqs. (3.21) and (3.22) form a linear system of equations that can be solved directly once we add suitable boundary conditions. Recall that the BE has Neumann boundary conditions with a GR rate at the source and drain contacts as described in Sect. 2.4.3. Furthermore, the PE has Dirichlet boundary conditions on the gate contacts and Neumann boundary conditions everywhere else as described in Sect. (2.2.2).

If we apply a small signal bias as in Eq. (3.1) to the gate contacts described by the sets  $\partial D_{\text{TG}}, \partial D_{\text{BG}} \subset \mathbb{R}^2$ , the Dirichlet boundary conditions of the linearized PE are given by

$$\underline{V}(\mathbf{r}) \Big|_{(\mathbf{r}) \in \partial D_{\text{TG/BG}}} = \underline{V}_{\text{appl}}^{\text{TG/BG}}, \quad (3.29)$$

which is nothing else but the trivial linearization of Eq. (2.4).

Let us turn to the boundary conditions of the BE on the source and drain contacts, respectively. Recall that we defined them in Sect. 2.4.3 as GR rates at  $y = y_{\text{S/D}}$ . Thus, in order to understand how a small signal bias affects them, we need to linearize  $\Gamma$  of Eq. (2.83) around its stationary value. This includes the linearization w.r.t. the distribution function and the potential but it also includes a response from the equilibrium distribution of carriers within the contact. This is because when we apply a small signal bias to the source or drain, we effectively change the Fermi level. The effect of the small signal bias on the equilibrium distribution function of Eq. (2.70) can be expressed via a Taylor expansion up to linear order:

$$\tilde{f}_{\text{eq}}(y_{\text{S/D}}, H, t) = \frac{1}{\exp\left(\frac{H + q\tilde{V}_{\text{appl}}^{\text{S/D}}(t)}{k_B T} - \log\left(\frac{n_i}{N_{3\text{D}}^{\text{eff}}}\right)\right) + 1}$$

$$\begin{aligned}
&= f_{\text{eq}}(y, H) + f_{\text{eq}}(y, H)(f_{\text{eq}}(y, H) - 1) \frac{1}{V_T} \text{Re} \left( \underline{V}_{\text{appl}}^{\text{S/D}} e^{i\omega t} \right) \\
&\quad + \mathcal{O} \left( \left( \underline{V}_{\text{appl}}^{\text{S/D}} \right)^2 \right). \tag{3.30}
\end{aligned}$$

It follows immediately that the phasor of the small signal equilibrium distribution function is given by

$$\underline{f}_{\text{eq}}(y, H) = f_{\text{eq}}(y, H)(f_{\text{eq}}(y, H) - 1) \frac{V_{\text{appl}}(y)}{V_T}$$

and therefore the total box-integrated phasor of the small signal GR rate on the contacts reads

$$\underline{\Gamma}_\alpha = \sum_\beta \frac{\partial \Gamma_\alpha}{\partial f_\beta} \underline{f}_\beta + \sum_b \frac{\partial \Gamma_\alpha}{\partial V_b} V_b + \sum_{C=\text{S,D}} \frac{\partial \Gamma_\alpha}{\partial V_{\text{appl}}^C} V_{\text{appl}}^C, \tag{3.31}$$

where the derivatives w.r.t.  $\underline{f}$  and  $\underline{V}$  are already contained in Eq. (3.12) in the derivatives of  $F_\alpha^{\text{BE}}$ . The remaining term represents the term proportional to the applied bias  $V_{\text{appl}}^C$  of the source and drain contacts and is given by

$$\begin{aligned}
\sum_{C=\text{S,D}} \frac{\partial \Gamma_\alpha}{\partial V_{\text{appl}}^C} V_{\text{appl}}^C &= \sum_{C=\text{S,D}} \frac{\partial \bar{\Gamma}_m^\nu(y_i, H_j)}{\partial V_{\text{appl}}^C} V_{\text{appl}}^C \\
&= v_{\text{GR}} Z^v \frac{\delta_{m,0}}{Y_0} \underline{f}_{\text{eq}}(y_i, H_j) (\delta_{y_i, y_S} + \delta_{y_i, y_D}) \Delta H^\nu(y_i, H_j). \tag{3.32}
\end{aligned}$$

Note that this term is only proportional to the small signal phasor of the applied bias  $V_{\text{appl}}$  and therefore it is located on the r.h.s. of Eq. (3.12).

### 3.4 Ramo-Shockley Theorem

Usually we are not interested in internal quantities of the device but rather such quantities as terminal currents that can be directly compared to measurements. Later on, we will see that the adjoint method of the small signal analysis also relies on a projection operator onto the terminal currents (cf. Sect. 3.6.2). Moreover, in Sect. 4.6 we will formulate the power spectral density of fluctuations of the terminal currents as well as the origin of noise that is seen in the terminal currents.

For all these purposes, it becomes necessary to have a sound definition of the terminal currents. The terminal current at a contact  $C$  is defined as the sum of the charge current and displacement current flowing over the contact:

$$\tilde{I}_C(t) = - \int_{\partial D_C} d\mathbf{A} \cdot \left[ q \tilde{\mathbf{J}}(\mathbf{r}, t) + \frac{\partial}{\partial t} \left( \kappa(\mathbf{r}) \nabla \tilde{V}(\mathbf{r}, t) \right) \right], \tag{3.33}$$

where  $\partial D_C$  is the set of vectors contained in the contact  $C$ ,  $d\mathbf{A}$  is the normal vector on the contact pointing into the device,  $\tilde{\mathbf{J}}$  is the electron current density,  $\kappa$  is the dielectric constant, and  $\tilde{V}$  is the electric potential. The numerical issue with this definition is that it relies on an integration over the flux through a contact, which cannot be expressed reliably on a grid where fluxes are on adjoint grid points and contacts are on direct grid points. Therefore we aim to transform the integration over a contact into a volume integration over the whole device using the Ramo-Shockley theorem [102, 103, 104] which is numerically more robust. The following derivation is valid for general devices where the BE is used to compute transport phenomena and the SE is used to compute the confinement in the directions perpendicular to the transport.

To this end, let us disassemble the current density into its constituents. Since the BE describes a multitude of subbands which conduct a current, we can separate the electron current density into its subbands as

$$\tilde{\mathbf{J}}(\mathbf{r}, t) = \sum_{\nu} \tilde{\mathbf{J}}^{\nu}(\mathbf{r}, t).$$

Furthermore, we can split up the current density into a part in transport direction and a part orthogonal to the transport direction as

$$\tilde{\mathbf{J}}^{\nu}(\mathbf{r}, t) = \tilde{\mathbf{J}}_y^{\nu}(\mathbf{r}, t) + \tilde{\mathbf{J}}_{\perp}^{\nu}(\mathbf{r}, t) \quad (3.34)$$

with

$$\tilde{\mathbf{J}}_y(\mathbf{r}, t) := \tilde{j}^{\nu}(y, t) |\tilde{\Psi}^{\nu}(\mathbf{r}, t)|^2 \mathbf{e}_y, \quad (3.35)$$

where  $\tilde{j}$  is the 1D current density described by the time-dependent BE in transport direction (cf. Eq. (A.6)),  $\tilde{\Psi}$  is the time-dependent wave function of the confined electron gas, and  $\mathbf{e}_y$  is the unit vector in transport direction. Integrating the BE for a single subband over  $\mathbf{k}$ -space yields the 1D continuity equation:

$$\frac{\partial}{\partial y} \tilde{j}^{\nu}(y, t) + \frac{\partial}{\partial t} \tilde{n}^{\nu}(y, t) = \tilde{S}^{\nu}(y, t) + \tilde{\Gamma}^{\nu}(y, t), \quad (3.36)$$

where  $\tilde{S}^{\nu}(y, t)$  and  $\tilde{\Gamma}^{\nu}(y, t)$  are the scattering rate and the boundary GR rate integrated over  $\mathbf{k}$ -space. Note that even when the scattering rate is charge conserving in total, it need not be charge conserving *per subband*.

Now let us derive the expression for the terminal current by using the Ramo-Shockley theorem. The Ramo-Shockley test functions  $h_C$  are chosen to obey the Laplace equation

$$\nabla \cdot (\kappa(\mathbf{r}) \nabla h_C(\mathbf{r})) = 0$$

with Dirichlet boundary conditions on all contacts

$$h_C(\mathbf{r}) = \delta_{C,C'}, \quad \mathbf{r} \in \partial D_{C'} \quad (3.37)$$

and Neumann boundary conditions everywhere else. Here,  $D \subset \mathbb{R}^2$  denotes the set containing all the vectors in the device,  $\partial D$  denotes the surface of the device, and  $\partial D_{C'}$  is the set containing all vectors of the  $C'$ -th contact.

Due to the boundary conditions of Eq. (3.37), we can simply multiply the integrand of Eq. (3.33) by the test function and extend the integral over the whole surface, yielding

$$\tilde{I}_C(t) = - \int_{\partial D} d\mathbf{A} \cdot \left\{ h_C(x, y) \left[ q \sum_{\nu} \tilde{\mathbf{J}}^{\nu}(\mathbf{r}, t) + \frac{\partial}{\partial t} \left( \kappa(\mathbf{r}) \nabla \tilde{V}(\mathbf{r}, t) \right) \right] \right\}.$$

Using the Gaussian integral theorem, we can already transform a part of the surface integral into a volume integral

$$\begin{aligned} \tilde{I}_C(t) &= -q \sum_{\nu} \underbrace{\int_{\partial D} d\mathbf{A} \cdot \left( h_C(\mathbf{r}) \tilde{\mathbf{J}}^{\nu}(\mathbf{r}, t) \right)}_{=: \mathcal{I}_1} \\ &\quad - \underbrace{\int_D dV h_C(\mathbf{r}) \frac{\partial}{\partial t} \left( \nabla \cdot \left( \kappa(\mathbf{r}) \nabla \tilde{V}(\mathbf{r}, t) \right) \right)}_{=: \mathcal{I}_2} \\ &\quad - \underbrace{\int_D dV \nabla h_C(\mathbf{r}) \cdot \frac{\partial}{\partial t} \left( \kappa(\mathbf{r}) \nabla \tilde{V}(\mathbf{r}, t) \right)}_{=: \mathcal{I}_3}, \end{aligned} \quad (3.38)$$

where we defined the abbreviations  $\mathcal{I}_1$  through  $\mathcal{I}_3$  for convenience.

Due to the dimensional structure of our problem, where source and drain contacts are perpendicular to the transport direction and where the gate contacts are outside of the semiconductor region (cf. Fig. 2.1), the only contribution to the surface integral over the current density stems from the component in transport direction at the source and drain contacts:

$$\mathcal{I}_1 = \int_{\partial D} d\mathbf{A} \cdot \left( h_C(\mathbf{r}) \tilde{\mathbf{J}}_y^{\nu}(\mathbf{r}, t) \right) = \int_{\partial D} d\mathbf{A} \cdot \mathbf{e}_y \left( h_C(\mathbf{r}) \tilde{j}^{\nu}(y, t) |\tilde{\Psi}^{\nu}(\mathbf{r}, t)|^2 \right),$$

where we used Eq. (3.35) in the second step. As was shown in Sect. 2.4.3, the current density at the contacts is solely determined by the GR term at the contacts. Thus, using Eqs. (2.39) and (2.40), we find

$$\mathcal{I}_1 = - \int_D dV h_C(\mathbf{r}) \tilde{\Gamma}^{\nu}(y, t) |\tilde{\Psi}^{\nu}(\mathbf{r}, t)|^2. \quad (3.39)$$

Note that the integration over the surface has been converted into a volume integration over the whole device, since  $\tilde{\Gamma}$  is defined using distribution functions on the contacts (see Eq. (2.38)).

The second term can be transformed using the PE of Eq. (2.1):

$$\mathcal{I}_2 = q \int_D dV h_C(\mathbf{r}) \frac{\partial}{\partial t} \left( \sum_{\nu} q \tilde{n}^{\nu}(y, t) |\tilde{\Psi}^{\nu}(\mathbf{r}, t)|^2 \right). \quad (3.40)$$

Finally, the third term can be simplified by noting that the potential can be split into two parts [104] as

$$\tilde{V}(\mathbf{r}, t) = \tilde{V}_{\text{sc}}(\mathbf{r}, t) + \sum_{C'} h_{C'}(\mathbf{r}) \tilde{V}_{\text{appl}}^{C'}(t),$$

where  $\tilde{V}_{\text{sc}}(\mathbf{r}, t)$  is the part of the potential relating to the space charge, i.e. satisfying the PE, but with Dirichlet boundary conditions on all contacts given by

$$\tilde{V}_{\text{sc}}(\mathbf{r}, t) = 0, \quad \text{for } \mathbf{r} \in \partial D_C.$$

Conversely, the other part of the potential satisfies the Laplace equation as is explicitly clear from its construction with the Ramo-Shockley test functions  $h_C$ , however, it has Dirichlet boundary conditions given by the applied biases  $\tilde{V}_{\text{appl}}^C$ . Therefore the third term yields

$$\begin{aligned} \mathcal{I}_3 &= \frac{\partial}{\partial t} \int_D dV \kappa(\mathbf{r}) \nabla h_C(\mathbf{r}) \cdot \nabla \left( \tilde{V}_{\text{sc}}(\mathbf{r}, t) + \sum_{C'} h_{C'}(\mathbf{r}) \tilde{V}_{\text{appl}}^{C'}(t) \right) \\ &= \frac{\partial}{\partial t} \left[ \underbrace{\int_{\partial D} d\mathbf{A} \cdot \left( \tilde{V}_{\text{sc}}(\mathbf{r}, t) \kappa(\mathbf{r}) \nabla h_C(\mathbf{r}) \right)}_{=0} + \int_D dV \tilde{V}_{\text{sc}}(\mathbf{r}, t) \underbrace{\nabla \cdot \left( \kappa(\mathbf{r}) \nabla h_C(\mathbf{r}) \right)}_{=0} \right] \\ &\quad + \sum_{C'} \int_D dV \kappa(\mathbf{r}) \nabla h_C(\mathbf{r}) \cdot \nabla h_{C'}(\mathbf{r}) \tilde{V}_{\text{appl}}^{C'}(t). \end{aligned}$$

With the definition of the capacitance matrix,

$$\mathcal{C}_{C,C'} := \int_D dV \kappa(\mathbf{r}) \nabla h_C(\mathbf{r}) \cdot \nabla h_{C'}(\mathbf{r}), \quad (3.41)$$

we obtain

$$\mathcal{I}_3 = \frac{\partial}{\partial t} \sum_{C'} \mathcal{C}_{C,C'} \tilde{V}_{\text{appl}}^{C'}(t). \quad (3.42)$$

With  $\mathcal{I}_1$ ,  $\mathcal{I}_2$ , and  $\mathcal{I}_3$  of Eq. (3.39), (3.40), and (3.42), respectively, the terminal current of Eq. (3.38) can be expressed as

$$\begin{aligned} \tilde{I}_C(t) = q \int_D dV h_C(\mathbf{r}) \sum_{\nu} \left[ \tilde{\Gamma}^{\nu}(y, t) |\tilde{\Psi}^{\nu}(\mathbf{r}, t)|^2 - \frac{\partial}{\partial t} \left( \tilde{n}^{\nu}(y, t) |\tilde{\Psi}^{\nu}(\mathbf{r}, t)|^2 \right) \right] \\ - \frac{\partial}{\partial t} \sum_{C'} \mathcal{C}_{C, C'} \tilde{V}_{\text{appl}}^{C'}(t) \end{aligned} \quad (3.43)$$

and using the continuity equation of Eq. (3.36), we find

$$\begin{aligned} \tilde{I}_C(t) = q \int_D dV h_C(\mathbf{r}) \sum_{\nu} \left[ \left( \frac{\partial}{\partial y} \tilde{j}^{\nu}(y, t) - \tilde{S}^{\nu}(y, t) \right) |\tilde{\Psi}^{\nu}(\mathbf{r}, t)|^2 \right. \\ \left. - \tilde{n}^{\nu}(y, t) \frac{\partial}{\partial t} |\tilde{\Psi}^{\nu}(\mathbf{r}, t)|^2 \right] - \frac{\partial}{\partial t} \sum_{C'} \mathcal{C}_{C, C'} \tilde{V}_{\text{appl}}^{C'}(t). \end{aligned} \quad (3.44)$$

In our case, we apply a small signal bias to the contacts and therefore, we can linearize the equation using the usual expansion into a stationary part and a small harmonic perturbation as was shown in Sect. 3.1. Then we find for the phasor of the small signal terminal current

$$\begin{aligned} \underline{I}_C = q \int_D dV h_C(\mathbf{r}) \sum_{\nu} \left[ \left( \frac{\partial}{\partial y} \underline{j}^{\nu}(y) - \underline{S}^{\nu}(y) \right) |\Psi^{\nu}(\mathbf{r})|^2 \right. \\ \left. + 2 \left( \frac{\partial}{\partial y} j^{\nu}(y) - S^{\nu}(y) \right) \Psi^{\nu}(\mathbf{r}) \underline{\Psi}^{\nu}(\mathbf{r}) \right. \\ \left. - 2i\omega n^{\nu}(y) \Psi^{\nu}(\mathbf{r}) \underline{\Psi}^{\nu}(\mathbf{r}) \right] - i\omega \sum_{C'} \mathcal{C}_{C, C'} \underline{V}_{\text{appl}}^{C'}, \end{aligned}$$

where we used that the wave functions are purely real. Finally, we can use the continuity equation for the stationary case,

$$\frac{\partial}{\partial y} j^{\nu}(y) = S^{\nu}(y) + \Gamma^{\nu}(y),$$

and the definition of the GR rate of Eq. (2.38) to simplify

$$\begin{aligned} \int_D dV h_C(\mathbf{r}) \left( \frac{\partial}{\partial y} j^{\nu}(y) - S^{\nu}(y) \right) \Psi^{\nu}(\mathbf{r}) \underline{\Psi}^{\nu}(\mathbf{r}) \\ = \int_D dV h_C(\mathbf{r}) \Gamma^{\nu}(y) \Psi^{\nu}(\mathbf{r}) \underline{\Psi}^{\nu}(\mathbf{r}) \end{aligned}$$

$$\begin{aligned}
&= - \int_{\partial D_S \cup \partial D_D} dA h_C(\mathbf{r}) v_{GR} (n^\nu(y) - n_{eq}) \Psi^\nu(\mathbf{r}) \underline{\Psi}^\nu(\mathbf{r}) \\
&= - (\delta_{C,S} + \delta_{C,D}) \int_{\partial D_C} dA v_{GR} (n^\nu(y) - n_{eq}) \Psi^\nu(\mathbf{r}) \underline{\Psi}^\nu(\mathbf{r}) = 0,
\end{aligned}$$

where we used that the source and drain contacts span the whole semiconductor region (cf. Fig. 2.1). This implies that the Ramo-Shockley test function  $h_C$  is unity over the whole surface integration region on the  $C$ -th contact while it is zero on the other contacts. The reason why this integral vanishes in the last step is that  $\Psi$  and  $\underline{\Psi}$  are orthogonal to each other. Note that we can only omit this term if our source and drain contacts cover the whole semiconductor region in confinement direction, otherwise we have to include it.

Thus, the final expression for the phasor of the small signal terminal current is given by

$$\begin{aligned}
\underline{I}_C = q \int_D dV h_C(\mathbf{r}) \sum_\nu \left[ \left( \frac{\partial}{\partial y} \underline{j}^\nu(y) - \underline{S}^\nu(y) \right) |\Psi^\nu(\mathbf{r})|^2 \right. \\
\left. - 2i\omega n^\nu(y) \Psi^\nu(\mathbf{r}) \underline{\Psi}^\nu(\mathbf{r}) \right] - i\omega \sum_{C'} \mathcal{C}_{C,C'} V_{\text{appl}}^{C'}. \quad (3.45)
\end{aligned}$$

and since the device is homogeneous in  $z$ -direction (cf. Fig. 2.1), we can trivially integrate over a box of size  $\Delta z$  which yields

$$\underline{I}_C = \underline{I}'_C \Delta z,$$

where  $\underline{I}'_C$  is the terminal current per length in  $z$ -direction.

Discretizing Eq. (3.45) is straightforward. With a spatial grid as presented in Sect. 2.4.5, the derivative of the current density can be expressed as

$$\frac{\partial \underline{j}^\nu}{\partial y}(y_i) = \frac{\underline{j}^\nu(y_{i+}) - \underline{j}^\nu(y_{i-})}{\Delta y_i},$$

where the current density vanishes outside the device, i.e.  $\underline{j}^\nu(y_{1-}) = \underline{j}^\nu(y_{N_y+}) = 0$ . All other quantities are located on direct spatial grid points and therefore we find

$$\begin{aligned}
\underline{I}'_C = q \sum_{k,i,\nu} \Delta x_k \Delta y_i h_C(x_k, y_i) \left[ \left( \frac{\underline{j}^\nu(y_{i+}) - \underline{j}^\nu(y_{i-})}{\Delta y_i} - \underline{S}^\nu(y_i) \right) |\Psi^\nu(x_k, y_i)|^2 \right. \\
\left. - 2i\omega n^\nu(y_i) \Psi^\nu(x_k, y_i) \underline{\Psi}^\nu(x_k, y_i) \right] - i\omega \sum_{C'} \mathcal{C}'_{C,C'} V_{\text{appl}}^{C'}. \quad (3.46)
\end{aligned}$$

### 3.5 Reciprocity

In most cases, as in ours, we expect reciprocity of a device in equilibrium without a magnetic field [51]. Reciprocity is a property in equilibrium which states that applying a small signal bias at contact  $C$  and measuring the small signal terminal current at contact  $C'$  will yield the same result as applying a small signal bias to contact  $C'$  and measuring the small signal terminal current at contact  $C$ . This property is valid irrespective of the sizes or shapes of the contacts or whether they are ohmic or metal contacts. Thus, reciprocity simply means that the admittance matrix  $Y$  has to be symmetric, i.e.

$$Y_{C,C'} = Y_{C',C},$$

with

$$Y_{C,C'} = \frac{\partial I_C}{\partial V_{\text{appl}}^{C'}} = \frac{I_C}{V_{\text{appl}}^{C'}}, \quad \text{with } \underline{V}_{\text{appl}}^{C''} = 0, \quad C'' \neq C'. \quad (3.47)$$

In this subsection, we will see that although reciprocity is fulfilled analytically, it is not fulfilled numerically for arbitrary discretization schemes in  $H$ -space. Violating reciprocity and therefore yielding unphysical results presents a major problem for a simulator and thus we will address and resolve this issue in this section.

#### 3.5.1 Proof up to the First Harmonic

In order to show that reciprocity must be fulfilled and where it breaks for our system of equations, we will consider the small signal BE without any scattering, i.e.  $\underline{S} = 0$ . Furthermore, we are taking the limit  $v_{\text{GR}} \rightarrow \infty$ , i.e. the GR-rate  $\underline{\Gamma}$  of Eq. (3.31) effectively becomes a Dirichlet boundary condition.

The discretization summarized by Eq. (3.12) does not yield reciprocal devices in equilibrium. The reason for this is a broken symmetry between BE and PE. To investigate and resolve this issue, let us write down the discretized ballistic equations up to first order in the harmonics, since this is sufficient to discover the origin of the non-reciprocity.

The linearized free streaming term in equilibrium up to the first Fourier harmonic, including the time-derivative of Eq. (3.20), reads

$$\underline{L}_0^\nu(y, H) = Z^v \frac{\partial}{\partial y} \left( a_{0,1}^\nu(y, H) \underline{f}_1^\nu(y, H) \right) + i\omega Z^v \underline{f}_0^\nu(y, H) + i\omega Z^v \frac{\partial f_0^\nu(y, H)}{\partial H} \underline{\varepsilon}^\nu(y),$$

$$\underline{L}_1^\nu(y, H) = Z^v a_{1,0}^\nu(y, H) \frac{\partial}{\partial y} \underline{f}_0^\nu(y, H) + i\omega Z^v \underline{f}_1^\nu(y, H).$$

Note that we do not need to include negative harmonics since they are composed of sine functions (cf. Eq. (2.55)) whose symmetry is consistent with an

acceleration in the  $z$ -direction (cf. Fig. 2.1). An example of this is the Lorentz force of a magnetic field, however, in this work we do not include such forces.

If we ignore scattering, the BE is given by the set of equations  $\underline{L}_0 = 0$  and  $\underline{L}_1 = 0$ . Eliminating the first harmonic of the distribution function yields

$$0 = \frac{i}{\omega} Z^v \frac{\partial}{\partial y} \left[ (a'_{0,1}(y, H))^2 \frac{\partial}{\partial y} f_0^\nu(y, H) \right] + i\omega Z^v \underline{f}_0^\nu(y, H) + i\omega Z^v \frac{\partial f_0^\nu(y, H)}{\partial H} \underline{\varepsilon}^\nu(y), \quad (3.48)$$

where we used that  $a'_{0,1} \equiv a'_{1,0}$ . Moreover, in equilibrium we have

$$f_0^\nu(y, H) = \frac{1}{Y_0} f_{\text{eq}}(y, H).$$

Multiplying Eq. (3.48) by  $1/Y_0$  and integrating over  $H$ -space yields the 1D continuity equation of Eq. (3.36) without the scattering or boundary terms, of course. From there, the current and density can be read off as

$$\begin{aligned} \underline{j}^\nu(y) &= \frac{iZ}{\omega Y_0} \int dH (a'_{0,1}(y, H))^2 \frac{\partial}{\partial y} f_0^\nu(y, H), \\ \underline{n}^\nu(y) &= \frac{Z}{Y_0} \int dH \left( f_0^\nu(y, H) + \frac{\partial f_0^\nu(y, H)}{\partial H} \underline{\varepsilon}^\nu(y) \right). \end{aligned} \quad (3.49)$$

Note that the expression for the current density of Eq. (3.49) is only valid in this simplified case.

For our proof of reciprocity, we will start from an equivalent expression of Eq. (3.33) for the small signal phasor of the terminal current:

$$\underline{I}_C = - \int_{\partial D_C} d\mathbf{A} \cdot \left[ \sum_\nu q \underline{j}^\nu(y) \mathbf{e}_y \delta(x) + i\omega \kappa(\mathbf{r}) \nabla V(\mathbf{r}) \right], \quad (3.50)$$

where the  $x$ -component of the conduction current was omitted since it cannot contribute to the source and drain contact currents due to its orthogonality on the measure  $d\mathbf{A}$  (cf. Fig. 2.1). Furthermore, since the gates are separated from the semiconductor region by oxides and we assume that electrons cannot penetrate into the oxide (see Sect. 2.3), the electron current cannot contribute to the gate terminal current. It is assumed that the delta-distribution  $\delta(x)$  restricts the 1D conduction current in transport direction in such a way that it originates and terminates at the contacts along the  $y$ -direction. Note that the equivalence to Eq. (3.33) is due to the surface measure  $d\mathbf{A}$  being oriented perpendicular to the  $x$ -direction.

To proceed, we need to define suitable basis functions – or fundamentals – similar to the Ramo-Shockley test functions that allow us to explicitly write down the admittance parameters and which are also suitable to manifest the symmetry of our terms. Since the distribution function and the potential have different dependencies, we need to define different fundamentals  $\underline{p}$  and  $\underline{q}$ :

$$\underline{f}_0^\nu(y, H) =: \sum_C \underline{p}_C^\nu(y, H) \Lambda^\nu(H) \underline{V}_{\text{appl}}^C, \quad (3.51)$$

$$\underline{V}(\mathbf{r}) =: \sum_C \underline{q}_C(\mathbf{r}) \underline{V}_{\text{appl}}^C, \quad (3.52)$$

where a differential equation for  $\underline{p}$  can be obtained straightforwardly by inserting Eq. (3.51) into Eq. (3.48) and  $\underline{q}$  fulfills the Laplace equation. Here,  $\Lambda$  contains the whole  $H$ -dependence of  $\underline{f}_0$  when a bias is applied to either source or drain, so that the fundamentals can fulfill the boundary conditions

$$\begin{aligned} \underline{p}_C(y, H) &= \delta_{C,C'}, \quad (x=0, y) \in \partial D_{C'}, \\ \underline{q}_C(\mathbf{r}) &= \delta_{C,C'}, \quad \mathbf{r} \in \partial D_{C'}, \end{aligned} \quad (3.53)$$

where the boundary condition for the fundamental  $\underline{p}$  means the Kronecker-delta only yields unity when  $\mathbf{r}$  is on one of the contacts in the semiconductor region. However, if a bias is applied to one of the contacts outside the semiconductor region, the fundamental  $\underline{p}$  can still be non-zero inside the device. In equilibrium, we can derive the boundary conditions of the small signal distribution functions. A small perturbation of the applied bias on the  $k$ -th contact yields (cf. Eq. (3.30))

$$\underline{f}_0^\nu(y_{S/D}, H) Y_0 = \frac{\partial}{\partial V_{\text{appl}}^{S/D}} f_{\text{eq}}^\nu(y_{S/D}, H) \underline{V}_{\text{appl}}^{S/D} = \frac{\partial}{\partial H} f_{\text{eq}}^\nu(y_{S/D}, H) q \underline{V}_{\text{appl}}^{S/D}. \quad (3.54)$$

Using Eqs. (3.51) and (3.53), we immediately find

$$\Lambda^\nu(H) = \frac{q}{Y_0} \frac{\partial}{\partial H} f_{\text{eq}}^\nu(y_{S/D}, H). \quad (3.55)$$

Note that  $\Lambda$  does not depend on the source or drain contact since we have  $f_{\text{eq}}^\nu(y_S, H) = f_{\text{eq}}^\nu(y_D, H)$  in equilibrium.

Quite similar to Sect. 3.4 about the Ramo-Shockley theorem, we can multiply the integrand of Eq. (3.50) with the fundamentals without changing the result

$$\begin{aligned} \underline{I}_C = - \int_{\partial D_C} d\mathbf{A} \cdot \left[ q \sum_\nu \int dH \left( \underline{p}_C^\nu(y, H) \mathcal{J}_f \underline{f}_0^\nu(y, H) \right) \mathbf{e}_y \delta(x) \right. \\ \left. + i\omega \underline{q}_C(\mathbf{r}) \kappa(\mathbf{r}) \nabla \underline{V}(\mathbf{r}) \right], \end{aligned}$$

where we used the current operator

$$\mathcal{J}_f f_0^\nu(y, H) := \frac{iZ^\nu}{\omega Y_0} (a_{0,1}^\nu(y, H))^2 \frac{\partial}{\partial y} f_0^\nu(y, H).$$

Then, extending the integration to the whole surface and using the Gaussian integration theorem, while expressing  $f_0$  and  $\underline{V}$  in terms of the fundamentals given by Eq. (3.51) and Eq. (3.52), respectively, we obtain

$$\begin{aligned} \underline{I}_C &= - \sum_{C'} \int_D dV \left[ q \sum_\nu \int dH \Lambda^\nu(H) \nabla \underline{p}_{C'}^\nu(y, H) \cdot \left( \mathcal{J}_f \underline{p}_{C'}^\nu(y, H) \mathbf{e}_y \delta(x) \right) \right. \\ &\quad + q \sum_\nu \int dH \Lambda^\nu(H) \underline{p}_{C'}^\nu(y, H) \nabla \cdot \left( \mathcal{J}_f \underline{p}_{C'}^\nu(y, H) \mathbf{e}_y \delta(x) \right) \\ &\quad + i\omega \kappa(\mathbf{r}) \nabla \underline{q}_{C'}(\mathbf{r}) \cdot \nabla \underline{q}_{C'}(\mathbf{r}) \\ &\quad \left. + i\omega \underline{q}_{C'}(\mathbf{r}) \nabla \cdot \left( \kappa(\mathbf{r}) \nabla \underline{q}_{C'}(\mathbf{r}) \right) \right] \underline{V}_{\text{appl}}^{C'} \\ &=: \sum_{C'} (\mathcal{I}_{1,C,C'} + \mathcal{I}_{2,C,C'} + \mathcal{I}_{3,C,C'} + \mathcal{I}_{4,C,C'}) \underline{V}_{\text{appl}}^{C'} \\ &=: \sum_{C'} Y_{C,C'} \underline{V}_{\text{appl}}^{C'}, \end{aligned}$$

where we defined an abbreviation of each term in the penultimate step. Furthermore, the definition of the admittance matrix  $Y$  arises naturally due to the proportionality to the applied bias. What is left to show is that reciprocity, i.e.  $Y_{C,C'} = Y_{C',C}$ , is fulfilled.

For the first term we find

$$\begin{aligned} \mathcal{I}_{1,C,C'} &= -q \sum_\nu \frac{iZ^\nu}{\omega Y_0} \int_D dV \int dH \Lambda^\nu(H) \frac{\partial \underline{p}_{C'}^\nu(y, H)}{\partial y} (a_{0,1}^\nu(y, H))^2 \\ &\quad \times \frac{\partial \underline{p}_{C'}^\nu(y, H)}{\partial y} \delta(x), \end{aligned}$$

which is clearly symmetric in interchanging  $C$  and  $C'$ . The third term  $\mathcal{I}_{3,C,C'}$  is trivially symmetric. However, the second and fourth term,  $\mathcal{I}_{2,C,C'}$  and  $\mathcal{I}_{4,C,C'}$ , are not individually symmetric but only their sum. To see this, let us split the second term as follows

$$\begin{aligned} \mathcal{I}_{2,C,C'} &= -q \sum_\nu \frac{iZ^\nu}{\omega Y_0} \int_D dV \int dH \Lambda^\nu(H) \underline{p}_{C'}^\nu(y, H) \\ &\quad \times \frac{\partial}{\partial y} \left[ (a_{0,1}^\nu(y, H))^2 \frac{\partial \underline{p}_{C'}^\nu(y, H)}{\partial y} \delta(x) \right] \end{aligned}$$

$$\begin{aligned}
&= q \sum_{\nu} \frac{i\omega Z^{\nu}}{Y_0} \int_D dV \int dH \delta(x) \underline{p}_{C'}^{\nu}(y, H) \left[ \Lambda^{\nu}(H) \underline{p}_{C'}^{\nu}(y, H) \right. \\
&\quad \left. + \frac{\partial f_0^{\nu}(y, H)}{\partial H} \int dx' \Delta_{\underline{\varepsilon}}^{\nu}(x', y) \underline{q}_{C'}(x', y) \right] \\
&= q \sum_{\nu} \frac{i\omega Z^{\nu}}{Y_0} \int_D dV \int dH \underline{p}_{C'}^{\nu}(y, H) \left[ \delta(x) \Lambda^{\nu}(H) \underline{p}_{C'}^{\nu}(y, H) \right. \\
&\quad \left. - q \frac{\partial f_0^{\nu}(y, H)}{\partial H} |\Psi^{\nu}(\mathbf{r})|^2 \underline{q}_{C'}(\mathbf{r}) \right] \\
&=: \mathcal{I}'_{2,C,C'} + \mathcal{I}''_{2,C,C'},
\end{aligned}$$

where we used Eqs. (3.48) and (3.8), and we executed the integration over  $x$  in the second term and subsequently renamed  $x' \rightarrow x$ . Obviously  $\mathcal{I}'_{2,C,C'}$  is symmetric by itself. The fourth term can be expressed using the small signal PE of Eq. (3.16) as

$$\begin{aligned}
\mathcal{I}_{4,C,C'} &= -q \sum_{\nu} \frac{i\omega Z^{\nu}}{Y_0} \int_D dV \underline{q}_{C'}(\mathbf{r}) \int dH \underline{p}_{C'}^{\nu}(y, H) \Lambda^{\nu}(H) |\Psi^{\nu}(\mathbf{r})|^2 \\
&\quad - 2iq\omega \sum_{\nu} \int_D dV \underline{q}_{C'}(\mathbf{r}) n^{\nu}(y) \Psi(\mathbf{r}) \int dx' \Delta_{\underline{\Psi}}^{\nu}(x, x', y) \underline{q}_{C'}(x', y) \\
&=: \mathcal{I}'_{4,C,C'} + \mathcal{I}''_{4,C,C'},
\end{aligned}$$

where we also used Eq. (3.9). Adding  $\mathcal{I}''_{2,C,C'}$  and  $\mathcal{I}'_{4,C,C'}$  leads to

$$\begin{aligned}
\mathcal{I}''_{2,C,C'} + \mathcal{I}'_{4,C,C'} &= -q \sum_{\nu} \frac{i\omega Z^{\nu}}{Y_0} \int_D dV \int dH \\
&\quad \times \left[ q \underline{p}_{C'}^{\nu}(y, H) \frac{\partial f_0^{\nu}(y, H)}{\partial H} |\Psi^{\nu}(\mathbf{r})|^2 \underline{q}_{C'}(\mathbf{r}) \right. \\
&\quad \left. + \underline{q}_{C'}(\mathbf{r}) \underline{p}_{C'}^{\nu}(y, H) \Lambda^{\nu}(H) |\Psi^{\nu}(\mathbf{r})|^2 \right] \\
&= -q^2 \sum_{\nu} \frac{i\omega Z^{\nu}}{Y_0} \int_D dV \int dH \frac{\partial f_0^{\nu}(y, H)}{\partial H} |\Psi^{\nu}(\mathbf{r})|^2 \\
&\quad \times \left[ \underline{p}_{C'}^{\nu}(y, H) \underline{q}_{C'}(\mathbf{r}) + \underline{p}_{C'}^{\nu}(y, H) \underline{q}_{C'}(\mathbf{r}) \right], \quad (3.56)
\end{aligned}$$

which is evidently symmetric. Note that we used that the energy dependency  $\Lambda$  of the boundary condition in equilibrium can be expressed as in Eq. (3.55).

The remaining term reads

$$\mathcal{I}_{4,C,C'}'' = -2iq\omega \sum_{\nu} \iint dy dz n^{\nu}(y) \iint dx dx' \underline{q}_C(\mathbf{r}) \Psi^{\nu}(\mathbf{r}) \Delta_{\underline{\Psi}}^{\nu}(x, x', y) \underline{q}_{C'}(x', y).$$

Showing that this is symmetric is the same as showing that

$$\begin{aligned} 0 &\stackrel{!}{=} \iint dx dx' \underline{q}_C(x, y) \Psi^{\nu}(x, y) \Delta_{\underline{\Psi}}^{\nu}(x, x', y) \underline{q}_{C'}(x', y) \\ &\quad - \iint dx dx' \underline{q}_{C'}(x, y) \Psi^{\nu}(x, y) \Delta_{\underline{\Psi}}^{\nu}(x, x', y) \underline{q}_C(x', y) \\ &= \iint dx dx' \underline{q}_C(x, y) \underline{q}_{C'}(x', y) [\Psi^{\nu}(x, y) \Delta_{\underline{\Psi}}^{\nu}(x, x', y) - \Psi^{\nu}(x', y) \Delta_{\underline{\Psi}}^{\nu}(x', x, y)]. \end{aligned}$$

holds. But this is always true since

$$\begin{aligned} &\Psi^{\nu}(x, y) \Delta_{\underline{\Psi}}^{\nu}(x, x', y) - \Psi^{\nu}(x', y) \Delta_{\underline{\Psi}}^{\nu}(x', x, y) \\ &= -q\Psi^{\nu}(x, y) \sum_{s \neq \nu} \frac{\Psi^s(x', y) \Psi^{\nu}(x', y)}{\varepsilon^{\nu}(y) - \varepsilon^s(y)} \Psi^s(x, y) \\ &\quad + q\Psi^{\nu}(x', y) \sum_{s \neq \nu} \frac{\Psi^s(x, y) \Psi^{\nu}(x, y)}{\varepsilon^{\nu}(y) - \varepsilon^s(y)} \Psi^s(x', y) \\ &= 0, \end{aligned}$$

where the sums run over all subbands  $s \neq \nu$  of the same valley.

Thus, we proved that our device has to be reciprocal in the continuum. At the very least up to first order in Fourier harmonics and with Dirichlet boundary conditions. However, our derivation shows that reciprocity relies on a consistent formulation of the BE and the PE. In Eq. (3.56)  $\mathcal{I}_{2,C,C'}''$  stems from inserting the BE into  $\mathcal{I}_{2,C,C'}$  and  $\mathcal{I}_{4,C,C'}'$  stems from inserting the PE into  $\mathcal{I}_{4,C,C'}$ . Thereafter we use Eq. (3.55) to prove the symmetry. If the BE and PE are not consistent with each other, reciprocity cannot hold in equilibrium.

### 3.5.2 Restoring Reciprocity in Discretized $H$ -Space

Reciprocity is violated once we discretize  $H$ -space since, after the box-integration, Eq. (3.54) is given by

$$\begin{aligned} \underline{f}_0^{\nu}(y_{S/D}, H_j) (H_{j+} - H_{j-}) &\stackrel{!}{=} \int_{H_{j-}}^{H_{j+}} dH \frac{\partial f_0(y_{S/D}, H)}{\partial H} qV_{\text{appl}}^{S/D} \\ &= (f_0(y_{S/D}, H_{j+}) - f_0(y_{S/D}, H_{j-})) qV_{\text{appl}}^{S/D}. \end{aligned} \quad (3.57)$$

We cannot fulfill Eq. (3.57) with the discretization described by Eq. (3.28) in Sect. 3.2. But when Eq. (3.57) does not hold, we cannot get reciprocity, since

this relation is required to make Eq. (3.56) symmetric. Thus, with a naive discretization, our device will not be reciprocal numerically.

In order to restore reciprocity, we need to ensure two things. First, we need to find some suitable representation of the r.h.s. of Eq. (3.57) that will also be reciprocal in discretized  $H$ -space. This representation then needs to be used in the time-derivative of the BE shown in Eq. (3.21) and (3.22). Second, we need to substitute this representation in the density of the PE as well, in order for Eq. (3.56) to be symmetric in discretized  $H$ -space.

To commence, let us define the distribution function in terms of an equilibrium part  $f_{\text{eq}}$  and a non-equilibrium part, i.e.

$$f_0^\nu(y, H) =: f_{\text{eq}}^\nu(y, H) f_{\text{ne},0}^\nu(y, H),$$

where  $f_{\text{ne},0}$  is the coefficient of the zeroth harmonic of the non-equilibrium part. Then we can formulate the box-integration over the derivative as

$$\begin{aligned} \int_{H_{j-}}^{H_{j+}} dH \frac{\partial f_0^\nu(y, H)}{\partial H} &\approx [f_{\text{ne},0}^\nu(y, H_{j+}) - f_{\text{ne},0}^\nu(y, H_{j-})] f_{\text{eq}}^\nu(y, H_j) \\ &+ \left. \frac{\partial f_{\text{eq}}^\nu(y, H)}{\partial H} \right|_{H=H_j} f_{\text{ne},0}^\nu(y, H_j) [H_{j+} - H_{j-}], \end{aligned} \quad (3.58)$$

where the distribution function  $f_{\text{ne},0}$  on the intermediate points  $H_{j\pm}$  is evaluated as in Eq. (3.28). Since the equilibrium distribution function is well known, we can execute the derivative w.r.t.  $H$  on the r.h.s. analytically and obtain the relation of Eq. (3.57) in equilibrium exactly.

The second condition can be met by replacing the density in the PE of Eq. (3.16) in such a way that the Eq. (3.56) is still valid in discretized  $H$ -space. To this end, let us write down the small signal density of Eq. (3.18) using Eq. (3.54) as

$$\underline{n}^\nu(y) = \frac{Z^\nu}{Y_0} \int_{\varepsilon^\nu(y)}^{\infty} dH \left[ f_0^\nu(y, H) + \frac{\partial f_0^\nu(y, H)}{\partial H} \underline{\varepsilon}^\nu(y) \right]. \quad (3.59)$$

Then, in discretized  $H$ -space, we use Eq. (3.58) to box-integrate the second term on the r.h.s. of Eq. (3.59).

Note that the procedure described in this section yields identical results in the continuum. Only upon discretization do we find differences such that the approach described here eliminates discretization artifacts which would lead to non-reciprocal results.

As a side note, we already used the equality of densities described by the PE and BE implicitly in the derivation of the Ramo-Shockley theorem of Sect. 3.4. Thus, the derivation of the terminal current for a discretized system also requires that we express the small signal density as in Eq. (3.59).

### 3.5.3 General Case

Although the previous two sections dealt with reciprocity in the special case of ballistic transport with Dirichlet boundary conditions and only up to the first Fourier harmonic, the results are equally applicable to the general case. However, if we extend the results to the general case, we will find some ambiguity in the way we formulate our restoration procedure.

For example, once we leave equilibrium conditions, the BE requires us to formulate the derivative of  $f_1^\nu(y, H)$  w.r.t.  $H$  in the time-derivative. But there is no mandate on how this term should be expressed since  $f_1$  vanishes in equilibrium anyway and therefore does not influence any conservation laws in equilibrium. We could in principle also formulate higher orders of harmonics using Eq. (3.58) but instead we choose to leave them as they are. However, any choice can be legitimized as convergence of both cases to the continuum values is expected for reasonably fine  $H$ -grids.

Therefore the zeroth harmonic of the even time derivative of Eq. (3.21) will be replaced with

$$\begin{aligned} \bar{T}_0^\nu(y_i, H_j) = i\omega Z^v \Delta y_i \left( \left[ [f_{\text{ne},0}^\nu(y_i, H_{j+}) - f_{\text{ne},0}^\nu(y_i, H_{j-})] f_{\text{eq}}^\nu(y_i, H_j) \right. \right. \\ \left. \left. + \frac{\partial f_{\text{eq}}^\nu(y, H)}{\partial H} \Big|_{H=H_j} f_{\text{ne},0}^\nu(y, H_j) \Delta H^\nu(y_i, H_j) \right] \underline{\varepsilon}^\nu(y_i) \right. \\ \left. + \Delta H^\nu(y_i, H_j) \underline{f}_0^\nu(y_i, H_j) \right), \end{aligned} \quad (3.60)$$

whereas the other harmonics remain unaltered. Note that  $f_{\text{ne},0}^\nu(y_i, H_{j\pm})$  has to be evaluated according to Eq. (3.28).

In any case, once we implement the restoration procedure for reciprocity of Sect. 3.5.2, we find that the device is reciprocal in equilibrium for arbitrary numbers of harmonics, for our boundary conditions of Sect. 3.3 with finite GR recombination velocities, for inhomogeneous devices, and for transport including scattering.

## 3.6 Solving for Small Signal Parameters

We can go about solving the small signal problem in two ways. Either we regard our set of equations comprising BE and PE – with the SE included via perturbation theory – in the same way as we did in the stationary case or we can transform the system of equations and solve for the admittance parameters directly. We will refer to the former case as the direct method while the latter case will be referred to as the adjoint method. Furthermore, we will also briefly

explain how to compute the admittance parameters of the common-source configuration which will be used exclusively in Chap. 5 about the simulation results.

Our goal is the determination of the admittance parameters:

$$Y'_{C,C'} = \frac{\underline{I}'_C[\underline{f}, \underline{V}, \underline{V}_{\text{appl}}^{C'}]}{\underline{V}_{\text{appl}}^{C'}}, \quad \text{with } \underline{V}_{\text{appl}}^{C''} = 0, \quad C'' \neq C',$$

where  $\underline{I}'_C[\underline{f}, \underline{V}, \underline{V}_{\text{appl}}^{C'}]$  is the small signal terminal current of the contact  $C$  per length in  $z$ -direction when a small signal bias is applied to the contact  $C'$ . Thus the admittance parameters  $Y'$  are also given per length in  $z$ -direction. We wrote down the small signal terminal current  $\underline{I}'$  as a functional of the small signal variables in order to bring across the following point: When a small signal bias is applied to the contact  $C'$ , the expression for the small signal terminal current is strictly linear in one of either  $\underline{f}$ ,  $\underline{V}$ , or  $\underline{V}_{\text{appl}}^{C'}$ , as was shown in Eq. (3.46). Therefore, we may write down the admittance parameter as

$$Y'_{C,C'} = \underline{I}'_C \left[ \frac{\underline{f}}{\underline{V}_{\text{appl}}^{C'}}, \frac{\underline{V}}{\underline{V}_{\text{appl}}^{C'}}, 1 \right] = \underline{I}'_C \left[ \underline{f}^{\text{norm},C'}, \underline{V}^{\text{norm},C'}, 1 \right], \quad (3.61)$$

where we defined the small signal distribution functions normalized by the small signal applied bias at the contact  $C'$ .

We will attempt to solve the small signal system of equations only ever with a small signal bias applied to one contact, which we will call  $C$ . Since the small signal BE of Eq. (3.12) and the small signal PE of Eq. (3.15) with its boundary condition of Eq. (3.29) are also strictly linear in either  $\underline{f}$ ,  $\underline{V}$ , or  $\underline{V}_{\text{appl}}^C$ , we can divide both the BE and PE by  $\underline{V}_{\text{appl}}^C$  and obtain equations for  $\underline{f}^{\text{norm},C}$  and  $\underline{V}^{\text{norm},C}$ . In the remainder of this section we will refer to the system of BE and PE in matrix form as

$$A^{\text{AC}} \underline{\mathbf{x}}^C = \underline{\mathbf{b}}^C, \quad (3.62)$$

where  $A^{\text{AC}} \in \mathbb{C}^{N \times N}$  contains the BE and PE, the vector

$$\underline{\mathbf{x}}^C = (\underline{f}^{\text{norm},C} \quad \underline{V}^{\text{norm},C})^t \quad (3.63)$$

contains the normalized distribution function and potential for a bias applied to the contact  $C$ , and  $\underline{\mathbf{b}}^C$  contains the term proportional to the applied bias of the GR rate of Eq. (3.31), as well as the small signal variant of the normalized PE boundary condition given by

$$\underline{V}^{\text{norm},C}(\mathbf{r}) \Big|_{\mathbf{r} \in \partial D_{\text{TG}}} = \delta_{C,\text{TG}}, \quad \underline{V}^{\text{norm},C}(\mathbf{r}) \Big|_{\mathbf{r} \in \partial D_{\text{BG}}} = \delta_{C,\text{BG}}.$$

### 3.6.1 Direct Method

We want to directly determine the solution  $\underline{\mathbf{x}}^C$  of Eq. (3.62) when a small signal bias is applied to the contact  $C$ . Since Eq. (3.62) is a linear equation, we can avail ourselves of any linear solver to obtain the solution directly. However, it is prudent to employ the same procedure shown in Sect. 2.5.2 to remove the odd distribution functions in order to reduce the computational load significantly. Using the same transformation matrix  $S$  and compression matrix  $C_e$  as in Sect. 2.5.2, we find

$$\underbrace{C_e S A^{\text{AC}} C_e^t}_{=:A_{e/e}^{\text{AC}}} \underbrace{C_e \underline{\mathbf{x}}^C}_{=: \underline{\mathbf{x}}_e^C} = \underbrace{C_e S \mathbf{b}^C}_{=: \mathbf{b}_e^C}, \quad (3.64)$$

where  $A_{e/e}^{\text{AC}} \in \mathbb{C}^{n \times n}$  only contains the even rows and columns of the BE as well as the full PE,  $\underline{\mathbf{x}}_e^C \in \mathbb{C}^n$  contains the even distribution functions and the whole potential, and  $\mathbf{b}_e^C \in \mathbb{C}^n$  contains the even rows of the BE but the full PE for a bias applied to the contact  $C$ . Equation (3.64) is a linear sparse array system of equations which we solve with ILUPACK [68].

Once we know the distribution functions with even harmonics and the potential, we can use

$$\left. \frac{f_{\alpha=(\nu, y_{i+}, H_j, m)}^{\text{norm}, C}}{m \text{ odd}} \right| = -\frac{1}{A_{\alpha, \alpha}^{\text{AC}}} \left( \sum_{\substack{\beta=1 \\ \beta \neq \alpha}}^{N_{\text{BE}}} A_{\alpha, \beta}^{\text{DC}} f_{\beta}^{\text{norm}, C} + \sum_{b=1}^{N_{\text{PE}}} A_{\alpha, N_{\text{BE}}+b}^{\text{AC}} V_b^{\text{norm}, C} \right), \quad (3.65)$$

to determine the distribution functions with odd harmonics. The reasoning behind Eq. (3.65) is completely analogous to Eq. (2.112), however, the boundary term  $b_{\alpha}^C$  does not contribute since it is zero except for coordinates at which boundary conditions apply. But there are no boundary conditions in odd equations (cf. Sect. 2.4.5).

Once the normalized distribution functions and the potential are known, we can use the expression for the small signal terminal current of Eq. (3.46) to compute the admittance parameters according to Eq. (3.61). In addition, we can also compute other quantities such as the small signal density or the small signal current density.

### 3.6.2 Adjoint Method

In case we are only interested in the admittance parameters, we can use the adjoint method to compute them directly [105]. But using the adjoint method has no advantage by itself. It even prevents us from computing other quantities

such as small signal densities – which would be possible with the direct method of the previous section. However, it will open up the possibility to compute the noise and admittance parameters in one step as will be shown later on in Sect. 4.8.

In order to derive a system of equations which can be solved directly for the admittance parameters, let us write down a representation for the admittance parameters with the help of a projection operator:

$$Y_{C',C} = \mathcal{P}_{\underline{I}_{C'}}^t \mathbf{x}^C - i\omega C'_{C',C}. \quad (3.66)$$

Here,  $\mathcal{P}_{\underline{I}_{C'}} \in \mathbb{C}^N$  is a vector chosen in such a way that Eq. (3.66) is equivalent to Eq. (3.61), i.e. the elements of  $\mathcal{P}_{\underline{I}_{C'}}$  are the coefficients multiplying the distribution function and potential in the terminal current of Eq. (3.46). Note that the capacitance matrix  $C'_{C',C}$  appears without a sum over all contacts since we assume that we only apply a small signal bias at the contact C. Furthermore, the applied bias does not appear since it is normalized to unity as in Eq. (3.61).

What complicates the matter in our case is that we want to solve the reduced system of equations which only contains the even equations. To this end, let us split the distribution functions into even and odd parts as

$$\mathbf{x}^C = C_e^t \underline{\mathbf{x}}_e^C + (C_o^{\text{BE}})^t \underline{\mathbf{f}}_o^{\text{norm},C}, \quad (3.67)$$

where  $C_e$  is given by Eq. (2.110), the odd compression matrix for the BE  $C_o^{\text{BE}} \in \mathbb{R}^{N_{\text{BE}}/2 \times N_{\text{BE}}}$  can be defined in analogy to Eq. (2.99) with

$$C_o^{\text{BE}} = \begin{pmatrix} & m=0 & m=1 & m=2 & m=3 & \dots \\ 0 & 1 & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & 1 & \dots & 0 \\ \vdots & & \ddots & & \vdots & \end{pmatrix} \begin{matrix} m=1 \\ m=3 \\ \vdots \end{matrix}, \quad (3.68)$$

and  $\underline{\mathbf{f}}_o^{\text{norm},C} \in \mathbb{C}^{N_{\text{BE}}/2}$  contains only the odd distribution functions. Note that the potentials are contained in  $\underline{\mathbf{x}}_e$ . The transposed compression matrices  $C_e^t$  and  $(C_o^{\text{BE}})^t$  in Eq. (3.67) insert zeros where the distribution functions of odd and even Fourier harmonics should be, respectively, such that the dimensions add up.

Furthermore, it is convenient to express Eq. (3.65) in matrix form as

$$\underline{\mathbf{f}}_o^{\text{norm},C} = -\text{diag} \left( A_{o/o}^{\text{AC, BE}} \right)^{-1} \text{offdiag} \left( A_{o/e}^{\text{AC, BE}} \right) \underline{\mathbf{x}}_e^C, \quad (3.69)$$

where  $A_{o/o}^{\text{AC, BE}} \in \mathbb{C}^{N_{\text{BE}}/2 \times N_{\text{BE}}/2}$  only contains the rows and columns of odd harmonics of the BE and  $\text{diag}(A_{o/o}^{\text{AC, BE}})$  contains only the main diagonal elements

of  $A_{o/o}^{\text{AC, BE}}$ . Moreover,  $A_{o/e}^{\text{AC, BE}} \in \mathbb{C}^{N_{\text{BE}}/2 \times n}$  contains only the rows of odd harmonics of the BE while the columns contain the BE of even harmonics as well as the PE. Lastly,  $\text{offdiag}\left(A_{o/e}^{\text{AC, BE}}\right)$  contains only the off-diagonal elements of  $A_{o/e}^{\text{AC, BE}}$ .

Inserting Eqs. (3.67) and (3.69) into Eq. (3.66) yields

$$Y_{C',C} = \mathcal{P}_{I_{C'}}^t \left[ C_e^t - (C_o^{\text{BE}})^t \text{diag}\left(A_{o/o}^{\text{AC, BE}}\right)^{-1} \text{offdiag}\left(A_{o/e}^{\text{AC, BE}}\right) \right] \mathbf{x}_e^C - i\omega \mathcal{C}'_{C',C}.$$

Using Eq. (3.64) to formally replace  $\mathbf{x}_e^C = \left(A_{e/e}^{\text{AC}}\right)^{-1} \mathbf{b}_e^C$ , we find

$$\begin{aligned} Y_{C',C} &= \mathcal{P}_{I_{C'}}^t \left[ C_e^t - (C_o^{\text{BE}})^t \text{diag}\left(A_{o/o}^{\text{AC, BE}}\right)^{-1} \text{offdiag}\left(A_{o/e}^{\text{AC, BE}}\right) \right] \left(A_{e/e}^{\text{AC}}\right)^{-1} \mathbf{b}_e^C \\ &\quad - i\omega \mathcal{C}'_{C',C} \\ &= (\mathbf{b}_e^C)^t \left( \left(A_{e/e}^{\text{AC}}\right)^t \right)^{-1} \\ &\quad \times \left[ C_e - \text{offdiag}\left(A_{o/e}^{\text{AC, BE}}\right)^t \text{diag}\left(A_{o/o}^{\text{AC, BE}}\right)^{-1} C_o^{\text{BE}} \right] \mathcal{P}_{I_{C'}} - i\omega \mathcal{C}'_{C',C} \\ &=: (\mathbf{b}_e^C)^t \mathbf{y}_e^{C'} - i\omega \mathcal{C}'_{C',C}. \end{aligned} \quad (3.70)$$

Thus, we can determine  $\mathbf{y}_e^{C'}$  by solving

$$\left(A_{e/e}^{\text{AC}}\right)^t \mathbf{y}_e^{C'} = \left[ C_e - \text{offdiag}\left(A_{o/e}^{\text{AC, BE}}\right)^t \text{diag}\left(A_{o/o}^{\text{AC, BE}}\right)^{-1} C_o^{\text{BE}} \right] \mathcal{P}_{I_{C'}}, \quad (3.71)$$

which utilizes the same matrix as the initial system of equations of Eq. (3.64), albeit in transposed form.

Once we solved Eq. (3.71) for  $\mathbf{y}_e^{C'}$  with a sparse linear equation solver such as ILUPACK [68], we can determine the admittance parameters using Eq. (3.70).

### 3.6.3 Common-Source Configuration

MOSFETs are usually operated in the common-source configuration. The device of Fig. 2.1 has four contacts: a source (S), a drain (D), a top gate (TG), and a bottom gate (BG). In the common-source configuration, we ground the source and short the two gates. Then, we are left with two contacts, the drain (D) and the shorted gates (G) and any voltage differences are measured w.r.t. the source contact.

Previously in this chapter, we showed how to compute the full  $4 \times 4$  admittance matrix which satisfies

$$\begin{pmatrix} \underline{I}'_{TG} \\ \underline{I}'_{BG} \\ \underline{I}'_D \\ \underline{I}'_S \end{pmatrix} = \begin{pmatrix} Y'_{TG,TG} & Y'_{TG,BG} & Y'_{TG,D} & Y'_{TG,S} \\ Y'_{BG,TG} & Y'_{BG,BG} & Y'_{BG,D} & Y'_{BG,S} \\ Y'_{D,TG} & Y'_{D,BG} & Y'_{D,D} & Y'_{D,S} \\ Y'_{S,TG} & Y'_{S,BG} & Y'_{S,D} & Y'_{S,S} \end{pmatrix} \begin{pmatrix} \underline{V}'_{appl}{}^{TG} \\ \underline{V}'_{appl}{}^{BG} \\ \underline{V}'_{appl}{}^D \\ \underline{V}'_{appl}{}^S \end{pmatrix}. \quad (3.72)$$

The goal of this section is to see how the full admittance matrix relates to the admittance matrix in common source configuration.

Grounding the source contact means  $\underline{V}'_{appl}{}^S = 0$  and we are not interested in the current that flows through the source contact. Therefore, we may drop the columns and rows associated with the source contact. Moreover, the gates are shorted which means  $\underline{V}'_{appl}{}^{TG} = \underline{V}'_{appl}{}^{BG} =: \underline{V}'_{appl}{}^G$  and therefore we find

$$\begin{aligned} \begin{pmatrix} \underline{I}'_{TG} \\ \underline{I}'_{BG} \\ \underline{I}'_D \end{pmatrix} &= \begin{pmatrix} Y'_{TG,TG} & Y'_{TG,BG} & Y'_{TG,D} \\ Y'_{BG,TG} & Y'_{BG,BG} & Y'_{BG,D} \\ Y'_{D,TG} & Y'_{D,BG} & Y'_{D,D} \end{pmatrix} \begin{pmatrix} \underline{V}'_{appl}{}^G \\ \underline{V}'_{appl}{}^G \\ \underline{V}'_{appl}{}^D \end{pmatrix} \\ &= \begin{pmatrix} Y'_{TG,TG} + Y'_{TG,BG} & Y'_{TG,D} \\ Y'_{BG,TG} + Y'_{BG,BG} & Y'_{BG,D} \\ Y'_{D,TG} + Y'_{D,BG} & Y'_{D,D} \end{pmatrix} \begin{pmatrix} \underline{V}'_{appl}{}^G \\ \underline{V}'_{appl}{}^D \end{pmatrix}. \end{aligned}$$

Since we want to know the total current that runs through the gates, we can add the two rows referring to the top and bottom gates as  $\underline{I}'_G := \underline{I}'_{TG} + \underline{I}'_{BG}$  which yields

$$\begin{aligned} \begin{pmatrix} \underline{I}'_G \\ \underline{I}'_D \end{pmatrix} &= \begin{pmatrix} Y'_{TG,TG} + Y'_{TG,BG} + Y'_{BG,TG} + Y'_{BG,BG} & Y'_{TG,D} + Y'_{BG,D} \\ Y'_{D,TG} + Y'_{D,BG} & Y'_{D,D} \end{pmatrix} \begin{pmatrix} \underline{V}'_{appl}{}^G \\ \underline{V}'_{appl}{}^D \end{pmatrix} \\ &=: \begin{pmatrix} Y'_{GG} & Y'_{GD} \\ Y'_{DG} & Y'_{DD} \end{pmatrix} \begin{pmatrix} \underline{V}'_{appl}{}^G \\ \underline{V}'_{appl}{}^D \end{pmatrix}, \end{aligned}$$

where we defined the admittance parameters  $Y'_{GG}$ ,  $Y'_{GD}$ ,  $Y'_{DG}$ , and  $Y'_{DD}$  of the common-source configuration.



# Chapter 4

## Noise

The main result of this work is the calculation of noise in a nanoscale nMOSFET. In this chapter, we will introduce the Langevin-source approach which is the basis to formulate a system of equations for the Green's functions. The Green's functions will then be utilized to compute the power spectral density (PSD) of the terminal currents, which is a measure for the amount of noise.

The implementation of the noise calculation has some unexpected twists to it – similar to the small signal analysis of Chap. 3. However, with the previous chapter, we already laid the groundwork to properly deal with noise. The derivation of the terminal current via the Ramo-Shockley theorem only needs to be modified minimally in order to apply to the framework of Green's functions. Furthermore, the considerations of Sect. 3.5.2, where we needed to take great care during the discretization of the time-derivative of the Boltzmann equation (BE) as well as the density in the Poisson equation (PE), can be reused in this chapter.

Additionally, caution has to be exercised in dealing with degeneracy. While the degeneracies of the BE in Sect. 2.4.6 can be dealt with on an intuitive level by considering only the dynamics of one state in a set of degenerate states and multiplying the resulting observables by appropriate multiplicities, we would run into trouble if we were to continue this kind of reasoning for the Green's function equations. Therefore, we have to derive the multiplicities more rigorously and we will find that they appear in different and unexpected places.

As was already hinted at in Sect. 3.6.2 about the adjoint method for the admittance parameters, we are in dire need of the adjoint method for the computation of the Green's functions of noise. With unlimited resources, we could simply compute the Green's functions of fluctuations in the distribution functions. However, the size of our problem would make this approach computationally prohibitively expensive. Instead, we use the adjoint method to derive a system of equations that solves for the Green's functions of terminal currents

directly. This reduces the computational cost to a level comparable to the small signal analysis.

In fact, due to the similarity of the linear systems for the admittance parameters and the Green's functions of terminal currents, we can reduce the total effort of characterizing the device even more by using the results for the Green's functions to compute the admittance parameters.

The Green's functions of terminal currents are all we need to compute the PSDs of terminal currents. And from the PSDs and the admittance parameters, we can compute all noise related quantities in the linear response regime.

## 4.1 Introduction

In this chapter we will introduce the methods to calculate the noise for a coupled PE, Schrödinger equation (SE), and BE system by a Langevin-source approach. But let us first understand what noise actually is. When we measure any observables of an electronic device, we can see that the observable fluctuates around an average value. Say, we measure the observable  $X$  of an ergodic system, over some time period much longer than any intrinsic relaxation times, then we will find that it has some average value

$$\langle X(t) \rangle = \lim_{\Delta t \rightarrow \infty} \frac{1}{\Delta t} \int_{-\Delta t/2}^{\Delta t/2} dt X(t).$$

And therefore we can define the fluctuations as the miniscule deviations around the average value:

$$\delta X(t) = X(t) - \langle X \rangle.$$

Circuit designers are particularly interested in the fluctuations of terminal currents and voltages. But since it is possible to compute the fluctuations of voltages from the terminal current fluctuations and vice versa, it suffices to compute either of the two. We will be aiming at the computation of the terminal current fluctuations because it suits our approach better.

A typical example of terminal current fluctuations is shown in Fig. 4.1, where the terminal current is measured over a time period. Since the behavior of the current is a stochastic process, we need to describe it in terms of its non-random statistical characteristics, i.e. averages and correlation functions. The PSD is a generally useful quantity and sufficient to compute noise related quantities like the minimum noise figure, the gate and drain excess noise factors, and the cross-correlation coefficient. It can be derived with the Wiener-Khinchin theorem [106, 107] which states that the PSD is twice the Fourier transform of

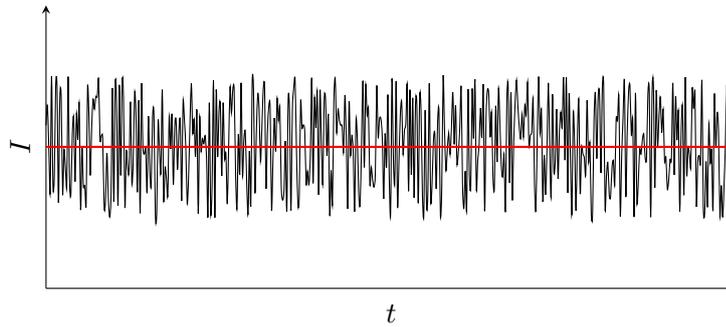


Figure 4.1: A terminal current (black line) fluctuating around an average value (red line).

the correlation function. Thus, if the correlation function of two macroscopic quantities  $X_\alpha$  and  $X_\beta$  is given by

$$\langle \delta X_\alpha(t_1) \delta X_\beta(t_2) \rangle = \lim_{\Delta t \rightarrow \infty} \frac{1}{\Delta t} \int_{-\Delta t/2}^{\Delta t/2} dt \delta X_\alpha(t_1 + t) \delta X_\beta(t_2 + t), \quad (4.1)$$

the PSD can be expressed as

$$P_{\alpha\beta}(\omega) = 2 \int_{-\infty}^{\infty} dt e^{i\omega t} \langle \delta X_\alpha(t) \delta X_\beta(0) \rangle, \quad (4.2)$$

where we used that  $\langle \delta X_\alpha(t_1) \delta X_\beta(t_2) \rangle = \langle \delta X_\alpha(t_1 - t_2) \delta X_\beta(0) \rangle$  is valid in a stationary system. When  $\alpha = \beta$ , Eq. (4.1) is often referred to as the *autocorrelation* function.

In order to understand what the PSD means, let us first try to make sense of the correlation function appearing on the r.h.s. of Eq. (4.2). Three possible cases are illustrated in Fig. 4.2, where we assumed  $t_0 > 0$  without loss of generality. If we have two random variables  $X_\alpha$  and  $X_\beta$ , which might be terminal currents or densities or even microscopic quantities like the distribution functions, they are considered correlated if a positive fluctuation in  $X_\beta$  at time  $t = 0$  – say a sudden increase above the average value – is followed by a similar positive fluctuation in  $X_\alpha$  at time  $t = t_0$  (Fig. 4.2A). In that case, the integral of Eq. (4.1) is positive and the correlation function itself becomes positive. Anticorrelation means that a positive fluctuation in  $X_\beta$  at time  $t = 0$  is usually followed by a negative fluctuation in  $X_\alpha$  at time  $t = t_0$ , or vice versa, in which case the correlation function becomes negative (Fig. 4.2B). The random variables are uncorrelated, i.e. the correlation function vanishes, if a fluctuation in  $X_\beta$  is usually not followed by a fluctuation in  $X_\alpha$  (Fig. 4.2C).

The PSD, as the Fourier transform of the correlation function, represents correlations occurring at certain frequencies. So if we have two positively correlated functions with a time of about  $t_0$  between the signals as in Fig. 4.2A, we will find that the PSD of Eq. (4.2) is positive at a frequency of about  $\omega = 2\pi/t_0$ .

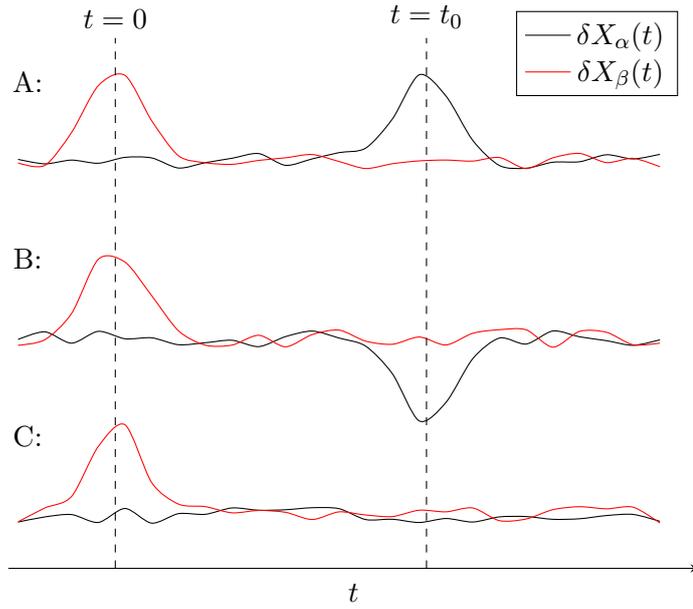


Figure 4.2: The correlation function  $\langle \delta X_\alpha(t_0) \delta X_\beta(0) \rangle$  of two random variables can be positive, negative, or zero when a typical response of a signal at  $t = 0$  looks as in A, B, or C, respectively.

Without going into too much detail, we want to mention some properties of the PSD of Eq. (4.2). An obvious property of the autocorrelation function is that it is symmetric under time-reversal in stationary conditions, i.e.

$$\langle \delta X_\alpha(t) \delta X_\alpha(0) \rangle = \langle \delta X_\alpha(-t) \delta X_\alpha(0) \rangle.$$

It follows that the PSD of Eq. (4.2) is strictly real and positive semidefinite,

$$P_{\alpha\alpha}(\omega) \geq 0. \quad (4.3)$$

A cross-correlation between some observables  $\alpha$  and  $\beta$  can give rise to real and imaginary parts of the PSD, but only in the case where the cross-correlation is totally antisymmetric as

$$\langle \delta X_\alpha(t) \delta X_\beta(0) \rangle = -\langle \delta X_\alpha(-t) \delta X_\beta(0) \rangle,$$

will we find a purely imaginary PSD.

In electronic devices, we are usually interested in the PSDs of the contact currents, e.g. the PSD of the drain or gate terminal currents of a MOSFET or the PSD of their cross-correlation. Thus, what determines the amount of noise we measure in a contact is governed by the random processes carriers experience within the device which only consist of scattering processes as well as generation

and recombination (GR) processes. Due to historical reasons, there exist many names for noise but they all arise microscopically from scattering or GR [51].

There are three main historical types of noise we want to introduce since we will refer to them later on. The first one is Johnson-Nyquist noise [108, 109], which is also known as *thermal noise*. Scattering in an electronic device can be macroscopically summarized as a conductivity. Near equilibrium we find that the PSD of the current  $I$  through some device is proportional to the conductivity and temperature as

$$P_{II} = 4k_B T \operatorname{Re}(Y), \quad (4.4)$$

where  $Y$  is the self-admittance of the contact the current  $I$  is flowing through. Carriers move around the device even if the net current vanishes. Their scattering with the surrounding lattice of atoms is dependent on the temperature since lattice vibrations increase with temperature. Every device exhibits thermal noise in equilibrium. Therefore thermal noise is basically the noise floor at a given temperature which cannot be avoided. It is often the predominant noise deteriorating the signal-to-noise ratio in analog circuits.

The second type is *shot noise* and it is the noise occurring when carriers surpass a potential barrier [51, 110, 111]. The idea behind shot noise is that current actually consists of individual carriers which need to surpass a potential barrier one by one. Such a process – also known as a Poisson process – leads to fluctuations and it can be derived that the PSD is given by

$$P_{II} = 2qI, \quad (4.5)$$

where  $I$  is the current flowing through the device. Shot noise is often drowned by thermal noise but it can also become significant, especially when the current is small such as in MOSFETs in the subthreshold. Equation (4.5) suggests that shot noise is white, but in real devices it actually starts decreasing at very high frequencies, similar to thermal noise.

The third and last type of noise we want to touch on is  $1/f$ -noise or *flicker noise*. As the name suggests, it is inversely proportional to the frequency and therefore it becomes the dominant noise source in the low frequency domain. Due to its ubiquity, a comprehensive explanation has remained elusive [51] and we will disregard it in this work. However, we want to note that we are going to compute frequency ranges which would typically involve flicker noise. We do this in order to reasonably compare different devices (cf. Sect. 5.4) but we urge the reader to keep in mind that real devices would exhibit flicker noise.

## 4.2 Langevin-Source Approach

We will use the Langevin-source approach (see e.g. [51]) to determine the noise in our device. The Langevin-source approach is based on the idea that we introduce

small noise sources into our equations. These Langevin-source terms represent small fluctuations that modulate the distribution functions and potential as

$$\begin{aligned} f^\nu(y, \mathbf{k}) &\longrightarrow f^\nu(y, \mathbf{k}) + \delta f^\nu(y, \mathbf{k}, t), \\ V(\mathbf{r}) &\longrightarrow V(\mathbf{r}) + \delta V(\mathbf{r}, t), \end{aligned}$$

where  $f$  and  $V$  are the solutions to the stationary problem which are assumed to be known quantities. Analogously to the linearization in the small signal analysis of Sect. 3.1, we find the linearized BE for the fluctuations in the distribution function to be given by its variation

$$\begin{aligned} [\delta F^{\text{BE}}]^\nu(y, \mathbf{k}, t) &= \frac{\partial}{\partial t} \delta f^\nu(y, \mathbf{k}, t) + \frac{1}{\hbar} \delta F^\nu(y, t) \frac{\partial}{\partial k_y} f(y, \mathbf{k}) \\ &\quad + \frac{1}{\hbar} F^\nu(y) \frac{\partial}{\partial k_y} \delta f(y, \mathbf{k}, t) + v^\nu(\mathbf{k}) \frac{\partial}{\partial y} \delta f(y, \mathbf{k}, t) \\ &\quad - \delta S^\nu(y, \mathbf{k}, t) - \delta \Gamma^\nu(y, \mathbf{k}, t) \\ &= \xi_{\text{BE}}^\nu(y, \mathbf{k}, t), \end{aligned} \tag{4.6}$$

where  $\xi_{\text{BE}}$  is the Langevin-source term in the BE. Since the Langevin-source elicits a small change in the distribution function, we will often refer to it as a fluctuation in the distribution function. The variation of the force is given by

$$\delta F^\nu(y, t) = -\frac{\partial}{\partial y} \delta \varepsilon^\nu(y, t) \tag{4.7}$$

and  $\delta \varepsilon^\nu(y, t)$  is assumed to be the quasistationary effect of a variation of the potential on the subband energy and therefore it is given by Eq. (2.10) even in the time-dependent case. Regarding the scattering term as a functional of the distribution function, subband energies, and wave functions, we can use the functional derivative to obtain

$$\begin{aligned} \delta S^\nu(y, \mathbf{k}, t) &= \sum_{\nu'} \int dy' \int \frac{d^2 k'}{(2\pi)^2} \frac{\delta S^\nu(y, \mathbf{k})}{\delta f^{\nu'}(y', \mathbf{k}')} \delta f^{\nu'}(y', \mathbf{k}', t) \\ &\quad + \sum_{\nu'} \int dy' \frac{\delta S^\nu(y, \mathbf{k})}{\delta \varepsilon^{\nu'}(y')} \delta \varepsilon^{\nu'}(y', t) \\ &\quad + \sum_{\nu'} \int dx' \int dy' \frac{\delta S^\nu(y, \mathbf{k})}{\delta \Psi^{\nu'}(\mathbf{r}, t)} \delta \Psi^{\nu'}(\mathbf{r}', t). \end{aligned}$$

Note that the functional derivative w.r.t. a function comprises the respective integrals and sums over all variables of the function. In the continuum, we would find that the integral w.r.t.  $y'$  drops out since all scattering terms are local. However, in the discretized case we need to use an averaging scheme for

the scattering term on the adjoint grid points (see Eq. (2.82)) and therefore multiple grid points contribute. For this reason, we will leave the integral over  $y'$  unevaluated.

Assuming a quasistationary perturbation of the wave function, we can use Eq. (2.11) for the variation of the wave function and therefore we get

$$\begin{aligned} \delta S^\nu(y, \mathbf{k}, t) &= \sum_{\nu'} \int dy' \int \frac{d^2 k'}{(2\pi)^2} \frac{\delta S^\nu(y, \mathbf{k})}{\delta f^\nu(y', \mathbf{k}')} \delta f^\nu(y', \mathbf{k}', t) \\ &\quad + \int dx' \int dy' \frac{\delta S^\nu(y, \mathbf{k})}{\delta V(\mathbf{r})} \delta V(\mathbf{r}', t). \end{aligned}$$

Finally, the variation of the GR rate can be treated similarly to find

$$\delta \Gamma^\nu(y, \mathbf{k}, t) = \sum_{\nu'} \int dy' \int \frac{d^2 k'}{(2\pi)^2} \frac{\delta \Gamma^\nu(y, \mathbf{k})}{\delta f^\nu(y', \mathbf{k}')} \delta f^\nu(y', \mathbf{k}', t).$$

Equation (4.6) is often referred to as the Langevin-Boltzmann equation (LBE).

Likewise, the PE reads

$$\delta F^{\text{PE}}(\mathbf{r}) = \nabla \cdot (\kappa(\mathbf{r}) \nabla \delta V(\mathbf{r}, t)) - q \delta n(\mathbf{r}, t) = \xi_{\text{PE}}(\mathbf{r}, t), \quad (4.8)$$

where  $\xi_{\text{PE}}$  is the Langevin-source term of the PE and the variation of the density is given by

$$\delta n(\mathbf{r}, t) = \sum_{\nu} \int \frac{d^2 k}{(2\pi)^2} [\delta f^\nu(y, \mathbf{k}, t) |\Psi^\nu(\mathbf{r})|^2 + 2f^\nu(y, \mathbf{k}) \Psi^\nu(\mathbf{r}) \delta \Psi^\nu(\mathbf{r}, t)] \quad (4.9)$$

with a real wave function. Since the Langevin-source in the PE causes a small change in the potential, we will refer to it as a fluctuation in the potential.

The Langevin-source terms  $\xi_{\text{BE}}$  and  $\xi_{\text{PE}}$  need not necessarily be known. In particular, we will see that since we are only interested in the PSD of the terminal currents, all we need to know is the PSD of these Langevin-sources. These are nothing else but the microscopic origin of the fluctuations which is given by the scattering processes. The details for the computation of the PSDs are given in Sect. 4.6.

In order to solve Eqs. (4.6) and (4.8), we are going to use the Green's function approach. The Green's function equation of the LBE can be obtained by replacing the variations of functions by their corresponding transfer functions and replacing the Langevin-source term by an appropriate delta-distribution as well as Fourier transforming the equation into the frequency domain:

$$i\omega (G^f)^{\nu, \nu'}(y, \mathbf{k}; y', \mathbf{k}') + \frac{1}{\hbar} (G^F)^{\nu, \nu'}(y; y', \mathbf{k}') \frac{\partial}{\partial k_y} f(y, \mathbf{k})$$

$$\begin{aligned}
& + \frac{1}{\hbar} F^\nu(y) \frac{\partial}{\partial k_y} (G^f)^{\nu,\nu'}(y, \mathbf{k}; y', \mathbf{k}') + v^\nu(\mathbf{k}) \frac{\partial}{\partial y} (G^f)^{\nu,\nu'}(y, \mathbf{k}; y', \mathbf{k}') \\
& - (G^S)^{\nu,\nu'}(y, \mathbf{k}; y', \mathbf{k}') - (G^\Gamma)^{\nu,\nu'}(y, \mathbf{k}; y', \mathbf{k}') = \delta_{\nu,\nu'} (2\pi)^2 \delta(\mathbf{k} - \mathbf{k}') \delta(y - y').
\end{aligned} \tag{4.10}$$

where  $G^f$  is the Green's function of the distribution function,  $G^F$  is the Green's function of the force, and so on. Note that the above equation describes Green's functions only at the frequency  $\omega$ , however, we omitted the frequency index to keep the notation less burdensome. Equation 4.10 is analogous to the Green's function equation of the LBE found in Ref. [60]. But in our case we only consider a 2D  $\mathbf{k}$ -space and our  $z$ -direction of the spatial coordinates is homogeneous, which means we can integrate it out. Thus, the Green's function of the distribution function has units [s m] where the seconds stem from the Fourier transform and meters due to the integration over some length  $\Delta z$  in the homogeneous  $z$ -direction. The meaning of  $G^f$  is therefore the cumulative response in an interval  $\Delta z$  to the appearance of a carrier.

We can further deconstruct the Green's functions until we arrive at an equation described solely by the Green's functions of the distribution function and potential. Take, for example, the Green's function of the force. It can be expressed using Eq. (4.7) as

$$(G^F)^{\nu,\nu'}(y; y', \mathbf{k}') = -\frac{\partial}{\partial y} (G^\varepsilon)^{\nu,\nu'}(y; y', \mathbf{k}'),$$

with the Green's function of the subband energy  $G^\varepsilon$ , which in turn can be expressed using Eq. (2.10) as

$$(G^\varepsilon)^{\nu,\nu'}(y; y', \mathbf{k}') = -q \int dx |\Psi^\nu(x)|^2 (G^V)^{\nu'}(\mathbf{r}; y', \mathbf{k}').$$

Note how the Green's function equations can be derived by simply following along the variations of functionals.

Green's functions quantify the response of a system. Let there be a miniscule fluctuation in the distribution function in the valley and subband  $\nu'$  at position  $y'$  and wave vector  $\mathbf{k}'$ . Then, the Green's function of the distribution function  $(G^f)^{\nu,\nu'}(y, \mathbf{k}; y', \mathbf{k}')$  will tell us the response of the distribution function in the valley and subband  $\nu$  at position  $y$  and wave vector  $\mathbf{k}$ . If, for example,  $(\nu, y, \mathbf{k}) = (\nu', y', \mathbf{k}')$ , the distribution function is given exactly by the fluctuation and therefore the Green's function is given by the divergence of delta-distributions w.r.t. each of the coordinates. Further away from the fluctuation, we usually also see a weaker response to the fluctuation and thus Green's functions become smaller.

Other Green's functions behave similarly. The Green's function of the potential  $(G^V)^{\nu'}(\mathbf{r}; y', \mathbf{k}')$  will tell us the response of the electric potential  $V$  at the position  $\mathbf{r}$  if there is a fluctuation in the distribution function at  $(\nu, y', \mathbf{k})$ . Moreover, the Green's function of the force  $G^F$  will tell us the response of the force to fluctuations in the distribution function, the Green's function of the scattering rate  $G^S$  will tell us the response of the scattering rate, and so on.

The Green's function equation for the PE can be obtained analogously as

$$\delta(\mathbf{r} - \mathbf{r}') = \nabla_{\mathbf{r}} \cdot (\kappa(\mathbf{r}) \nabla_{\mathbf{r}} G^V(\mathbf{r}; \mathbf{r}')) - q \sum_{\nu} (G^n)^{\nu}(y; \mathbf{r}') |\Psi^{\nu}(\mathbf{r})|^2 \quad (4.11)$$

$$- 2q \sum_{\nu} n^{\nu}(y) \Psi^{\nu}(\mathbf{r}) (G^{\Psi})^{\nu}(\mathbf{r}; \mathbf{r}'). \quad (4.12)$$

Note that the Green's functions shown here are different as they describe how quantities respond to a Langevin-source in the PE at position  $\mathbf{r}'$ . However, what we are actually interested in is the response to fluctuations in the distribution function. Recall that the PE is coupled through the density to the BE, therefore the PE of Green's functions is coupled to the LBE of Green's functions via the Green's function of distribution functions. This means the density responds to fluctuations at  $(\nu', y', \mathbf{k}')$  in the distribution function as in Eq. (4.9) and thus  $(G^n)^{\nu'}(\mathbf{r}; y', \mathbf{k}')$  must obey the PE as

$$0 = \nabla_{\mathbf{r}} \cdot (\kappa(\mathbf{r}) \nabla_{\mathbf{r}} (G^V)^{\nu'}(\mathbf{r}; y', \mathbf{k}')) - q \sum_{\nu} (G^n)^{\nu, \nu'}(y; y', \mathbf{k}') |\Psi^{\nu}(\mathbf{r})|^2 \\ - 2q \sum_{\nu} n^{\nu}(y) \Psi^{\nu}(\mathbf{r}) (G^{\Psi})^{\nu, \nu'}(\mathbf{r}; y, \mathbf{k}'),$$

where there is no Langevin-source in the PE, since it is contained in the LBE. Conversely, the PE influences the LBE which means a PE with a Langevin-source will elicit responses of the distribution functions. Hence, we also have a BE of Green's functions where the Langevin-sources are located in the PE.

Here, the Green's function of the wave function is given by perturbation theory using Eq. (2.11) and the Green's function of the density for either a Langevin-source in the LBE or in the PE:

$$(G^n)^{\nu}(y; \cdot) = \int \frac{d^2 k}{(2\pi)^2} (G^f)^{\nu}(y, \mathbf{k}; \cdot),$$

where  $\cdot$  denotes either the coordinates  $(\nu', y', \mathbf{k}')$  of a Langevin-source in the LBE or the coordinates  $\mathbf{r}'$  of a Langevin-source in the PE. It is important to keep in mind that the Green's functions represent the response of the system to an insertion of a carrier into a single state. Therefore, the Green's function of the density needs to acknowledge that this state may, for example, either be a

spin up or a spin down state, but never both at the same time. It follows that the Green's function of the density – and also of all other quantities – does not contain multiplicities in the case of degeneracy. We will postpone the detailed discussion and possible simplifications of the LBE and PE system in the case of degeneracy to Sect. 4.7.

Now let us set up a system of equations that can be solved in order to find the Green's functions  $G^f$  and  $G^V$ . To this end, we need to use the Herring-Vogt transformation, the projection onto equi-energies and Fourier harmonics, the  $H$ -transformation, and the discretization (see Sect. 2.4.4). Bear in mind that the  $H$ -transformation and the subsequent box-integration introduces additional fluctuating quantities, since the lowest  $H$ -boxes depend on the subband energy. Therefore, we have to perform the aforementioned transformations and the box-integration before we compute the variation of the LBE. However, from the derivation of the PE and LBE Green's function equations, it is obvious our problem is analogous to the linearization around a stationary state described in Sect. 3.1 for the small signal analysis. Thus, we can treat the first set of variables of the Green's functions – the unprimed ones – completely identical to the way we treated the small signal BE and PE of Eqs. (3.12) and (3.15). The second set of variables of the Green's functions can also be Herring-Vogt transformed, projected onto equienergy surfaces and Fourier harmonics,  $H$ -transformed, and discretized. But this amounts to trivially replacing  $(y', \mathbf{k}') \rightarrow (y_{i'}, H_{j'}, m')$  and  $\mathbf{r}' \rightarrow (x_{k'}, y_{i'})$  in all the Green's functions.

Thus, reusing the derivations on the time-derivative of Sect. 3.2 and the restoration of reciprocity of Sect. 3.5.2, we obtain the system of equations for the Green's functions analogous to Eqs. (3.12) and (3.15):

$$G_{\alpha\gamma}^T + \sum_{\beta} \frac{\partial F_{\alpha}^{\text{BE}}}{\partial f_{\beta}} G_{\beta\gamma}^f + \sum_b \left( \sum_k \frac{\partial F_{\alpha}^{\text{BE}}}{\partial \varepsilon_k} \frac{\partial \varepsilon_k}{\partial V_b} + \sum_{\ell} \frac{\partial F_{\alpha}^{\text{BE}}}{\partial \Psi_{\ell}} \frac{\partial \Psi_{\ell}}{\partial V_b} \right) G_{b\gamma}^V = \delta_{\alpha\gamma}, \quad (4.13)$$

$$\sum_{\beta} \frac{\partial F_a^{\text{PE}}}{\partial f_{\beta}} G_{\beta\gamma}^f + \sum_b \frac{\partial F_a^{\text{PE}}}{\partial V_b} G_{b\gamma}^V = \delta_{a\gamma}, \quad (4.14)$$

where  $G^T$  is the Green's function of the time-derivative defined in analogy to the time-derivative in the small signal case of Sect. 3.2, complete with the reciprocity conserving discretization of Sect. 3.5.2, where the density in the PE is defined analogously to Eq. (3.59). The indices  $\alpha$  and  $\beta$  are defined as in Eq. (2.86) as aggregate indices of the BE and the indices  $a$  and  $b$  are defined as in Eq. (2.105) as aggregate indices of the PE. The index  $\gamma$  designates the location of Langevin-sources in a Green's function and since we've seen that there can be Langevin-sources in both the LBE and the PE, the index must run over both the BE

indices and the PE indices as

$$\gamma = \begin{cases} \alpha', & \text{for } \gamma \leq N_{\text{BE}}, \\ a', & \text{for } \gamma > N_{\text{BE}}, \end{cases} \quad (4.15)$$

where  $\alpha'$  denotes an index like  $\alpha$  running over all variables of the BE and  $a'$  denotes an index like  $a$  running over all variables of the PE. Hence, the Kronecker-delta on the r.h.s. of Eq. (4.13) can only yield unity when  $\gamma \leq N_{\text{BE}}$ , i.e. when the Langevin-source is in the LBE. Conversely, the Kronecker-delta on the r.h.s. of Eq. (4.14) can only yield unity when the Langevin-source is in the PE.

For later convenience, we want to represent Eqs. (4.13) and (4.14) in matrix form as

$$A^{\text{AC}} G = B, \quad (4.16)$$

where  $G \in \mathbb{C}^{N \times N}$  is the matrix containing all Green's functions,  $B \in \mathbb{R}^{N \times N}$  is the matrix containing the Kronecker-deltas of the Langevin-sources, and  $A^{\text{AC}} \in \mathbb{C}^{N \times N}$  is the *same* matrix as in the small signal case of Eq. (3.62). Note that Eq. (4.16) also contains the boundary conditions discussed in the following section and therefore  $B$  is not a unit matrix – although it is quite close to being one.

Note that the system of Green's function equations of Eq. (4.16) needs to be set up in an identical way to the small signal case. If we did not use  $A^{\text{AC}}$  in the Green's function equation but forego the restoration procedure for reciprocity of Sect. 3.5.2, we would not be able to verify the consistency of the noise and small signal analyses by the Nyquist-theorem as will be shown in Sect. 5.4.2.

In theory, we could solve Eq. (4.16) directly and compute the Green's functions of all observables from  $G^f$  and  $G^V$ . But this approach is computationally too expensive since the number of variables  $G$  in the discretized system of the Green's functions is the number of distribution functions and potentials squared. It is also wildly inefficient when all we are interested in are the Green's functions of terminal current fluctuations to compute the PSD of terminal currents (cf. Sect. 4.6). Therefore, we will use the adjoint method discussed in Sect. 4.5, in order to solve for the Green's functions of terminal currents directly. This method is quite similar to the adjoint method of the small signal analysis of Sect. 3.6.2. However, before that, we need to discuss the boundary conditions of the Green's function system and how the Green's functions of terminal current fluctuations can be computed via the Ramo-Shockley theorem.

### 4.3 Boundary Conditions

The boundary conditions of Green's function equations are chosen such that there cannot be a response to Langevin-sources where Dirichlet boundary conditions apply. For our case, this means that we have to set the Green's functions

in the PE to zero where Dirichlet boundary conditions apply, i.e. on the top and bottom gates.

Recall that we used Eq. (3.29) to set the top and bottom gate small signal biases. Analogously, we can use

$$(G^V)^{\nu'}(\mathbf{r}; y', \mathbf{k}') \Big|_{\mathbf{r} \in \partial D_{\text{TG/BG}}} = 0, \quad (4.17)$$

$$G^V(\mathbf{r}; \mathbf{r}') \Big|_{\mathbf{r} \in \partial D_{\text{TG/BG}}} = 0, \quad (4.18)$$

which translates to the discretized equations as

$$G_{a\gamma}^V \Big|_{a=(x_k, y_i) \in \partial D_{\text{TG/BG}}} = 0. \quad (4.19)$$

The source and drain contacts use Neumann boundary conditions in the PE which are automatically applied by the finite volume discretization approach.

The LBE, just as the BE also only uses Neumann boundary conditions which are automatically applied by the box-integration method. The GR term  $\Gamma$  in the small signal case contains a term proportional to  $\underline{V}_{\text{appl}}^C$  that goes to the r.h.s. of Eq. (3.12). However, since no small signal bias is applied in the noise calculation, this term does not exist.

Therefore, only the rows of coordinates on the top and bottom gates of the PE in Eq. (4.16) need to be replaced to satisfy Eq. (4.19). But this has also been done in the small signal case and therefore  $A^{\text{AC}}$  for the Green's function equation is truly identical to the small signal equation. On the r.h.s.,  $B$  is essentially a unit matrix due to the Langevin-sources on the main-diagonal, except for where the Dirichlet boundary conditions of the PE apply.

In the remainder of this work, we will always assume that the boundary conditions are contained in the Green's function system of Eq. (4.16).

## 4.4 Ramo-Shockley Theorem

Once again we aim to use the Ramo-Shockley theorem to express the terminal current as a volume integration over the device. In Sect. 3.4 we already showed how this can be achieved for the small signal current. In this section, we apply the same ideas to obtain the corresponding expression for the Green's functions of terminal currents.

Let us start out from the Green's function expression corresponding to Eq. (3.33) for the terminal current in the frequency domain:

$$G_{\zeta}^{I_C} = - \int_{\partial D_C} d\mathbf{A} \cdot \left[ q \sum_{\nu} \mathbf{G}_{\zeta}^{\mathbf{J}^{\nu}}(\mathbf{r}) + i\omega \kappa(\mathbf{r}) \nabla_{\mathbf{r}} G_{\zeta}^V(\mathbf{r}) \right]. \quad (4.20)$$

While the first set of coordinates of the Green's functions remain explicit, we abbreviated the second set of coordinates with the index  $\zeta \in \{(\nu', y', \mathbf{k}')\} \cup \{\mathbf{r}'\}$ . The Green's function of the terminal current  $G_\zeta^{IC}$  is the response of the terminal current  $I_C$  caused by a fluctuation in the distribution function or a fluctuation in the potential, depending on the value of  $\zeta$ , which runs over both BE and PE indices.

The Green's function of the 2D conduction current is given analogous to Eq. (3.34) as

$$\mathbf{G}_\zeta^{\mathbf{J}^\nu}(\mathbf{r}) = G_\zeta^{J_y^\nu}(\mathbf{r})\mathbf{e}_y + \mathbf{G}_\zeta^{J_\perp^\nu}(\mathbf{r}),$$

where the  $y$ -component can be derived in analogy to Eq. (3.35) as

$$G_\zeta^{J_y^\nu}(\mathbf{r}) = G_\zeta^{j^\nu}(y)|\Psi^\nu(\mathbf{r})|^2 + 2j^\nu(y)\Psi^\nu(\mathbf{r})G_\zeta^{\Psi^\nu}(\mathbf{r}) \quad (4.21)$$

with the Green's function of the wave function given similar to Eq. (3.9) by perturbation theory as

$$G_\zeta^{\Psi^\nu}(x, y) = \int dx' \Delta_{\Psi}^\nu(x, x', y)G_\zeta^V(x', y).$$

Since the Green's function equations behave exactly like the linearization around the stationary state for the small signal terminal current, we will obtain an analogous result, except for one distinction: The Green's functions obey a different continuity equation due to the presence of the Langevin-source. Integrating Eq. (4.10) over  $\mathbf{k}$ -space, we obtain

$$\frac{\partial}{\partial y} G_\zeta^{j^\nu}(y) + i\omega G_\zeta^{m^\nu}(y) - G_\zeta^{S^\nu}(y) - G_\zeta^{\Gamma^\nu}(y) = \delta_{\nu, \nu'} \delta(y - y'),$$

where the primed coordinate  $y'$  belongs to the BE part of the second set of coordinate  $\zeta$ . Bear in mind that the Langevin-source in the BE only exists as long as  $\zeta$  runs over BE indices.

The Langevin-source impacts the derivation of the terminal current twofold. First, since it is singular, it contributes to the current generated in the source and drain contacts as described by Eq. (3.39). Here, we find in the case of the Green's functions

$$\begin{aligned} & \int_{\partial D} d\mathbf{A} \cdot \left( h_C(\mathbf{r}) G_\zeta^{J_y^\nu}(\mathbf{r}) \mathbf{e}_y \right) \\ &= \int_{\partial D} d\mathbf{A} \cdot \mathbf{e}_y h_C(\mathbf{r}) \left( G_\zeta^{j^\nu}(y) |\Psi^\nu(\mathbf{r})|^2 + 2j^\nu(y) \Psi^\nu(\mathbf{r}) G_\zeta^{\Psi^\nu}(\mathbf{r}) \right) \\ &= - \int dV h_C(\mathbf{r}) \left\{ \left( G_\zeta^{\Gamma^\nu}(y) \right. \right. \end{aligned}$$

$$\begin{aligned}
& + \delta_{\nu,\nu'} \left[ \delta(y - y_S) \Delta(y', y_S) + \delta(y - y_D) \Delta(y', y_D) \right] |\Psi^\nu(\mathbf{r})|^2 \\
& + 2\Gamma^\nu(y) \Psi^\nu(\mathbf{r}) G_\zeta^{\Psi^\nu}(\mathbf{r}) \Big\},
\end{aligned}$$

where the integration over a Langevin-source is given analogously to the integration over the GR term in Eqs. (2.39) and (2.40) as

$$\Delta(y', y_{S/D}) = \lim_{\ell \rightarrow 0} \int_{y_{S/D} - \ell/2}^{y_{S/D} + \ell/2} d\tilde{y} \delta(\tilde{y} - y'). \quad (4.22)$$

Note that  $\Delta(y', y_{S/D})$  vanishes unless  $y' = y_{S/D}$ , in which case it is equal to unity. Usually we can simply ignore such terms, however, when we discretize, we will find Langevin-source terms located precisely on the source and drain contacts and therefore we need to include their impact on the generation and recombination of current at the contacts. Hence, we need to keep track of the Langevin-sources using  $\Delta(y', y_{S/D})$ .

Second, when replacing the boundary term  $G_\zeta^{\Gamma^\nu}$  using the continuity equation – analogous to the step from Eq. (3.43) to Eq. (3.44) in the small signal case – we will obtain an additional Langevin-source term. Therefore the Green's function of the terminal current is given by

$$\begin{aligned}
G_\zeta^{IC} = q \int_D dV h_C(\mathbf{r}) \sum_\nu \left[ \left( \frac{\partial}{\partial y} G_\zeta^{j^\nu}(y) - G_\zeta^{S^\nu}(y) \right. \right. \\
\left. \left. - \delta_{\nu,\nu'} \left[ \delta(y - y') - \delta(y - y_S) \Delta(y', y_S) - \delta(y - y_D) \Delta(y', y_D) \right] \right) |\Psi^\nu(\mathbf{r})|^2 \right. \\
\left. - 2i\omega n^\nu(y) \Psi^\nu(\mathbf{r}) G_\zeta^{\Psi^\nu}(\mathbf{r}) \right]. \quad (4.23)
\end{aligned}$$

After integrating the  $z$ -direction out and after the usual transformations and the discretization, we find for the terminal current expression where the Langevin-source is in the BE:

$$\begin{aligned}
G_\alpha^{IC} = q \sum_{k,i,\nu} \Delta x_k h_C(x_k, y_i) \left[ \left( G_\alpha^{j^\nu}(y_{i+}) - G_\alpha^{j^\nu}(y_{i-}) - G_\alpha^{S^\nu}(y_i) \Delta y_i \right. \right. \\
\left. \left. - \frac{1}{Y_0} \delta_{m',0} \delta_{\nu,\nu'} \left[ \delta_{y_i, y_{i'}} - \delta_{y_i, y_S} \delta_{y_{i'}, y_S} - \delta_{y_i, y_D} \delta_{y_{i'}, y_D} \right] \right) |\Psi^\nu(x_k, y_i)|^2 \right. \\
\left. - 2i\omega n^\nu(y_i) \Psi^\nu(x_k, y_i) G_\alpha^{\Psi^\nu}(x_k, y_i) \Delta y_i \right] \quad (4.24)
\end{aligned}$$

with  $\alpha \in \{(\nu', y_{i'}, H_{j'}, m')\}$  being the aggregate index of the BE. Here, the Kronecker-delta  $\delta_{y_i, y_{i'}}$  is unity when  $i = i'$  while the Kronecker-deltas  $\delta_{y_i, y_{S/D}}$  are unity when the index  $i$  describes a  $y_i$  on the source and drain contact, respectively, and likewise for  $\delta_{y_{i'}, y_{S/D}}$ .

If, on the other hand, the Langevin-source is in the PE, we need to include it when we use the PE of the Green's functions analogously to Eq. (3.40). After the discretization we arrive at

$$G_a^{I'C} = q \sum_{k,i} h_C(x_k, y_i) \times \left[ \sum_{\nu} (G_a^{j\nu}(y_{i+}) - G_a^{j\nu}(y_{i-}) - G_a^{S\nu}(y_i) \Delta y_i) |\Psi^{\nu}(x_k, y_i)|^2 \Delta x_k - i\omega \delta_{x_k, x_{k'}} \delta_{y_i, y_{i'}} - 2i\omega \sum_{\nu} n^{\nu}(y_i) \Psi^{\nu}(x_k, y_i) G_a^{\Psi\nu}(x_k, y_i) \Delta x_k \Delta y_i \right] \quad (4.25)$$

with  $a \in \{x_{k'}, y_{i'}\}$ .

In order to use the above Green's functions of terminal currents later on, we will cast them in matrix form. To this end, note that Eqs. (4.24) and (4.25) have only terms which are either constant or linear in either of  $G^f$  or  $G^V$ . Therefore, we may write the discretized Green's function of the terminal current as

$$(\mathbf{G}^{I'C})^t := \begin{pmatrix} (G_{\alpha}^{I'C}) \\ (G_a^{I'C}) \end{pmatrix}^t = \mathcal{P}_{I'C}^t G + \mathbf{\Xi}^t, \quad (4.26)$$

where  $\mathcal{P}_{I'C} \in \mathbb{C}^N$  is the same current projector as in Eq. (3.66)<sup>1</sup> and the constant terms comprise the contributions of Langevin-sources which are given by

$$\mathbf{\Xi} = \begin{pmatrix} \Xi_{\alpha} \\ \Xi_a \end{pmatrix},$$

with

$$\Xi_{\alpha=(\nu', y_{i'}, H_{j'}, m')} = -\frac{q}{Y_0} \sum_k \Delta x_k h_C(x_k, y_{i'}) \times |\Psi^{\nu'}(x_k, y_{i'})|^2 \delta_{m', 0} [1 - \delta_{y_{i'}, y_S} - \delta_{y_{i'}, y_D}] \quad (4.27)$$

<sup>1</sup>Note that the projector for the Green's function of the terminal current is different from the small signal case if we consider degeneracy. In that case, we will find that the projector for the small signal terminal current of Eq. (3.66) contains multiplicities for the spin and valley. In contrast, the Green's functions are responses to a fluctuation in a *single true state* and therefore they do not contain multiplicities. This will be discussed in more detail in Sect. 4.7.

if the Langevin-source is in the BE and

$$\Xi_{a=(x_{k'}, y_{l'})} = -i\omega q h_C(x_{k'}, y_{l'})$$

if the Langevin-source is in the PE. Note that the Langevin-sources in Eq. (4.27) do not contribute on the source and drain contacts which follows from keeping track of the Langevin-source as a GR term on the contact as described by Eq. (4.22).

## 4.5 Adjoint Method

As was already mentioned in this chapter, solving Eq. (4.16) for the Green's functions of the distribution functions and the potential is computationally infeasible since we would have to solve a linear equation that has  $N$  right hand sides, compared to just four in the small signal case of Eq. (3.62).

But if all we are interested in are the Green's functions of the terminal current fluctuations, we need not know the Green's functions of the distribution function and the potential fluctuations. For this reason, we aim to transform Eq. (4.16) in such a way that we can compute the Green's functions of terminal currents of Eq. (4.26) directly, similar to how the adjoint method for the admittance parameters of Sect. 3.6.2 works.

In analogy to Sect. 3.6.2, we want to remove the odd distribution functions from the LBE. Using the linear transformation matrix  $S \in \mathbb{C}^{N \times N}$  of Eq. (2.109) to remove the odd distribution functions in the even equations of the BE and the compression matrix  $C_e \in \mathbb{R}^{n \times N}$  of Eq. (2.110), we can reduce the system of Eq. (4.16) to

$$\underbrace{C_e S A^{\text{AC}} C_e^t}_{=: A_{e/e}^{\text{AC}}} \underbrace{C_e G}_{=: G_e} = \underbrace{C_e S B}_{=: B_e}, \quad (4.28)$$

where  $A_{e/e}^{\text{AC}} \in \mathbb{C}^{n \times n}$  is identical to the matrix defined in in the small signal case of Eq. (3.64) and  $G_e \in \mathbb{C}^{n \times N}$  as well as  $B_e \in \mathbb{R}^{n \times N}$  have no rows of odd Fourier harmonics in the BE but contain them in their columns. The odd rows of  $G$  can be recovered from the even rows  $G_e$  analogously to Eq. (3.69) as

$$G_o^{\text{BE}} := \text{diag} \left( A_{o/o}^{\text{AC, BE}} \right)^{-1} \left[ B_o^{\text{BE}} - \text{offdiag} \left( A_{o/e}^{\text{AC, BE}} \right) G_e \right], \quad (4.29)$$

where  $G_o^{\text{BE}} \in \mathbb{C}^{N_{\text{BE}}/2 \times N}$  contains only the odd rows of the BE.

Using that the Green's function of the terminal current can be expressed as in Eq. (4.26), we find

$$\left( \mathbf{G}'_C \right)^t = \mathcal{P}'_C{}^t \left( C_e^t G_e + (C_o^{\text{BE}})^t G_o^{\text{BE}} \right) + \Xi^t$$

$$\begin{aligned}
&= \mathcal{P}_{I'_C}^t \left( C_e^t G_e \right. \\
&\quad \left. + (C_o^{\text{BE}})^t \text{diag} \left( A_{o/o}^{\text{AC, BE}} \right)^{-1} \left[ B_o^{\text{BE}} - \text{offdiag} \left( A_{o/e}^{\text{AC, BE}} \right) G_e \right] \right) \\
&\quad + \Xi^t,
\end{aligned}$$

where we also used the odd compression matrix of Eq. (3.68).

Replacing formally  $G_e = \left( A_{e/e}^{\text{AC}} \right)^{-1} B_e$  and transposing the equation yields

$$\begin{aligned}
\mathbf{G}_{I'_C}^t &= B_e^t \left( \left( A_{e/e}^{\text{AC}} \right)^t \right)^{-1} \left[ C_e - \text{offdiag} \left( A_{o/e}^{\text{AC, BE}} \right)^t \text{diag} \left( A_{o/o}^{\text{AC, BE}} \right)^{-1} C_o^{\text{BE}} \right] \mathcal{P}_{I'_C}^t \\
&\quad + (B_o^{\text{BE}})^t \text{diag} \left( A_{o/o}^{\text{AC, BE}} \right)^{-1} C_o^{\text{BE}} \mathcal{P}_{I'_C}^t + \Xi \\
&=: B_e^t \mathbf{y}_e + (B_o^{\text{BE}})^t \text{diag} \left( A_{o/o}^{\text{AC, BE}} \right)^{-1} C_o^{\text{BE}} \mathcal{P}_{I'_C}^t + \Xi.
\end{aligned} \tag{4.30}$$

Thus, we can determine  $\mathbf{y}_e$  using the equation

$$\left( A_{e/e}^{\text{AC}} \right)^t \mathbf{y}_e = \left[ C_e - \text{offdiag} \left( A_{o/e}^{\text{AC, BE}} \right)^t \text{diag} \left( A_{o/o}^{\text{AC, BE}} \right)^{-1} C_o^{\text{BE}} \right] \mathcal{P}_{I'_C}^t \tag{4.31}$$

and subsequently insert it into Eq. (4.30) in order to compute  $\mathbf{G}_{I'_C}^t$ . Using this approach is computationally much more inexpensive than the determination of the full Green's function matrix in Eq. (4.16). The time to solve Eq. (4.31) is comparable to one Newton-Raphson iteration in the determination of the stationary solution.

Note that as long as we disregard any degeneracies, Eq. (4.31) is identical to Eq. (3.71) and therefore we can reuse  $\mathbf{y}_e$  in order to compute the admittance parameters as well. Thus, with a single solution of (4.31), we can determine both the Green's functions of terminal currents with Eq. (4.30) and the admittance parameters with Eq. (3.70). However, additional complications occur when we want to avoid the explicit computation of all degenerate states. The discussion of this is postponed to Sect. 4.8 after multiplicities in the Green's function equations have been discussed.

## 4.6 Power Spectral Density

We already discussed the meaning of the PSD in Sect. 4.1. In this section, we are going to show how we can compute the PSD of terminal currents, i.e. the noise power contained in drain and gate currents as well as their cross-correlation.

There are two contributions to noise in devices: scattering as well as generation and recombination of carriers. Both are treated as stochastic events in the LBE. Thus, whenever a carrier scatters, its time of arrival at the terminal is modulated which manifests in a fluctuation of the terminal current. A generation or recombination adds or subtracts a charged carrier to or from the terminal current and thus leads to fluctuations as well. Moreover, whenever a scattering event or a GR event occurs, the electric field changes and thus the displacement current in the terminals changes instantaneously (cf. the definition of the terminal current of Eq. (4.20)).

Note that throughout this section, we only used the Green's functions describing a Langevin-source in the LBE. In this work, we will not concern ourselves with Langevin-sources in the PE.

### 4.6.1 Scattering

Let us start with the noise due to scattering events. We know that the total rate of scattering of carriers in our LBE is given by the integral over either the in-scattering or out-scattering part of the scattering term of Eq. (2.15). Which one we use ultimately does not matter since we will integrate over initial and final states anyway. Thus the one-sided PSD of the fluctuations can be expressed with the white noise of the scattering rate [51, 60]

$$P_{\xi\xi}^{\nu,\nu'}(y, \mathbf{k}; y', \mathbf{k}') = 2\Omega \delta(y - y') \sum_{\eta} \left( (1 - f^{\nu}(y, \mathbf{k})) S_{\eta}^{\nu,\nu'}(y; \mathbf{k}, \mathbf{k}') f^{\nu'}(y, \mathbf{k}') \right), \quad (4.32)$$

where the factor of two stems from the fact that scattering noise is a Poisson process, and  $\Omega$  is the system area. In other words,  $P_{\xi\xi}$  is the PSD of Langevin-sources and it tells us the amount of fluctuations related to carriers scattering from the state  $(\nu', \mathbf{k}')$  to the state  $(\nu, \mathbf{k})$  at the position  $y$ .

Our goal is to use the Wiener-Lee theorem [111] to work our way towards the PSD of terminal currents. It states that once we know the transfer function – or Green's function in the frequency domain – from one variable to another, we can compute the PSD of other quantities. The transfer function of the distribution function  $H^f$  lets us compute the fluctuations of the distribution function by integrating over all Langevin-sources as

$$\delta f^{\nu}(y, \mathbf{k}, t) = \sum_{\nu'} \int dy' \int \frac{d^2 k'}{(2\pi)^2} \int dt' (H^f)^{\nu,\nu'}(y, \mathbf{k}, t; y', \mathbf{k}', t') \xi_{\text{BE}}^{\nu'}(y', \mathbf{k}', t'). \quad (4.33)$$

Since we are only interested in noise in stationary systems, the transfer function  $H^f$  can be expressed such that it only depends on the time difference  $t - t'$  as

$$(H^f)^{\nu,\nu'}(y, \mathbf{k}, t; y', \mathbf{k}', t') = (H^f)^{\nu,\nu'}(y, \mathbf{k}, 0; y', \mathbf{k}', t' - t)$$

and therefore we can Fourier transform it to the Green's function of the distribution as

$$(G^f)^{\nu,\nu'}(y, \mathbf{k}; y', \mathbf{k}'; \omega) = \int d(t' - t) (H^f)^{\nu,\nu'}(y, \mathbf{k}, 0; y', \mathbf{k}', t' - t) e^{i\omega(t' - t)},$$

where we wrote down the frequency argument of the Green's function explicitly. Bear in mind that although we did not write it down everywhere, all of the Green's functions in this chapter are frequency-dependent. The fluctuations of the distribution function in the frequency domain could in principle be computed with the above Green's function similar to how the transfer function is used in the time domain. However,  $\xi_{\text{BE}}^{\nu'}(y', \mathbf{k}', t')$  is usually not square-integrable and therefore its Fourier transform does not exist. But the Wiener-Khinchin theorem states that despite its non-existence, the PSD is still well defined. And in that sense we can use Green's functions to compute PSDs related to whichever quantity we want in frequency space.

This idea will be used to compute the PSD of terminal currents starting from the PSD of Langevin-sources formulated in Eq. (4.32) which is the essence of the Wiener-Lee theorem. But first we need to understand what the Green's function of a whole scattering process is. Recall that the Green's function  $G^f$  was defined with Eq. (4.10), where the Langevin-source was replaced by a delta-distribution, suggesting the generation of a single electron in some state. Therefore  $G^f$  is the response of the distribution function if there were an electron created *ex nihilo*. With this interpretation, it makes sense that we can use  $G^f$  multiplied by the size of a fluctuation in order to calculate the fluctuation of the distribution function.

But this is not what is happening in a scattering process. Each scattering event is a perfectly correlated annihilation and subsequent creation of an electron. Say we have an electron in the initial state  $(\nu'_0, y'_0, \mathbf{k}'_0)$  which is involved in some kind of scattering event that changes it to the final state  $(\nu_0, y_0, \mathbf{k}_0)$ . Then we can define the response of the distribution function at  $(\nu, y, \mathbf{k})$  to such an annihilation and creation event as

$$(G^f)^{\nu,\nu_0}(y, \mathbf{k}; y_0, \mathbf{k}_0; \omega) - (G^f)^{\nu,\nu'_0}(y, \mathbf{k}; y'_0, \mathbf{k}'_0; \omega), \quad (4.34)$$

where the Green's function with the negative sign is the response to the annihilation.

Equipped with this combination of Green's functions for a scattering event, we can use the Wiener-Lee theorem to compute the PSD of distribution functions

due to the scattering events:

$$\begin{aligned}
Q_{ff}^{\nu,\nu'}(y, \mathbf{k}; y', \mathbf{k}'; \omega) &= \sum_{\nu_0, \nu'_0} \iint dy_0 dy'_0 \iint \frac{d^2 k_0}{(2\pi)^2} \frac{d^2 k'_0}{(2\pi)^2} \\
&\times \left( (G^f)^{\nu, \nu_0}(y, \mathbf{k}; y_0, \mathbf{k}_0; \omega) - (G^f)^{\nu, \nu'_0}(y, \mathbf{k}; y'_0, \mathbf{k}'_0; \omega) \right) \\
&\times \left( (G^f)^{\nu, \nu_0}(y', \mathbf{k}'; y_0, \mathbf{k}_0; \omega) - (G^f)^{\nu, \nu'_0}(y', \mathbf{k}'; y'_0, \mathbf{k}'_0; \omega) \right)^* \\
&\times P_{\xi\xi}^{\nu_0, \nu'_0}(y_0, \mathbf{k}_0; y'_0, \mathbf{k}'_0).
\end{aligned} \tag{4.35}$$

Ultimately, we do not want to compute the PSD shown in Eq. (4.35) but rather the PSD of terminal currents. The Green's function of the terminal current  $G^{I_C}$  of Eq. (4.23) expresses the response of the terminal current in the frequency domain as

$$\delta I_C(\omega) = \sum_{\nu'} \int dy' \int \frac{d^2 k'}{(2\pi)^2} (G^{I_C})^{\nu'}(y', \mathbf{k}'; \omega) \xi_{\text{BE}}^{\nu'}(y', \mathbf{k}', \omega).$$

And using the Wiener-Lee theorem, we find the expression for the PSD of terminal currents

$$\begin{aligned}
Q_{I_C, I_{C'}}(\omega) &= \sum_{\nu_0, \nu'_0} \iint dy_0 dy'_0 \iint \frac{d^2 k_0}{(2\pi)^2} \frac{d^2 k'_0}{(2\pi)^2} \\
&\times \left( (G^{I_C})^{\nu_0}(y_0, \mathbf{k}_0; \omega) - (G^{I_C})^{\nu'_0}(y'_0, \mathbf{k}'_0; \omega) \right) \\
&\times \left( (G^{I_{C'}})^{\nu_0}(y_0, \mathbf{k}_0; \omega) - (G^{I_{C'}})^{\nu'_0}(y'_0, \mathbf{k}'_0; \omega) \right)^* \\
&\times P_{\xi\xi}^{\nu_0, \nu'_0}(y_0, \mathbf{k}_0; y'_0, \mathbf{k}'_0).
\end{aligned} \tag{4.36}$$

#### 4.6.2 Generation and Recombination

The stochastic generation and recombination of carriers represents a source for fluctuations, similar to the scattering processes. In our case we only have generation and recombination of carriers at the source and drain contacts (see Sect. 2.4.3) where they are injected into or extracted from the semiconductor region. Recall that we introduced the generation and recombination in Eq. (2.36) as a scattering process in between the subbands of the semiconductor and a fictional thermal bath representing a contact. This way, the GR rate of Eq. (2.36) is structurally identical to a scattering rate and therefore, we may compute the PSD in the same way as for a scattering process, i.e. just as if the thermal baths of the contacts were trap levels [61].

In analogy to the scattering processes, we may define the white noise of the GR rate as

$$\begin{aligned}
R_*^{\nu,\nu'}(y, \mathbf{k}; y', \mathbf{k}') &= 2\Omega \delta_{\nu,\nu'} \delta(y - y') \sum_{C=S,D} (1 - f^\nu(y, \mathbf{k})) \gamma^{\nu,C}(y, \mathbf{k}; \mathbf{k}') f_{\text{eq}}^C(\mathbf{k}'), \\
R_\dagger^{\nu,\nu'}(y, \mathbf{k}; y', \mathbf{k}') &= 2\Omega \delta_{\nu,\nu'} \delta(y - y') \sum_{C=S,D} (1 - f_{\text{eq}}^C(\mathbf{k}')) \gamma^{C,\nu}(y, \mathbf{k}'; \mathbf{k}) f^\nu(y, \mathbf{k}),
\end{aligned}
\tag{4.37}$$

where  $\gamma^{\nu,C}$  is given by Eq. (2.37). Here,  $R_*$  is the rate of creation and  $R_\dagger$  is the rate of annihilation. Using these kernels, we can express the PSD of terminal currents using the Wiener-Lee theorem but this time acknowledging that fluctuations in the fictional thermal baths of the contacts do not contribute. Therefore we only include Green's functions for the creation and annihilation in the semiconductor region while the Green's functions for creation and annihilation events in the contact thermal bath vanish. With Eq. (4.37) we find two contributions:

$$\begin{aligned}
R_{I_C, I_{C'}}^{\nu_0, \nu'_0} &= \sum_{\nu_0, \nu'_0} \iint dy_0 dy'_0 \iint \frac{d^2 k_0}{(2\pi)^2} \frac{d^2 k'_0}{(2\pi)^2} \\
&\quad \times \left[ (G^{I_C})^{\nu_0}(y_0, \mathbf{k}_0) \left( (G^{I_{C'}})^{\nu'_0}(y_0, \mathbf{k}_0) \right)^* R_*^{\nu_0, \nu'_0}(y_0, \mathbf{k}_0; y'_0, \mathbf{k}'_0) \right. \\
&\quad \left. + \left( - (G^{I_C})^{\nu'_0}(y'_0, \mathbf{k}'_0) \right) \left( - (G^{I_{C'}})^{\nu_0}(y'_0, \mathbf{k}'_0) \right)^* R_\dagger^{\nu_0, \nu'_0}(y_0, \mathbf{k}_0; y'_0, \mathbf{k}'_0) \right] \\
&= 2 \sum_{C_0=S,D} \sum_{\nu_0} \int \frac{d^2 k_0}{(2\pi)^2} (G^{I_C})^{\nu_0}(y_{C_0}, \mathbf{k}_0) \left( (G^{I_{C'}})^{\nu_0}(y_{C_0}, \mathbf{k}_0) \right)^* \\
&\quad \times v_{\text{GR}} \left[ f_{\text{eq}}^{\nu_0}(y_{C_0}, \mathbf{k}_0) + f^{\nu_0}(y_{C_0}, \mathbf{k}_0) - 2f^{\nu_0}(y_{C_0}, \mathbf{k}_0) f_{\text{eq}}^{\nu_0}(y_{C_0}, \mathbf{k}_0) \right],
\end{aligned}
\tag{4.38}$$

where we omitted the frequency argument for brevity. Note that – somewhat surprisingly – the actual GR rate of Eq. (2.38) is linear in either the distribution function or the equilibrium distribution function but the PSD has terms proportional to their product. This is because in the GR rate, the terms proportional to their product in the in- and out-scattering terms cancel each other since they have opposite signs. That means that the GR rate does in fact include the Pauli principle despite being linear. On the other hand, both the in- and out-scattering contribute positively to the total noise and therefore there is no cancellation and products of the distribution functions appear in the PSD.

### 4.6.3 Discretization

Let us discretize the total PSD consisting of the scattering noise  $Q$  of Eq. (4.36) and the GR noise  $R$  on the contacts of Eq. (4.38). Since we compute the Green's functions of the terminal currents using Eq. (4.23) in the Herring-Vogt transformed  $\mathbf{k}$ -space with a subsequent projection onto equienergy circles and Fourier harmonics as well as a transformation to the total energy  $H$  (cf. Sect. 2.4.4), we need to transform the PSD in the same way.

In Sect. 2.4.4, we mentioned that we will simply assume that the expressions for the scattering rates are already in the Herring-Vogt transformed space from which it follows that the same must hold true for the scattering rate in the PSD. The rest of the expression for the PSD remains invariant and therefore we can simply assume that  $\mathbf{k}$ -space is already Herring-Vogt transformed.

Inserting Eq. (4.32) into Eq. (4.36) and expanding the Green's functions and distribution functions in Fourier harmonics and transforming to  $H$ -space, we find for the PSD of the scattering processes

$$\begin{aligned}
Q_{I'_C, I'_C}(\omega) &= 2\Omega \sum_{\nu, \nu'} Z^\nu Z^{\nu'} \int dy \int dH \int d\phi \int dH' \int d\phi' \\
&\times \left( \sum_m (G^{I'_C})^\nu_m(y, H; \omega) Y_m(\phi) - \sum_{m'} (G^{I'_C})^{\nu'}_{m'}(y, H'; \omega) Y_{m'}(\phi') \right) \\
&\times \left( \sum_n (G^{I'_C})^\nu_n(y, H; \omega) Y_n(\phi) - \sum_{n'} (G^{I'_C})^{\nu'}_{n'}(y, H'; \omega) Y_{n'}(\phi') \right)^* \\
&\times \left( \left( 1 - \sum_\ell f'_\ell(y, H) Y_\ell(\phi) \right) \sum_\eta S_\eta^{\nu, \nu'}(y; H, H') \sum_{\ell'} f'_{\ell'}(y, H') Y_{\ell'}(\phi') \right).
\end{aligned} \tag{4.39}$$

Working out the product of Fourier harmonics is tedious but straightforward. Here we can also use that the scattering processes fixes the difference between the initial and final energy with a delta-distribution as shown in Eq. (2.21) and therefore we can straightforwardly integrate over  $H'$ . Due to its lengthiness the resulting expression is shown in the appendix in Eq. (C.1). Nevertheless, we want to mention a detail that is important for consistency. When discretizing Eq. (4.39), we need to integrate over products of quantities which may be defined on the direct or on the adjoint grid. The box-integration method suggest that we cannot average quantities individually but we must average the whole product. In Sect. 2.4.5, we already discussed what that means for the scattering term and here it is the same approach: consider a product of some functions  $d_n$  on the direct grid and some functions  $a_m$  on the adjoint grid. Then the integral over  $y$

needs to be computed as

$$\begin{aligned} \int dy \left( \prod_n d_n(y) \right) \left( \prod_m a_m(y) \right) &= \Delta y_1 \left( \prod_n d_n(y_1) \right) \left( \prod_m a_m(y_{1+}) \right) \\ &+ \sum_{i=2}^{N_y-1} \Delta y_i \left( \prod_n d_n(y_i) \right) \frac{\prod_m a_m(y_{i+}) + \prod_m a_m(y_{i-})}{2} \\ &+ \Delta y_{N_y} \left( \prod_n d_n(y_{N_y}) \right) \left( \prod_m a_m(y_{N_y-}) \right), \end{aligned}$$

where the first and last boxes,  $\Delta y_1$  and  $\Delta y_{N_y}$ , are only half-sized (see Eq. (2.74)). If we would not use the above averaging procedure, we could not be consistent with the way the scattering rate is discretized in Sect. 2.4.5 and therefore the PSD would be inconsistent with the small signal admittance parameters (cf. Sect. 5.4.2).

The PSD  $R$  of the GR rate at the contacts can be transformed and its constituents expanded in the same way. Since the GR rate is only defined on the contacts, only even harmonics can contribute and therefore, after the  $H$ -transformation and the expansion in Fourier harmonics, Eq. (4.38) yields

$$\begin{aligned} R_{I'_C, I'_{C'}} &= 2 \sum_{C_0=S,D} \sum_{\nu} Z^{\nu} \int dH \int d\phi \\ &\times \left( \sum_{m \text{ even}} (G^{I'_C})_m^{\nu}(y_{C_0}, H) Y_m(\phi) \right) \left( \sum_{n \text{ even}} (G^{I'_{C'}})_n^{\nu}(y_{C_0}, H) Y_n(\phi) \right)^* \\ &\times v_{\text{GR}} \left[ f_{\text{eq}}^{\nu}(y_{C_0}, H) + (1 - 2f_{\text{eq}}^{\nu}(y_{C_0}, H)) \sum_{\ell \text{ even}} f_{\ell}^{\nu}(y_{C_0}, H) Y_{\ell}(\phi) \right], \end{aligned} \quad (4.40)$$

The result of the discretization is shown in the appendix in Eq. (C.2).

The total PSD is then obviously given by

$$P_{I'_C, I'_{C'}}(\omega) = Q_{I'_C, I'_{C'}}(\omega) + R_{I'_C, I'_{C'}}(\omega). \quad (4.41)$$

Since we are only going to compute the PSD of terminal currents, we are often going to refer to it as

$$P_{C, C'}(\omega) := P_{I'_C, I'_{C'}}(\omega) \quad (4.42)$$

in the remainder of this work. Moreover, in order to inspect the noise in a device in more detail, we want to understand where in the device it originates. Therefore we define the density  $K_{C, C'}$  of the PSD as

$$P_{C, C'}(\omega) =: \int dy K_{C, C'}(y; \omega), \quad (4.43)$$

where  $K_{C,C'}(y; \omega)$  is given by Eq. 4.41 but without the integration over  $y$ . That means that  $K_{C,C'}$  is the contribution to the PSD per position in transport direction or – in the discretized case – per grid point in transport direction. Similarly, we can also define the contribution to the PSD per position in  $y$ -direction *and* per total energy  $H$  as

$$P_{C,C'}(\omega) =: \int dy \int dH \mathcal{K}_{C,C'}(y, H; \omega), \quad (4.44)$$

where  $\mathcal{K}_{C,C'}(y, H; \omega)$  is the result of Eq. 4.41 omitting both the  $y$ - and the  $H$ -integration. Note that there is an ambiguity in defining  $\mathcal{K}_{C,C'}$  since the scattering process is non-local in  $H$ -space. In principle there are two densities, one w.r.t. the initial energy  $H'$  and one w.r.t. the final energy  $H$  of Eq. (4.39). In this work, we choose to compute the density as a function of the initial energy  $H'$ . Lastly, we may also define a PSD density per subband as

$$\mathcal{K}_{C,C'}(y, H; \omega) =: \sum_{\nu} \mathcal{K}_{C,C'}^{\nu}(y, H; \omega). \quad (4.45)$$

Bear in mind that in practice we will compute the PSD of systems which have some level degeneracy. In that case, we will find multiplicities in the PSD. The discussion of this is postponed to Sect. 4.7.3

#### 4.6.4 Common-Source Configuration

When we compute the PSDs of the terminal currents later on in Chap. 5, we will apply biases in the common-source configuration of the device. As was already mentioned in Sect. 3.6.3, the common-source configuration is given when the gates are shorted and the voltages are measured relative to the source contact. In terms of the PSD this means that we need to understand how a fluctuation inside the device elicits a fluctuation in the terminal currents when contacts are shorted. This question can be answered entirely by noting that the Green's function of the terminal current of Eq. (4.20) is additive in the sense that

$$G_{\zeta}^{IG} = G_{\zeta}^{ITG} + G_{\zeta}^{IBG},$$

since the integral over a shorted gate (G) is merely the sum of an integral over the top gate (TG) and the bottom gate (BG) of the device in Fig. 2.1. This entails that the PSD of Eq. (4.41) for a shorted gate can be expanded into a sum of all possible top and bottom gate contributions, including the cross-correlations between the gates as

$$P_{GG} = P_{TG,TG} + P_{TG,BG} + P_{BG,TG} + P_{BG,BG}.$$

Likewise, the cross-correlation between the shorted gates and the drain can be expanded as

$$P_{GD} = P_{TG,D} + P_{BG,D}.$$

## 4.7 Degeneracy

Treating degeneracy in the framework of the Green's functions of fluctuations is somewhat different from the case where we solve the BE. We will use the same terminology as in Sect. 2.4.6, where a single individual state is referred to as a *true state*. If a set of true states are indistinguishable, they belong to a *degenerate group*. A single true state, which is part of a degenerate group, is called a *degenerate state*.

The states of a degenerate group share all properties like the band structure and the scattering rates. For example, in our current framework, it is impossible to distinguish between a spin up and a spin down electron. Furthermore, the valley multiplicity introduced in Eq. 2.93 also means that we cannot distinguish between electrons occupying one or the other valley on the same axis in  $\mathbf{k}$ -space.

In order to treat the degeneracy in the Green's function equations of Sect. 4.16, it would be tempting to proceed analogously to the stationary case shown in Sect. 2.4.6 which is equally valid for the small signal BE. However, there is a difference that we must account for: the Langevin-source terms on the r.h.s. of Eq. (4.6) represent fluctuations of each individual true state. Therefore, the Green's functions represent the response to a fluctuation in a single true state. This is also the reason why there are no multiplicities in the Green's function of the terminal current of Eq. (4.23), even when degeneracies are considered.

In order to understand where multiplicities appear, we will reduce the system of Green's functions of Eq. (4.16) to avoid computing degenerate states several times. Then we can use the adjoint method of Sect. 4.5 unaltered in order to compute the Green's functions of the terminal currents.

### 4.7.1 One Degenerate Group

Before we consider the case of multiple degenerate groups, let us first take a look at the LBE as if it only contained a single degenerate group of  $\mu$ -times degenerate states. Once we understand this case, we can reuse results later on. The system of equations consisting of the LBE with a single degenerate group and the PE can be written down as

$$\begin{pmatrix} & & L_V \\ & L_f^S & \vdots \\ & & L_V \\ A_f & \cdots & A_f & A_V \end{pmatrix} \underbrace{\begin{pmatrix} G_{11} & \cdots & \cdots & G_{1\mu} & G_{1V} \\ \vdots & \ddots & & \vdots & \vdots \\ \vdots & & \ddots & \vdots & \vdots \\ G_{\mu 1} & \cdots & \cdots & G_{\mu\mu} & G_{\mu V} \\ G_{V1} & \cdots & \cdots & G_{V\mu} & G_{VV} \end{pmatrix}}_{=G} = \underbrace{\begin{pmatrix} \mathbb{I} & 0 \\ 0 & \mathbb{I} \end{pmatrix}}_{=B} \quad (4.46)$$

with

$$L_f^S = \begin{pmatrix} L_f & S & \cdots & S \\ S & \ddots & & \vdots \\ \vdots & & \ddots & S \\ S & \cdots & S & L_f \end{pmatrix}. \quad (4.47)$$

Since we have  $\mu$  degenerate states, we can split the LBE into  $\mu$  identical submatrices  $L_f$ . The scattering submatrices  $S$  connecting the degenerate states must be identical or else we would allow for an imbalance in the occupation of the states which would render them non-degenerate. Since the  $\mu$  degenerate states are indistinguishable they must also contribute in the same way to the charge in the potential via  $A_f$  and conversely the potential must contribute in the same way to the LBE via  $L_V$ . We also split the Green's function matrix  $G$  into the same subspaces as the LBE and PE matrix. Note that due to notational difficulties, the boundary conditions of Sect. 4.3 are not explicitly shown but they are assumed to be present nonetheless.

Our goal is to reduce Eq. (4.46) to a system of equations where the degenerate group is represented by only one true state of the degenerate group, hence, reducing the computational load significantly. To this end, consider the terminal current expressed as in Eq. (4.26) with the terminal current operator  $\mathcal{P}_{I_C}$ . The terminal current operator can be split into operators that act upon the true states of the degenerate groups as

$$\begin{aligned} (\mathbf{G}^{I_C})^t &= \mathcal{P}_{I_C}^t G + \Xi^t = ((\mathcal{P}_1)^t \cdots (\mathcal{P}_\mu)^t (\mathcal{P}_V)^t) G + \Xi^t \\ &= ((\mathcal{P}_1)^t \cdots (\mathcal{P}_1)^t (\mathcal{P}_V)^t) G + \Xi^t, \end{aligned}$$

where  $\mathcal{P}_1$  through  $\mathcal{P}_\mu$  are identical since the states are indistinguishable. Now let us introduce a transformation of the Green's function matrix,

$$G = T \tilde{G} = \begin{pmatrix} & & 0 \\ & T' & \vdots \\ & & 0 \\ 0 & \cdots & 0 & \mathbb{I} \end{pmatrix} \begin{pmatrix} \tilde{G}_{11} & \cdots & \cdots & \tilde{G}_{1\mu} & \tilde{G}_{1V} \\ \vdots & \ddots & & \vdots & \vdots \\ \vdots & & \ddots & \vdots & \vdots \\ \tilde{G}_{\mu 1} & \cdots & \cdots & \tilde{G}_{\mu\mu} & \tilde{G}_{\mu V} \\ \tilde{G}_{V1} & \cdots & \cdots & \tilde{G}_{V\mu} & \tilde{G}_{VV} \end{pmatrix}, \quad (4.48)$$

where the unit matrix is in the subspace of the PE. The transformation  $T$  has full rank and fulfills

$$\mathcal{P}_{I_C}^t T = ((\mathcal{P}_1)^t \ 0 \ \cdots \ 0 \ (\mathcal{P}_V)^t),$$

i.e. only the projection operator of the first set of true states of the degenerate group does not vanish. Thus, we may choose the transformation as

$$(T')_{k\ell} = \begin{cases} \frac{1}{\mu}\mathbb{I}, & \text{for } \ell = 1 \vee \ell = k, \\ -\frac{1}{\mu(\mu-1)}\mathbb{I}, & \text{else,} \end{cases} \quad (4.49)$$

where  $k, \ell = 1, \dots, \mu$ . With  $T$ , we can compute the Green's function of the terminal current as

$$\begin{aligned} (\mathbf{G}'_C)^t &= \mathcal{P}'_C{}^t T \tilde{G} + \Xi^t \\ &= ((\mathcal{P}_1)^t \quad (\mathcal{P}_V)^t) \begin{pmatrix} \tilde{G}_{11} & \cdots & \cdots & \tilde{G}_{1\mu} & \tilde{G}_{1V} \\ \tilde{G}_{V1} & \cdots & \cdots & \tilde{G}_{V\mu} & \tilde{G}_{VV} \end{pmatrix} + \Xi^t. \end{aligned} \quad (4.50)$$

Note that in Sect. 4.5 we showed how we can compute the Green's functions of terminal currents directly from the general Green's function equations. Here, we showed that by using  $\tilde{G}$  instead of  $G$ , it is sufficient to know only the first row and the last row of the matrix  $\tilde{G}$  in order to determine the Green's function of the terminal current.

Inserting Eq. (4.48) into Eq. (4.46), we obtain

$$\begin{pmatrix} & & & & L_V \\ & L_f^S T' & & & \vdots \\ & & & & L_V \\ A_f & 0 & \cdots & 0 & A_V \end{pmatrix} \tilde{G} = B \quad (4.51)$$

with

$$\begin{aligned} (L_f^S T')_{k\ell} &= L_f (T')_{k\ell} + \sum_{\substack{n=1 \\ n \neq k}}^{\mu} S (T')_{n\ell} \\ &= \begin{cases} \frac{1}{\mu}(L_f - S) + S, & \text{for } \ell = 1, \\ \frac{1}{\mu}(L_f - S), & \text{for } \ell \neq 1 \wedge \ell = k, \\ -\frac{1}{\mu(\mu-1)}(L_f - S), & \text{else.} \end{cases} \end{aligned}$$

Now, we can use equivalence transformations on Eq. (4.51) to remove the dependencies on all but the first set of true states. To this end, let us add the rows  $2, \dots, \mu$  to the first row of each column. The computation boils down to a simple sum:

$$\ell = 1 : \quad \sum_{k=1}^{\mu} (L_f^S T')_{k\ell} = L_f + (\mu - 1)S,$$

$$\ell \neq 1 : \sum_{k=1}^{\mu} (L_f^S T')_{k\ell} = \frac{1}{\mu} (L_f - S) - (\mu - 1) \frac{1}{\mu(\mu - 1)} (L_f - S) = 0.$$

Therefore, the first row only depends on  $\tilde{G}_{1\ell}$  and  $\tilde{G}_{V\ell}$  with  $\ell = 1, \dots, \mu, V$ . Furthermore, we do not need to compute them since – according to Eq. (4.50) – we can compute the Green's function of terminal currents with only  $\tilde{G}_{1\ell}$  and  $\tilde{G}_{V\ell}$ . Dropping all equations but the ones necessary for the terminal current, we find that Eq. (4.51) yields

$$\begin{pmatrix} L_f + (\mu - 1)S & \mu L_V \\ A_f & A_V \end{pmatrix} \begin{pmatrix} \tilde{G}_{11} & \cdots & \tilde{G}_{1\mu} & \tilde{G}_{1V} \\ \tilde{G}_{V1} & \cdots & \tilde{G}_{V\mu} & \tilde{G}_{VV} \end{pmatrix} = \begin{pmatrix} \mathbb{I} & \cdots & \mathbb{I} & 0 \\ 0 & \cdots & 0 & \mathbb{I} \end{pmatrix}. \quad (4.52)$$

Note that the factor  $\mu$  in front of  $L_V$  as well as the unit matrices in the first row of the r.h.s. stem from the addition of the rows. Since the columns  $\ell = 1, \dots, \mu$  of Green's function matrix in Eq. (4.52) obey the same equations, we can trivially conclude that

$$\tilde{G}_{1\ell} = \tilde{G}_{11}, \quad \tilde{G}_{V\ell} = \tilde{G}_{V1}, \quad \text{with } \ell = 1, \dots, \mu, \quad (4.53)$$

and therefore reduce Eq. (4.52) even further to

$$\begin{pmatrix} L_f + (\mu - 1)S & \mu L_V \\ A_f & A_V \end{pmatrix} \begin{pmatrix} \tilde{G}_{11} & \tilde{G}_{1V} \\ \tilde{G}_{V1} & \tilde{G}_{VV} \end{pmatrix} = \begin{pmatrix} \mathbb{I} & 0 \\ 0 & \mathbb{I} \end{pmatrix}. \quad (4.54)$$

Using Eq. (4.53), we can also see how the Green's functions of the terminal currents of Eq. (4.50) are identical for each of the  $\tilde{G}_{1\ell}$  and  $\tilde{G}_{V\ell}$  because they are degenerate. Therefore, it is sufficient to know the Green's function terminal current w.r.t. only one true state of the degenerate group given by

$$\left( \left( \mathbf{G}_1^{I_C} \right)^t \quad \left( \mathbf{G}_V^{I_C} \right)^t \right)^t = \left( (\mathcal{P}_1)^t \quad (\mathcal{P}_V)^t \right) \begin{pmatrix} \tilde{G}_{11} & \tilde{G}_{1V} \\ \tilde{G}_{V1} & \tilde{G}_{VV} \end{pmatrix} + \left( \Xi_1^t \quad \Xi_V^t \right)^t. \quad (4.55)$$

The Green's function of the terminal current obviously does not contain any multiplicities since it is the response to a fluctuation in a single true state.

In summary, it is sufficient to solve Eq. (4.54) instead of the whole system of Eq. (4.46). Moreover, in the presence of a single degenerate group, the terminal current response is identical for fluctuations in degenerate states. Hence it is sufficient to determine the Green's function of the terminal current of only one true state as shown in Eq. (4.55).

## 4.7.2 Multiple Degenerate Groups

Let us now turn to the case where we have  $m$  degenerate groups, where the  $i$ -th group consists of  $\mu_i$  true states. The total number of true states will be

referred to as  $\mu = \sum_i \mu_i$ . Again, true states may be indistinguishable spin up and spin down states or indistinguishable states due to symmetries in the valley structure. Considering this composition of degeneracy, we can write down the LBE and PE system of Eq. (4.16) as

$$\underbrace{\begin{pmatrix} L_f^{S,1} & S^{12} & \cdots & \cdots & S^{1m} & L_V^1 \\ S^{21} & L_f^{S,2} & \cdots & \cdots & S^{2m} & L_V^2 \\ \vdots & & \ddots & & \vdots & \vdots \\ \vdots & & & \ddots & S^{m-1,m} & \vdots \\ S^{m1} & \cdots & \cdots & S^{m,m-1} & L_f^{S,m} & L_V^m \\ A_f^1 & A_f^2 & \cdots & \cdots & A_f^m & A_V \end{pmatrix}}_{=L^S} \times \underbrace{\begin{pmatrix} G^{11} & \cdots & \cdots & G^{1m} & G^{1V} \\ \vdots & \ddots & & \vdots & \vdots \\ \vdots & & \ddots & \vdots & \vdots \\ G^{m1} & \cdots & \cdots & G^{mm} & G^{mV} \\ G^{V1} & \cdots & \cdots & G^{Vm} & G^{VV} \end{pmatrix}}_{=G} = \underbrace{\begin{pmatrix} \mathbb{I} & & 0 \\ & \ddots & \\ 0 & & \mathbb{I} \end{pmatrix}}_{=B}, \quad (4.56)$$

where the submatrices  $L_f^{S,i}$  describe a degenerate group of states and can be represented as

$$L_f^{S,i} = \begin{pmatrix} L_f^i & S^i & \cdots & S^i \\ S^i & \ddots & & \vdots \\ \vdots & & \ddots & S^i \\ S^i & \cdots & S^i & L_f^i \end{pmatrix}. \quad (4.57)$$

The submatrices  $L_V^i$  are the derivatives of the BE w.r.t. the potential, the submatrices  $A_f^i$  are the derivatives of the PE w.r.t. the distribution functions of the respective distinguishable states, and  $A_V$  is the derivative of the PE w.r.t. the potential.

The Green's function matrix in Eq. (4.56) has been split into submatrices of the distinguishable states  $G^{ij}$ , where the submatrices related to indices of the PE use the superscript  $V$ . The r.h.s. of Eq. (4.56) contains the delta-distributions of the Green's functions approach which are effectively a unit matrix after discretization. Again, we didn't explicitly write down the boundary conditions of Sect. 4.3 due to the notational difficulties but they are assumed to be contained in the equation.

The submatrix  $L_f^{S,i}$  of Eq. (4.57) represents the BE operator of a set of degenerate states. Therefore, each  $L_f^{S,i}$  is equivalent to the set of degenerate states of Eq. (4.47).

The scattering submatrices  $S^{ij}$  in between degenerate groups are in general  $\mathbb{R}^{\mu_i \times \mu_j}$  matrices since the initial and final degenerate groups may have different levels of degeneracy. However, since each initial true state of the initial degenerate group is indistinguishable from any other true state in that degenerate group – and the same is valid for the final states – the submatrices of  $S^{ij}$  referring to the scattering between individual true states of the degenerate groups must be identical. Therefore, we can write down  $S^{ij}$  down as

$$(S^{ij})_{k\ell} = s^{ij}, \quad \text{with } i, j = 1, \dots, m, \quad k = 1, \dots, \mu_i, \quad \ell = 1, \dots, \mu_j.$$

Once again our goal is to reduce Eq. (4.56) to a system of equations where each degenerate group is represented by only one true state of the degenerate group, hence, reducing the computational load significantly. To this end, we will proceed analogously to Sect. 4.7.1, only this time, we need to transform  $m$  degenerate groups instead of one. Analogous to Eq. (4.48), the transformation on the Green's function reads

$$G := T \tilde{G} := \begin{pmatrix} T^1 & & 0 & 0 \\ & \ddots & & \vdots \\ 0 & & T^m & 0 \\ 0 & \dots & 0 & \mathbb{I} \end{pmatrix} \tilde{G}, \quad (4.58)$$

where the diagonal blocks  $T^i \in \mathbb{R}^{\mu_i \times \mu_i}$  in the subspaces of the  $m$  degenerate groups are given analogously to Eq. (4.49) as

$$T^i = \begin{cases} \frac{1}{\mu_i} \mathbb{I}, & \text{for } \ell = 1 \vee \ell = k, \\ -\frac{1}{\mu_i(\mu_i-1)} \mathbb{I}, & \text{else.} \end{cases}$$

Dividing the terminal current projection operator into  $m$  subspaces, we obtain results analogous to Eq. (4.50) in every subspace of a degenerate group:

$$\begin{aligned} (\mathbf{G}^{I'C})^t &= \mathcal{P}^t G + \Xi^t \\ &= ((\mathcal{P}^1)^t T^1 \quad \dots \quad (\mathcal{P}^m)^t T^m \quad (\mathcal{P}^V)^t) \begin{pmatrix} \tilde{G}^{11} & \dots & \dots & \tilde{G}^{1m} & \tilde{G}^{1V} \\ \vdots & \ddots & & \vdots & \vdots \\ \vdots & & \ddots & \vdots & \vdots \\ \tilde{G}^{m1} & \dots & \dots & \tilde{G}^{mm} & \tilde{G}^{mV} \\ \tilde{G}^{V1} & \dots & \dots & \tilde{G}^{Vm} & \tilde{G}^{VV} \end{pmatrix} \\ &\quad + \Xi^t. \end{aligned} \quad (4.59)$$

The projection operators of each subspace reduce to

$$(\mathcal{P}^i)^t T^i = \underbrace{\left( (\mathcal{P}_1^i)^t \quad 0 \quad \cdots \quad 0 \right)}_{\mu_i \text{ elements}}, \quad (4.60)$$

where  $(\mathcal{P}_1^i)^t$  is the projection operator on the first true state of the  $i$ -th degenerate group.

Using Eq. (4.58) in Eq. (4.56), we find

$$L^S T \tilde{G} = B \quad (4.61)$$

with

$$L^S T = \begin{pmatrix} L_f^{S,1} T^1 & S^{12} T^2 & \cdots & \cdots & S^{1m} T^m & L_V^1 \\ S^{21} T^1 & L_f^{S,2} T^2 & \cdots & \cdots & S^{2m} T^m & L_V^2 \\ \vdots & & \ddots & & \vdots & \vdots \\ \vdots & & & \ddots & S^{m-1,m} T^m & \vdots \\ S^{m1} T^1 & \cdots & \cdots & S^{m,m-1} T^{m-1} & L_f^{S,m} T^m & L_V^m \\ A_f^1 T^1 & A_f^2 T^2 & \cdots & \cdots & A_f^m T^m & A_V \end{pmatrix}.$$

Here, the diagonal terms are given by

$$\begin{aligned} \left( L_f^{S,i} T^i \right)_{kl} &= \left( L_f^i \right)_{kk} \left( T^i \right)_{kl} + \sum_{n \neq k} \left( L_f^i \right)_{kn} \left( T^i \right)_{nl} \\ &= L_f^i \left( T^i \right)_{kl} + \sum_{n \neq k} S^i \left( T^i \right)_{nl} \\ &= \begin{cases} \frac{1}{\mu_i} \left( L_f^i - S^i \right) + S^i, & \text{for } l = 1, \\ \frac{1}{\mu_i} \left( L_f^i - S^i \right), & \text{for } l \neq 1 \wedge i = j, \\ -\frac{1}{\mu_i(\mu_i-1)} \left( L_f^i - S^i \right), & \text{else,} \end{cases} \end{aligned}$$

while the off-diagonal terms read

$$\left( S^{ij} T^j \right)_{kl} = \sum_{n=1}^{\mu_j} \left( S^{ij} \right)_{kn} \left( T^j \right)_{nl} = \sum_{n=1}^{\mu_j} s^{ij} \left( T^j \right)_{nl} = \begin{cases} s^{ij}, & \text{for } l = 1, \\ 0, & \text{else.} \end{cases}$$

Lastly, the operator  $A_f^j$  of the PE acts upon the  $\mu_j$  states of the degenerate group  $j$ . Therefore, it can be divided into  $\mu_j$  identical subspaces, each with an operator  $a_f^j$  acting on a single true state. Thus, multiplying with the operator  $T^j$  yields

$$A_f^j T^j = \underbrace{\left( a_f^j \quad a_f^j \quad \cdots \quad a_f^j \right)}_{\mu_j \text{ times}} T^j = \left( a_f^j \quad 0 \quad \cdots \quad 0 \right).$$

As in the case of a single degenerate group, we can use equivalence transformations on Eq. (4.61). This time, we need to treat every subspace of a degenerate group analogously to the case of the previous section. That means, we take the subspace of the first true state of each degenerate group and add the equations for the other true states to the first one. Furthermore, we drop all duplicate equations from each subspace. Then Eq. (4.61) reduces similarly to Eq. (4.54) to a system of equations where each degenerate group is represented by one true state:

$$\begin{pmatrix} L_f^1 + (\mu_1 - 1)S^1 & \mu_1 s^{12} & \cdots & \cdots & \mu_1 s^{1m} & \mu_1 L_V^1 \\ \mu_2 s^{21} & L_f^2 + (\mu_2 - 1)S^2 & \cdots & \cdots & \mu_2 s^{2m} & \mu_2 L_V^2 \\ \vdots & & \ddots & & \vdots & \vdots \\ \vdots & & & \ddots & & \vdots \\ \mu_m s^{m1} & \cdots & \cdots & \mu_m s^{m,m-1} & L_f^m + (\mu_m - 1)S^m & \mu_m L_V^m \\ a_f^1 & a_f^2 & \cdots & \cdots & a_f^m & A_V \end{pmatrix} \times \begin{pmatrix} (\tilde{G}^{11})_{11} & \cdots & \cdots & (\tilde{G}^{1m})_{11} & \tilde{G}^{1V} \\ \vdots & \ddots & & \vdots & \vdots \\ \vdots & & \ddots & \vdots & \vdots \\ (\tilde{G}^{m1})_{11} & \cdots & \cdots & (\tilde{G}^{mm})_{11} & \tilde{G}^{mV} \\ (\tilde{G}^{V1})_{11} & \cdots & \cdots & (\tilde{G}^{Vm})_{11} & \tilde{G}^{VV} \end{pmatrix} = \begin{pmatrix} \mathbb{I} & 0 \\ & \ddots \\ 0 & \mathbb{I} \end{pmatrix}, \quad (4.62)$$

where we used that the columns of the first row of Green's functions within a degenerate group obey the same equations and are therefore identical, i.e.

$$(\tilde{G}^{ij})_{1\ell} = (\tilde{G}^{ij})_{11}, \quad (\tilde{G}^{Vj})_{1\ell} = (\tilde{G}^{Vj})_{11}, \quad \text{with } \ell = 1, \dots, \mu_j, \quad i, j = 1, \dots, m. \quad (4.63)$$

Finally, using Eq. (4.60) and Eq. (4.63) in Eq. (4.59), we can express the Green's function of the terminal current as

$$\begin{pmatrix} \mathbf{G}_1^{I_C} \\ \vdots \\ \mathbf{G}_m^{I_C} \\ \mathbf{G}_V^{I_C} \end{pmatrix}^t := \left( (\mathcal{P}_1^1)^t \quad \cdots \quad (\mathcal{P}_1^m)^t \quad (\mathcal{P}^V)^t \right) \times \begin{pmatrix} (\tilde{G}^{11})_{11} & \cdots & \cdots & (\tilde{G}^{1m})_{11} & \tilde{G}^{1V} \\ \vdots & \ddots & & \vdots & \vdots \\ \vdots & & \ddots & \vdots & \vdots \\ (\tilde{G}^{m1})_{11} & \cdots & \cdots & (\tilde{G}^{mm})_{11} & \tilde{G}^{mV} \\ (\tilde{G}^{V1})_{11} & \cdots & \cdots & (\tilde{G}^{Vm})_{11} & \tilde{G}^{VV} \end{pmatrix} + \begin{pmatrix} \mathbb{E}_1 \\ \vdots \\ \mathbb{E}_m \\ \mathbb{E}_V \end{pmatrix}^t, \quad (4.64)$$

where we only kept one true state per degenerate group. Again, as is evident the Green's function of the terminal current does not contain multiplicities.

When we compare the reduced system of Green's functions of Eq. (4.62) and the reduced BE in the stationary case of Eq. (2.91), we can see that they are not the same. This is because the Green's functions are responses to fluctuations in a single true state while in the stationary case we simply consider the states indistinguishable and sum up their contributions. Therefore the conjecture expressed at the end of Sect. 4.5 that we could solve the system of Green's function equations and the small signal analysis in one step only holds true unconditionally if we do not reduce systems due to their degeneracy. In the case of a reduction, as in this section, it depends on the kind of degeneracy. We will defer the in depth discussion of this to Sect. 4.8.

### 4.7.3 Power Spectral Density

Since the whole set of Green's function equations contains redundancy in the case of degeneracy, any quantity derived from the Green's functions can be reduced to a representation where a single true state out of each degenerate group represents the whole degenerate group. This is equally true for the PSD.

We will approach the problem in matrix form, i.e. we assume the PSD of Eq. (4.41) has already been discretized. The PSD of some quantities  $X$  and  $Y$  whose Green's functions are assumed to be known – for example the Green's functions of terminal currents –, can be written down using the Wiener-Lee theorem [111] as

$$P_{XY} = (\mathbf{G}_X)^t P_{\xi\xi} (\mathbf{G}_Y)^*, \quad (4.65)$$

where  $P_{\xi\xi}$  is the matrix containing the fluctuations due to scattering and GR rates. Note that Eq. (4.65) is equivalent to the discretized version of Eq. (4.39) if the Green's function  $\mathbf{G}_X$ , expressing the effect of a fluctuation due to a scattering event on the quantity  $X$ , contains both creation and annihilation parts as in Eq. (4.34). Moreover,  $P_{\xi\xi}$  is the PSD of the Langevin-sources in the LBE, therefore the Green's functions  $\mathbf{G}_X$  and  $\mathbf{G}_Y$  only contain the responses of  $X$  and  $Y$ , respectively, to a fluctuation in the distribution function. In other words, if we use the terminal current Green's function of Eq. (4.64), the Green's function only contains the subspaces 1 through  $m$ .

Since  $P_{\xi\xi}$  can be expressed as the sum of all scattering rates and GR rates (cf. Eq. (4.41)), it shares any symmetry properties arising from the structure of these rates, e.g. symmetries due to degeneracy. Therefore we can split it

analogously to the LBE into subspaces of degenerate groups, i.e.

$$P_{\xi\xi} = \begin{pmatrix} P_{\xi\xi}^{11} & \cdots & P_{\xi\xi}^{1m} \\ \vdots & \ddots & \vdots \\ P_{\xi\xi}^{m1} & \cdots & P_{\xi\xi}^{mm} \end{pmatrix},$$

where each of the diagonal blocks is given a rate of a true state scattering to itself and rates in between degenerate true states

$$(P_{\xi\xi}^{ii})_{k\ell} =: \begin{cases} p_{\text{diag}}^{ii}, & \text{for } k = \ell, \\ p_{\text{offdiag}}^{ii}, & \text{for } k \neq \ell, \end{cases} \quad \text{with } k, \ell = 1, \dots, \mu_i.$$

The off-diagonal blocks of  $P_{\xi\xi}$  connect true states of two degenerate groups. Therefore, all rates connecting individual true states of these degenerate groups must be identical:

$$(P_{\xi\xi}^{ij})_{k\ell} =: p^{ij}, \quad \text{with } i \neq j, \quad k = 1, \dots, \mu_i, \quad \ell = 1, \dots, \mu_j.$$

Using this notation to express the PSD of degenerate groups and noting, once more, that the Green's function of the terminal current has identical elements per degenerate group, we can simplify the PSD of Eq. (4.65) as

$$\begin{aligned} P_{XY} &= \sum_{i=1}^m \sum_{k=1}^{\mu_i} \sum_{\ell=1}^{\mu_i} (\mathbf{G}_X^i)_k^t (P_{\xi\xi}^{ij})_{k\ell} (\mathbf{G}_Y^j)_\ell^* \\ &= \sum_{i=1}^m \left[ \sum_{k=1}^{\mu_i} \sum_{\ell=1}^{\mu_i} (\mathbf{G}_X^i)_k^t (P_{\xi\xi}^{ii})_{k\ell} (\mathbf{G}_Y^i)_\ell^* + \sum_{\substack{j=1 \\ j \neq i}}^{\mu_j} \sum_{k=1}^{\mu_i} \sum_{\ell=1}^{\mu_j} (\mathbf{G}_X^i)_k^t (P_{\xi\xi}^{ij})_{k\ell} (\mathbf{G}_Y^j)_\ell^* \right] \\ &= \sum_{i=1}^m \left[ \mu_i (\mathbf{G}_X^i)_1^t (p_{\text{diag}}^{ii} + (\mu_i - 1) p_{\text{offdiag}}^{ii}) (\mathbf{G}_Y^i)_1^* \right. \\ &\quad \left. + \sum_{\substack{j=1 \\ j \neq i}}^{\mu_j} \mu_i \mu_j (\mathbf{G}_X^i)_1^t p^{ij} (\mathbf{G}_Y^j)_1^* \right]. \end{aligned} \quad (4.66)$$

#### 4.7.4 Application

With the general derivation of multiplicities in the Green's function equations and the derivation of the PSD including degenerate groups, we have all the set pieces to reduce the Green's function equations for the silicon case.

We do not consider any spin-dependent scattering rates such as spin-flipping transitions. Therefore spin up and spin down states of carriers are degenerate as was already mentioned in Sect. 2.4.6. Moreover, we consider the six  $X$ -valleys

of silicon (cf. Fig. 2.6), where each pair of valleys on the same axis in  $\mathbf{k}$ -space is indistinguishable. Hence, it is sufficient to consider three valleys, each being twice degenerate (cf. Fig. 2.15).

In the notation of Sect. 2.4.6, which includes the spin orientations  $\uparrow$  and  $\downarrow$ , we can write down the system of equations with the matrix

$$L^S = \begin{pmatrix} L_f^{S,1,\uparrow} & S^{12,\uparrow} & S^{13,\uparrow} & 0 & 0 & 0 & L_V^{1,\uparrow} \\ S^{21,\uparrow} & L_f^{S,2,\uparrow} & S^{23,\uparrow} & 0 & 0 & 0 & L_V^{2,\uparrow} \\ S^{31,\uparrow} & S^{32,\uparrow} & L_f^{S,3,\uparrow} & 0 & 0 & 0 & L_V^{3,\uparrow} \\ 0 & 0 & 0 & L_f^{S,1,\downarrow} & S^{12,\downarrow} & S^{13,\downarrow} & L_V^{1,\downarrow} \\ 0 & 0 & 0 & S^{21,\downarrow} & L_f^{S,2,\downarrow} & S^{23,\downarrow} & L_V^{2,\downarrow} \\ 0 & 0 & 0 & S^{31,\downarrow} & S^{32,\downarrow} & L_f^{S,3,\downarrow} & L_V^{3,\downarrow} \\ A_f^{1,\uparrow} & A_f^{2,\uparrow} & A_f^{3,\uparrow} & A_f^{1,\downarrow} & A_f^{2,\downarrow} & A_f^{3,\downarrow} & A_V \end{pmatrix},$$

where we already split up the product space of indices into the three valley superscripts 1, 2, and 3 as well as the possible spin orientations  $\uparrow$  and  $\downarrow$ . Since the LBE and the PE do not discriminate between spin orientations, both diagonal blocks are identical and we can simply drop the spin index and apply the result of Sect. 4.7.1, i.e. Eq. (4.54), to find the system of equations

$$\begin{pmatrix} L_f^{S,1} & S^{12} & S^{13} & 2L_V^1 \\ S^{21} & L_f^{S,2} & S^{23} & 2L_V^2 \\ S^{31} & S^{32} & L_f^{S,3} & 2L_V^3 \\ A_f^1 & A_f^2 & A_f^3 & A_V \end{pmatrix} \begin{pmatrix} \tilde{G}^{11} & \tilde{G}^{12} & \tilde{G}^{13} & \tilde{G}^{1V} \\ \tilde{G}^{21} & \tilde{G}^{22} & \tilde{G}^{23} & \tilde{G}^{2V} \\ \tilde{G}^{31} & \tilde{G}^{32} & \tilde{G}^{33} & \tilde{G}^{3V} \\ \tilde{G}^{V1} & \tilde{G}^{V2} & \tilde{G}^{V3} & \tilde{G}^{VV} \end{pmatrix} = \begin{pmatrix} \mathbb{I} & 0 & 0 & 0 \\ 0 & \mathbb{I} & 0 & 0 \\ 0 & 0 & \mathbb{I} & 0 \\ 0 & 0 & 0 & \mathbb{I} \end{pmatrix}.$$

Keep in mind that the boundary conditions are included but are not explicitly written down. Each of the valleys one through three is twice degenerate and has the  $g$ -type scattering of the inter-valley phonon scattering of Sect. 2.4.2 as transitions within each degenerate group, i.e.

$$L_f^{S,i} = \begin{pmatrix} L_f^i & S^i \\ S^i & L_f^i \end{pmatrix},$$

where  $S^i$  is the  $g$ -type transition between the opposing  $i$ -th valleys. Using Eq. (4.62), we find that our system of equations can be expressed as

$$\begin{pmatrix} L_f^1 - S^1 & 2s^{12} & 2s^{13} & 4L_V^1 \\ 2s^{21} & L_f^2 - S^2 & 2s^{23} & 4L_V^2 \\ 2s^{31} & 2s^{32} & L_f^3 - S^3 & 4L_V^3 \\ a_f^1 & a_f^2 & a_f^3 & A_V \end{pmatrix} \begin{pmatrix} (\tilde{G}^{11})_{11} & (\tilde{G}^{12})_{11} & (\tilde{G}^{13})_{11} & \tilde{G}^{1V} \\ (\tilde{G}^{21})_{11} & (\tilde{G}^{22})_{11} & (\tilde{G}^{23})_{11} & \tilde{G}^{2V} \\ (\tilde{G}^{31})_{11} & (\tilde{G}^{32})_{11} & (\tilde{G}^{33})_{11} & \tilde{G}^{3V} \\ (\tilde{G}^{V1})_{11} & (\tilde{G}^{V2})_{11} & (\tilde{G}^{V3})_{11} & \tilde{G}^{VV} \end{pmatrix}$$

$$= \begin{pmatrix} \mathbb{I} & 0 & 0 & 0 \\ 0 & \mathbb{I} & 0 & 0 \\ 0 & 0 & \mathbb{I} & 0 \\ 0 & 0 & 0 & \mathbb{I} \end{pmatrix}, \quad (4.67)$$

where the Green's functions  $(\tilde{G}^{ij})_{11}$  are the Green's function of, say, spin up electrons and only one of each of the valleys per axis  $\mathbf{k}$ -space.

It follows that the Green's function of terminal currents can be computed using

$$\begin{aligned} (\mathbf{G}^{I_C})^t &= \left( (\mathcal{P}_1^{1,\uparrow})^t \quad (\mathcal{P}_1^{2,\uparrow})^t \quad (\mathcal{P}_1^{3,\uparrow})^t \quad (\mathcal{P}^V)^t \right) \\ &\times \begin{pmatrix} (\tilde{G}^{11})_{11} & (\tilde{G}^{12})_{11} & (\tilde{G}^{13})_{11} & \tilde{G}^{1V} \\ (\tilde{G}^{21})_{11} & (\tilde{G}^{22})_{11} & (\tilde{G}^{23})_{11} & \tilde{G}^{1V} \\ (\tilde{G}^{31})_{11} & (\tilde{G}^{32})_{11} & (\tilde{G}^{33})_{11} & \tilde{G}^{1V} \\ (\tilde{G}^{V1})_{11} & (\tilde{G}^{V2})_{11} & (\tilde{G}^{V3})_{11} & \tilde{G}^{VV} \end{pmatrix} + \Xi^t, \end{aligned}$$

knowing that the terminal current response for a fluctuation of spin down electrons, or a fluctuation in the other valley of a degenerate group, will result in an identical response of the terminal current.

The multiplicities in the PSD due to the spin degeneracy can be computed with the PSD of fluctuations

$$P_{\xi\xi} = \begin{pmatrix} P_{\xi\xi}^\uparrow & 0 \\ 0 & P_{\xi\xi}^\downarrow \end{pmatrix}$$

with

$$P_{\xi\xi}^\uparrow = P_{\xi\xi}^\downarrow.$$

Due to the degeneracy, the Green's functions of spin up and spin down are indistinguishable as well and therefore, with Eq. (4.66), we obtain that the PSD of the quantities  $X$  and  $Y$  reads

$$P_{XY} = \mu_{\text{spin}} (\mathbf{G}_X^\uparrow)^t P_{\xi\xi}^\uparrow \mathbf{G}_Y^\uparrow,$$

where the spin multiplicity  $\mu_{\text{spin}}$  of Eq. (2.92) is used. The valley degeneracy can be included by applying Eq. (4.66) once more. The scattering rates in between the three degenerate valleys give us a PSD of fluctuations for the spin up electrons which decomposes as

$$P_{\xi\xi}^\uparrow = \begin{pmatrix} P_{\xi\xi}^{11} & P_{\xi\xi}^{12} & P_{\xi\xi}^{13} \\ P_{\xi\xi}^{21} & P_{\xi\xi}^{22} & P_{\xi\xi}^{23} \\ P_{\xi\xi}^{31} & P_{\xi\xi}^{32} & P_{\xi\xi}^{33} \end{pmatrix},$$

where the submatrices of the PSD are given by

$$P_{\xi\xi}^{ii} = \begin{pmatrix} p_{\text{diag}}^{ii} & p_{\text{offdiag}}^{ii} \\ p_{\text{offdiag}}^{ii} & p_{\text{diag}}^{ii} \end{pmatrix}, \quad P_{\xi\xi}^{ij} = \begin{pmatrix} p^{ij} & p^{ij} \\ p^{ij} & p^{ij} \end{pmatrix}, \quad \text{with } i \neq j, \text{ and } i, j = 1, 2, 3.$$

Hence, the PSD of terminal currents is given by

$$P_{XY} = \mu_{\text{spin}} \sum_{i=1}^3 \left[ \mu_{\text{val}} (\mathbf{G}_X^{i,\uparrow})_1^t (p_{\text{diag}}^{ii} + p_{\text{offdiag}}^{ii}) (\mathbf{G}_Y^{i,\uparrow})_1^* + \sum_{\substack{j=1 \\ j \neq i}}^2 \mu_{\text{val}}^2 (\mathbf{G}_X^{i,\uparrow})_1^t p^{ij} (\mathbf{G}_X^{j,\uparrow})_1^* \right],$$

where  $(\mathbf{G}_X^{i,\uparrow})_1$  denotes the Green's function of the terminal current at contact  $X$  with a fluctuation in the first true state of the  $i$ -th valley of a spin up electron. Moreover, the valley multiplicity  $\mu_{\text{val}}$  is given by Eq. (2.93).

## 4.8 Simultaneous Small Signal Analysis

In Sect. 3.6.2 about the adjoint method for the small signal analysis, we found that we need to solve Eq. (3.71) in order to find the admittance parameters. When we derived the adjoint method for the Green's functions in Sect. 4.5, we found that we need to solve Eq. (4.31), which is identical to Eq. (3.71) of the small signal method. That means that we can solve, say, Eq. (4.31) to find  $\mathbf{y}_e$  and then insert it into Eq. (3.70) to compute the admittance parameters and insert it into Eq. (4.30) to compute the Green's functions of the terminal currents.

This approach fails when we reduce our systems using their inherent degeneracies. The reduction was shown in Sect. 2.4.6 for the BE, which is identical in the case of the small signal analysis, and in Sect. 4.7 for the Green's function equation of the Langevin source approach. Ultimately, we will end up with two distinct matrices given by something analogous to Eq. (2.91) in the small signal case (but including the PE) and given by Eq. (4.62) in the Green's function case. Note how the multiplicities multiply the columns in the small signal case, while they multiply the rows in the Green's function case. Furthermore, the small signal terminal current contains multiplicities since it is the sum of all contributions of the degenerate states, while the Green's function of the terminal current does not contain multiplicities since it is the response to a fluctuation in a single true state.

This means that for a general set of multiplicities, we need to choose between two options. We can solve the system of equations only once to compute both

the admittance parameters and the Green's functions of terminal currents but then we need to disregard reductions in the size of the system from degeneracies. On the other hand, we can reduce the size by considering degeneracies but then we need to solve two distinct systems for the admittance parameters and for the Green's functions.

Fortunately, there are special cases in which we can both exploit degeneracies and compute both admittance parameters and Green's functions by only solving one system of equations. Assume that we have multiple degenerate groups, each comprising the same number  $\mu$  of true states. Then, we can write down Eq. (3.71) for the admittance parameters as

$$\begin{pmatrix} L_f & L_V \\ \mu a_f & A_V \end{pmatrix}^t \begin{pmatrix} \mathbf{y}_f \\ \mathbf{y}_V \end{pmatrix} = \begin{pmatrix} L_f^t & \mu a_f^t \\ L_V^t & A_V^t \end{pmatrix} \begin{pmatrix} \mathbf{y}_f \\ \mathbf{y}_V \end{pmatrix} = \begin{pmatrix} \mu \mathbf{r}_f \\ \mu \mathbf{r}_V \end{pmatrix},$$

where we explicitly wrote down the multiplicities. Note that  $L_f$ ,  $a_f$ , and  $L_V$  might contain several degenerate groups analogous to Eq. (4.62), each with the same multiplicity  $\mu$ , but we do not need to keep track of them individually and therefore we lumped them together. Furthermore, the r.h.s. of Eq. (3.71) contains the projection operator on the terminal current and therefore it needs to be multiplied by the multiplicities if degeneracy is considered. To keep track of this multiplicity we wrote it down explicitly on the r.h.s. and defined the vectors  $\mathbf{r}_f$  and  $\mathbf{r}_V$  which contain the r.h.s. but only for the true states considered.

If we substitute,

$$\mathbf{y}_f = \mu \mathbf{y}'_f \tag{4.68}$$

and then divide the upper equation of the system by  $\mu$ , we find

$$\begin{pmatrix} L_f^t & a_f^t \\ \mu L_V^t & A_V^t \end{pmatrix} \begin{pmatrix} \mathbf{y}'_f \\ \mathbf{y}_V \end{pmatrix} = \begin{pmatrix} L_f & \mu L_V \\ a_f & A_V \end{pmatrix}^t \begin{pmatrix} \mathbf{y}'_f \\ \mathbf{y}_V \end{pmatrix} = \begin{pmatrix} \mathbf{r}_f \\ \mu \mathbf{r}_V \end{pmatrix}. \tag{4.69}$$

Now recall that the matrix on the l.h.s. is the same as in Eq. (4.62), where the multiplicity appears in front of  $L_V$ . Finally, consider the r.h.s. of Eq. (4.31) to see that the term  $\text{offdiag} \left( A_{o/e}^{\text{AC, BE}} \right)^t$  contains the multiplicity  $\mu$  in the rows running over PE indices and therefore, compared to the single particle current projector of the Green's function, we will find an additional multiplicity in the PE related part of the r.h.s. of Eq. (4.69).

Thus, Eq. (4.69) is the system of equations we need to solve in order to find the Green's functions of terminal currents, which only differs by a single substitution. We can repeat the procedure for arbitrary levels of degeneracies with identical multiplicities for each degenerate group, such as in the present case, where we consider spin and valley degeneracy.

Let us summarize what this means for us. First, we can set up the system of equations for the Green's functions with the degeneracies spin and valleys Eq. (4.67). Then we use the resulting matrix to compute  $\mathbf{y}_e$  with Eq. (4.31). Thereafter,  $\mathbf{y}_e$  can be used directly to compute the Green's functions of the terminal current with Eq. (4.30). Finally, we multiply the elements enumerated by indices of the BE by  $\mu_{\text{spin}} \mu_{\text{val}} = 4$ , as in Eq. (4.68), and insert the resulting vector into Eq. (3.70) to compute the admittance parameters.



# Chapter 5

## Results

In this chapter we apply the self-consistent and deterministic solver of the PE, SE, and BE to a nanoscale double gate nMOSFET. We start out in Sect. 5.1 by introducing the device and the associated spatial and energy grids as well as parameters involved in the computation of scattering rates. In Sect. 5.2 matters of stability, convergence, and the time to solve the system are discussed. Furthermore, we examine stationary solutions involving quantities such as electron densities and currents as well as mobilities and velocities at various operating voltages. Results concerning the small signal behavior, i.e. quantities such as admittance parameters and cut-off frequencies, are presented in Sect. 5.3. The last section is concerned with results involving quantities derived from the power spectral density (PSD) or the Green's functions of terminal currents.

The focus of this chapter is to understand how short-channel devices work. However, we will also simulate devices with longer channels and show their behavior wherever it is appropriate to provide a context for the results of the short-channel device.

### 5.1 Device

We simulate the double gate nMOSFET with a silicon channel depicted in Fig. 5.1 with three different channel lengths given by  $L_G = 16$  nm, 100 nm, and 500 nm. In the remainder, we will refer to the nMOSFETs with the three channel lengths as the 16 nm, 100 nm, and 500 nm devices. The shades of the silicon region indicate the donor doping concentration shown in Fig. 5.2. The doping density of the channel and the contact regions will stay the same throughout this chapter – we will only vary the length of the channel.

We will operate the device in the common-source configuration described in Sects. 3.6.3 and 4.6.4. In the stationary case this means that we measure voltages relative to the source and apply the same bias to both gates.

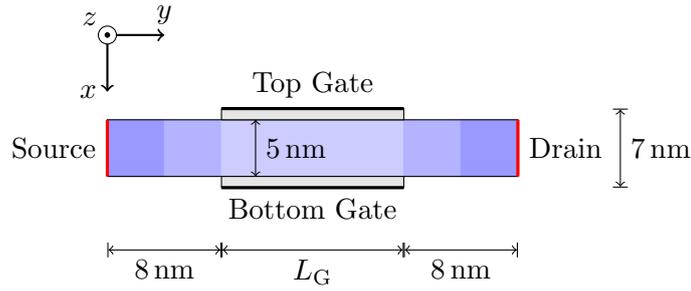


Figure 5.1: Nanoscale double gate nMOSFET with a silicon channel. The shades indicate the doping density given in Fig. 5.2.  $L_G$  is a free parameter that varies the length of the gate and the lightly doped channel region beneath it.

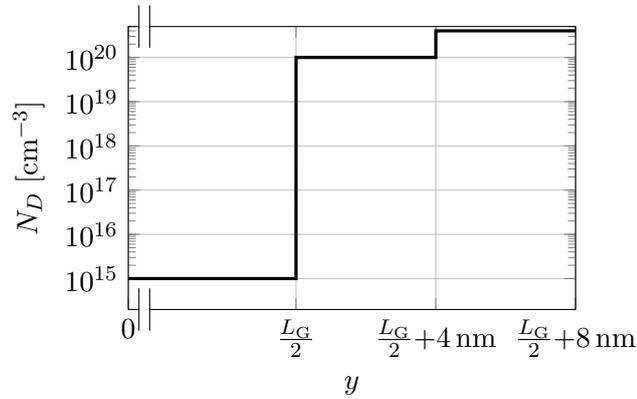


Figure 5.2: One side of the symmetric donor doping density.

### 5.1.1 Scattering Parameters

We include the scattering processes introduced in Sect. 2.4.2, i.e. the elastic acoustic phonon scattering, the inter-valley acoustic phonon scattering, and the surface roughness scattering. For the reasons explained in Sect. 2.4.2, we omit impurity scattering but adjust the deformation potential  $D_{ac}$  of the elastic acoustic phonon scattering in the contact regions such that we recover approximately correct mobilities in these highly doped regions.

While the phonon energies of the inter-valley phonon scattering are considered fixed, the deformation potential of the elastic acoustic phonon scattering and the root mean square height and correlation length of the surface roughness scattering are usually considered to be fitting parameters. It is already known that the deformation potential seems to change its value depending on the width of the channel [79]. Therefore we need to find a value suitable for the

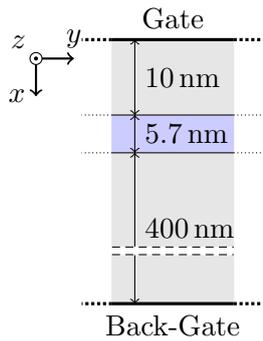


Figure 5.3: Infinitely long device to approximate the SOI MOSFET of Ref. [113]. Transport is in  $y$ -direction, where we apply periodic boundary conditions. In order to mimic an SOI device, we only apply a voltage at the top gate.

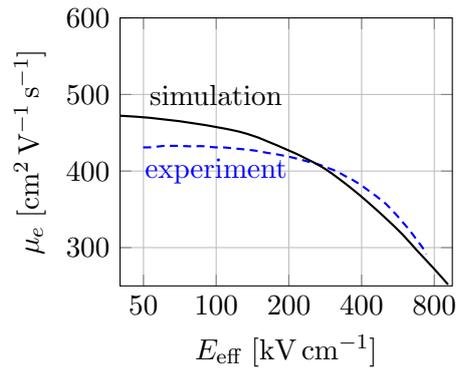


Figure 5.4: Simulated electron mobility vs. effective field of the device in Fig. 5.3 (black solid line) as well as experimental results taken from Ref. [113] for SOIs with a 5.7 nm channel width (blue dashed line).

thinness of the channel of our device. Concerning the parameters of the surface roughness scattering we must acknowledge that their values depend on which kind of model is employed. Thus, in practice, the surface roughness parameters are usually fitted to the mobility at high inversion densities, where the surface roughness scattering limits the mobility [112].

For the purposes of the present work, we are not interested to fit the scattering rates to within minute errors of experimental data. Instead, we are satisfied with the approximate reproduction of the mobilities in the operating range of our device. To this end, we will simulate the infinitely long silicon-on-insulator (SOI) device depicted in Fig. 5.3 in order to compute low-field mobilities which can be compared to the experimental data published in Ref. [113]. Note that mobilities of nanoscale devices strongly depend on the channel width [86] and therefore it is imperative to gauge our scattering parameters with experimental results of devices with similar channel properties.

We can construct an infinitely long device with the device simulator described in Chap. 2 by using three direct grid points – and two adjoint grid points – in transport direction and applying periodic boundary conditions. The driving electric field in transport direction is constructed with a constant gradient of the potential in between the three direct grid points.

The low-field drift mobility of electrons can be extracted straightforwardly from the current density  $j$  and the driving electric field  $E_{el}$  as

$$\mu_e = \frac{j}{qnE_{el}},$$

where  $n$  is the electron sheet density computed via Eq. (A.2) and  $j$  is the current density computed via Eq. (A.8).

In order to compare the mobilities to the results of Ref. [113], we need to determine the effective field in confinement direction as

$$E_{\text{eff}} = \frac{q}{\kappa_{\text{Si}}} \left( \frac{1}{2}n + w_{\text{Si}}N_D \right), \quad (5.1)$$

where  $\kappa_{\text{Si}}$  is the permittivity of silicon and  $w_{\text{Si}} = 5.7 \text{ nm}$  is the width of the silicon channel. Note that for such a device the average electric field in confinement direction is not the same as the effective field [69].

We choose the deformation potential as  $D_{\text{ac}} = 20 \text{ eV}$ ; the surface roughness scattering parameters as  $\gamma = 1.5$  [89],  $\Delta = 0.3 \text{ nm}$ ,  $L = 1 \text{ nm}$ ; and the inter-valley phonon scattering parameters as in Tab. 2.1. Then we find that the mobility conforms well to the experimental data of Ref. [113] as is shown in Fig. 5.4.

As is explained in Ref. [112], the mobility is limited by surface roughness scattering for high effective fields. At low effective fields, Coulomb scattering, i.e. impurity scattering, limits the mobility. Since we did not include any impurity scattering we already expected that the mobility is overestimated at low effective fields. However, the threshold voltage of the double gate nMOSFET of Fig. 5.1 is at effective fields of about  $\approx 200 \text{ kV cm}^{-1}$  and therefore the mobility fits reasonably in the on-state of the double gate nMOSFET.<sup>1</sup>

Since impurity scattering is missing, we need another way to account for the mobility reduction in the highly doped contact regions. Therefore, we are going to adjust the deformation potential to reduce the mobilities to values in accord with Ref. [114]. Note that Ref. [114] specifically explains how the doping density becomes irrelevant to the mobility past a certain threshold since inactive dopants only marginally reduce the mobility. Hence, the mobility should be expressed through the carrier density rather than the doping density. For our simulations, we use the assumption that the dopants are always fully ionized (see Sect. 2.2 on the PE). Therefore we find that the carrier density in the highly doped regions is the same as the doping density (cf. Fig. 5.18). Moreover, Ref. [114] shows that the type of dopant influences the mobility – for  $n$ -type doping, Arsenic leads to smaller mobilities than Phosphorus. But Ref. [114] is about bulk mobilities which are generally larger than the mobilities inside a device. We are once again not interested in the details and will be satisfied to have our mobility in the approximate range of the experimental data. Thus, we choose the deformation potentials as given by Tab. 5.1. Note, that simulations were performed for an effective field of  $E_{\text{eff}} = 400 \text{ kV cm}^{-1}$  which would be the on-state of the double gate nMOSFET.

---

<sup>1</sup>Generally, we cannot directly compare the effective fields of an SOI MOSFET and a double gate MOSFET since the shapes of the confining fields are different. The effective field in the double gate MOSFET was determined by calculating the effective field per gate and then multiplying by a factor of two, which is approximately the same as using Eq. A.10.

$N_D$ [cm <sup>-3</sup> ]	$\mu_e^{\text{exp}}$ (Ref. [114]) [cm <sup>2</sup> V <sup>-1</sup> s <sup>-1</sup> ]	$D_{\text{ac}}$ [eV]	$\mu_e$ [cm <sup>2</sup> V <sup>-1</sup> s <sup>-1</sup> ]
$1 \times 10^{20}$	$\approx 75$	50	77
$4 \times 10^{20}$	$\approx 40$	65	47

Table 5.1: Adjustment of deformation potential in highly doped contact regions to approximate the mobilities  $\mu_e^{\text{exp}}$  given in Ref. [114] for Phosphorus (left column). The chosen deformation potentials and the simulator mobilities are shown on the right. Simulations were done for an effective field of  $E_{\text{eff}} = 400 \text{ kV cm}^{-1}$ .

### 5.1.2 Contacts

As was shown in Sect. 2.4.3 our source and drain contacts are represented by generation and recombination (GR) rates relative to fictional thermal baths representing the source and drain contacts. The GR rates are characterized by the recombination velocity  $v_{\text{GR}}$  which is an input parameter that determines the resistance of the contacts. A high recombination velocity reduces the resistance and in the limit of  $v_{\text{GR}} \rightarrow \infty$ , we recover Dirichlet boundary conditions at the contacts since in that case the GR rate dwarves all other contributions from the BE.

Conversely, choosing  $v_{\text{GR}}$  small increases the resistance of the contacts and therefore also the noise contributed by the contacts. For this work, we choose a recombination velocity of

$$v_{\text{GR}} = 10^6 \text{ m s}^{-1}, \quad (5.2)$$

which is neither so small that the contact noise drowns the channel noise, nor so large that the contacts' resistance disappears. When fitting to real devices, we would need to ensure that we match the actual contact resistances.

### 5.1.3 Discretization

With the scattering parameters determined, we can turn our attention back to the device shown in Fig. 5.1. Before we can start actual simulations, we need to estimate the resolution of our discretization described throughout Chap. 2. In total, we have grid parameters in five dimensions we need to adjust. As always, the resolution of the grids is a trade-off between accuracy and simulation time as well as memory requirements. As a measure of accuracy, we use the drain current in the on-state, i.e.  $V_{\text{GS}} = V_{\text{DS}} = 0.7 \text{ V}$ , of the 16 nm device. If the drain current does not change significantly when the grid is chosen finer, we consider the resolution to be sufficient.

Let us start with the spatial dimensions. In confinement direction, i.e.  $x$ -direction, we solve the Schrödinger equation and obtain a set of wave functions. Here, we need to make sure that we have sufficiently many grid points to resolve

the highly localized wave functions during strong inversion. Moreover, too few grid points will be unable to resolve wave functions of higher subbands. We use a spacing of  $\Delta x = 0.2 \text{ nm}$  which is sufficient since halving the spacing would change the drain current by less than 1 %.

The transport direction, i.e.  $y$ -direction, has to be discretized depending on the channel length  $L_G$  because – clearly – we cannot afford the same spatial resolution for a 500 nm device as for a 16 nm device. The rationale behind the resolution in transport direction is that the potential energy may not change significantly between any two grid points. However, the built-in fields due to the doping profile still exist which means that we have to resolve the contact regions with the same resolution for all devices. Nevertheless, in the homogeneously doped channel, the gradient of the potential decreases with the device length. Thus, we keep the spatial resolution of the highly doped contact regions at a spacing of  $\Delta y = 0.5 \text{ nm}$  while in the lowly doped channel region between the gates we use a grid spacing of 0.5 nm, 3 nm, and 18 nm for the 16 nm, 100 nm, and 500 nm devices, respectively. Doubling the resolution in transport direction changes the drain current by only about 1.5 % in the case of the 16 nm device and 0.5 % for the 500 nm device.

We also need to limit the number of subbands we consider in the BE. A higher number of subbands is particularly expensive to consider because each additional subband involves scattering processes to all other subbands. Since the subbands are associated with progressively higher eigenvalues and the distribution function drops off exponentially with higher energies, we can truncate the series of subbands when the eigenvalues are far away from the lowest eigenvalue. Following this notion, we omit all subbands that are more than about 0.5 eV away from the lowest subband and thus we have 10 subbands in total: six subbands in the energetically lowest  $x$ -valleys on the  $\mathbf{k}$ -axis in confinement direction and two subbands in each of the  $y$ - and  $z$ -valleys. Increasing the total number of subbands to 17 only changes the drain current by about 1.5 %.

Due to the Fourier expansion described in Sect. 2.4.4, we do not have an angular dependence but a dependence on the number of harmonics. Naturally, a higher number of Fourier harmonics coefficients is better suited to represent anisotropic quantities as for example the distribution function far away from equilibrium. In equilibrium, however, we find that the distribution function becomes isotropic in  $\mathbf{k}$ -space and therefore only the radially symmetric zeroth harmonic is necessary to describe the solution. The further we are away from equilibrium the more harmonics we need to describe the anisotropy. Fortunately, the Fourier harmonics series converges rapidly enough that even in the on-state of our device, using the zeroth to seventh harmonic is sufficient. Going up to the 15th harmonic would only change the drain current by 0.8 %. For the purposes of this work, we can ignore negative Fourier harmonics since they are proportional to the sine of the angle relative to the transport direction (cf. Eq. (2.55)).

Therefore the appearance of negative harmonics of the distribution function would require a force perpendicular to the transport direction, e.g. the Lorentz force exerted on a carrier by a magnetic field. However, we do not contemplate such forces in this work.

For the energy grid we usually use a spacing which is a fraction of the largest inelastic phonon energy. This is due to the mapping of the phonon energies onto a multiple of the grid spacing (cf. Sect. 2.4.2). The phonon with the largest energy contributes the strongest to the total scattering rate. Thus, by choosing the energy grid such that the largest phonon energy can be resolved exactly, we reduce the error in the inter-valley phonon scattering rate. In the case of the energy grid, or  $H$ -grid, we find that we cannot judge the quality of our results solely based on the error in the drain current. If it were so, we might be satisfied with an energy grid of 1/12th the largest phonon energy, which is  $\Delta H = 5.17$  meV. However, as we will see in Sect. 5.3.2 the  $H$ -transform leads to discontinuities in the derivatives w.r.t. the potential which can be reduced by increasing the density of the  $H$ -grid. Therefore we settle for a grid spacing of  $\Delta H = 2.583$  meV – 1/24th of the largest phonon energy. The error in the on-state current between the two grid spacings is only about 0.01 %. In principle, for a given subband, energy space extends from the subband energy to infinity but since the distribution function declines exponentially with higher energies, we truncate the energy grid about 0.5 eV above the maximum of the energetically highest subband.

Using these grid densities we find that the Jacobian of the Newton-Raphson approach for the stationary solution or the linear systems of the small signal or noise analyses have approximately 1.3 to 1.9 million variables with about 0.01 % to 0.02 % non-zero elements – depending on the operating point.

## 5.2 Stationary Solutions

### 5.2.1 Numerics and Solver Requirements

In order to solve the linear systems of the BE and PE, we use ILUPACK [68], which – although single-threaded – is a fast, memory efficient, and reliable sparse linear system solver based on an incomplete LU-decomposition [115].

As was mentioned in Sect. 2.5.1, the SE needs to be solved at each iteration of the Gummel loop or the Newton-Raphson approach. The one-dimensional SE used in our work is not particularly demanding and can be fully solved in a matter of seconds, i.e. for all subband energies and wave functions. We use the eigenvalue solver routine DSTEVR for real symmetric and tridiagonal matrices contained in the LAPACK functions of the Intel Math Kernel Library [72].

The deterministic solver described in Chap. 2 is generally well behaved and converges reliably. However, there are some caveats. In order to use the Newton-

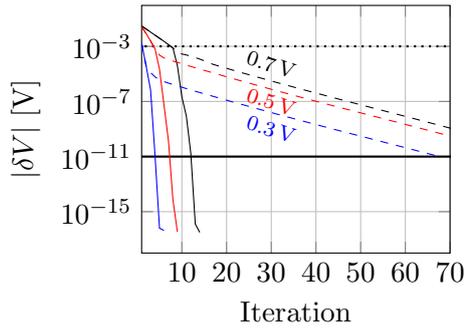


Figure 5.5: Convergence given by the maximum in absolute change of the potential at each iteration step for  $V_{GS} = 0.3$  V (blue),  $0.5$  V (red),  $0.7$  V (black) of the Newton-Raphson approach (solid lines) and the Gummel type iteration (dashed lines) for  $V_{DS} = 0.7$  V. The threshold where the Newton-Raphson approach is turned on is indicated by the dotted horizontal line.

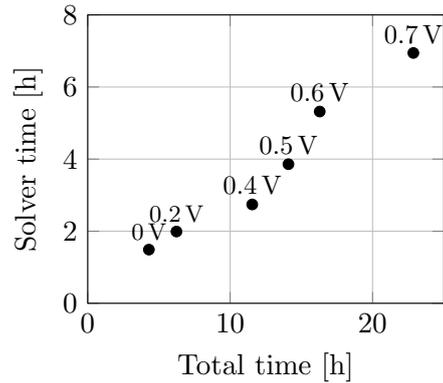


Figure 5.6: The total runtime of the simulator, i.e. setting up the system and solving, vs. the solver time for a stationary operating point running through an initial Gummel loop and then using the Newton-Raphson approach. The drain bias is  $V_{DS} = 0.7$  V while the gate bias  $V_{GS}$  is chosen as indicated at the data points.

Raphson approach for the PE, SE, and BE described in Sect. 2.5.2, we need to be close to the solution. If our initial guess is too far off, the Newton-Raphson approach will – in general – not converge. On the other hand, the Gummel type iteration described in Sect. 2.5.1 works for a much wider range of initial guesses. Therefore, we usually use a Gummel type iteration to come close to the solution and then switch over to the Newton-Raphson approach to make use of the quadratic convergence. Figure 5.5 shows the convergence of the solver in terms of the magnitude of the maximum change in the potential. The Gummel type iteration converges linearly which takes an excessive amount of iterations in the on-state of the device. Switching over to the Newton-Raphson approach as soon as we are somewhat close to the solution, we can see how the Newton-Raphson approach converges rapidly after just a few iterations even for high bias operating conditions. The accuracy of the solution is only limited by the machine precision.

Prior to our own publication on the full Newton-Raphson approach [63], Ref. [29] already improved upon the convergence of the Gummel loop by implementing a partial Newton-Raphson scheme, i.e. they used an approximation to the derivative of the subband energy w.r.t. the potential and omitted perturbations to the wave functions. In that case, convergence is still linear although somewhat faster than the Gummel type iteration. Nevertheless, it cannot possibly compete with the quadratic convergence of the full Newton-Raphson scheme. More importantly, the Newton-Raphson approach is an essential ingredient to

the setup of self-consistent small signal and noise analyses.

The quadratic convergence is useful in another way: we can solve our system of equations up to errors within the machine precision without excessive iterations. For stationary solutions it is not necessary to solve our equations up to errors of  $10^{-16}$  V in the potential, however, the small signal and noise analyses have operating frequencies in which sensitivity to errors in the stationary solution has an impact and therefore a proper convergence is essential (cf. Sect. 5.3.1).

In order to solve the system of equations with the discretization described in Sect. 5.1.3, we use computers with two 12-core Intel Xeon E5-2690 v3 CPUs at 2.6 GHz and with 384GB of RAM. For certain simulations, e.g. for doubling the grid density of one dimension, the memory requirements were higher and machines with up to 768GB of RAM were used.

Depending on the operating point the system of equations with the grid densities described in Sect. 5.1.3 can be solved within a day or less. Figure 5.6 shows total time of running the simulator – which includes the time to set up the system of equations – and the solver time until full convergence of the Newton-Raphson based approach with an initial Gummel type iteration. Note that although solving the system of equations takes a significant amount of time, we need a similar time simply to set up the whole matrix for each iteration. Bear in mind that this is the time for the computation of a *single* operating point. The full characterization of a device needs up to two orders of magnitude more CPU time.

## 5.2.2 Distribution Function

The direct result of our solver is the distribution function  $f_m^\nu(y, H)$ . In order to visualize it for each subband, we need to compute

$$f^\nu(y, \mathbf{k}(E, \phi)) = \sum_m f_m^\nu(y, H - \varepsilon^\nu(y)) Y_m(\phi)$$

with  $E = H - \varepsilon^\nu(y)$ . In particular, we are interested in the distribution function in the channel, where the electron velocity is high, and at the drain contact, in order to inspect how our contacts work. Figure 5.7 shows the two positions A and B where we will visualize  $f$ .

First, let us take a look at the device with  $L_G = 500$  nm in the on-state, i.e.  $V_{GS} = V_{DS} = 0.7$  V. Figure 5.8 shows the distribution function of the energetically lowest subband in the center of the device at position A. The energetically lowest subband contains most of the electrons and therefore it is representative of how the distribution function behaves. It is evident that the distribution function is so close to equilibrium that it seems completely radially symmetric.

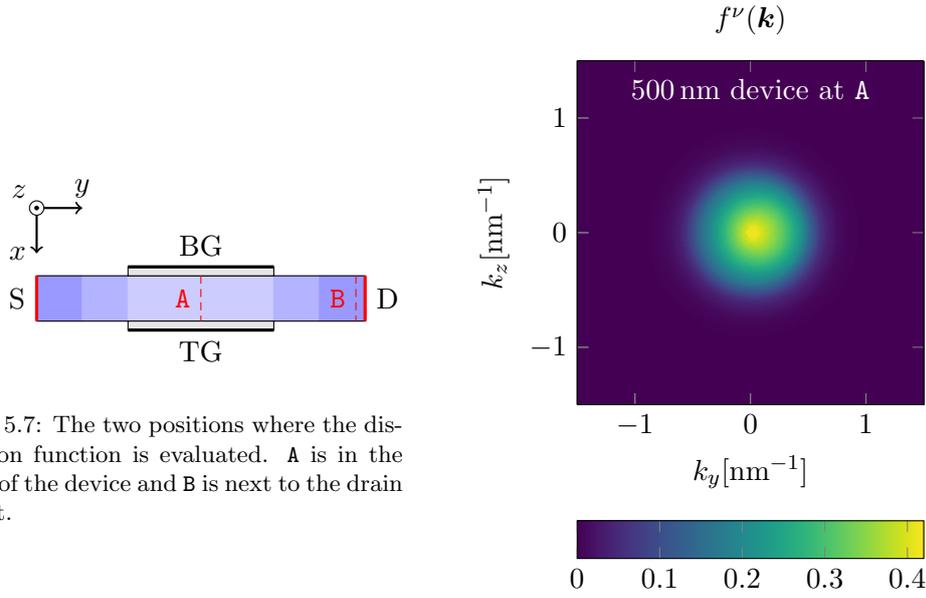


Figure 5.7: The two positions where the distribution function is evaluated. A is in the center of the device and B is next to the drain contact.

Figure 5.8: Distribution function of the energetically lowest subband at position A of the 500 nm device for  $V_{GS} = V_{DS} = 0.7V$ .

This is typical for scattering dominated transport and is due to the length of the device being much larger than the mean free path of electrons.

Evaluating the distribution function of the 16 nm device in the center yields Fig. 5.9 which paints another picture. The crescent moon shape is typical for ballistic transport far from equilibrium and it manifests due to the strong acceleration in transport direction (positive  $y$ -direction) with only little scattering back to smaller energies which would randomize the direction of electron momenta. Note that we simulated the distribution function shown in Fig. 5.9 using up to and including the 15th Fourier harmonic in order to better visualize the non-equilibrium distribution function. Even with this many Fourier harmonics, we can observe how the distribution function becomes slightly negative, which should not happen in principle. The negative parts of the distribution function are an artifact of the truncation of the Fourier series and they become smaller with an increasing number of harmonics. Nevertheless, in practice it is acceptable that the distribution function becomes slightly negative in the ballistic regime since the BE remains stable and the associated errors in the observables remain negligible. In our case, the difference between using up the 7th Fourier harmonic and using up to the 15th Fourier harmonic is a mere 0.8% in the stationary drain current of the on-state.

From the distribution function of the 16 nm device in the channel it is immediately clear that a model based on the drift-diffusion equations cannot describe

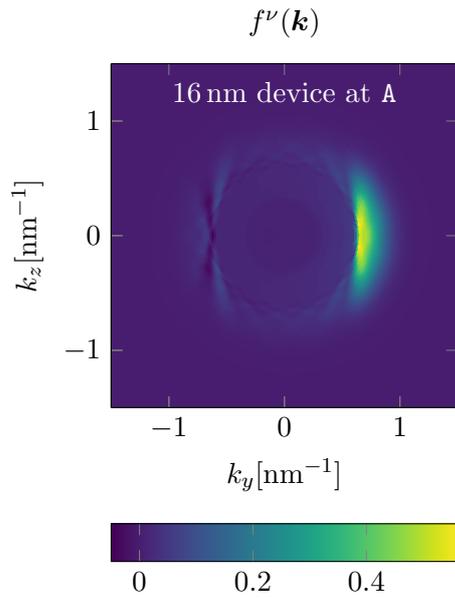


Figure 5.9: Distribution function of the energetically lowest subband at position **A** of the 16 nm device for  $V_{GS} = V_{DS} = 0.7V$ . This plot was generated using up to the 15th Fourier harmonic.

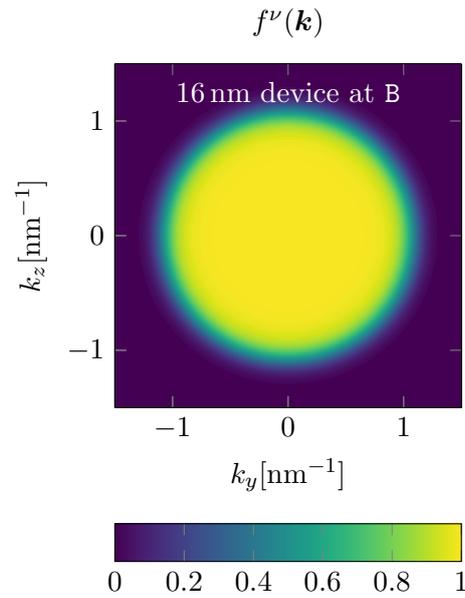


Figure 5.10: Distribution function of the energetically lowest subband at position **B** of the 16 nm device for  $V_{GS} = V_{DS} = 0.7V$ .

the device adequately as the derivation of the drift-diffusion model relies on the assumption that the distribution function is radially symmetric, i.e. close to equilibrium [9, 10].

As was mentioned in Sect. 5.1.1, we adjust the deformation potential in the contact regions to account for the mobility reduction in the highly doped contact regions. Figure 5.10 shows the distribution function of the energetically lowest subband at the drain contact of the 16 nm device at  $V_{GS} = V_{DS} = 0.7V$ . Since the distribution function is radially symmetric at these mobilities, scattering acts as a global resistance. Therefore the adjustment of the deformation potential in the contact regions is a reasonable approach to approximate the mobility reduction.

Figure 5.10 also makes it obvious how important the Pauli principle is. A large portion of the low-energy region is fully occupied and therefore no scattering into these states is possible. Since the drift-diffusion model does not contain the Pauli principle, this is another argument as to why the BE is necessary to model the transport of such devices.

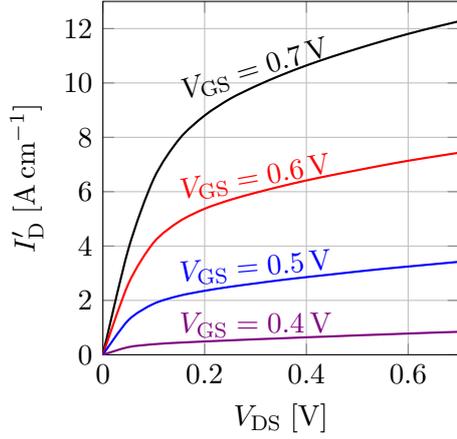


Figure 5.11: Drain current  $I'_D$  vs. drain voltage  $V_{DS}$  of the 16 nm device for  $V_{GS} = 0.4$  V (violet line),  $V_{GS} = 0.5$  V (blue line),  $V_{GS} = 0.6$  V (red line),  $V_{GS} = 0.7$  V (black line).

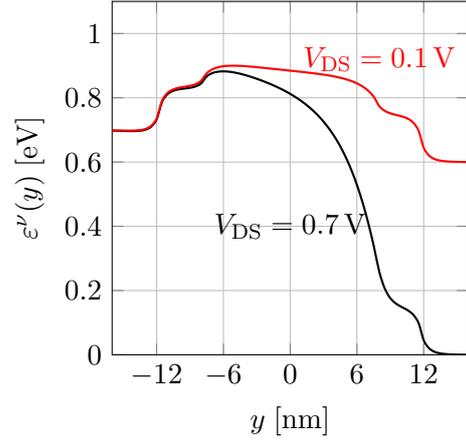


Figure 5.12: Subband energy of the energetically lowest subband of the 16 nm device for  $V_{GS} = 0.7$  V and two different drain biases,  $V_{DS} = 0.1$  V (red line),  $V_{DS} = 0.7$  V (black line). DIBL becomes apparent at the maximum of the subband energy.

### 5.2.3 Current and Subband Energies

In the stationary case, the continuity equation shown in Eq. (3.23) implies that the current throughout the device is constant. Therefore the stationary drain current equals the current density

$$\begin{aligned} I'_D &\equiv qj(y) = q \mu_{\text{spin}} \mu_{\text{val}} \sum_{\nu} \int_{\varepsilon^{\nu}(y)}^{\infty} \frac{d^2k}{(2\pi)^2} v^{\nu}(\mathbf{k}) f^{\nu}(y, \mathbf{k}) \\ &= q \mu_{\text{spin}} \mu_{\text{val}} \sum_{\nu} (T_{\text{HV}}^{\nu})_{yy} Z^{\nu} \int_{\varepsilon^{\nu}(y)}^{\infty} dH (v_y^{\nu})_1(y, H) f_1^{\nu}(y, H), \end{aligned}$$

where  $I'_D$  is defined as the charge current from drain to source per length in  $z$ -direction, i.e. the current at the drain pointing inwards of the device (see Eq. (A.9) for a discretization of the current density). Note that since our  $\mathbf{k}$ -space is Herring-Vogt transformed, the group velocity carries the transformation matrix with itself (cf. Eq. (2.50)). Equivalently we can also evaluate the terminal current through the Ramo-Shockley theorem as if there were no small signal perturbation (cf. Eq. (3.44)).

The drain current vs. the drain voltage of the 16 nm device is shown in Fig. 5.11. Due to the short channel, the device is strongly affected by drain-induced barrier lowering (DIBL) [5, 116]. Figure 5.12 shows the subband energy of the energetically lowest subband for  $V_{GS} = 0.7$  V and two different drain biases  $V_{DS} = 0.1$  V and 0.7 V. When the drain bias increases, the subband energy on the drain side decreases and pulls down the maximum of the subband energy in

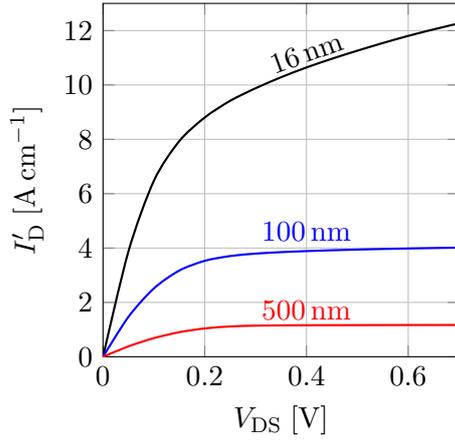


Figure 5.13: Drain currents  $I_D$  vs. drain bias  $V_{DS}$  at  $V_{GS} = 0.7$  V of the 16 nm (black line), 100 nm (blue line), and 500 nm (red line) devices.

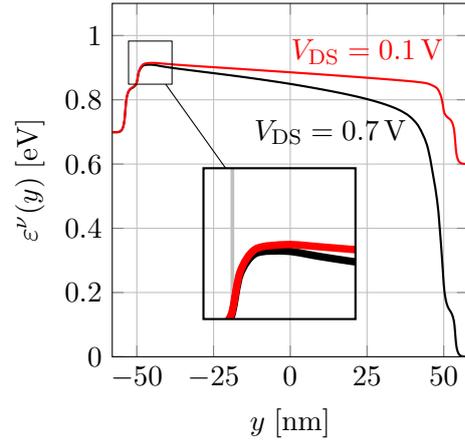


Figure 5.14: Subband energy of the energetically lowest subband of the 100 nm device for  $V_{GS} = 0.7$  V and two different drain biases,  $V_{DS} = 0.1$  V (red line),  $V_{DS} = 0.7$  V (black line). DIBL is barely noticeable.

the channel. Then the potential energy barrier that electrons need to surpass decreases and therefore more current can flow from source to drain.

In contrast to that, the current in the 100 nm and 500 nm devices shown in Fig. 5.13 saturates since the potential of the drain is too far from the subband maximum to influence the energy barrier in the channel as is shown in Fig. 5.14. The drain current vs. the gate voltage of the 16 nm device, on both a linear and logarithmic scale, is shown in Fig 5.15. The device's threshold voltage can be read off as about  $V_{th} \approx 0.35$  V. Note how in all these plots, even well below the threshold voltage, the current does not show any numerical noise which is a testament to the deterministic nature of our simulator.

With the deterministic solver, we can also plot the current density against position and energy space and hence resolve which carrier states contribute to the transport. The density of the current in energy space is defined as

$$j(y) =: \int dH \mathcal{J}(y, H),$$

which, in discretized  $H$ -space, is the contribution of each energy grid point to the current density. Figure 5.16 shows a contour plot of  $\mathcal{J}$  in  $y$ - and  $H$ -space of the 500 nm device. The white region in the plot represents inaccessible energies below the energetically lowest subband energy. Any conducting electron states must have higher energies. Thus the lowest subband energy constitutes a barrier that needs to be surpassed. Carriers which were injected with high enough energies will then cascade down the slope to the drain side, always staying close to the conduction band edge. Since transport in the 500 nm device is scattering

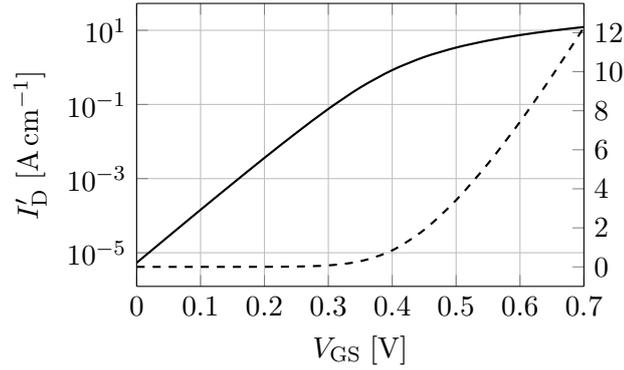


Figure 5.15: Drain current  $I_D$  vs. gate voltage  $V_{GS}$  of the 16 nm device at  $V_{DS} = 0.7$  V on both a logarithmic (solid; left axis) and linear scale (dashed; right axis).

dominated, carriers will scatter to lower energies as soon as they picked up enough kinetic energy.

The same plot for the 16 nm device is shown in Fig. 5.17. Here, electrons seem to move mostly along horizontal lines – along constant  $H$  values – from source to drain, since the mean free path is on the order of magnitude of the device length. This is what is called ballistic transport where energy is only converted between potential and kinetic types but none of it is dissipated through interactions with the environment by scattering processes. Naturally, since there is still some degree of scattering left, we can observe that the current density is not a perfectly horizontal band but smears out to lower energies in the vicinity of the drain contact.

#### 5.2.4 Electron Density, Velocity, and Mean Energy

The electron sheet density is given by

$$n(y) = \mu_{\text{spin}} \mu_{\text{val}} \sum_{\nu} n^{\nu}(y) = \mu_{\text{spin}} \mu_{\text{val}} \frac{1}{Y_0} \sum_{\nu} \int_{\varepsilon^{\nu}(y)}^{\infty} dH Z^{\nu} f_0^{\nu}(y, H) \quad (5.3)$$

and it is depicted in Fig. 5.18 for the 16 nm device at  $V_{DS} = 0.7$  V (see Eq. (A.5) for the discretization of the density). In the contact regions, the electron density is high due to the high doping density. In the channel, we can control the density with the gate voltage over many orders of magnitude. Note that even for  $V_{GS} = -0.7$  V, the deterministic solver is able to compute a perfectly smooth electron density even though the difference between the density in the channel and at the contacts is more than 18 orders of magnitude.

In order to plot the density in confinement direction, we need to multiply the density of a subband by the distribution of electrons in confinement direction,

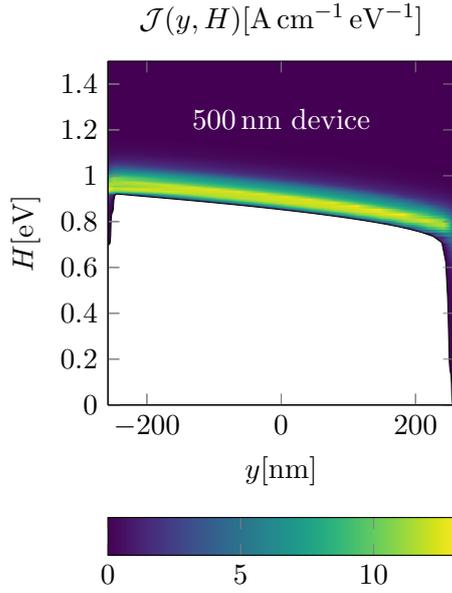


Figure 5.16: Drain current density  $\mathcal{J}$  as a function of  $y$  and  $H$  of the 500 nm device at  $V_{\text{DS}} = V_{\text{GS}} = 0.7 \text{ V}$ . The white region represents inaccessible energies below the energetically lowest subband.

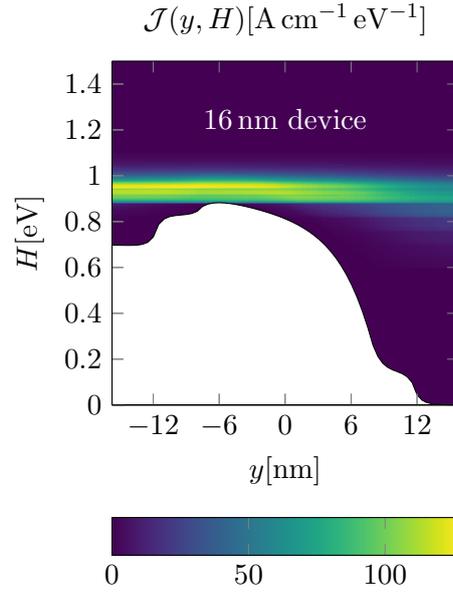


Figure 5.17: Drain current density  $\mathcal{J}$  as a function of  $y$  and  $H$  of the 16 nm device at  $V_{\text{DS}} = V_{\text{GS}} = 0.7 \text{ V}$ . The white region represents inaccessible energies below the energetically lowest subband.

which is given by the probability of finding an electron, i.e. the wave function squared:

$$n(\mathbf{r}) = \mu_{\text{spin}} \mu_{\text{val}} \sum_{\nu} n^{\nu}(y) |\Psi^{\nu}(\mathbf{r})|^2.$$

Figure 5.19 shows the electron density throughout the device for  $V_{\text{DS}} = 0.7 \text{ V}$  and  $V_{\text{GS}} = 0 \text{ V}$ , i.e. in the off-state. We can clearly see how the doping density increases the density in the vicinity of the source and drain contacts (cf. Fig. 5.1). In the channel the electron density is reduced by about ten orders of magnitude due to the low bias applied to the gates. On the other hand, applying a gate bias of  $V_{\text{GS}} = 0.7 \text{ V}$  to both gates, electrons are accumulated in the channel region as is shown in Fig. 5.20, where the density only drops by about two orders of magnitudes.

The average electron velocity is given by

$$v(y) = \frac{j(y)}{n(y)}$$

and it is shown in Fig. 5.21 for the 16 nm device in the on-state resulting from a simulation with up to the 7th Fourier harmonic and a simulation with up to the

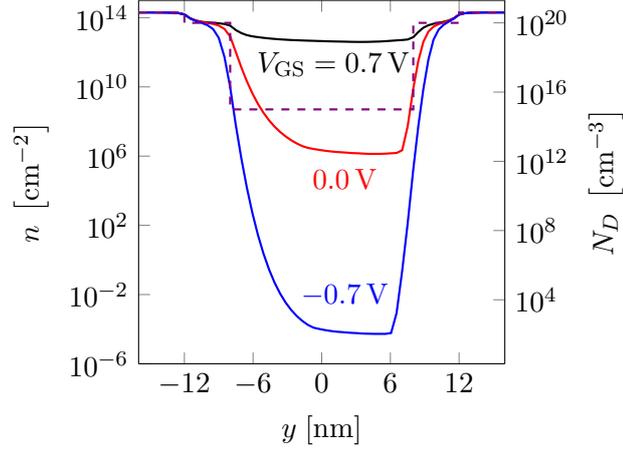


Figure 5.18: Electron sheet density  $n(y)$  of the 16 nm device at  $V_{DS} = 0.7$  V and  $V_{GS} = 0.7$  V (black line; left axis),  $V_{GS} = 0.0$  V (red line; left axis),  $V_{GS} = -0.7$  V (blue line; left axis). The donor doping density is given for reference (violet dashed line; right axis).

15th harmonic. Note how the velocity is small in the contact regions and then increases in the channel until it surpasses the electron velocity of bulk silicon of about  $10^7$   $\text{cm s}^{-1}$ . This is called a velocity overshoot and it is typical for the electron behavior in devices. Usually the velocity's maximum is lower than what we see in our case. This is because we approximate the band structure by the simplistic parabolic model (see Sect. 2.4.1). A correction through a non-parabolicity factor [54, 70] or a model based on a more general interpolation approach as in Ref. [117] would reduce the maximum velocity. We choose to use the parabolic band structure since it is the fastest method and it has sufficient accuracy for our needs. Note that velocity overshoot cannot be reproduced with a drift-diffusion model because it is tantamount to carrier heating which is only described by the higher moments contained in the hydrodynamic approach.

From Fig. 5.21 we can also learn that the velocity for a simulation with up to the 15th harmonic is only marginally different from our usual simulations using up to the 7th Fourier harmonic. Therefore, we conclude that simulations using up to the 7th harmonic accurately represent the actual solution of our system.

The average energy of electrons as measured relative to the energetically lowest subband  $\varepsilon^{\nu_{\min}}$  is given by (see Eq. (A.11) for the discretization)

$$\begin{aligned} \langle E \rangle(y) &= \frac{\mu_{\text{spin}} \mu_{\text{val}} \sum_{\nu} \int \frac{d^2 k}{(2\pi)^2} [E^{\nu}(\mathbf{k}) + \varepsilon^{\nu}(y) - \varepsilon^{\nu_{\min}}(y)] f^{\nu}(y, \mathbf{k})}{\mu_{\text{spin}} \mu_{\text{val}} \sum_{\nu} \int \frac{d^2 k}{(2\pi)^2} f^{\nu}(y, \mathbf{k})} \\ &= \frac{\mu_{\text{spin}} \mu_{\text{val}}}{Y_0 n(y)} \sum_{\nu} Z^{\nu} \int dH [H - \varepsilon^{\nu_{\min}}(y)] f_0^{\nu}(y, H). \end{aligned}$$

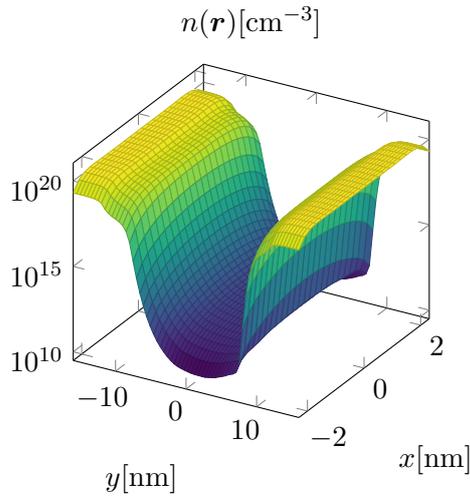


Figure 5.19: Electron density  $n(\mathbf{r})$  as a function of  $x$  and  $y$  throughout the 16 nm device for  $V_{DS} = 0.7$  V and  $V_{GS} = 0$  V.

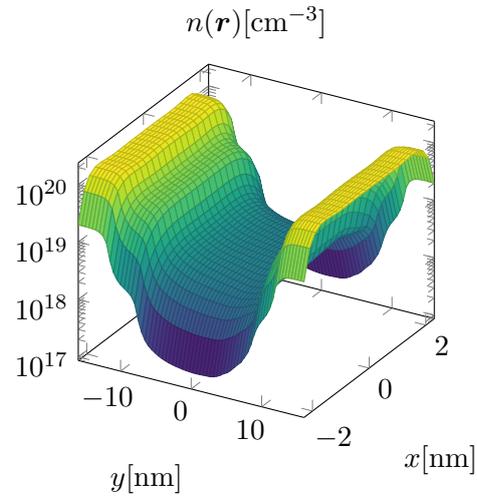


Figure 5.20: Electron density  $n(\mathbf{r})$  as a function of  $x$  and  $y$  throughout the 16 nm device for  $V_{DS} = V_{GS} = 0.7$  V.

The average energy of the 16 nm device is shown in Fig. 5.22 with and without the Pauli principle. Note that the mean energy in the highly doped contact regions is higher when the Pauli principle is included since it prohibits carriers to lose energy (cf. Fig. 5.10).

### 5.3 Small Signal Analysis

In this section, we will compute the small signal parameters of the device shown in Fig. 5.1. Although we will use the full admittance matrix of Eq. (3.72) in order to show that our device satisfies conservation laws, for the actual results we will only refer to the device in common-source configuration. How to arrive at the admittance parameters of the common-source configuration is straightforward and has been shown in Sect. 3.6.3.

We usually find that the small signal results are numerically sound, albeit with some caveats. First, we will see in Sect. 5.3.1 that conservation laws and symmetries are not perfectly satisfied but with errors that depend on the applied biases and on the frequency of the small signal bias. But barring extreme conditions, we can compute small signal quantities over a wide range of operating conditions, including the deep sub-threshold region and from a few Hz up to and beyond the THz range.

Second, we will see in Sect. 5.3.2 that due to the  $H$ -transformation the derivatives of any quantity w.r.t. the potential or w.r.t. an applied bias is nec-

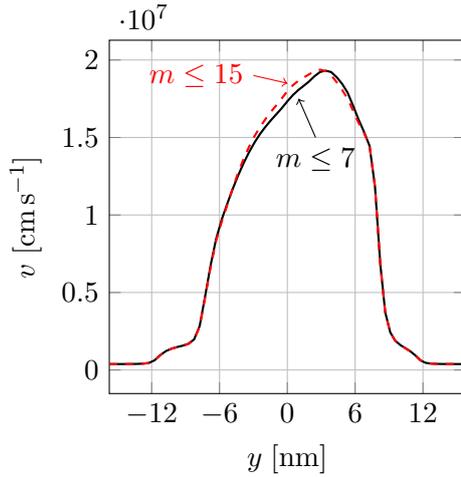


Figure 5.21: Average electron velocity  $v(y)$  in the 16 nm device at  $V_{DS} = V_{GS} = 0.7$  V computed using up to the 7th Fourier harmonic (black solid line) and using up to the 15th Fourier harmonic (red dashed line).

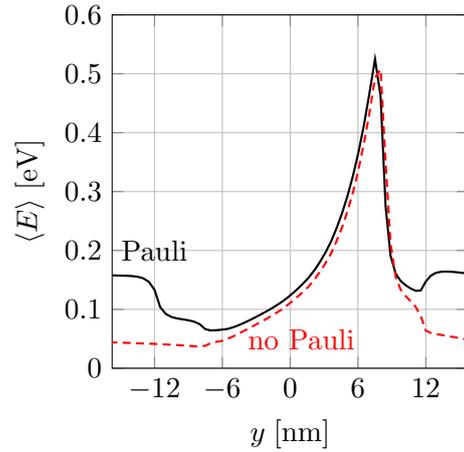


Figure 5.22: Average electron energy  $\langle E \rangle(y)$  in the 16 nm device at  $V_{DS} = V_{GS} = 0.7$  V including the Pauli principle (black solid line) and excluding the Pauli principle (red dashed line).

essarily discontinuous. This is an artifact of the  $H$ -transformation that cannot be avoided but it can be ameliorated by choosing a fine  $H$ -grid spacing.

Solving for the small signal parameters involves the solution of the system of equations shown in Eq. (3.64) or in the adjoint form of Eq. (3.71) with four different vectors on the right hand side, corresponding to small signal biases applied at the four contacts of our device. Since Eqs. (3.64) and (3.71) are linear systems, we can solve them in one step and, moreover, we only need to perform a single LU-decomposition for the four different contact bias vectors on the right hand side. Thus, for a single operating point,  $V_{GS}$  and  $V_{DS}$ , and a single frequency,  $f$ , the system takes about two hours to completely set up and an additional half hour to solve. When computing multiple frequencies, most of the set up time reduced since most terms are not frequency dependent. Bear in mind that for a full characterization of a device, we need to know the admittance parameters at various frequencies and operating points for which we also need the respective stationary solutions. The total time for a full characterization is therefore at least two to three orders of magnitude larger (cf. Fig. 5.6 for the total time to solve for the stationary solution).

In the following we will first make sure that our device fulfills essential conservation laws and symmetries and we will estimate the size of numerical errors we make. Then we will proceed to show the admittance parameters as well as cutoff frequencies and maximum oscillation frequencies. Note that any small signal quantity shown here relates to the intrinsic properties of the device. In order to make a comparison to a real device, we would need to embed the device

in a circuit, adding various capacitances and resistances which would modify the small signal quantities.

### 5.3.1 Conservation Laws and Symmetries

We want to verify that our solver yields physically sound solutions. This can be checked by looking for conservation laws or basic properties of the device that must be fulfilled. The most obvious property is that the small signal current is conserved, i.e.

$$\underline{I}'_{\text{TG}} + \underline{I}'_{\text{BG}} + \underline{I}'_{\text{D}} + \underline{I}'_{\text{S}} = 0.$$

Applying a small signal bias to one contact at a time, we find that the conservation of current implies

$$\sum_C Y'_{C,C'} = 0, \quad C, C' \in \{\text{TG}, \text{BG}, \text{D}, \text{S}\}.$$

Note that due to the calculation of the small signal terminal current via the Ramo-Shockley theorem, we expect the current to be conserved by construction. Figure 5.23 shows the cumulative relative error defined as

$$\Delta Y'_{\text{curr}} := \sum_{C'} \frac{|\sum_C Y'_{C,C'}|}{\sum_C |Y'_{C,C'}|}$$

for different operating conditions and frequencies. For high gate biases the error is quite close to the accuracy of double precision floating point numbers. But we can also see that for low gate biases and low frequencies the relative errors increase. They are still small enough to be manageable but it shows us the ceiling of accuracy we can expect at these operating conditions. It is not clear what exactly leads to the deterioration of the numerical properties but we speculate that we would need to expand the system of equations around zero frequency in order to reduce numerical errors in the low frequency domain.

In Sect. 3.5 a lot of work has been put into understanding and restoring reciprocity numerically. To recall, in equilibrium a device is reciprocal if the current flowing through a contact  $C$  when a small signal bias is applied to a contact  $C'$  is the same as the current flowing through the contact  $C'$  if a small signal bias is applied to contact  $C$ , irrespective of the shape or other properties of the contacts [51]. Thus, reciprocity is a statement about the symmetry of the admittance matrix in equilibrium:

$$Y'_{C,C'} \stackrel{\text{equilibrium}}{=} Y'_{C',C}, \quad C, C' \in \{\text{TG}, \text{BG}, \text{D}, \text{S}\}.$$

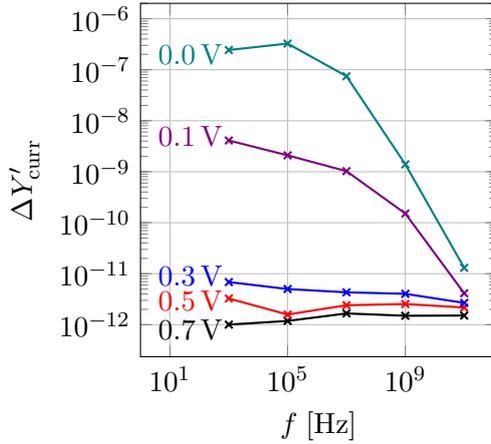


Figure 5.23: Cumulative relative error  $\Delta Y'_{\text{curr}}$  of the 16 nm device for  $V_{\text{DS}} = 0.7$  V and  $V_{\text{GS}} = 0$  V, 0.1 V, 0.3 V, 0.5 V, and 0.7 V as indicated.

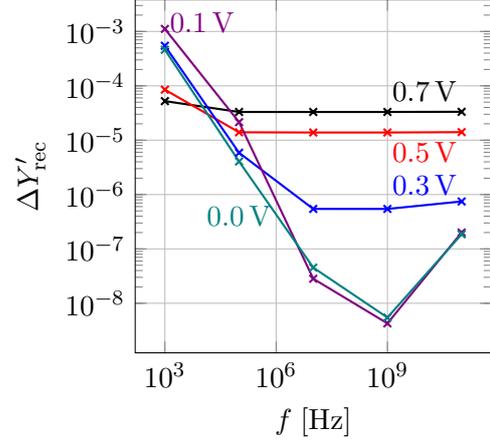


Figure 5.24: Cumulative relative error  $\Delta Y'_{\text{rec}}$  of the 16 nm device for  $V_{\text{DS}} = 0$  V and  $V_{\text{GS}} = 0$  V, 0.1 V, 0.3 V, 0.5 V, and 0.7 V as indicated.

Let us express the cumulative relative error or the deviation from reciprocity as

$$\Delta Y'_{\text{rec}} := \frac{1}{2} \sum_{C, C'} \frac{|Y_{C, C'} - Y_{C', C}|}{|Y_{C, C'}| + |Y_{C', C}|},$$

where we divided by two since the sum over both contacts counts its symmetric summands twice. The relative errors of reciprocity are shown in Fig. 5.24 proving that the procedure to restore reciprocity derived in Sect. 3.5.2 is effective. Bear in mind that without the procedure, reciprocity is violated at least in the tens of percent, therefore a reduction to a relative error in the range of  $10^{-3}$  to  $10^{-8}$  is a noteworthy improvement.

The errors in the reciprocity give us a better estimate for the numerical accuracy of our results. Any result close to equilibrium cannot be more accurate than the error we find in the reciprocity. This is also true for the passivity of the device in equilibrium. A device is considered passive if the matrix  $Y + (Y^*)^t$  is positive definite, i.e. it has positive eigenvalues. Note that  $Y + (Y^*)^t$  is Hermitian and therefore its eigenvalues are strictly real. In order to see whether the device is passive, we want to compute the relative minimum eigenvalue defined as

$$\Delta \lambda_{\min} := \frac{\min(\{\lambda_i\})}{\sum_i |\lambda_i|},$$

where  $\lambda_i$  denotes the  $i$ -th eigenvalue and  $\min(\{\lambda_i\})$  denotes the smallest eigenvalue. Thus, if  $\Delta \lambda$  becomes negative, the device becomes active. Prior to the

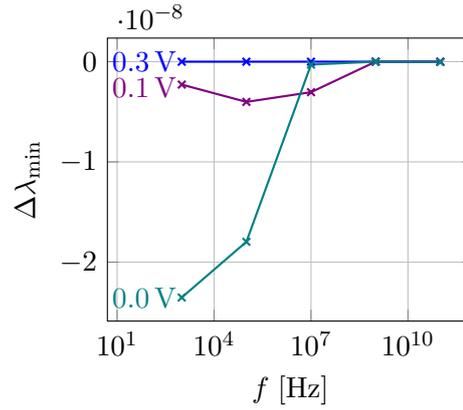


Figure 5.25: Relative error in minimum eigenvalue  $\Delta\lambda_{\min}$  of the 16 nm device for  $V_{DS} = 0$  V and  $V_{GS} = 0$  V, 0.1 V, and 0.3 V as indicated. Curves for gate biases above 0.3 V have even smaller values and would lie on top of the 0.3 V curve. Therefore they were omitted in this plot.

reciprocity restoration procedure of Sect. 3.5.2 the device would become active for certain combinations of operating points and frequencies, but afterwards Fig. 5.25 shows that although the minimal eigenvalue may become negative, its value is in agreement with the value zero considering our previous estimates for the numerical error.

### 5.3.2 Admittance Parameters

The admittance parameters are the small signal response of our device to an applied small signal bias. They tell us about capacitances and conductances inherent to our device. To understand the meaning of each of the parameters, we can express a MOSFET in common-source configuration with a simple small signal equivalent circuit as in Fig. 5.26. Then the admittance parameters read

$$\hat{Y}^{\text{equiv}} = \begin{pmatrix} \frac{i\omega(C_{GS}+C_{DG})}{1+i\omega(C_{GS}+C_{DG})R_G} & \frac{-i\omega C_{DG}}{1+i\omega(C_{GS}+C_{DG})R_G} \\ \frac{g_m-i\omega C_{DG}}{1+i\omega(C_{GS}+C_{DG})R_G} & \frac{i\omega C_{DG}(1+g_m R_G+i\omega C_{GS}R_G)}{1+i\omega(C_{GS}+C_{DG})R_G} + \frac{1}{R_{DS}} \end{pmatrix}, \quad (5.4)$$

which makes it apparent what role each of the capacitances and resistances plays. Here,  $C_{GS}$  is the gate-source capacitance,  $C_{DG}$  is the drain-gate capacitance,  $R_G$  is the gate resistance,  $R_{DS}$  is the channel resistance, and  $g_m$  is the transconductance.

Obviously the admittance parameters of the equivalent circuit are only a crude approximation to reality and equivalent circuits will never be able to capture the complexities of an actual BE solver. However, they enable us to develop an intuition for the most important aspects of a device. Keep in mind that our simulator describes the naked device. To fit such a device to experimental data,

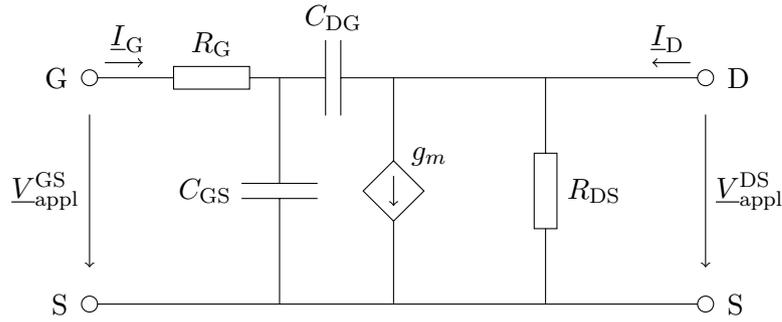


Figure 5.26: Simple small signal equivalent circuit of a MOSFET in common-source configuration.

we would have to embed it by adding additional gate, source, and drain resistances as well as external parasitic capacitances due to the way the device is contacted.

Let us take a look at the real and imaginary parts of the gate self-admittance  $Y'_{GG}$ . From Eq. (5.4) it is clear that up to linear order in the frequency, i.e. at very low frequencies, we will only find an imaginary part to  $Y'_{GG}$  which is proportional to the sum of the capacitances of the gate w.r.t. the other contacts. Indeed, Fig. 5.27 shows the real and imaginary parts for the 16 nm device of the simulated gate self-admittance which reflect this behavior at lower frequencies. In the THz region however, we find that  $Y_{GG}^{\text{equiv}}$  is insufficient to describe the MOSFET's behavior. The bumps of the admittance parameter beyond a THz are not a numerical problem but they occur due to plasma oscillations within the device. At frequencies lower than about a MHz, we observe that the real part is drowned out by numerical noise but since the real part around these frequencies is orders of magnitudes smaller than the imaginary part, we do not need to be concerned by this.

Similarly, Fig. 5.28, which shows the real and imaginary parts of the simulated  $Y'_{GD}$  reflects the capacitance between the drain and gate at lower frequencies (see Eq. (5.4)). At frequencies beyond a THz, we see the impact of plasma oscillations. Similar to the gate self-admittance, we also see how the real part of  $Y'_{GD}$  is governed by numerical noise but we can also note in this case that the real part for these frequencies is negligible compared to the imaginary part.

The admittance parameter  $Y'_{DG}$  expresses the change in drain current when the gate voltage is changed. The gradient in drain current is encoded in the transconductance  $g_m$  and it is the dominant contribution to zeroth order in frequency. Figure 5.29 shows the simulated real and imaginary parts of  $Y'_{DG}$  which – as is obvious from  $Y_{DG}^{\text{equiv}}$  – has a constant real part at lower frequencies due to the transconductance, while its imaginary part is linear in the frequency. Also note the strong impact of plasma oscillations on this parameter in the THz

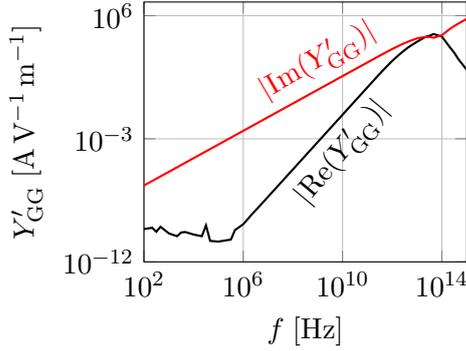


Figure 5.27: Real and imaginary parts of admittance parameter  $Y'_{GG}$  vs. frequency of the 16 nm device at  $V_{DS} = V_{GS} = 0.7$  V.

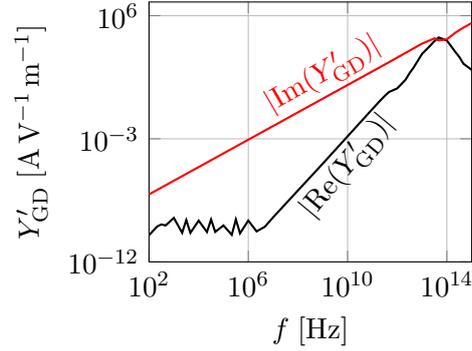


Figure 5.28: Real and imaginary parts of admittance parameter  $Y'_{GD}$  vs. frequency of the 16 nm device at  $V_{DS} = V_{GS} = 0.7$  V.

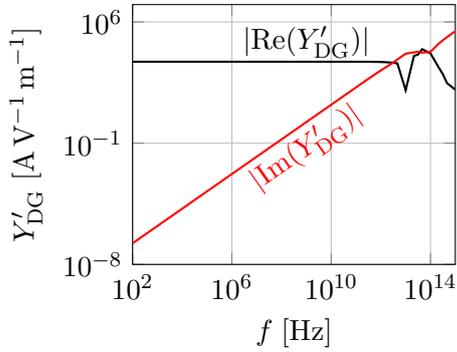


Figure 5.29: Real and imaginary parts of admittance parameter  $Y'_{DG}$  vs. frequency of the 16 nm device at  $V_{DS} = V_{GS} = 0.7$  V.

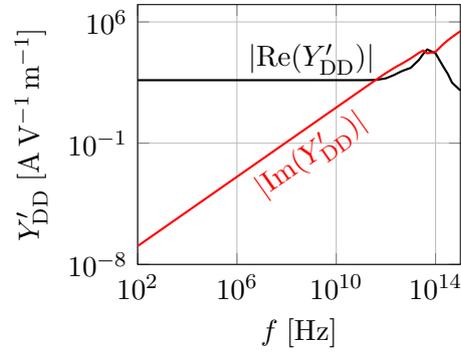


Figure 5.30: Real and imaginary parts of admittance parameter  $Y'_{DD}$  vs. frequency of the 16 nm device at  $V_{DS} = V_{GS} = 0.7$  V.

region.

The drain self-admittance  $Y'_{DD}$  has a constant contribution to the real part due to the channel resistance  $R_{DS}$  while the imaginary part is, again, linear at lower frequencies, as shown in Fig. 5.30. Both  $Y'_{DG}$  and  $Y'_{DD}$  are unaffected by numerical noise at low frequencies, however, plasma oscillations are visible in both admittance parameters at frequencies beyond a THz.

At this point we want to draw attention to the upper frequency limit of our approach. In Figs. 5.27, 5.28, 5.29, and 5.30, we have plotted the admittance parameters up until a frequency of  $10^{15}$  Hz but while our solver can easily handle these and even higher frequencies, our approach cannot yield reasonable results anymore. Frequencies beyond  $10^{14}$  Hz fall within the visible spectrum and the energies associated with these waves are high enough to excite carriers into higher subbands by absorption. Since the physics of such processes are not included in

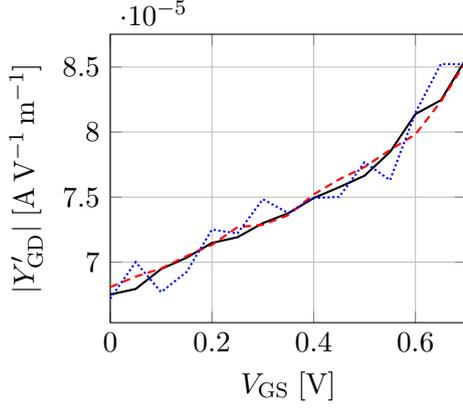


Figure 5.31: Absolute value of  $Y'_{GD}$  vs.  $V_{GS}$  of the 16 nm device at  $V_{DS} = 0.7$  V and  $f = 100$  kHz for the usual grid spacing (black solid line), for double the  $H$ -grid spacing (red dashed line), and for quadruple the  $H$ -grid spacing.

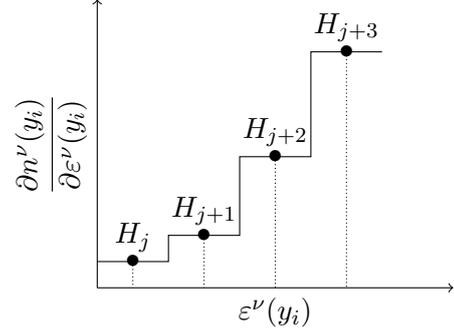


Figure 5.32: Illustration of the discontinuities in the derivative of the density  $n^\nu$  w.r.t. the subband energy  $\epsilon^\nu$ . When the subband energy changes from the box surrounding the grid point  $H_j$  to the box surrounding  $H_{j+1}$ , the derivative changes discontinuously.

our approach, we will only consider frequencies up to  $10^{14}$  Hz in the remainder of this work.

In order to fully understand the numerical properties of the small signal analysis, we want to point out how the admittance parameters behave as functions of contact biases. To this end, we plotted the absolute value of  $Y'_{DG}$  vs. the gate bias for  $V_{DS} = 0.7$  V and for a frequency of 100 kHz in Fig. 5.31. In contrast to the plots w.r.t. the frequency, this curve is not smooth. In fact the curve shown in Fig. 5.31 has a discontinuous derivative w.r.t. the gate bias. These discontinuities are a direct consequence of the  $H$ -transformation which can be illustrated using any function that contains an integral over  $\mathbf{k}$ -space – or  $H$ -space – like the terminal current and thus the admittance parameters. What happens when we change the gate bias is that the potential in the device changes and therefore the subband energies change. But after the  $H$ -transformation, the lowest box in  $H$ -space depends on the subband energy and therefore we find for the derivative of a quantity like, e.g., the density:

$$\frac{\partial n^\nu(y_i)}{\partial \epsilon^\nu(y_i)} = \frac{\partial}{\partial \epsilon^\nu(y_i)} \left[ \frac{1}{Y_0} \sum_j Z^v f^\nu(y_i, H_j) \Delta H^\nu(y_i, H_j) \right] = -\frac{1}{Y_0} Z^v f^\nu(y_i, H_{j_{\min}}),$$

where the  $H$ -box is defined as in Eq. (2.75) and  $j_{\min}$  is the index of the  $H$ -grid denoting the lowest non-zero  $H$ -box, i.e. the box where the lower boundary is the subband energy. Figure 5.32 illustrates the discontinuous derivative. Say we have some terminal bias which we increase a little bit, then we will see that the potential will increase a little bit and the subband energies will decrease. But

the small decrease could just be enough for the subband energy to move from one box to another which will change  $j_{\min}$  and with it the  $H$ -grid point where the rest of the density is evaluated. Therefore a small change in the contact bias can change the derivative of the density w.r.t. that contact bias abruptly.

This does not only apply to the density but to all quantities derived via an integration over  $\mathbf{k}$ -space and therefore we find that the admittance parameters also suffer from these discontinuities as shown in Fig. 5.31. The strength of this numerical issue depends on various circumstances. First of all, if we increase the scattering, the discontinuities become smaller. Second, the dimensionality of  $\mathbf{k}$ -space affects the discontinuities strongly since the density of states  $Z^v$  at the subband energy is different for 1D, 2D, and 3D electron gases. The discontinuities present themselves the worst in the 1D case since the relative volume close to  $\mathbf{k} = 0$  is largest compared to the rest of the integration. In other words, the density of states becomes large for small energies and therefore it enhances the discontinuities.

Fortunately, the illustration in Fig. 5.32 also shows us how to mitigate the effect of the discontinuities. Decreasing the spacing of the  $H$ -grid to  $\Delta H = 2.585$  meV as was explained in Sect. 5.1.3 is sufficient to make the effects of the discontinuities on observables less noticeable. As a side note we should bring to attention that these discontinuities are a reason why large signal simulations using the  $H$ -transformation are numerically not feasible.

### 5.3.3 Stability

We want to investigate the stability of our device in order to show that the plasma oscillations occurring in the THz frequency range cannot be used for self-amplifying THz wave generation. The three conditions that must be satisfied for a device to be unconditionally stable are given by [118]

$$\operatorname{Re}(Y'_{GG}) > 0, \quad \operatorname{Re}(Y'_{DD}) > 0, \quad K > 1,$$

where the Rollet factor is given by

$$K = \frac{2\operatorname{Re}(Y'_{GG})\operatorname{Re}(Y'_{DD}) - \operatorname{Re}(Y'_{DG}Y'_{GD})}{|Y'_{DG}Y'_{GD}|}.$$

Conversely, a device that violates any of these conditions at some operating point and frequency can become unstable and generate plasma waves by itself for some combination of passive generator and load admittances.

Figure 5.33 shows the real parts of the gate and drain self-admittances in the on-state of the device for the THz frequency range. Note that both are positive over the whole frequency range. The Rollet factor is shown in Fig. 5.34 and thus we find that the device is unconditionally stable beyond 3 THz in the on-state.

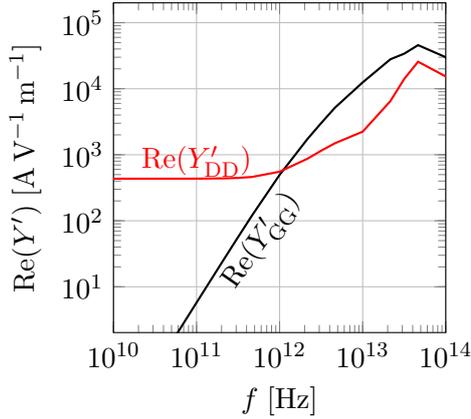


Figure 5.33: Real parts of  $Y'_{GG}$  and  $Y'_{DD}$  vs. frequency for  $V_{GS} = V_{DS} = 0.7$  V of the 16 nm device.

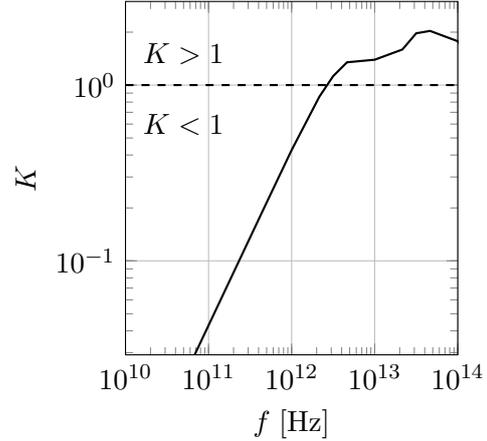


Figure 5.34: Rollet factor  $K$  vs. frequency of the 16 nm device for  $V_{GS} = V_{DS} = 0.7$  V. The device is unconditionally stable above the dashed line.

### 5.3.4 Cutoff and Maximum Oscillation Frequency

From the admittance parameters, we can directly calculate the hybrid parameters. One particularly interesting hybrid parameter is

$$H'_{DG} := \left. \frac{Y'_{DG}}{Y'_{GG}} \right|_{V_{\text{appl}}^D=0},$$

which is the ratio of the small signal currents of the drain and gate contacts when a small signal bias is applied to the gate. Thus, when  $H'_{DG}$  – shown in Fig. 5.35 – drops off to unity, there is no small signal current amplification anymore. The frequency at which this happens is defined as the cutoff frequency  $f_T$ . It can either be computed directly by finding the frequency at which  $|H'_{DG}(f_T)| = 1$  or by noting that  $H'_{DG}$  can be extrapolated from a known frequency  $f_0$  as

$$f_T \approx |H'_{DG}(f_0)| f_0. \quad (5.5)$$

For this work, we compute the cutoff frequency using the approximation of Eq. (5.5) since iterating until we approached the cutoff frequency is too time consuming. In any case, the extrapolated cutoff frequency is sufficiently accurate for all practical purposes. Figure 5.36 shows the cutoff frequencies for the three devices. Note that these are for the intrinsic devices, i.e. no embedding into a circuit has been performed. If we embedded the device into a real circuit, we would find, e.g., a larger gate self-admittance which would reduce the cutoff frequencies.

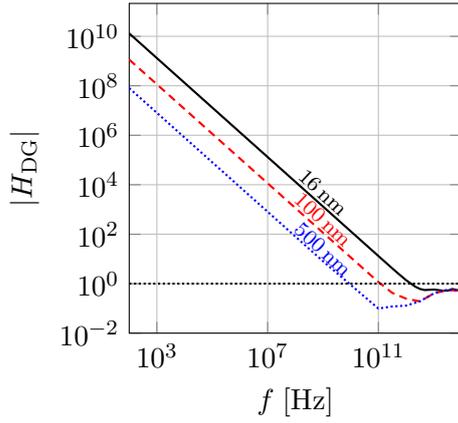


Figure 5.35: Absolute value of the hybrid parameter  $H'_{DG}$  of the 16 nm (black solid line), 100 nm (red dashed line), and 500 nm (blue dotted line) devices at  $V_{DS} = V_{GS} = 0.7$  V.

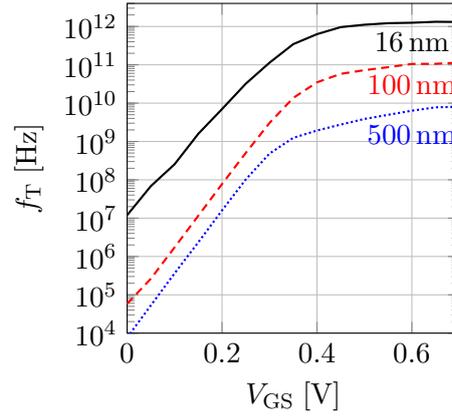


Figure 5.36: Cutoff frequency  $f_T$  vs.  $V_{GS}$  of the 16 nm (black solid line), 100 nm (red dashed line) and 500 nm (blue dotted line) devices at  $V_{DS} = 0.7$  V.

Figure 5.36 shows exactly what is to be expected from the cutoff frequencies: the shorter the channel, the higher the cutoff frequency. Nevertheless, we can still observe the high fidelity of the results even in the deep sub-threshold. Only in the 16 nm device in the deep sub-threshold region can we see the consequence of the discontinuities in the derivatives w.r.t. the gate bias but only if we look carefully.

Another figure of merit that is used to compare device characteristics is Mason's invariant or the unilateral gain [119]. It is defined as

$$U = \frac{|Y'_{DG} - Y'_{GD}|^2}{4 \left( \text{Re}(Y'_{GG})\text{Re}(Y'_{DD}) - \text{Re}(Y'_{GD})\text{Re}(Y'_{DG}) \right)}$$

and it is shown in Fig. 5.37 for the three devices. Recall that at low frequencies the real parts of  $Y'_{GG}$  and  $Y'_{GD}$  were drowned by numerical noise (cf. Figs. 5.27 and 5.28) and therefore the unilateral gain will be meaningless in this frequency range.

The maximum oscillation frequency  $f_{\max}$  is the frequency at which  $U(f_{\max}) = 1$  holds. In the same way as the cutoff frequency, it can be extrapolated from some lower frequency  $f_0$  using

$$f_{\max}^2 \approx U(f_0) f_0^2.$$

Bear in mind that  $f_0$  needs to be chosen above the aforementioned numerical noise in the admittance parameters. In the case of  $V_{GS} = V_{DS} = 0.7$  V shown in Fig. 5.37 this is somewhere above  $10^8$  Hz. Figure 5.38 shows the maximum oscillation frequency for the three devices. Once again, the largest values are

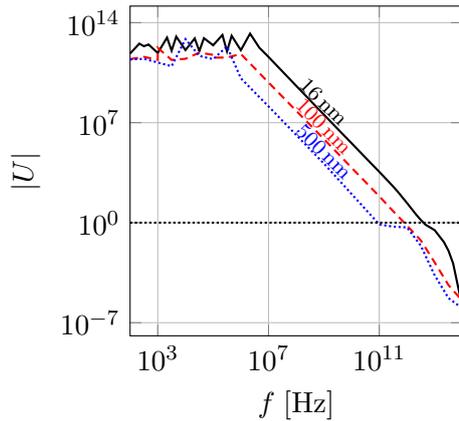


Figure 5.37: Absolute value of the unilateral gain  $U$  of the 16 nm (black solid line), 100 nm (red dashed line), and 500 nm (blue dotted line) devices at  $V_{DS} = V_{GS} = 0.7$  V.

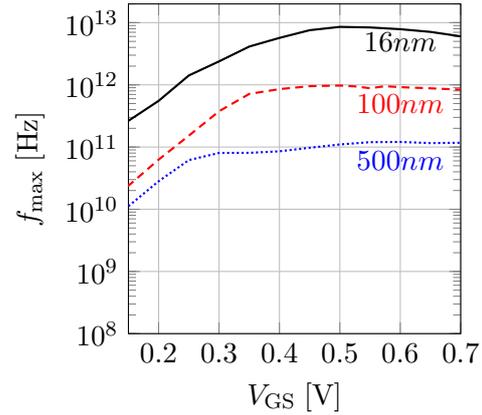


Figure 5.38: Maximum oscillation frequency  $f_{\max}$  vs.  $V_{GS}$  of the 16 nm (black solid line), 100 nm (red dashed line), and 500 nm (blue dotted line), devices at  $V_{DS} = 0.7$  V.

found for the device with the shortest channel. An interesting observation for the 16 nm device is that the maximum oscillation frequency has a maximum around  $V_{GS} = 0.5$  V.

## 5.4 Noise Analysis

The main results of this work concern the self-consistent noise solved with the system of equations comprising PE, SE, and BE. First, we want to present the resulting PSDs in Sect. 5.4.1 in order to get an understanding of what the noise of the terminal currents looks like and what the quality of our results is. Then, in Sect. 5.4.2, we want to verify that our approach to the calculation of noise is indeed reasonable by checking that the simulator fulfills the Nyquist theorem, which is a special form of the fluctuation-dissipation theorem. Once we established the veracity of our model, we compute the usual figures of merit in Sects. 5.4.3 through 5.4.5 consisting of the Fano factor, drain and gate excess noise factors, as well as the cross-correlation.

We will also investigate the qualitative nature of noise in our 16 nm nanoscale double gate MOSFET and find its origin within the device in Sect. 5.4.6. And finally we will investigate and understand the nature of noise in our nanoscale device in Sect. 5.4.7.

We will continually compare our results to the 100 nm and 500 nm devices and reference similar findings in literature in order to demonstrate the integrity of our work.

Note that we can compute admittance parameters and Green's functions of the terminal current at the same time (see Sect. 4.8). Therefore computing the

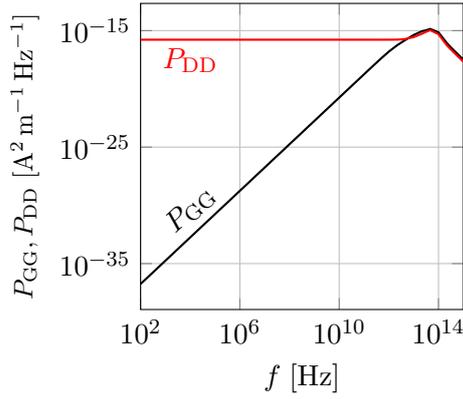


Figure 5.39: Power spectral densities  $P_{GG}$  (black line) and  $P_{DD}$  (red line) vs. frequency of the 16 nm device at  $V_{DS} = V_{GS} = 0.7$  V.

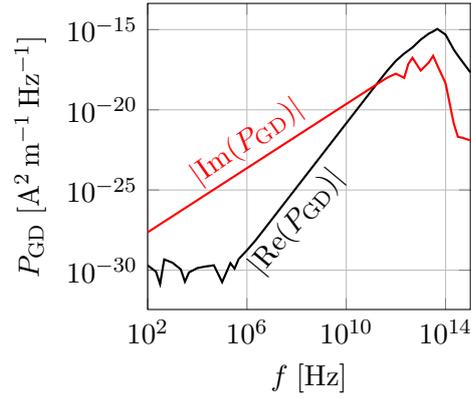


Figure 5.40: Absolute values of real (black line) and imaginary (red line) parts of  $P_{GD}$  vs. frequency of the 16 nm device at  $V_{DS} = V_{GS} = 0.7$  V.

PSDs only takes the additional time needed by Eqs. (4.36) and (4.38). Since this entails the integration over all coordinates, the time we measure is about an hour for a single operating point and frequency.

#### 5.4.1 Power Spectral Density

To get a sense for the noise computed in our devices, we want to take a look at the PSDs. Recall that the PSD is twice the Fourier transform of the correlation function as shown in Eq. 4.2. That means that the PSDs  $P_{GG}$  and  $P_{DD}$  express the variance of the gate and drain terminal current in frequency space, respectively, or – in other words – they quantify the amount of noise contained in the terminal currents. In Eq. (4.3) we noted that  $P_{GG}$  and  $P_{DD}$  need to be real. In the numerical computation we find this to be true – within numerical errors around the machine precision – and therefore we will assume that the PSDs  $P_{GG}$  and  $P_{DD}$  are purely real in the remainder of this work.

Figure 5.39 shows the gate and drain PSDs in the on-state of the 16 nm device. The PSDs exhibit a high numeric fidelity from very low frequencies to beyond the THz. We do not encounter a numerical error floor as in the gate self-admittance shown in Fig. 5.27. In a real device we would encounter flicker noise at low frequencies, but since we do not include the physical processes leading to such noise in our simulations, our PSDs cannot include it either.

Since the two cross PSDs are related through

$$P_{GD} = P_{DG}^*,$$

we will only ever show one cross PSD, i.e.  $P_{GD}$ , in the remainder of this work. The cross PSD is defined as twice the Fourier transform of the correlation func-

tion between the gate and drain terminal currents. Assuming that there is a fluctuation occurring with some frequency somewhere in the device, the cross PSD  $P_{GD}$  will tell us how the gate and drain terminal current fluctuations will behave. Since the coupling between the gate and the channel is based on the involved capacitances (cf. Eq. (5.4)), we expect that a fluctuation in the channel will impact the drain terminal current directly while the resulting gate terminal current is phase shifted since it is mediated by the displacement current. It follows that at low frequencies the resulting cross PSD  $P_{GD}$  should contain the purely real drain terminal current and the purely imaginary gate terminal current Green's function (cf. Eqs. (4.20), (4.36), and (4.38)) and since the transition rate is purely real, the overall PSD is expected to be purely imaginary to leading order.

Indeed, at low frequencies we observe the real and imaginary parts of  $P_{GD}$  as shown in Fig. 5.40 where the total PSD is dominated by its imaginary part. Also note that – similar to Fig. 5.28 – the real part of the cross PSD at frequencies below  $10^5$  Hz is solely governed by numerical errors but since the imaginary part is larger by a few orders of magnitude, we can safely ignore its values in the low frequency domain. As we approach the cutoff frequency of the 16 nm device (cf. Fig. 5.36), we see that the above argument on the real and imaginary parts of the  $P_{GD}$  falls apart. Close to the cutoff frequency both the Green's functions of the gate and drain terminal currents have real and imaginary parts stemming from the displacement currents but also from the time-derivative in the LBE due to the non-quasistationary time-evolution.

### 5.4.2 Nyquist Theorem

The response of a system to fluctuations and the response of the system to external perturbations are governed by the same physical processes and therefore there exists a relation between the two which is compiled in the fluctuation-dissipation theorem [120, 121, 122, 123]. But for our present purposes we only need a special form of this relation which historically preceded the fluctuation-dissipation theorem and is often referred to as the Nyquist theorem [108]. In our current notation, it states that in equilibrium, the noise in a system is related to the small signal response by

$$P_{C,C} \stackrel{\text{equilibrium}}{=} 4k_B T \operatorname{Re}(Y'_{C,C}). \quad (5.6)$$

Note that the admittance  $Y$  contains the effects on terminal currents of the system to an *external* small signal bias applied at a contact while the PSD  $P$  represents the correlation of terminal currents in frequency space due to *internal* fluctuations caused by scattering processes as well as GR processes.

Our simulator conforms excellently with the Nyquist theorem as is shown in

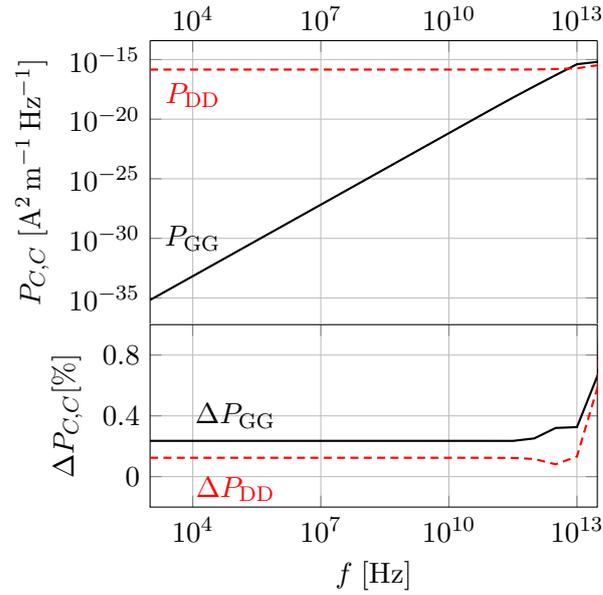


Figure 5.41: Verification of the Nyquist theorem. PSDs (top) and relative error of the noise as given by the Nyquist theorem (bottom) vs. frequency for  $V_{GS} = 0.7\text{ V}$  and  $V_{DS} = 0\text{ V}$  of the 16 nm device.

Fig. 5.41. The maximum error, defined as

$$\Delta P_{C,C} := \left| \frac{P_{C,C} - 4k_B T \operatorname{Re}(Y_{C,C})}{P_{C,C}} \right|,$$

is about 0.3% at low frequencies, which is a remarkably good agreement considering the difference in the computations of the PSD in Sect. 4.6 and the small signal parameters in Sect. 3.6. Note that this level of accuracy in the reproduction of the Nyquist theorem result is only possible due to the careful considerations regarding the numerical restoration of reciprocity of Sect. 3.5.2, the averaging scheme shown in Sect. 4.6.3, and the correct treatment of the GR rate in the PSD as was discussed in Sect. 4.6.2. If one of these had not been considered properly, the Nyquist theorem would have had an error at least an order of magnitude larger. However, beyond a frequency of 10 THz the error in the Nyquist theorem grows and reaches about 6% at a frequency of 100 THz, calling into question the reliability of our simulator at these frequencies and beyond. But as was already discussed in Sect. 5.3.2, these frequencies reach beyond the physical phenomena included in our simulation approach and therefore we cannot simulate these frequencies reasonably anyway.

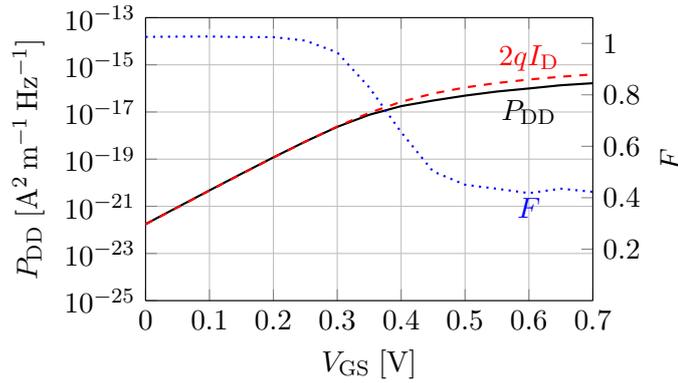


Figure 5.42: PSD of drain current fluctuations (black solid line; left axis) vs.  $V_{GS}$  for  $V_{DS} = 0.7\text{ V}$  and  $f = 100\text{ kHz}$  of the 16 nm device together with the equivalent shot noise  $2qI_D$  (red dashed line; left axis) and the resulting Fano factor  $F$  (blue dotted line; right axis).

### 5.4.3 Fano Factor

We want to take a look at the Fano factor which will be helpful in explaining aspects of our devices' excess noise. The Fano factor is the ratio of the drain current noise and the noise if it were pure shot noise:

$$F := \frac{P_{DD}}{2qI_D}. \quad (5.7)$$

Thus, a Fano factor  $F = 1$  implies pure shot noise, while  $F < 1$  implies that the noise is smaller than pure shot noise – an effect that we will refer to as *suppression of noise*.

In Fig. 5.42 we show the drain current noise  $P_{DD}$  as well as the equivalent shot noise  $2qI_D$  vs. the gate bias. Note how they only deviate at gate biases above the threshold voltage. The associated Fano factor  $F$  has been plotted on the right axis of Fig. 5.42. At low gate biases it is close to unity with a deviation of only about 2.5%, which is an acceptable numerical error. Above the threshold voltage the Fano factor shows the strong suppression of noise at higher gate biases.

Figure 5.43 is the analogous plot for the 500 nm device. Here we see that we obtain shot noise to within less than 1% numerical error and at higher gate biases the noise is suppressed as well. In fact, the suppression is even stronger than in the 16 nm device. However, as we will find out later on, the type of noise in these devices is not the same and therefore the reason for suppression is different. In the 500 nm device we will find noise qualitatively similar to thermal noise (cf. Sect. 5.4.6) while in the 16 nm matters are more complicated as will be discussed in Sect. 5.4.7.

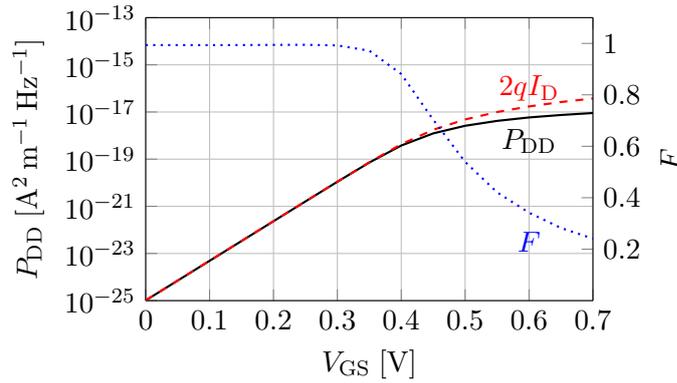


Figure 5.43: PSD of drain current fluctuations (black solid line; left axis) vs.  $V_{GS}$  for  $V_{DS} = 0.7$  V and  $f = 100$  kHz of the 500 nm device together with the equivalent shot noise  $2qI_D$  (red dashed line; left axis) and the resulting Fano factor  $F$  (blue dotted line; right axis).

#### 5.4.4 Excess Noise

We want to compare the excess noise of the 16 nm, 100 nm, and 500 nm devices. To this end, we need to pick a frequency at which we can compare them where they are all in the same mode of operation. For the present section and also for the remainder of this work, we choose the frequency 100 kHz as a trade-off between two competing bounds: First, we need it to be below the cutoff frequencies of all three devices for all relevant operating points (cf. Fig. 5.36) since it does not make sense to compare a small device below the cutoff frequency with a large device above it. Second, we need the frequency to be high enough to avoid impact of the low-frequency numerical errors (cf. Figs. 5.36, 5.27, 5.28) in all quantities we are going to compute. Also note that the reason why we can pick such a low frequency for comparisons is that our simulations do not contain flicker noise.

The drain excess noise factor is given by [124]

$$\gamma = \frac{P_{DD}}{4k_B T g_{D0}}, \quad (5.8)$$

where  $g_{D0}$  is the drain self-admittance  $Y'_{DD}$  but at zero drain bias and zero frequency, i.e.

$$g_{D0} = Y'_{DD} \Big|_{V_{DS}=0 \text{ V}, f=0 \text{ Hz}}.$$

Note that the denominator is essentially the noise as given by the Nyquist theorem of Eq. (5.6) which is the PSD of thermal noise [108, 109]. Thus the drain excess noise is the ratio of the actual drain noise to the thermal noise in equilibrium and at zero frequency. In other words, it measures the deviation from

the thermal noise floor. In practice, we do not need to evaluate  $g_{D0}$  exactly at  $f = 0$  Hz since it is constant in the low frequency domain (cf. Fig. 5.30), therefore the exact frequency is irrelevant as long as it is low enough.

Figure 5.44 shows the drain excess noise factor against the drain bias for the three devices at 100 kHz. Evidently, the drain excess noise seems to be strongly affected by the discontinuities in the derivative w.r.t. the contact bias as was already explained in Sect. 5.3.2. This is a numerical issue that we cannot avoid when using the  $H$ -transformation but its impact is already greatly reduced by reducing the spacing of the  $H$ -grid points.

The drain excess noise factor of the 16 nm device first appears to drop slightly below 1 but then rises and saturates somewhere around  $\gamma = 1.15$  when the drain bias is increased. On the other hand, the drain excess noise of the 500 nm device tapers off and approaches the drain excess noise of the long channel model based on a charge sheet density approach of Ref. [125]. In between, for the 100 nm device, we find that the drain excess noise initially drops as for the 500 nm device but then starts to slowly increase with increasing drain bias. These results are qualitatively in line with the findings of Ref. [126].

In Figure 5.44, we included the plot for the drain excess noise of the long channel mode of van der Ziel found in Ref. [125]. Note that there is no particular reason why our simulator should approach the long channel model of a bulk MOSFET since we consider an actual double gate MOSFET and include many effects ignored by van der Ziel's oversimplified model. However, obtaining similar results in the long channel limit for the drain excess noise indicates that the drain excess noise is mostly only dependent on the channel length and not on the particularities of the confinement.

The drain excess noise vs. the gate bias is shown in Fig. 5.45. For the 500 nm device, the drain excess noise is about 1/2 in the sub-threshold and it rises to about 2/3 above the threshold voltage, which is in accordance with the long channel model of Ref. [125] (see also Ref. [127]). For the 100 nm device, we find the same behavior in the sub-threshold but above the threshold voltage,  $\gamma$  increases above the value of the long channel model as is already evident from Fig. 5.44. In contrast, the 16 nm device looks strikingly different at lower gate biases where its value is significantly larger than for the larger devices. This behavior can be explained with DIBL which is present in the 16 nm device but not in the devices with longer channels (cf. Figs. 5.12 and 5.14). The effect of DIBL is that it increases the current and thus the noise in the numerator of the drain excess noise factor of Eq. 5.8 while  $g_{D0}$  in the denominator remains unaffected by DIBL since it is evaluated at zero drain bias. The enhancement of  $\gamma$  is particularly prominent in the sub-threshold since – as shown in Fig. 5.42 – the device is generating pure shot noise which is directly proportional to the current. For higher gate biases, the noise is suppressed and therefore generated by another mechanism than shot noise – as we will see in Sect. 5.4.7. Thus the

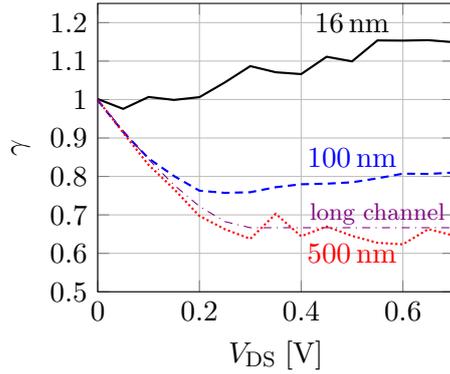


Figure 5.44: Drain excess noise factor  $\gamma$  vs. drain bias  $V_{DS}$  at  $V_{GS} = 0.7V$  and  $f = 100$  kHz for the 16 nm (black solid), 100 nm (blue dashed line), 500 nm (red dotted line) devices. In addition the long channel model of van der Ziel [125] is included (violet dash-dotted line).

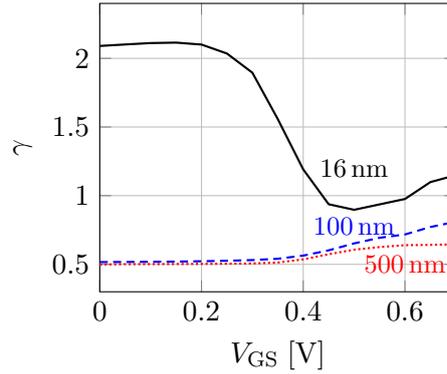


Figure 5.45: Drain excess noise factor  $\gamma$  vs. gate bias  $V_{GS}$  at  $V_{DS} = 0.7V$  and  $f = 100$  kHz for the 16 nm (black solid), 100 nm (blue dashed), 500 nm (red dotted) devices.

drain excess noise factor becomes smaller above the threshold voltage. However, even above the threshold voltage, the drain excess noise comprises DIBL but it is counteracted by the suppression.

This calls into question the utility of the drain excess noise in small devices with DIBL since it directly enhances  $\gamma$ . A comparison of  $\gamma$  using devices of different channel lengths amounts in part to a comparison of DIBL of the devices rather than a comparison of only their noise. In these cases a reasonable approach would be the inspection of quantities like the Fano factor of Eq. (5.7) where both the numerator and denominator are evaluated at the same drain voltage and therefore the enhancement due to DIBL is cancelled. Nevertheless, since  $\gamma$  is often reported by experiments, it is useful to calculate anyway. The finding that the drain excess noise rises for short channel devices and the increase due to DIBL in the sub-threshold is corroborated by Ref. [128] which uses the drift-diffusion and hydrodynamic models.

The gate excess noise is usually defined as [125]

$$\beta = \frac{P_{GG}}{\frac{4}{5}k_B T \frac{(\omega C_{GS,sat})^2}{g_{D0}}}, \quad (5.9)$$

where  $C_{GS,sat}$  is the gate-source capacitance in saturation. The gate excess noise has a similar interpretation as the drain excess noise. If we assume a simple charge sheet model of a MOSFET as in Ref. [125], we find for the PSD of the gate terminal current in saturation:

$$P_{GG}^{ziel} \Big|_{V_{DS} > V_{GS} - V_{th}} = \frac{16}{15} k_B T \frac{(\omega C_{GS,sat})^2}{g_{D0}}. \quad (5.10)$$

In the charge sheet model, we find that the drain self-admittance is the same in equilibrium and in saturation:

$$g_{D0} = g_{D,\text{sat}} := Y'_{DD} \Big|_{V_{DS} > V_{GS} - V_{th}}. \quad (5.11)$$

It follows that  $\beta = \frac{4}{3}$  in saturation if a device's gate noise is accurately described by the simple charge sheet model [129, 130]. Conversely, the charge sheet model yields in equilibrium

$$P_{GG}^{\text{ziel}} \Big|_{V_{DS}=0} = \frac{3}{4} k_B T \frac{(\omega C_{GS,\text{sat}})^2}{g_{D0}}, \quad (5.12)$$

leading to a gate excess noise of  $\beta = \frac{15}{16}$ .

Figure 5.46 shows the gate excess noise vs. the drain bias of the three devices for  $V_{GS} = 0.7\text{ V}$  and  $f = 100\text{ kHz}$ , where the gate-source capacitance has been approximated using the equivalent circuit representation of Eq. (5.4) as

$$\omega C_{GS,\text{sat}} \approx \text{Im}(Y_{GG} + Y_{GD}) \Big|_{V_{DS}=0.7V},$$

which is valid in the low-frequency regime, i.e. as long as  $Y_{GG}$  and  $Y_{GD}$  are linear in the frequency.

Figure 5.46 shows the gate excess noise vs. the drain bias. Before we begin the discussion, let us once again note that the gate excess noise is affected by the discontinuities introduced through the  $H$ -transformation as discussed in Sect. 5.3.2. From Fig. 5.46 it becomes apparent that not even the 500 nm device behaves according to the charge sheet model of Ref. [125] predicting values of the gate excess noise in equilibrium and in saturation. But this is not entirely unexpected since the present description of our nMOSFETs involves a very thin channel where the gate noise is mediated through a response of the SE to potential fluctuations. Therefore it stands to reason that the gate noise *cannot* be plausibly described by Eqs. (5.10) and (5.12). Indeed, we can easily show that the underlying assumptions of the charge sheet model are severely violated in our transport model by showing that Eq. (5.11) does not hold. To this end, we show the drain self-admittance vs. the drain bias for  $V_{GS} = 0.7\text{ V}$  and  $f = 1\text{ kHz}$  in Fig. 5.47. Keep in mind that  $\text{Im}(Y'_{DD}) \ll \text{Re}(Y'_{DD})$  at low frequencies (cf. Fig. 5.30 and Eq. (5.4)) and therefore  $Y'_{DD}$  can be approximated to leading order as purely real. As is evident, for none of the devices, not even for the 500 nm device, does Eq. (5.11) actually hold.

We can conclude that the gate excess noise  $\beta$  does not adhere to its usual interpretation where it can be seen as the deviation from the thermal noise floor. Nevertheless, the gate excess noise is a quantity often quoted as a device

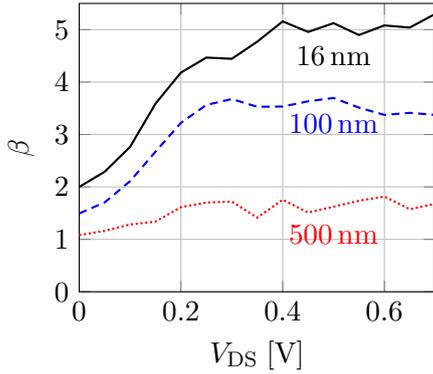


Figure 5.46: Gate excess noise factor  $\beta$  vs. drain bias  $V_{DS}$  at  $V_{GS} = 0.7V$  and  $f = 100$  kHz for the 16 nm (black solid), 100 nm (blue dashed), 500 nm (red dotted) devices.

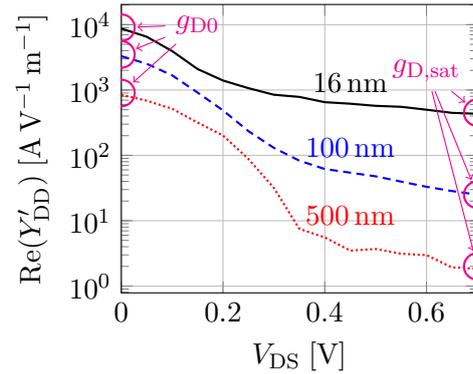


Figure 5.47: Drain self-admittance  $Y'_{DD}$  vs. drain bias at  $V_{GS} = 0.7V$  and  $f = 1$  kHz for the 16 nm (black solid line), 100 nm (blue dashed line), 500 nm (red dotted line) device. The respective values for  $g_{D0}$  and  $g_{D,sat}$  have been marked with magenta circles.

characteristic and therefore computing it has value for direct comparisons with experiments although its interpretation is not straightforwardly conclusive.

In this spirit, we observe that the gate excess noise shown in Fig. 5.46 is strictly larger for shorter channels, which by itself is difficult to interpret due to the denominator of  $\beta$  in Eq. (5.9) being different for each device. What can be said, though, is that the smaller the device, the higher  $\beta$  will rise in saturation compared to its equilibrium value. Since the denominator of  $\beta$  is independent of the drain bias, this is a direct consequence of a differing behavior of  $P_{GG}$  in the three devices indicating a different quality of noise in the devices.

Figure 5.48 shows the gate excess noise factor against the gate bias on both a linear and a logarithmic scale where it becomes apparent that the gate excess noise for the short channel device is not strictly larger than in the long channel device. Indeed, around the threshold voltage, we can observe how the gate excess noise for the 500 nm device exceeds even the 16 nm device. The results for the gate excess noise shown here are in qualitative agreement with Ref. [128].<sup>2</sup>

### 5.4.5 Cross Correlation

The correlation coefficient of noise is another defining quantity for the noise in a MOSFET. It is defined as [125]

$$c = \frac{P_{GD}}{\sqrt{P_{GG}P_{DD}}}.$$

<sup>2</sup>Note that the gate excess noise in Ref. [128] is evaluated with the gate-source capacitance  $C_{GS}$  at the actual bias point and not in saturation. This changes the absolute values of  $\beta$  below saturation somewhat but not as significantly as to change the qualitative appearance.

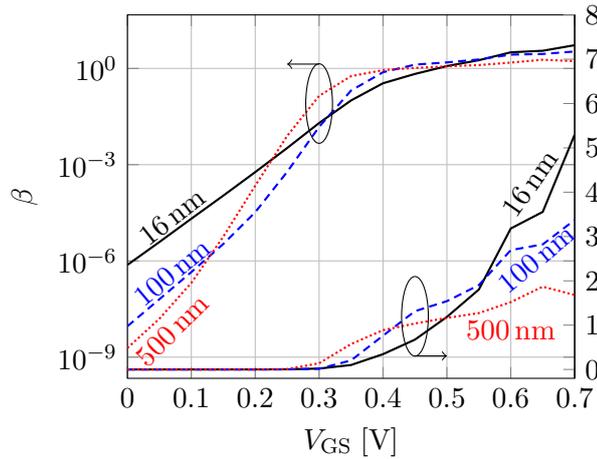


Figure 5.48: Gate excess noise factor  $\beta$  vs. gate bias  $V_{GS}$  at  $V_{DS} = 0.7V$  and  $f = 100$  kHz for the 16 nm (black solid), 100 nm (blue dashed), 500 nm (red dotted) devices.

As is explained in Ref. [131], the real part of the correlation coefficient in a MOSFET at low frequencies is usually negligible compared to the imaginary part and therefore it can be ignored for most practical purposes. We can corroborate this by explicitly calculating the real and imaginary parts of the correlation coefficient depicted in Fig. 5.49 for the 16 nm device and in Fig. 5.50 for the 500 nm device. Note how the real part is approximately zero everywhere. The imaginary part of the correlation coefficient of the 16 nm device in Fig. 5.49 rises up until the threshold voltage, where it reaches its maximum, and then tapers off towards higher gate biases. The same trend shows for the 500 nm device plotted in Fig. 5.50, however, for this device the peak is much wider and at high gate biases the correlation coefficient does not decrease as significantly as for the 16 nm device.

The value obtained from the charge sheet model of Ref. [125] is given by  $\text{Im}(c) = 0.395$  but we cannot observe it in our simulations. However, similar to the gate excess noise, the approximations of the charge sheet model of Ref. [125] are nowhere near of being capable of capturing the electron confinement of the simulated double gate MOSFET and thus any correlations mediated by the confined electron gas from gate to drain would be difficult to capture without self-consistent solutions such as the ones of this work.

The correlation coefficient expresses the correlation between the fluctuations in the gate and drain contacts as shown in Eq. (4.2). If a fluctuation occurs within the device due to a scattering event, we expect that the fluctuation will show itself in the terminal currents. If it shows in both the gate and drain terminal currents, the noise is correlated and we obtain a non-zero correlation

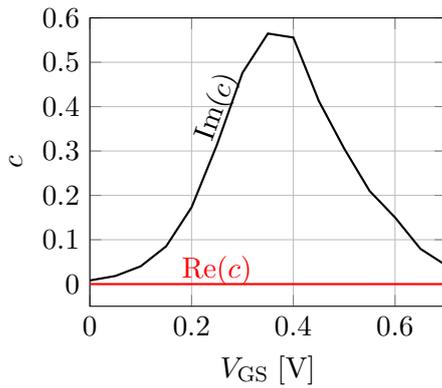


Figure 5.49: Real and imaginary parts of the correlation coefficient  $c$  vs. gate bias  $V_{GS}$  for  $V_{DS} = 0.7$  V and  $f = 100$  kHz of the 16 nm device.

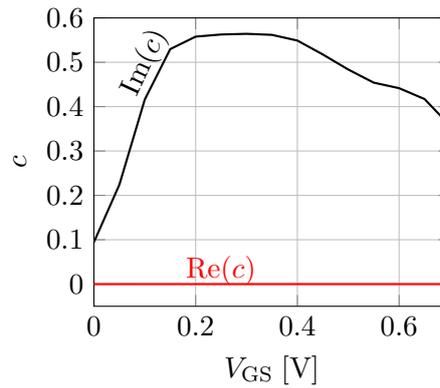


Figure 5.50: Real and imaginary parts of the correlation coefficient  $c$  vs. gate bias  $V_{GS}$  for  $V_{DS} = 0.7$  V and  $f = 100$  kHz of the 500 nm device.

coefficient as is obvious from Eq. (4.1). To better understand the reason why the correlation coefficient has a maximum around the threshold voltage, we can inspect the origin of the correlation coefficient within the device by plotting the contribution to the correlation coefficient per grid point in transport direction. Figure 5.51 shows the origin of the correlation coefficient for the 16 nm device using the spatial density of the PSD as defined by Eq. (4.43). Thus, if the plot is non-zero at some spatial location, it means that a fluctuation occurring at this position will elicit a correlated response in both the gate and drain terminal currents. In the sub-threshold at  $V_{GS} = 0.2$  V – at voltages lower than the peak of  $c$  in Fig. 5.49 – the fluctuations which generate a correlated terminal current response mainly occur on the source side in front of the channel region. As the gate voltages increases up to the threshold voltage, the current flowing through the device increases as well which increases the amount of fluctuations. The noise at the threshold voltage and in the sub-threshold is qualitatively the same, as in both cases it is generated on the source side like shot noise, only their absolute values are different since the stationary current is different. However, when the gate voltage is turned up beyond the threshold voltage, we can see that the quality of the noise changes. Fluctuations on the source side have less impact and the channel noise starts to contribute to the correlation coefficient. Moreover, on the drain side, we see that fluctuations cause a negative correlation coefficient.

Therefore the reason for the low correlation at low gate biases is due to the small effect of fluctuations. However, at high gate biases the terminal currents become uncorrelated not due to the weak effect of fluctuations on the terminal currents but because the fluctuations on the drain side counteract the fluctuations on the source side. This cancellation was previously noted with a

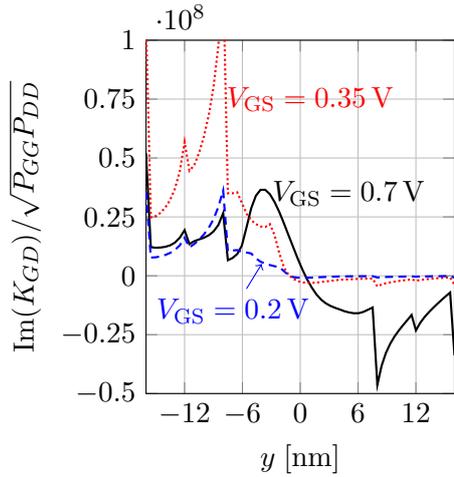


Figure 5.51: Imaginary part of the correlation coefficient density in transport direction of the 16 nm device for  $V_{DS} = 0.7$  V,  $f = 100$  kHz, and three different gate biases  $V_{GS} = 0.2$  V (blue dashed line),  $0.35$  V (red dotted line),  $0.7$  V (black dashed line).

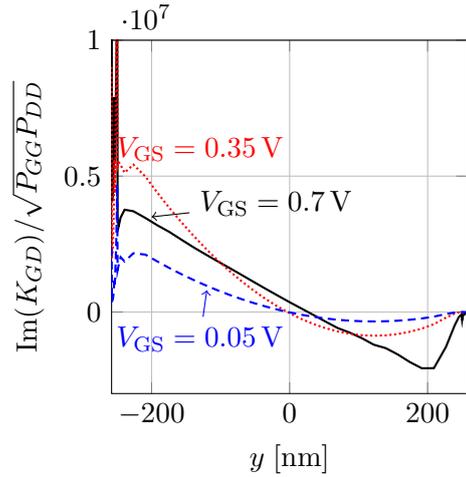


Figure 5.52: Imaginary part of the correlation coefficient density in transport direction of the 500 nm device for  $V_{DS} = 0.7$  V,  $f = 100$  kHz, and three different gate biases  $V_{GS} = 0.05$  V (blue dashed line),  $0.35$  V (red dotted line),  $0.7$  V (black solid line).

drift-diffusion and a hydrodynamic model in Ref. [132].

Figure 5.52 shows a similar plot for the 500 nm device but we chose the gate bias in the sub-threshold region to be  $V_{GS} = 0.05$  V in order to be at a voltage lower than the maximum in  $c$  shown in Fig. 5.50. Here, we find a similar behavior but the correlated noise in the sub-threshold is generated within the channel on the source side rather than in the contact region. Once the gate bias is turned up, we can see that the channel region close to the drain counteracts the correlation on the source side and therefore the overall cross correlation decreases.

#### 5.4.6 Origin of Noise

The deterministic solver allows us to directly compute the origin of noise as has already been done for the correlation coefficient in Figs. 5.51 and 5.52. The various definitions of densities of the PSD are given in the end of Sect. 4.6.3 by Eqs. (4.43), (4.44), and (4.45). Figure 5.53 shows the noise contributions to the drain terminal current throughout the 500 nm device in the on-state. It is evident that most of the noise is generated throughout the channel. As was shown in Sect. 5.2.2, the 500 nm device exhibits scattering dominated transport, i.e. a lot of scattering events occur in the channel when electrons cascade down the potential created by the subband energies from source to drain, making the distribution of carriers appear thermal (cf. Fig. 5.8). And since scattering events

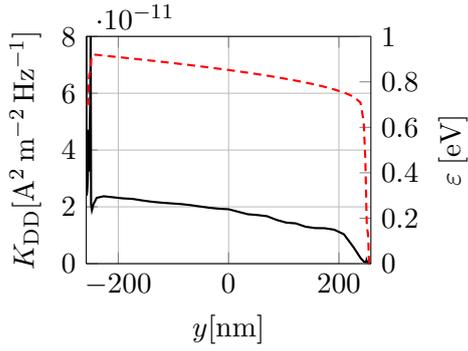


Figure 5.53: PSD of drain current fluctuations per grid point in transport direction of the 500 nm device at  $V_{GS} = V_{DS} = 0.7$  V and  $f = 100$  kHz (black solid line; left axis). For reference, the energy of the energetically lowest subband (red dashed line; right axis) has been plotted.

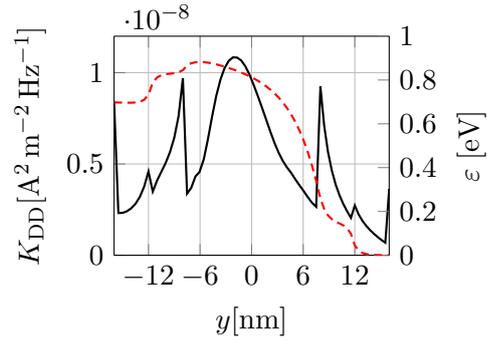


Figure 5.54: PSD of drain current fluctuations per grid point in transport direction of the 16 nm device at  $V_{GS} = V_{DS} = 0.7$  V and  $f = 100$  kHz (black solid line; left axis). For reference, the energy of the lowest subband (dashed red line; right axis) has been plotted.

are fluctuations we find that most of the noise is generated within the channel. Thus, the noise in the long channel device behaves similar to thermal noise although its magnitude is significantly larger than the Johnson-Nyquist noise of Eq. (4.4) and it is skewed somewhat towards the source side. We also observe a high contribution from the highly doped contact region but it is negligible compared to the integral over the noise of the channel region.

Compare this to the origin of noise in the 16 nm device shown in Fig. 5.54 in the on-state. We may note that the GR rate contributes noise on the source and drain contacts at  $y = \pm 16$  nm which is dependent on the size of the recombination velocity in Eq. (5.2). Moreover, the highly doped contact regions contribute a significant part of noise and we see discontinuities when the deformation potential is adjusted due to the abruptly changing doping density at  $y = -12$  nm,  $-8$  nm,  $8$  nm,  $12$  nm (cf. Fig. 5.2). As was explained in Sect. 5.1.1, we adjust the deformation potential of the elastic acoustic phonon scattering in order to mimic impurity scattering in the highly doped contact regions. Most of the noise in the 16 nm device stems from the channel region where we find a maximum in noise generation a few nanometers past the maximum in the energy barrier formed by the subband. It is clear that the noise in the channel is highly inhomogeneous and therefore it stands to reason that there is a more complicated underlying process at play that leads to this appearance.

In order to gain more insight into how noise emerges in the devices, we can plot the PSD not only as a spatial density but also as a density w.r.t.  $H$ -space as defined in Eq. 4.44. The resulting noise density of the drain current noise for the 500 nm device is shown in Fig. 5.55. The white region in this plot is

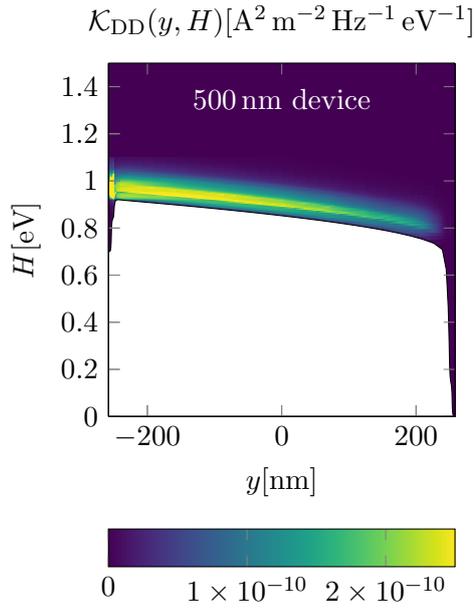


Figure 5.55: PSD of drain current fluctuations vs.  $y$  and  $H$  of the 500 nm device at  $V_{GS} = V_{DS} = 0.7$  V and  $f = 100$  kHz. Values have been truncated at  $2.5 \times 10^{-10} \text{ A}^2 \text{ m}^{-2} \text{ Hz}^{-1} \text{ eV}^{-1}$  for a better visualization.

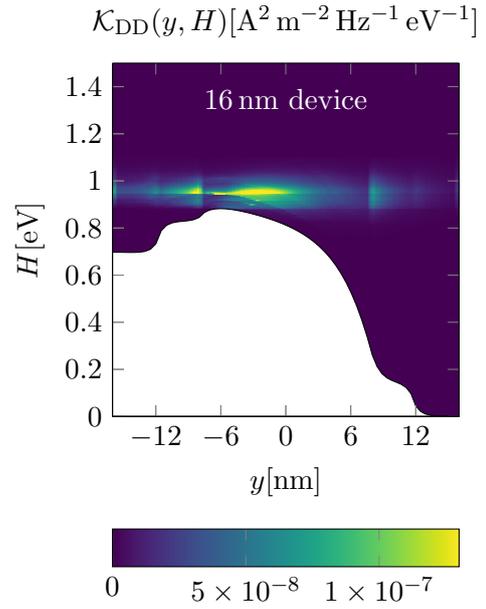


Figure 5.56: PSD of drain current fluctuations vs.  $y$  and  $H$  of the 16 nm device at  $V_{GS} = V_{DS} = 0.7$  V and  $f = 100$  kHz. Values have been truncated at  $1.3 \times 10^{-7} \text{ A}^2 \text{ m}^{-2} \text{ Hz}^{-1} \text{ eV}^{-1}$  for a better visualization.

kinematically inaccessible since it is below the lowest subband energy. Note how most of the noise is generated in the channel region just above the energetically lowest subband energy. Compare this to the plot of the current density in Fig. 5.16 which shows that the current flows in the same region. The obvious conclusion is that the fluctuations in the PSD of the terminal current can only originate where the current flows within the device. To be sure, an energy region that does not contribute to the drain terminal current can fluctuate equally well but it cannot influence the drain terminal current. Since the transport in the 500 nm device is scattering dominated, a lot of scattering events occur in the channel, leading to the accumulation of carriers near the lowest subband energy. These scattering events in the channel are fluctuations and thus they manifest in the PSD of the drain terminal current.

A careful inspection of Fig. 5.55 reveals that there appear to be multiple levels parallel to the lowest subband energy at which the PSD contribution increases stepwise. These levels correspond to the various subbands in the device. As soon as an additional subband is energetically available, scattering processes to and from it suddenly become possible and the contribution of these additional scattering processes is immediately noticeable in the PSD.

Let us now turn to where the noise originates in the 16 nm device. Recall that the current density was a horizontal band above the maximum of the energetically lowest subband as was shown in Fig. 5.17. Any fluctuation perceptibly impacting the drain terminal current must occur within that horizontal band to be effective. However, compared to the 500 nm device, we see that in Fig. 5.56 fluctuations in the channel have a peculiar way of contributing to the drain terminal current. The maximum in the PSD per spatial and energy grid point is located in the channel and shaped like the subband energy but strangely it is shifted to higher energies, away from the minimum subband energy. To be sure, the PSD does not look like shot noise which is generated before the potential energy maximum that carriers must surpass. From Fig. 5.42 we already know that in the sub-threshold, we find pure shot noise. Hence, we may conclude that whatever is suppressing the noise when the gate bias is high, is responsible for the appearance of the PSD in Fig. 5.56. In the next section, we will comprehensively investigate the suppression of noise and determine the nature of noise in the 16 nm device.

#### 5.4.7 Suppression of Noise

In light of the appearance of the noise in the 16 nm device shown in Figs. 5.54 and 5.56 as well as the Fano factor of Fig. 5.42, it stands to reason that the noise in the short channel device cannot be characterized by either shot noise or something with the qualitative appearance of thermal noise. As the Fano factor in Fig. 5.42 shows, the noise at lower gate voltages is pure shot noise, while at higher gate voltages it becomes suppressed. In this section, we will come to understand the dominant mechanism of noise generation in the short channel device leading to this suppression.

The suppression of noise has been reported in literature to have at least two origins which we will investigate. The first reason for suppressed noise is the Pauli principle (see e.g. Ref. [51]). Shot noise is the noise originating from carriers surpassing a potential barrier. If the rate of transfer over the barrier is known – e.g. in terms of the current –, we can directly relate it to the PSD. However, such a simple relation between noise and current can only be derived if the Pauli principle is ignored. Once we incorporate it, scattering events terminating in already occupied states are forbidden and thus the total scattering rate is reduced. Since scattering events are noise, we can conclude that in a system where the predominant noise source is from carriers overcoming a potential barrier and where the Pauli principle is important, we will find a PSD smaller than for pure shot noise.

The second reason for the suppression of noise is due to Coulomb interactions which was already reported in Refs. [133, 134] in the context of MOSFETs. The idea is that a local accumulation of carriers increases the energy necessary to

$F$	Pauli	no Pauli
Coulomb	0.42	0.41
no Coulomb	0.71	1.01

Table 5.2: Fano factor  $F$  of the 16 nm device at  $V_{GS} = 0.7$  V,  $V_{DS} = 0.7$  V, and  $f = 100$  kHz with and without noise suppressing mechanisms.

add further carriers while a depletion increases the energy to remove further carriers. Then a scattering event not only modifies the distribution functions of each subband but also the respective electric potentials the carriers will be subject to. Completely homogeneous carrier distributions do not exhibit this phenomenon but since our device has a strong spatial dependence on the density, we expect to find some suppression.

Note that the term *Coulomb interactions* in this context does *not* mean Coulomb scattering but the interactions of carriers through the electric potential computed via the PE. The way we compute the noise via the self-consistent system of PE, SE, and BE, we include how a fluctuation in the distribution function influences the electric potential and hence the subband energies and wave functions which in turn have an impact on the distribution functions. As long as this feedback loop is contained in a simulation, Coulomb interactions in the above sense are included.

We can make sure that there are no other mechanisms suppressing the noise by turning both the Pauli principle and the Coulomb interactions off. The Pauli principle can be omitted by ignoring the final state distribution functions in Eq. (2.15), i.e. using Eq. (2.16) as our scattering rate in the BE and LBE. The Coulomb interactions disappear once we ignore the self-consistency of the problem and solve the Green's function equation only for the LBE. In practice this can be achieved by setting the off-diagonal blocks of Eq. (4.16) to zero, i.e. remove the derivatives of the LBE w.r.t. the potential and remove the derivatives of the PE w.r.t. the distribution functions in Eqs. (4.13) and (4.14).

Table 5.2 shows the Fano factor if either or both of the suppressing mechanisms are turned off. Evidently both mechanisms impact the Fano factor but only when both are turned off do we obtain pure shot noise.<sup>3</sup> Thus we can conclude that there are no other mechanisms suppressing the noise.

### Pauli Principle

The Pauli principle is quite important in the 16 nm device as can already be seen from the distribution function in Fig. 5.10. Moreover, turning off the Pauli principle changes the mean energy significantly as was shown in Fig. 5.22. The Pauli

<sup>3</sup>Note that the 1% deviation of the Fano factor from unity is within our margin of error, see Sect. 5.4.3.

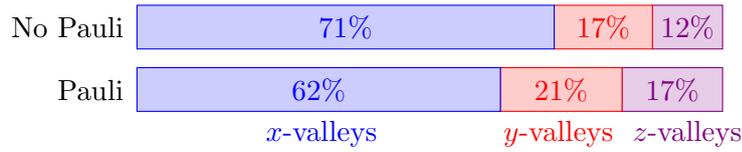


Figure 5.57: Occupation of valleys in the channel region at the position of maximum average velocity of the 16 nm device (cf. Fig. 5.21) for  $V_{GS} = V_{DS} = 0.7$  V. The  $x$ -,  $y$ -, and  $z$ -valleys refer to the valleys aligned with the respective axis in  $\mathbf{k}$ -space. Confinement is in  $x$ -direction and transport is in  $y$ -direction.

principle is also responsible for a substantial redistribution of carriers among the valleys in the band structure of silicon. Figure 5.57 shows the valley occupation in the channel of the 16 nm device with and without the Pauli principle in the on-state. Here,  $x$ -,  $y$ -, and  $z$ -valleys refer to the valleys aligned with the respective axis in  $\mathbf{k}$ -space (see Fig. 2.6). Recall that transport from the source to the drain contact happens in  $y$ -direction and the confinement is in  $x$ -direction (cf. Fig. 5.1). Thus, the  $x$ -valleys have a larger mass in confinement direction and therefore their subband energies are energetically lower, from which immediately follows that they hold the most carriers. Note how the Pauli principle prohibits about 12% of the total number of carriers to reside in the  $x$ -valleys and redistributes them to the energetically higher  $y$ - and  $z$ -valleys.

Another curiosity relating to the valley occupation is that the  $y$ -valleys contain more electrons than the  $z$ -valleys because the driving electric field is applied in  $y$ -direction. Since the  $y$ - and  $z$ -valleys are energetically equivalent, the only remaining difference is the mass of carriers in transport direction which is larger for  $y$ -valleys than for  $z$ -valleys (cf. Fig. 2.6). Thus, electrons in  $y$ -valleys are more inert when an electric field is applied in transport direction while electrons in  $z$ -valleys accelerate quicker and can therefore be scattered more readily. This creates an imbalance in scattering rates between the valleys which leads to an imbalance in the occupation.

Despite its impact on the electron configuration, the Pauli principle is not responsible for the surprising appearance of the PSD of the drain current fluctuations of Fig. 5.56. Turning off the Pauli principle does in fact not change the qualitative appearance of Fig. 5.56. Hence, the unexpected shape of the PSD must originate in the suppression due to the Coulomb interaction.

### Coulomb Interaction

To understand the appearance of the PSD of the drain current of Fig. 5.56, let us inspect the PSD per subband  $\mathcal{K}_{DD}^{\nu}(y, H)$  of Eq. (4.45). Figure 5.58 depicts the PSD vs.  $y$ - and  $H$ -space of the 500 nm device for the lowest subband of each of the three distinguishable valleys. The lowest subband of each valley is representative

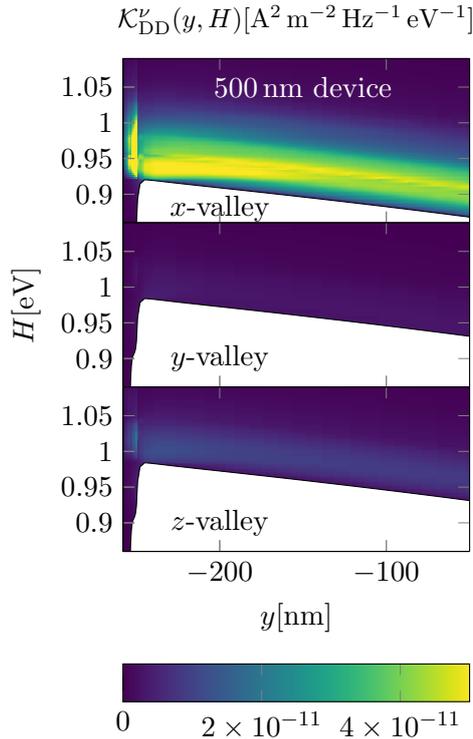


Figure 5.58: PSD of drain current fluctuations vs.  $y$  and  $H$  for the lowest subband of each valley of the 500 nm device for  $V_{GS} = V_{DS} = 0.7$  V and  $f = 100$  kHz. Values have been truncated at  $5 \times 10^{-11} \text{ A}^2 \text{ m}^{-2} \text{ Hz}^{-1} \text{ eV}^{-1}$  for a better visualization.

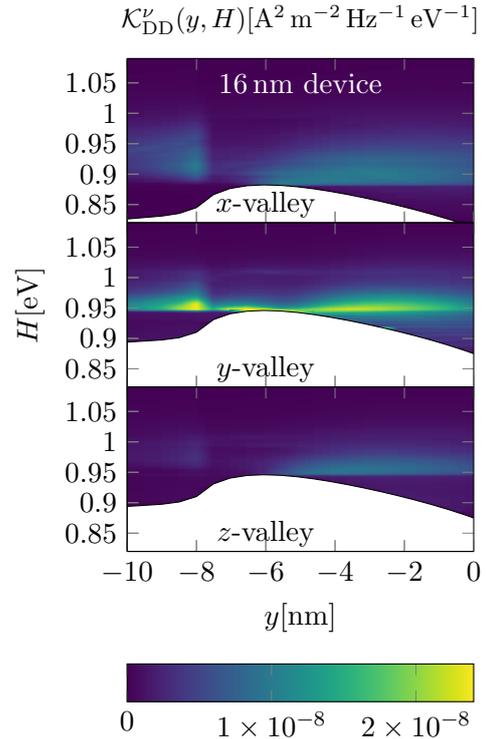


Figure 5.59: PSD of drain current fluctuations vs.  $y$  and  $H$  for the lowest subband of each valley of the 16 nm device at  $V_{GS} = V_{DS} = 0.7$  V and  $f = 100$  kHz. Values have been truncated at  $2.4 \times 10^{-8} \text{ A}^2 \text{ m}^{-2} \text{ Hz}^{-1} \text{ eV}^{-1}$  for a better visualization.

of the total noise generated in that valley since higher subbands contribute exponentially less. Note how practically all the noise is generated in the  $x$ -valleys. These are energetically lower than the other valleys and therefore they contain more electrons and carry the largest portion of the current. Thus, the intuitive and straightforward conclusion applies to the 500 nm device: We will find more scattering events in the  $x$ -valleys and thus more fluctuations impacting the terminal current and therefore the  $x$ -valleys contribute stronger to the PSD. Furthermore, if you look carefully, you can see that the  $z$ -valleys contribute more noise than the  $y$ -valleys. This is due to the aforementioned lighter mass of the  $z$ -valleys in  $y$ -direction. Carriers in the  $z$ -valleys are accelerated more quickly and therefore they scatter more readily, leading to the imbalance in valley occupation shown in Fig. 5.57, but also leading to more noise generation in the  $z$ -valleys.

If we plot the PSD per valley for the 16 nm device, we find the surprising

appearance of Fig. 5.59. The strongest contributions to the PSD stem from the region around the maximum of the subband energy in the  $y$ -valleys. The strong contributions from the  $y$ -valley are directly responsible for the appearance of Fig. 5.56. To be sure, it is still true that the  $x$ -valleys contain significantly more electrons than the  $y$ - or  $z$ -valleys (cf. Fig. 5.57) and it is also still true that the  $x$ -valleys carry significantly more current. In fact, in the channel only about 5% of the total current is transported through the  $y$ -valleys while about 77% is carried in the  $x$ -valleys. Thus, despite having more scattering events happen in the  $x$ -valleys, we see that they are not the predominant source of fluctuations in the drain terminal current of the 16 nm.

To make sure we are on the right track in investigating the origin of this behavior, we plot the PSD for the lowest subband of each valley when we turn off the Coulomb interactions. The resulting PSD is plotted in Fig. 5.60 and as is obvious, the origin of noise changes completely. Without the Coulomb interaction, we find that the  $x$ -valley contributes most of the noise to the drain terminal current, just as would have been expected from a long channel device. Indeed, the qualitative appearance of Fig. 5.56 without Coulomb interaction would change to something one could extrapolate from the long channel case. The reason why the  $z$ -valleys contribute more to the noise than the  $y$ -valleys is once again due to the differing masses in transport direction which also lead to the imbalance in the occupation shown in Fig. 5.57.

However, Coulomb interactions *do* exist and as we can see, their impact is significant, changing not only the absolute values of the noise but also the origin and thus the quality and interpretation of noise in a nanoscale device. Thus, the above interpretation for the origin of noise in the case of no Coulomb interactions does not apply once we include how fluctuations in the density perturb the electric potential.

To see how exactly the fluctuations in the 16 nm device elicit a drain current response, we plot the Green's function of the drain terminal current vs.  $y$ - and  $H$ -space shown in Fig. 5.61. Recall that the Green's function of the terminal current introduced in Eq. (4.20) is the response of the terminal current to a fluctuation in the distribution function in some subband at some position, energy, and with a certain Fourier harmonic. Thus, since the zeroth harmonic of the distribution function constitutes the electron density (see Eq. 5.3), the Green's function of the drain terminal current of the zeroth harmonic tells us what happens to the drain terminal current when we put a charge into some subband at some position and energy. However, keep in mind that while this is a valid interpretation of the Green's function, all our scattering processes are actually charge conserving which can be seen by the pairwise appearance of creation and annihilation Green's functions in the PSD of Eq. (4.34), i.e. a scattering process is tantamount to an annihilation of charge in the initial state and a creation of charge in the final state.

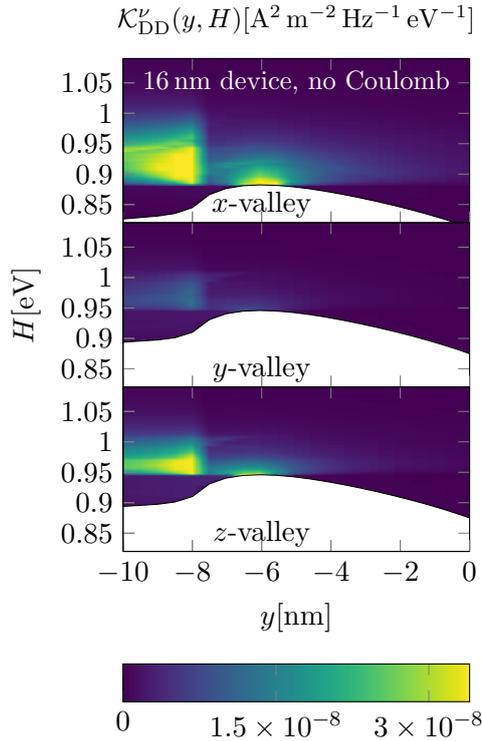


Figure 5.60: PSD of drain current fluctuations vs.  $y$  and  $H$  for the lowest subband of each valley of the 16 nm device at  $V_{GS} = V_{DS} = 0.7$  V and  $f = 100$  kHz without Coulomb interactions. Values have been truncated at  $3.4 \times 10^{-8} \text{ A}^2 \text{ m}^{-2} \text{ Hz}^{-1} \text{ eV}^{-1}$  for a better visualization.

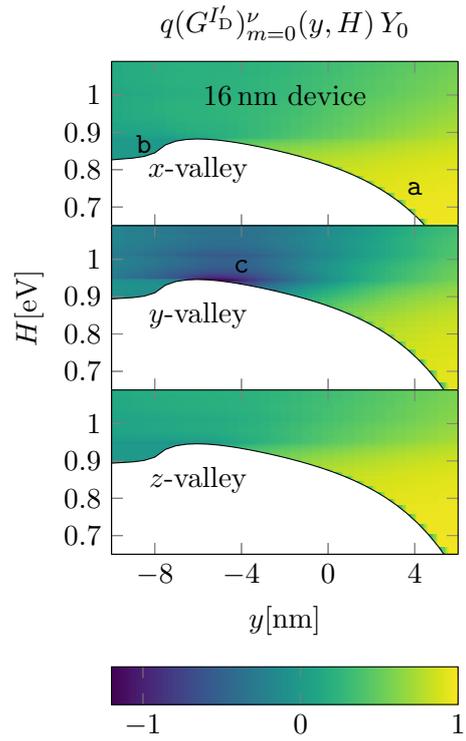


Figure 5.61: Zeroth harmonic of the Green's function of the drain terminal current for the lowest subband of each valley of the 16 nm device at  $V_{GS} = V_{DS} = 0.7$  V and  $f = 100$  kHz. Regions of interest have been marked with the letters a, b, and c.

Figure 5.61 shows that if a charge is placed in an  $x$ -valley in the region around a – close to the drain and at energies below the subband maximum –, the Green's function of the drain terminal current is '1'. This implies that the charge is completely converted into a current measurable in the drain terminal. The same argument holds for fluctuations, i.e. a fluctuation occurring at position a would be measurable as a fluctuation in the drain terminal current.

On the other hand, the Green's function is close to zero on the source side in the region around b – at energies lower than the subband maximum. This is because a charge placed there is unlikely to scatter to higher energies in order to surpass the energy barrier to reach the drain contact. It is far more likely that the charge moves directly towards the source contact or that it is reflected at the energy barrier and then returns to the source contact. Thus, any fluctuations occurring around b has no impact on the drain current.

The most interesting feature of Fig. 5.61 is the negative area around  $c$ , which is right next to the subband energy maximum of the  $y$ -valley. Here, the negative values imply that placing a charge in that region *reduces* the drain terminal current. Moreover, the Green's function reaches values of less than  $-1.2$  which means that the reduction of the terminal current is more effective than the increase this charge could elicit if it were placed directly on the drain contact. Since we already know that the noise in this device is strongly affected by the Coulomb interaction and we just learnt that it must be an indirect effect due to the effectiveness of a single charge in reducing rather than increasing the drain current, it stands to reason that fluctuations strongly affect the electric potential and thus the subband energies.

Recall that the current in the 16 nm device flows as in Fig. 5.17 as a horizontal band just above the subband energy barrier. Then, an increase in the maximum of the subband energy directly impacts the flow of electrons from source to drain. A charge placed somewhere around the maximum of the subband energy in the  $y$ -valley increases the subband energy and therefore decreases the drain current by reflecting back a larger portion of the electrons coming from the source contact. This explains both why a charge decreases the total drain current and why the charge is so effective in the reduction of the drain current.

What remains is the explanation of why the  $y$ -valley and  $z$ -valley do not contribute equally, i.e. why a charge placed into the  $z$ -valleys does not lead to such a reduction in the drain current as a charge placed into the  $y$ -valleys. The only difference between these valleys are the effective electron masses in transport direction. Since the mass in transport direction of electrons in the  $y$ -valleys is significantly larger than the one in the  $z$ -valleys, the electrons are more inert and therefore an electron in the  $y$ -valley can affect the subband energy more effectively while an electron in the  $z$ -valley is more ephemeral in that it is easily pushed to the drain contact by the driving field.

To summarize, we have shown that in a short-channel device with ballistic transport, the predominant effect of noise generation comes from the impact of fluctuations on the energy barrier carriers must surpass. This effect is only included if simulations are self-consistent. Omitting self-consistency and thus the Coulomb interactions, not only leads to different absolute values of noise but also misrepresents the way noise is generated within the device.



## Chapter 6

# Conclusion

We built upon the existing work on deterministic solvers using spherical or Fourier harmonics expansions of  $\mathbf{k}$ -space and proceeded to construct the full Newton-Raphson approach for the stationary system of Poisson equation (PE), Schrödinger equation (SE), and Boltzmann equation (BE) in Chap. 2 of this work. We set up the small signal system of equations in what followed in Chap. 3 and we identified difficulties in the definition of the time-derivative as well as in the discretization. Through careful analysis, we found a way to mitigate these issues such that our final system of equations for the small signal analysis conserves essential symmetries of real devices. Furthermore, we derived a useful form of the Ramo-Shockley theorem for confined electron gases which is also applicable to related cases. We continued to derive the equations to compute the self-consistent noise with the Langevin-source approach in Chap. 4. We also derived general equations to treat degeneracy in the Green's function equations and in the power spectral densities of noise.

The aforementioned theoretical work has lead to the first ever fully self-consistent and deterministic solver for small signal and noise analyses for the system of PE, SE, and BE in a nanoscale nMOSFET. In Chap. 5, we demonstrated our implementation of the solver and showed that it can produce results with unprecedented precision in a wide range of operating conditions and frequencies. In particular, due to the deterministic nature of the solver, we find that we can determine solutions in the low frequency domain and in the sub-threshold where traditional Monte Carlo (MC) based solvers are computationally infeasible. We showed that we can compute all relevant figures of merit concerning the small signal and noise behavior in the linear response regime of a nanoscale nMOSFET. We also showed that our simulations are consistent with literature wherever similar findings were available.

The absolute advantage of deterministic solvers over MC approaches is that it is possible to understand the inner workings of the device which we exploited

to find out where noise in a nanoscale nMOSFET originates. By using this property of our deterministic solver, we found that in devices where transport is essentially ballistic, noise in the drain current is predominantly generated by Coulomb interactions which influence the potential barrier the drain current needs to surpass. This means that the self-consistency of the PE, SE, and BE is vital during the determination of noise.

The work compiled in this manuscript is the product of a multi-year effort which yielded several publications. In Ref. [63], the full Newton-Raphson approach for the stationary solution of the self-consistent system of PE, SE, and BE is presented. In Ref. [58] the theoretical ground work for the self-consistent small signal and noise problems is elaborated. Finally, in Refs. [64] and [135] the results for a nanoscale nMOSFET are shown.

The implementation as well as the underlying ideas of the simulator used in this work have already been used in related research. References [117] and [136] use the self-consistent solver for an investigation of GaAs and graphene devices, respectively. Moreover, Ref. [137] uses parts of the code for an investigation of plasma waves. A self-consistent simulation of a silicon nanowire based on the same principles can be found in Ref. [41].

For future work, there are still some low-hanging fruits which include the noise calculation with different crystal orientations, non-parabolic band structures, and the inclusion of impurity scattering and screening. The simulator developed during this work would also need experimental validation which involves a non-trivial amount of fitting of device dimensions and material properties. Future work could also incorporate devices with different semiconductor materials or devices with altogether different shapes or even the full inclusion of holes. Furthermore, the current work could inform and improve the workings of compact models for noise.

# Appendix A

## Observables

In the following, we want to provide a list of all important quantities and their respective definitions.

### A.1 Stationary

#### Density

The electron sheet density in transport direction per subband and per spin is defined as

$$n^\nu(y) = \int \frac{d^2k}{(2\pi)^2} f^\nu(y, \mathbf{k}) = \frac{1}{Y_0} Z^\nu \int dH f_0^\nu(y, H). \quad (\text{A.1})$$

The total sheet density is given by

$$n(y) = \mu_{\text{spin}} \mu_{\text{val}} \sum_{\nu} n^\nu(y). \quad (\text{A.2})$$

The 3D density per subband and spin reads

$$n_{3\text{D}}^\nu(\mathbf{r}) = n^\nu(y) |\Psi^\nu(\mathbf{r})|^2 \quad (\text{A.3})$$

and its total is

$$n_{3\text{D}}(\mathbf{r}) = \mu_{\text{spin}} \mu_{\text{val}} \sum_{\nu} n_{3\text{D}}^\nu(\mathbf{r}), \quad (\text{A.4})$$

with  $\mathbf{r} = (x \ y)^t$ .

The discretization of the sheet density is given by

$$n^\nu(y_i) = \frac{1}{Y_0} Z^\nu \sum_j f_0^\nu(y_i, H_j) \Delta H^\nu(y_i, H_j) \quad (\text{A.5})$$

and the discretized 3D density is

$$n_{3\text{D}}^\nu(x_k, y_i) = n^\nu(y_i) |\Psi^\nu(x_k, y_i)|^2$$

### Current Density

The electron sheet current density in transport direction per subband and per spin is given by

$$\begin{aligned} j^\nu(y) &= (T_{\text{HV}}^v)_{yy} \int \frac{d^2k}{(2\pi)^2} v_y^v(\mathbf{k}) f^\nu(y, \mathbf{k}) \\ &= (T_{\text{HV}}^v)_{yy} \int dH Z^v (v_y^v)_1(y, H) f_1^\nu(y, H). \end{aligned} \quad (\text{A.6})$$

Note that  $\mathbf{k}$ -space is Herring-Vogt transformed and therefore the  $y$ -component of the group velocity needs to be transformed (cf. Eq. (2.50)). Here,  $(v_y^v)_1(y, H)$  denotes the Fourier coefficient of the first harmonic (cf. Eq. (2.61)). The 3D current density in transport direction per subband and spin is given by

$$\mathbf{J}_y^\nu(\mathbf{r}) = j^\nu(y) |\Psi^\nu(\mathbf{r})|^2 \mathbf{e}_y. \quad (\text{A.7})$$

Hence, the total electron sheet current density reads

$$j(y) = \mu_{\text{spin}} \mu_{\text{val}} \sum_\nu j^\nu(y) \quad (\text{A.8})$$

and likewise for the 3D current density

$$\mathbf{J}_y(\mathbf{r}) = \mu_{\text{spin}} \mu_{\text{val}} \sum_\nu \mathbf{J}_y^\nu(\mathbf{r}).$$

Discretization is a bit more involved since we have to be careful with the integration of the group velocity in  $H$ -space. We use the same analytical integration as in the free streaming term, which is given by Eq. (2.78). Hence, we find for the sheet current density

$$j^\nu(y_{i+}) = (T_{\text{HV}}^v)_{yy} Z^v \sum_j f_1^\nu(y_{i+}, H_j) \int_{\Delta H^\nu(y_{i+}, H_j)} dH (v_y^v)_1(y_{i+}, H) \quad (\text{A.9})$$

and for the 3D current density in transport direction

$$\mathbf{J}_y^\nu(x_k, y_{i+}) = j^\nu(y_{i+}) \frac{|\Psi^\nu(x_k, y_{i+1})|^2 + |\Psi^\nu(x_k, y_i)|^2}{2},$$

where the wave function needs to be averaged since it is defined on the direct grid points.

### Terminal Current

The terminal current is defined as the charge current flowing into the device. Therefore we find that the drain terminal current per length in the homogeneous  $z$ -direction in the stationary case is simply given by

$$I_D' = qj(y).$$

There is no need for the Ramo-Shockley theorem in the stationary case since the current is constant anyway (cf. the continuity equation (3.23)).

### Distribution Function

In order to visualize the distribution function, we need to sum over all Fourier components and invert the  $H$ -transformation as

$$f^\nu(y, \mathbf{k}(E, \phi)) = \sum_m f_m^\nu(y, H - \varepsilon^\nu(y)) Y_m(\phi).$$

Note that this is the distribution function per subband.

### Effective Field

The effective field in confinement direction is defined in our case as

$$E_{\text{eff}} = \frac{q}{\kappa_{\text{Si}}} \left( \frac{1}{2}n + w_{\text{Si}}N_D \right), \quad (\text{A.10})$$

where  $\kappa_{\text{Si}}$  is the permittivity of silicon,  $w_{\text{Si}}$  is the width of the silicon channel, and  $N_D$  is the donor density. Note that there exists a relation between the average electric field in confinement direction and the above expression but it only applies to special cases [69].

### Low-Field Mobility

The low-field mobility is given by

$$\mu_e = \frac{j}{qnE_{\text{el}}},$$

where  $E_{\text{el}}$  is the driving electric field. This expression only makes sense when diffusion can be neglected, i.e. in homogeneous devices without built-in fields.

### Velocity

The average electron velocity is given by

$$v(y) = \frac{j(y)}{n(y)}.$$

### Mean Energy

The average energy of electrons as measured relative to the energetically lowest subband  $\varepsilon^{\nu\min}$  is given by

$$\begin{aligned}\langle E \rangle(y) &= \frac{\mu_{\text{spin}} \mu_{\text{val}} \sum_{\nu} \int \frac{d^2k}{(2\pi)^2} [E^{\nu}(\mathbf{k}) + \varepsilon^{\nu}(y) - \varepsilon^{\nu\min}(y)] f^{\nu}(y, \mathbf{k})}{\mu_{\text{spin}} \mu_{\text{val}} \sum_{\nu} \int \frac{d^2k}{(2\pi)^2} f^{\nu}(y, \mathbf{k})} \\ &= \frac{\mu_{\text{spin}} \mu_{\text{val}}}{Y_0 n(y)} \sum_{\nu} Z^{\nu} \int dH [H - \varepsilon^{\nu\min}(y)] f_0^{\nu}(y, H).\end{aligned}$$

The discretization is given by

$$\langle E \rangle(y_i) = \frac{\mu_{\text{spin}} \mu_{\text{val}}}{Y_0 n(y_i)} \sum_{\nu} Z^{\nu} \sum_j \left[ \frac{1}{2} (H_{j+} - \varepsilon^{\nu\min}(y_i))^2 - \frac{1}{2} (H_{j-} - \varepsilon^{\nu\min}(y_i))^2 \right] \quad (\text{A.11})$$

$$\times f_0^{\nu}(y_i, H_j). \quad (\text{A.12})$$

## A.2 Small Signal

### Density

The straightforward linearization of the sheet density in Eq. (A.1) yields

$$\underline{n}^{\nu}(y) = \int \frac{d^2k}{(2\pi)^2} \underline{f}^{\nu}(y, \mathbf{k}) \quad (\text{A.13})$$

$$= \frac{Z^{\nu}}{Y_0} \int_{\varepsilon^{\nu}(y)}^{\infty} dH \underline{f}_0^{\nu}(y, H) - \frac{Z^{\nu}}{Y_0} f_0^{\nu}(y, \varepsilon^{\nu}(y)) \underline{\varepsilon}^{\nu}(y). \quad (\text{A.14})$$

Note the additional contribution of the small signal subband energy due to the  $H$ -transformation. In the continuum the sheet density can also be expressed as

$$\underline{n}^{\nu}(y) = \frac{Z^{\nu}}{Y_0} \int_{\varepsilon^{\nu}(y)}^{\infty} dH \left[ \underline{f}_0^{\nu}(y, H) + \frac{\partial f_0^{\nu}(y, H)}{\partial H} \underline{\varepsilon}^{\nu}(y) \right]. \quad (\text{A.15})$$

The 3D density of Eq. (A.3) can also be linearized as

$$\underline{n}_{3\text{D}}^{\nu}(\mathbf{r}) = \underline{n}^{\nu}(y) |\Psi^{\nu}(\mathbf{r})|^2 + 2n^{\nu}(y) \Psi^{\nu}(\mathbf{r}) \underline{\Psi}^{\nu}(\mathbf{r}), \quad (\text{A.16})$$

where it was assumed that the stationary wave function  $\Psi$  is real. The total densities can be obtained by a sum over all subbands and by multiplying with the valley and spin multiplicities.

Eqs. (A.13) and (A.15) are not the same in discretized  $H$ -space and for the reasons discussed in Sect. 3.5, we use the expression of Eq. (A.15) in the small

signal density of the PE. Therefore the small signal density as seen from the perspective of the small signal BE and PE is given by

$$\underline{n}^\nu(y_i) = \frac{Z^v}{Y_0} \sum_j \left[ \Delta H^\nu(y_i, H_j) \underline{f}_0^\nu(y_i, H_j) + \int_{\Delta H^\nu(y_i, H_j)} dH \frac{\partial f_0^\nu(y_i, H)}{\partial H} \underline{\varepsilon}^\nu(y_i) \right].$$

with the integral over the  $H$ -derivative given by Eq. (3.58).

### Current Density

The small signal current density is the linearization of Eq. (A.6) and therefore it reads

$$\begin{aligned} \underline{j}^\nu(y) &= (T_{\text{HV}}^v)_{yy} \int \frac{d^2k}{(2\pi)^2} v_y^v(\mathbf{k}) \underline{f}^\nu(y, \mathbf{k}) \\ &= (T_{\text{HV}}^v)_{yy} Z^v \left[ \int dH (v_y^v)_1(y, H) \underline{f}_1^\nu(y, H) + \int dH (\underline{v}_y^v)_1(y, H) f_1^\nu(y, H) \right. \\ &\quad \left. - (v_y^v)_1(y, \varepsilon^\nu(y)) f_1^\nu(y, \varepsilon^\nu(y)) \underline{\varepsilon}^\nu(y) \right], \end{aligned} \quad (\text{A.17})$$

where the first Fourier harmonic of the small signal group velocity is given by

$$(\underline{v}_y^v)_1(y, H) = \frac{\partial (v_y^v)_1(y, H)}{\partial \varepsilon^\nu(y)} \underline{\varepsilon}^\nu(y) = -\frac{1}{2} \sqrt{\frac{2\pi}{m_d^v(H - \varepsilon^\nu(y))}} \underline{\varepsilon}^\nu(y).$$

Then, the 3D current density can be obtained by linearization of Eq. (A.7) as

$$\underline{\mathbf{J}}_y^\nu(\mathbf{r}) = \underline{j}^\nu(y) |\Psi^\nu(\mathbf{r})|^2 \mathbf{e}_y + 2j^\nu(y) \Psi^\nu(\mathbf{r}) \underline{\Psi}^\nu(\mathbf{r}) \mathbf{e}_y,$$

where it is assumed that the stationary wave function is real. The total current densities can be found by summing over all subbands and multiplying by valley and spin multiplicities. The discretization reads

$$\begin{aligned} \underline{j}^\nu(y_{i+}) &= (T_{\text{HV}}^v)_{yy} Z^v \left[ \sum_j \underline{f}_1^\nu(y_{i+}, H_j) \int_{\Delta H^\nu(y_{i+}, H_j)} dH (v_y^v)_1(y_{i+}, H) \right. \\ &\quad + \sum_j f_1^\nu(y_{i+}, H_j) \int_{\Delta H^\nu(y_{i+}, H_j)} dH (\underline{v}_y^v)_1(y_{i+}, H) \\ &\quad \left. - f_1^\nu(y_i, H_{j_{\min}}) \underline{\varepsilon}^\nu(y_{i+}) \frac{\partial}{\partial \varepsilon^\nu(y_{i+})} \int_{\Delta H^\nu(y_{i+}, H_{j_{\min}})} dH (v_y^v)_1(y_{i+}, H) \right], \end{aligned}$$

where all integrals need to be evaluated analytically as shown in Eq. (2.78) and  $j_{\min}$  is the index of the lowest non-zero  $H$ -box. Note that the term proportional to the small signal subband energy needs to be evaluated in such a way that it is consistent with integral of the group velocity over the lowest box.

### Terminal Current

In Sect. 3.4, the small signal terminal current has been computed via the Ramo-Shockley theorem. The resulting discretized terminal current at contact  $C$  per length in  $z$ -direction is given by

$$\begin{aligned} \underline{I}'_C &= q\mu_{\text{spin}} \mu_{\text{val}} \sum_{k,i,\nu} \Delta x_k \Delta y_i h_C(x_k, y_i) \\ &\times \left[ \left( \frac{j^\nu(y_{i+}) - j^\nu(y_{i-})}{\Delta y_i} - \underline{S}^\nu(y_i) \right) |\Psi^\nu(x_k, y_i)|^2 \right. \\ &\quad \left. - 2i\omega n^\nu(y_i) \Psi^\nu(x_k, y_i) \underline{\Psi}^\nu(x_k, y_i) \right] - i\omega \sum_{C'} C'_{C,C'} \underline{V}_{\text{appl}}^{C'}. \end{aligned}$$

### Admittance Parameters

Admittance parameters are given by

$$Y'_{C,C'} = \frac{\partial \underline{I}'_C}{\partial \underline{V}_{\text{appl}}^{C'}} = \frac{\underline{I}'_C}{\underline{V}_{\text{appl}}^{C'}}, \quad \text{with } \underline{V}_{\text{appl}}^{C''} = 0, \quad C'' \neq C',$$

where  $\underline{V}_{\text{appl}}^{C'}$  is the small signal bias applied to the contact  $C'$ .

### Hybrid Parameters

For a set of admittance parameters  $Y'_{\text{GG}}, Y'_{\text{GD}}, Y'_{\text{DG}}$  and  $Y'_{\text{DD}}$  in common-source configuration (see Sect. 3.6.3), the hybrid parameters are defined as

$$\begin{aligned} H'_{\text{GG}} &= \frac{1}{Y'_{\text{GG}}}, & H'_{\text{GD}} &= -\frac{Y'_{\text{GD}}}{Y'_{\text{GG}}}, \\ H'_{\text{DG}} &= \frac{Y'_{\text{DG}}}{Y'_{\text{GG}}}, & H'_{\text{DD}} &= \frac{\det(Y')}{Y'_{\text{GG}}}. \end{aligned}$$

### Cutoff Frequency

The cutoff frequency is defined as the frequency at which

$$|H'_{\text{DG}}(f_{\text{T}})| \Big|_{\underline{V}_{\text{appl}}^{\text{D}}=0} = 1$$

holds. It can be approximated from the hybrid parameter measured at frequency  $f_0$  as

$$f_T \approx |H'_{DG}(f_0)|f_0.$$

### Unilateral Gain

The unilateral gain is defined as

$$U = \frac{|Y'_{DG} - Y'_{GD}|^2}{4(\operatorname{Re}(Y'_{GG})\operatorname{Re}(Y'_{DD}) - \operatorname{Re}(Y'_{GD})\operatorname{Re}(Y'_{DG}))}.$$

### Rollet Factor

The Rollet factor is given by

$$K = \frac{2\operatorname{Re}(Y'_{GG})\operatorname{Re}(Y'_{DD}) - \operatorname{Re}(Y'_{DG}Y'_{GD})}{|Y'_{DG}Y'_{GD}|}.$$

### Maximum Oscillation Frequency

The maximum oscillation frequency is the frequency at which

$$U(f_{\max}) = 1$$

holds. It can be extrapolated from a known frequency  $f_0$  using

$$f_{\max}^2 \approx U(f_0)f_0^2.$$

## A.3 Noise

### Fano Factor

The Fano factor is the ratio of the drain current noise to the noise as if it were pure shot noise. Thus, it is given by

$$F = \frac{P_{DD}}{2qI_D}.$$

### Drain Excess Noise Factor

The drain excess noise is defined as the ratio of the actual drain current noise to the thermal noise floor:

$$\gamma = \frac{P_{DD}}{4k_B T g_{D0}},$$

where  $g_{D0}$  is the drain self-admittance  $Y'_{DD}$  but at zero drain bias and zero frequency, i.e.

$$g_{D0} = Y'_{DD} \Big|_{V_{DS}=0V, f=0Hz}.$$

### Gate Excess Noise Factor

The gate excess noise is the ratio of gate noise to the gate noise in a simple charge sheet model. It is defined as

$$\beta = \frac{P_{GG}}{\frac{4}{5}k_B T \frac{(\omega C_{GS,sat})^2}{g_{D0}}},$$

where the gate-source capacitance can be approximated using the equivalent circuit representation of Eq. (5.4) which yields

$$\omega C_{GS,sat} \approx \text{Im}(Y'_{GG} + Y'_{GD}) \Big|_{V_{DS}=0.7V},$$

in the low-frequency regime, i.e. as long as  $Y'_{GG}$  and  $Y'_{GD}$  are linear in the frequency.

### Cross-Correlation Coefficient

The cross-correlation coefficient is the cross power spectral density, normalized by the gate and drain noise:

$$c = \frac{P_{GD}}{\sqrt{P_{GG}P_{DD}}}.$$

# Appendix B

## Normalization

It is generally a good idea to normalize all quantities in such a way that their order of magnitude is about the same in order to avoid numerical issues in the implementation of the simulator. To this end, we use a temperature dependent normalization. The procedure is as follows: A quantity, say  $X$  has units, say  $[U]$ . We then use the normalization  $U_0$  to determine  $X$  in simulator units as

$$X^{\text{sim}} = \frac{X}{U_0}.$$

Any calculations in the implementation are conducted in simulator units. Let us say that we compute some quantity  $Y^{\text{sim}}$  with units  $[V]$  in the end of our simulation.  $Y^{\text{sim}}$  could be a current or a density. In order to obtain the actual result, we would need to multiply with the normalization  $V_0$ , i.e.

$$Y = Y^{\text{sim}} V_0.$$

In the following we will list the temperature dependent normalization, where  $T$  is the temperature,  $m_0$  is the electron rest mass,  $q$  is the positive electron charge,  $\hbar$  is the Planck constant, and  $k_B$  is the Boltzmann constant. All other units can be derived directly from these.

<b>Unit Normalization</b>	<b>Description</b>
$K_0 = T$	kelvin
$C_0 = q$	coulomb
$\text{kg}_0 = m_0$	kilogram
$\text{eV}_0 = k_B T$	electron volt
$J_0 = q k_B T$	joule
$m_0 = \hbar q k_B T / m_0$	meter
$s_0 = \hbar / k_B T$	second
$A_0 = q k_B T / \hbar$	ampere
$V_0 = k_B T$	volt



## Appendix C

# Discretized Power Spectral Density

Due to the lengthiness of the equation for the discretized power spectral density (PSD), we did not explicitly write it down in Sect. 4.6.3. But for completeness, we still want to include the whole expression at this point. Recall that the total PSD was given by the sum of the PSD of scattering processes and the PSD of generation and recombination (GR) processes, i.e.

$$P_{I'_C, I'_{C'}}(\omega) = Q_{I'_C, I'_{C'}}(\omega) + R_{I'_C, I'_{C'}}(\omega).$$

### Scattering

The PSD of scattering processes is given by Eq. (4.39) and its discretization reads

$$\begin{aligned} Q_{I'_C, I'_{C'}} &= 2 \sum_{\nu, \nu'} Z^\nu Z^{\nu'} \sum_i \Delta y_i \sum_{\eta, \sigma} \sum_j \min(\Delta H^\nu(y_i, H_j), \Delta H^{\nu'}(y_i, H_j + \sigma \hbar \omega_\eta)) \\ &\quad \times c_\eta^{\nu, \nu'}(y_i, H_j, H_j + \sigma \hbar \omega_\eta) \\ &\quad \times \left[ \sum_m Q_m^{\nu, \nu'}(y_i, H_j, H_j + \sigma \hbar \omega_\eta) + \sum_{m, n} Q_{m, n}^{\nu, \nu'}(y_i, H_j, H_j + \sigma \hbar \omega_\eta) \right. \\ &\quad \left. + \sum_{n, n, \ell} Q_{m, n, \ell}^{\nu, \nu'}(y_i, H_j, H_j + \sigma \hbar \omega_\eta) \right] \end{aligned} \quad (\text{C.1})$$

In the above expression, we need to use the very same box-integration method in  $H$ -space, as in the Boltzmann equation, where the actual integrated volume in  $H$ -space is the *smaller* of the two  $H$ -boxes of the initial and final states (see Sect. 2.4.5):

$$\min(\Delta H^\nu(y_i, H_j), \Delta H^{\nu'}(y_i, H_j + \sigma \hbar \omega_\eta)).$$

The sum over  $\eta$  represents the sum over all scattering mechanisms with  $\sigma = \pm 1$  for the inelastic inter-valley scattering with the associated energy transfers  $\hbar\omega_\eta$  and  $\sigma = 0$  for the elastic acoustic phonon scattering and surface roughness scattering. Note that we assume that  $\hbar\omega_\eta$  has been mapped onto a multiple of the  $H$ -grid spacing, such that  $H_j + \hbar\omega_\eta$  lies on an  $H$ -grid point.

Moreover, we defined multiple quantities depending on the Fourier coefficients. The first one is given by

$$\begin{aligned} Q_m^{\nu,\nu'}(y_i, H_j, H_{j'}) \Big|_{m \text{ even}} &= \frac{1}{Y_0} \left\{ (G^{Ic})_m^\nu(y_i, H_j) \left( (G^{Ic'})_m^\nu(y_i, H_j) \right)^* f_0^{\nu'}(y_i, H_{j'}) \right. \\ &\quad - \left[ (G^{Ic})_0^\nu(y_i, H_j) \left( (G^{Ic'})_m^{\nu'}(y_i, H_{j'}) \right)^* \right. \\ &\quad \left. \left. + (G^{Ic})_m^{\nu'}(y_i, H_{j'}) \left( (G^{Ic'})_0^\nu(y_i, H_j) \right)^* \right] f_m^{\nu'}(y_i, H_{j'}) \right\} \end{aligned}$$

when  $m$  is an even number. For an odd  $m$ , we need to average the quantities defined on the adjoint grid as explained in Sect. 4.6.3, i.e. each product needs to be averaged as a whole. For the above term this yields

$$\begin{aligned} Q_m^{\nu,\nu'}(y_i, H_j, H_{j'}) \Big|_{m \text{ odd}} &= \frac{1}{Y_0} \frac{1}{2} \left\{ (G^{Ic})_m^\nu(y_{i-}, H_j) \left( (G^{Ic'})_m^\nu(y_{i-}, H_j) \right)^* f_0^{\nu'}(y_i, H_{j'}) \right. \\ &\quad - \left[ (G^{Ic})_0^\nu(y_i, H_j) \left( (G^{Ic'})_m^{\nu'}(y_{i-}, H_{j'}) \right)^* \right. \\ &\quad \left. \left. + (G^{Ic})_m^{\nu'}(y_{i-}, H_{j'}) \left( (G^{Ic'})_0^\nu(y_i, H_j) \right)^* \right] f_m^{\nu'}(y_{i-}, H_{j'}) \right. \\ &\quad + (G^{Ic})_m^\nu(y_{i+}, H_j) \left( (G^{Ic'})_m^\nu(y_{i+}, H_j) \right)^* f_0^{\nu'}(y_i, H_{j'}) \\ &\quad - \left[ (G^{Ic})_0^\nu(y_i, H_j) \left( (G^{Ic'})_m^{\nu'}(y_{i+}, H_{j'}) \right)^* \right. \\ &\quad \left. \left. + (G^{Ic})_m^{\nu'}(y_{i+}, H_{j'}) \left( (G^{Ic'})_0^\nu(y_i, H_j) \right)^* \right] f_m^{\nu'}(y_{i+}, H_{j'}) \right\}. \end{aligned}$$

Bear in mind that on the first and last grid point, we only integrate over half the box and there is no averaging necessary. For example, for the first grid point the box  $\Delta y_1$  goes from  $y_1$  to  $y_{1+}$ , therefore evaluating the even harmonics on  $y_1$  and the odd harmonics on  $y_{1+}$  is sufficient without any averaging.

The term proportional to two harmonics in Eq. (C.1) is given by

$$Q_{m,n}^{\nu,\nu'}(y_i, H_j, H_{j'}) \Big|_{m,n \text{ even}} = \left[ (G^{I_c})_m^\nu(y_i, H_j) \left( (G^{I_{c'}})_n^{\nu'}(y_i, H_{j'}) \right)^* \right. \\ \left. + (G^{I_c})_n^{\nu'}(y_i, H_{j'}) \left( (G^{I_{c'}})_m^\nu(y_i, H_j) \right)^* \right] \\ \times f_m^\nu(y_i, H_j) f_n^{\nu'}(y_i, H_{j'})$$

If one or both of the harmonics is odd, we average as before. If, say,  $m$  is even and  $n$  is odd, we find

$$Q_{m,n}^{\nu,\nu'}(y_i, H_j, H_{j'}) \Big|_{\substack{m \text{ even} \\ n \text{ odd}}} = \frac{1}{2} \left[ (G^{I_c})_m^\nu(y_i, H_j) \left( (G^{I_{c'}})_n^{\nu'}(y_{i-}, H_{j'}) \right)^* \right. \\ \left. + (G^{I_c})_n^{\nu'}(y_{i-}, H_{j'}) \left( (G^{I_{c'}})_m^\nu(y_i, H_j) \right)^* \right] \\ \times f_m^\nu(y_i, H_j) f_n^{\nu'}(y_{i-}, H_{j'}) \\ + \frac{1}{2} \left[ (G^{I_c})_m^\nu(y_i, H_j) \left( (G^{I_{c'}})_n^{\nu'}(y_{i+}, H_{j'}) \right)^* \right. \\ \left. + (G^{I_c})_n^{\nu'}(y_{i+}, H_{j'}) \left( (G^{I_{c'}})_m^\nu(y_i, H_j) \right)^* \right] \\ \times f_m^\nu(y_i, H_j) f_n^{\nu'}(y_{i+}, H_{j'}),$$

and completely analogous if  $m$  is odd and  $n$  is even or if both are odd.

The term in Eq. (C.1) with three Fourier harmonic coefficients reads

$$Q_{m,n,\ell}^{\nu,\nu'}(y_i, H_j, H_{j'}) \Big|_{m,n,\ell \text{ even}} = \frac{1}{Y_0} \int d\phi Y_m(\phi) Y_n(\phi) Y_\ell(\phi) \\ \times \left\{ (G^{I_c})_m^\nu(y_i, H_j) \left( (G^{I_{c'}})_n^{\nu'}(y_i, H_{j'}) \right)^* \right. \\ \times f_\ell^{\nu'}(y_i, H_{j'}) \left( \frac{1}{Y_0} - f_0^\nu(y_i, H_j) \right) \\ \left. - (G^{I_c})_m^\nu(y_i, H_j) \left( (G^{I_{c'}})_n^{\nu'}(y_i, H_{j'}) \right)^* \right. \\ \left. \times f_\ell^{\nu'}(y_i, H_{j'}) f_0^{\nu'}(y_i, H_{j'}) \right\}.$$

The averaging when one or more of the harmonics are odd proceeds analogously to the previous cases. The integral over the three Fourier harmonics is given by Eq. (2.62).

### Generation and Recombination

The discretized PSD of the GR term on the contacts of Eq. (4.40) is given by

$$\begin{aligned}
R_{I_C, I_{C'}} &= 2 \sum_{C_0=S,D} \sum_{\nu} Z^{\nu} \sum_j \Delta H^{\nu}(y_C, H_j) \\
&\times \sum_{m,n \text{ even}} (G^{I_C})_m^{\nu}(y_{C_0}, H_j) \left( (G^{I_{C'}})_n^{\nu}(y_{C_0}, H_j) \right)^* \\
&\times v_{\text{GR}} \left[ f_{\text{eq}}^{\nu}(y_{C_0}, H_j) \delta_{m,n} \right. \\
&\quad \left. + (1 - 2f_{\text{eq}}^{\nu}(y_{C_0}, H_j)) \sum_{\ell \text{ even}} f_{\ell}^{\nu}(y_{C_0}, H_j) \int d\phi Y_m(\phi) Y_n(\phi) Y_{\ell}(\phi) \right],
\end{aligned} \tag{C.2}$$

where the indices  $m$ ,  $n$ , and  $\ell$  only run over *even* numbers because the GR rate is defined only on the direct grid points of the source and drain contacts. Furthermore, the integral over the three Fourier harmonics is given by Eq. (2.62).

Bear in mind considerations relating degeneracy affect the GR noise just as well. If we describe only one true state of a  $\mu$ -times degenerate group, the total noise is a factor of  $\mu$  larger. Hence, when we consider our usual degeneracy of  $\mu_{\text{spin}} = \mu_{\text{val}} = 2$ , we need to multiply the resulting PSD of the GR noise by a factor of  $\mu_{\text{spin}} \mu_{\text{val}} = 4$ .

Obviously the GR rate only contributes at the positions  $y_S$  and  $y_D$  to the noise, therefore the contribution of the GR term to the PSD per grid point in transport direction (see Eq. (4.43)) is given by the respective term on the source or drain contact grid point divided by  $\Delta y$  of the source or drain grid point.

# Nomenclature

An alphabetic list of acronyms and symbols used in this work. Symbols are sorted according to their English spelling, e.g.  $\rho$  can be found at the position of ‘rho’ and  $\Gamma$  can be found at the position of ‘Gamma.’

<b>Acronym</b>	<b>Expansion</b>
BE	Boltzmann equation
BG	bottom gate
CMOS	complementary metal-oxide-semiconductor
D	drain
DD	drift-diffusion
DIBL	drain-induced barrier lowering
GR	generation and recombination
HD	hydrodynamic
LA	longitudinal acoustic
LBE	Langevin-Boltzmann equation
LO	longitudinal optical
MC	Monte Carlo
MOSFET	metal-oxide-semiconductor field-effect transistor
PE	Poisson equation
PSD	power spectral density
RF	radio frequency
S	source
SE	Schrödinger equation
TG	top gate
TA	transverse acoustic
TO	transverse optical

Symbol	Description
$a$	aggregate index for the Poisson equation, see Eq. (2.105)
$a_0$	silicon crystal lattice spacing, $a_0 \approx 0.54$ nm
$A^{\text{AC}}$	matrix of the full linear small signal system containing both the linearized time-dependent Boltzmann equation and the linearized quasistationary Poisson equation, see Eq. (3.62)
$A_{e/e}^{\text{AC}}$	the matrix $A^{\text{AC}}$ but without the odd rows and columns of the Boltzmann equation subspace, see Eq. (3.64)
$A_{o/e}^{\text{AC, BE}}$	the matrix containing the Boltzmann equation subspace $A^{\text{AC}}$ but only the odd harmonic rows and even harmonic columns, Sect. 3.6.2
$A_{o/o}^{\text{AC, BE}}$	the matrix containing the Boltzmann equation subspace $A^{\text{AC}}$ but only the odd harmonic rows and columns, Sect. 3.6.2
$A^{\text{BE}}$	matrix containing the discretized Boltzmann equations, see Eq. (2.95)
$A_{e/e}^{\text{BE}}$	only the even rows and even columns of $A^{\text{BE}}$ , see Eq. (2.100)
$A^{\text{DC}}$	Jacobian of the full Newton-Raphson approach in the stationary case, see Eq. (2.107)
$A_{e/e}^{\text{DC}}$	$A^{\text{DC}}$ but only with the even rows and even columns in the BE subspace, see Eq. (2.111)
$\alpha$	aggregate index of the Boltzmann equation, see Eq. (2.86)
$b$	aggregate index of the Poisson equation, see Eq. (2.105)
$\mathbf{b}^C$	r.h.s. of the full small signal system with an applied bias at contact $C$ , see Eq. (3.62)
$\mathbf{b}_e^C$	same as $\mathbf{b}^C$ but with the rows of odd harmonics removed, see Eq. (3.64)
$B$	matrix containing the Langevin-sources in the discretized Green's function system, see Eq. (4.16)
$B_e$	same as $B$ but does not contain the rows of odd harmonics in the Boltzmann equation subspace, see Eq. (4.28)
$B_o^{\text{BE}}$	same as $B$ but only includes rows of odd harmonics in the Boltzmann equation subspace
$\beta$	aggregate index of the Boltzmann equation, see Eq. (2.86)
$C, C'$	indices running over the contacts, $C, C' \in \{\text{TG, BG, S, D}\}$

$C_e$	same as $C_e^{\text{BE}}$ but for the whole Boltzmann and Poisson equation system, see Eq. (2.110)
$C_e^{\text{BE}}$	compression matrix to remove all odd rows in the matrix of the Boltzmann equation, see Eq. (2.99)
$C_o^{\text{BE}}$	compression matrix to remove all even rows in the matrix of the Boltzmann equation, see Eq. (3.68)
$c_{\eta}^{\nu,\nu'}(y; \mathbf{k}, \mathbf{k}')$	transition rate coefficient, see Eqs. (2.28), (2.31), (2.35)
$\mathcal{C}_{C,C'}$	capacitance matrix, see Eq. (3.41)
$D$	set containing all 2D vectors in the device, $\mathbf{r} \in D \subset \mathbb{R}^2$
$D_{\text{ac}}$	effective deformation potential of acoustic phonons, see Eq. (2.26)
$D_{\eta}$	phonon deformation potential of inter-valley transitions, see Tab. 2.1
$\delta \mathbf{x}$	solution of the full Newton-Raphson system containing the distribution function and the potential, see Eq. (2.108)
$\delta \mathbf{x}_e$	same as $\delta \mathbf{x}$ but only even equations in the Boltzmann equation part, see Eq. (2.111)
$\Delta H^{\nu}(y, H_j)$	box in $H$ -space surrounding the $H$ -grid point $H_j$ , see Eq. (2.75)
$\Delta x_k$	box in $x$ -direction surrounding the grid point $x_k$
$\Delta y_i$	box in $y$ -direction surrounding the grid point $y_i$ , see Eq. (2.74)
$\Delta z$	arbitrary length in the homogeneous $z$ -direction
$\text{diag}(d_1, \dots, d_n)$	$n \times n$ matrix with the diagonal elements $d_1, \dots, d_n$ , and all other elements zero
$\text{diag}(A)$	matrix with the same number of rows and columns as $A$ , but only containing the elements of $A$ on the main diagonal
$E^{\nu}(\mathbf{k})$	kinetic energy, see Eqs. (2.19), (2.49)
$E_C$	conduction band energy, see Eq. (2.42)
$E_{\text{el}}$	electric field in $y$ -direction
$E_{\text{F}}^{\text{S/D}}$	Fermi energy of source or drain contact thermal baths, see Eq. (2.46)
$\mathbf{e}_x, \mathbf{e}_y, \mathbf{e}_z$	unit vectors in $x$ -, $y$ -, and $z$ -direction, respectively
$\varepsilon^{\nu}(y)$	subband energy in the valley and subband $\nu$ and at position $y$ , see Eq. (2.7)
$\tilde{\varepsilon}^{\nu}(y, t)$	time-dependent subband energy in the valley and subband $\nu$ and at position $y$ , see Eq. (3.6)
$\underline{\varepsilon}^{\nu}(y)$	small signal phasor of subband energy in the valley and subband $\nu$ and at position $y$ , see Eq. (3.8)

$f$	frequency
$f^\nu(y, \mathbf{k})$	distribution function of the stationary Boltzmann equation, see Sect. 2.4.1
$\tilde{f}^\nu(y, \mathbf{k}, t)$	distribution function of the time-dependent Boltzmann equation, see Eq. (3.4)
$\underline{f}^\nu(y, \mathbf{k})$	small signal phasor of the distribution function, see Eq. (3.4)
$\underline{f}_{-\alpha}^{\text{norm}, C}$	normalized small signal phasor of the distribution function, see Eq. (3.61)
$f_{\text{eq}}^\nu(y, \mathbf{k})$	equilibrium distribution function
$\tilde{f}_{\text{eq}}^\nu(y_{\text{S/D}}, H, t)$	time-dependent equilibrium distribution function, see Eq. (3.30)
$F$	force acting on electrons, see Eq. (2.14)
$\mathbf{F}$	r.h.s. of full Newton-Raphson system, containing the Boltzmann and Poisson equations, see Eq. (2.108)
$F^{\text{BE}}$	stationary Boltzmann equation, see Eq. (2.12)
$\tilde{F}^{\text{BE}}$	full Boltzmann equation, see Eq. (3.2)
$\underline{F}^{\text{BE}}$	small signal phasor of Boltzmann equation, see Eq. (3.10)
$\mathbf{F}_e$	same as $\mathbf{F}$ but only even equations in the Boltzmann equation part, see Eq. (2.111)
$F^{\text{PE}}$	stationary Poisson equation, see Eq. (2.5)
$\tilde{F}^{\text{PE}}$	quasistationary Poisson equation with time-dependent variables
$\underline{F}^{\text{PE}}$	small signal phasor of Poisson equation, see Eq. (3.13)
$G$	matrix consisting of the Green's functions of the distribution function and the Green's functions of the potential, see Eq. (4.16)
$G_e$	same as $G$ but does not contain rows of odd harmonics in the Boltzmann equation subspace, see Eq. (4.28)
$G_o^{\text{BE}}$	same as $G$ but only contains the rows of odd harmonics in the Boltzmann equation subspace, see Eq. (4.29)
$(G^f)^{\nu, \nu'}(y, \mathbf{k}; y', \mathbf{k}')$	Green's function of the distribution function with a Langevin-source in the Langevin-Boltzmann equation, see Eq. (4.10)
$G^V(\mathbf{r}; \mathbf{r}')$	Green's function of the potential with a Langevin-source in the Poisson equation, see Eq. (4.11)
$G_\zeta^{IC}$	Green's function of the terminal current at contact $C$ for a Langevin-source at the coordinates $\zeta$ , see Eq. (4.23)
$G_\zeta^{I'C}$	same as $G_\zeta^{IC}$ but for the terminal current per length in $z$ -direction

$\mathbf{G}^{I_C}$	vector with the elements consisting of the Green's functions, $G_{\zeta}^{I_C}$ , see Eq. (4.26)
$\gamma$	aggregate index running over both the Boltzmann equation and the Poisson equation, see Eq. (4.15)
$\gamma^{\nu,C}(y; \mathbf{k}, \mathbf{k}')$	transition rate of the boundary generation and recombination term, see Eq. 2.37
$\Gamma^{\nu}(y, \mathbf{k})$	boundary generation and recombination rate, see Sect. 2.4.3
$\tilde{\Gamma}^{\nu}(y, \mathbf{k}, t)$	boundary term of time-dependent Boltzmann equation, see Eq. (3.2)
$\tilde{\Gamma}^{\nu}(y, t)$	boundary term of time-dependent Boltzmann equation integrated over $\mathbf{k}$ -space, see Eq. (3.36)
$\underline{\Gamma}_{\alpha}$	small signal phasor of boundary term of time-dependent Boltzmann equation, see Eq. (3.31)
$H$	total energy, variable in $H$ -transformed energy space, see Eq. (2.65)
$\hat{H}$	Hamilton operator, see Eq. (2.7)
$\hbar$	Planck's constant divided by $2\pi$
$I_C$	stationary terminal current at contact $C$
$\tilde{I}_C(t)$	time-dependent terminal current at contact $C$ , see Eq. (3.33)
$\underline{I}_C$	small signal phasor of terminal current at contact $C$ , see Eq. (3.45)
$\underline{I}'_C$	small signal phasor of terminal current per length in $z$ -direction at contact $C$ , see Eq. (3.46)
$\mathbb{I}$	unit matrix
$j(y)$	electron sheet current density in $y$ -direction at position $y$ , see Eq. (A.8)
$j^{\nu}(y)$	electron sheet current density in $y$ -direction in the subband $\nu$ at position $y$ , see Eq. (A.6)
$\tilde{j}^{\nu}(y)$	time-dependent electron sheet current density in $y$ -direction
$\underline{j}^{\nu}(y)$	small signal phasor of electron sheet current density in $y$ -direction, see Eq. (A.17)
$\tilde{\mathbf{J}}^{\nu}(\mathbf{r}, t)$	time-dependent current density in the subband $\nu$ , see Eq. (3.34)
$\tilde{\mathbf{J}}_y^{\nu}(\mathbf{r}, t)$	$y$ -component of time-dependent current density in the subband $\nu$ , see Eq. (3.35)
$j_{\min}$	$H$ -space index of the energetically lowest non-zero $H$ -box

$K_{C,C'}(y)$	power spectral density of terminal currents per grid point in $y$ -direction, see Eq. (4.43)
$\mathcal{K}_{C,C'}(y, H)$	power spectral density of terminal currents per grid point in $y$ - and $H$ -direction, see Eq. (4.44)
$\mathcal{K}_{C,C'}^\nu(y, H)$	power spectral density of terminal currents per grid point in $y$ - and $H$ -direction and subband $\nu$ , see Eq. (4.45)
$k_B$	Boltzmann constant
$\kappa(\mathbf{r})$	dielectric constant, see Eq. (2.1)
$L^\nu(y, \mathbf{k})$	stationary free streaming term, see Sect. 2.4.1
$\tilde{L}^\nu(y, \mathbf{k}, t)$	free streaming term of time-dependent Boltzmann equation, see Eq. (3.2)
$\underline{L}_m^\nu(y, H)$	small signal phasor of free streaming term of time-dependent Boltzmann equation
$m_0$	electron rest mass
$m_d$	Herring-Vogt mass, see Eq. (2.48)
$m_x$	effective mass in $x$ -direction, see Sect. 2.4.1
$m_y$	effective mass in $y$ -direction, see Sect. 2.4.1
$m_z$	effective mass in $z$ -direction, see Sect. 2.4.1
$\mu$	multiplicity, see Sect. 2.4.6
$\mu_{\text{spin}}$	spin multiplicity, see Eq. 2.92
$\mu_{\text{val}}$	valley multiplicity, see Eq. 2.93
$n$	total dimension of the discretized full Newton-Raphson Jacobian but only considering even harmonics, $n = N_{\text{BE}}/2 + N_{\text{PE}}$
$n(y)$	electron sheet density at position $y$ , see Eq. (A.2)
$n^\nu(y)$	electron sheet density in the subband $\nu$ at position $y$ , see Eq. (A.1)
$\tilde{n}(y, t)$	time-dependent electron sheet density, see Eq. (3.24)
$\underline{n}(y)$	small signal phasor of electron sheet density, see Eq. (3.24)
$n_{3\text{D}}(\mathbf{r})$	3D electron density, see Eq. (A.4)
$n_{3\text{D}}^\nu(\mathbf{r})$	3D electron density per subband, see Eq. (A.3)
$\underline{n}_{3\text{D}}(\mathbf{r})$	small signal phasor of the 3D electron density, see Eq. (3.17)
$n_{\text{eq}}^\nu(y)$	electron sheet density in the subband $\nu$ in equilibrium
$n_i$	intrinsic carrier density
$n_\lambda$	phonon number, see Eq. (2.25)
$N_{3\text{D}}^{\text{eff}}$	effective density of states of silicon in 3D, see Eq. (2.44)
$N$	total dimension of the discretized full Newton-Raphson Jacobian, $N = N_{\text{BE}} + N_{\text{PE}}$

$N_A$	acceptor doping density, see Eq. (2.2)
$N_{BE}$	dimension of the discretized Boltzmann equation system
$N_D$	donor doping density, see Eq. (2.2)
$N_{PE}$	dimension of the discretized Poisson equation system
$N_x$	highest index of grid in $x$ -direction, see Sect. 2.3.3
$N_y$	highest index of grid in $y$ -direction, see Eq. (2.72)
$N_H$	highest index of grid in $H$ -direction, see Eq. (2.73)
$\nabla_{\mathbf{r}}$	Nabla-operator in real space
$\nabla_{\mathbf{k}}$	Nabla-operator in $\mathbf{k}$ -space
$\nu$	combined valley and subband index, $\nu = (v, s)$ , see Sect. 2.4.1
$\text{offdiag}(A)$	matrix identical to $A$ except the elements on the main diagonal are set to zero
$\omega$	angular frequency
$P_{I'_C, I'_C}(\omega)$	power spectral density of the terminal currents, see Eq. (4.41)
$P_{C, C'}(\omega)$	same as $P_{I'_C, I'_C}(\omega)$ , see Eq. (4.42)
$\mathcal{P}_{I'_C}$	terminal current projection operator, see Eq. (3.66)
$\partial D$	boundary of the device set $D$
$\partial D_C$	set comprising only the contact $C$
$\varphi_{MS}$	metal-semiconductor work function difference
$\Psi^\nu(\mathbf{r})$	wave function of the Schrödinger equation, see Eq. (2.7)
$\tilde{\Psi}^\nu(\mathbf{r}, t)$	time-dependent wave function of the Schrödinger equation, see Eq. (3.7)
$\underline{\Psi}^\nu(\mathbf{r})$	small signal phasor of the wave function of the Schrödinger equation, see Eq. (3.9)
$q$	positive elementary charge
$Q_{I'_C, I'_C}(\omega)$	power spectral density of terminal currents from scattering processes, see Eq. (4.36)
$\mathbf{r}$	2D spatial vector, $\mathbf{r} = (x \ y)^t$
$R^{\nu, \nu'}(y)$	overlap integral, see Eq. (2.27)
$R_{I'_C, I'_C}(\omega)$	power spectral density of terminal currents from generation and recombination processes, see Eq. (4.38)
$\rho(\mathbf{r})$	space charge density, see Eq. (2.2)
$S^\nu(y, \mathbf{k})$	scattering term, see Sect. 2.4.2
$\tilde{S}^\nu(y, \mathbf{k}, t)$	scattering term of time-dependent Boltzmann equation, see Eq. (3.2)
$\tilde{S}^\nu(y, t)$	scattering term of time-dependent Boltzmann equation integrated over $\mathbf{k}$ -space, see Eq. (3.36)
$\underline{S}^\nu(y)$	small signal phasor of scattering term of Boltzmann equation integrated over $\mathbf{k}$ -space, see Eq. (3.2)

$S_{\eta}^{\nu,\nu'}(y; \mathbf{k}, \mathbf{k}')$	transition rate of the scattering term, see Eq. (2.20)
$S$	same as $S^{\text{BE}}$ but for the full system, see Eq. (2.109)
$S^{\text{BE}}$	linear transformation to eliminate the odd distribution functions in the even equations of the BE, see Eqs. (2.96), (2.97)
$s$	subband index, see Eq. (2.7)
$T$	temperature, 300 K throughout this work
$\tilde{T}^{\nu}(y, \mathbf{k}, t)$	time-derivative of the Boltzmann equation, see Eq. (3.3)
$\underline{T}_m^{\nu}(y, H)$	$m$ -th harmonic of small signal phasor of time-derivative of the Boltzmann equation, see Eq. (3.20)
$T_{\text{HV}}^v$	Herring-Vogt transformation matrix, see Eq. (2.48)
$(T_{\text{HV}}^v)_{yy}$	$yy$ -component of Herring-Vogt transformation matrix, see Eq. (2.48)
$\Theta(\cdot)$	Heaviside step-function
$v$	valley index of the silicon band structure, see Sect. 2.4.1
$\mathbf{v}^{\nu}(\mathbf{k})$	electron group velocity, see Eq. (2.50)
$v_y^{\nu}(\mathbf{k})$	electron group velocity in $y$ -direction
$v_{\text{GR}}$	recombination velocity, see Eq. (2.37)
$V(\mathbf{r})$	quasistatic electric potential, see Eq. (2.1)
$\tilde{V}(\mathbf{r}, t)$	time-dependent electric potential, see Eq. (3.5)
$\underline{V}(\mathbf{r})$	small signal phasor of electric potential, see Eq. (3.5)
$\underline{V}_a^{\text{norm},C}$	normalized small signal phasor of electric potential, see Eq. (3.61)
$\mathcal{V}(\mathbf{r})$	potential energy, see Eq. (2.8)
$V_{\text{appl}}^C$	applied bias at contact $C$
$\tilde{V}_{\text{appl}}^C(t)$	applied bias including small signal perturbation, see Eq. (3.1)
$\underline{V}_{\text{appl}}^C$	small signal phasor of applied bias, see Eq. (3.1)
$V_T$	thermal voltage, $V_T = k_B T / q$
$x$	confinement direction, see Fig. 2.1
$\underline{\mathbf{x}}^C$	solution of the full small signal system for a small signal bias applied to contact $C$ , see Eq. (3.63)
$\underline{\mathbf{x}}_e^C$	same as $\underline{\mathbf{x}}^C$ but only containing phasors of even distribution functions see Eq. (3.64)
$x_k$	$k$ -th grid point in $x$ -direction
$\xi_{\text{BE}}^{\nu}(y, \mathbf{k}, t)$	Langevin-source term in the Boltzmann equation, see Eq. (4.6)
$\xi_{\text{PE}}^{\nu}(\mathbf{r}, t)$	Langevin-source term in the Poisson equation, see Eq. (4.8)
$y$	transport direction, see Fig. 2.1
$y_{\text{D}}$	position of the drain contact in $y$ -direction

$y_i$	$i$ -th grid point in $y$ -direction
$y_S$	position of the source contact in $y$ -direction
$Y$	admittance matrix
$Y'$	admittance matrix per length in $z$ -direction
$Y_m(\phi)$	$m$ -th Fourier harmonic, see Eq. (2.55)
$z$	homogeneous direction, see Fig. 2.1
$Z^v$	2D density of states, see Eq. (2.57)
$\zeta$	location of the Langevin-source; either a Boltzmann equation index like $\alpha$ or a Poisson equation index like $a$ ; see Sect. 4.4



# Bibliography

- [1] International Roadmap Committee, “The international technology roadmap for semiconductors,” <http://www.itrs2.net/>, 2017.
- [2] Q. Huang, F. Piazza, P. Orsatti, and T. Ohguro, “The impact of scaling down to deep submicron on CMOS RF circuits,” *IEEE J. Solid-State Circuits*, vol. 33, no. 7, pp. 1023–1036, 1998.
- [3] D. Esseni, C. Jungemann, J. Lorenz, P. Palestri, E. Sangiorgi, and L. Selmi, “Technology computer aided design,” in *Guide to State-of-the-Art Electron Devices*, J. N. Burghartz, Ed. John Wiley & Sons Inc., 2013, ch. 8.
- [4] H. C. de Graaff and F. M. Klaassen, *Compact Transistor Modelling for Circuit Design*, ser. Computational Microelectronics, S. Selberherr, Ed. Wien, New York: Springer-Verlag, 1990.
- [5] N. Arora, *MOSFET DC Model*. Vienna: Springer Vienna, 1993. [Online]. Available: <http://dx.doi.org/10.1007/978-3-7091-9247-4.6>
- [6] M. Lundstrom, *Fundamentals of Carrier Transport*, ser. Modular Series on Solid State Devices. New York: Addison-Wesley, 1990, vol. 10.
- [7] W. Hänsch, C. Mazure, A. Lill, and M. K. Orłowski, “Hot carrier hardness analysis of submicrometer LDD devices,” *Electron Devices, IEEE Transactions on*, vol. 38, no. 3, pp. 512–517, 1991.
- [8] A. Schenk, *Advanced Physical Models for Silicon Device Simulation*, ser. Computational Microelectronics, S. Selberherr, Ed. Wien, New York: Springer, 1998.
- [9] W. van Roosbroeck, “Theory of the flow of electrons and holes in germanium and other semiconductors,” *Bell System Technical Journal*, pp. 561–607, 1950.
- [10] W. Shockley, *Electrons and Holes in Semiconductors*. Princeton, New Jersey: van Nostrand, 1950.

- [11] M. Lundstrom, “Drift-diffusion and computational electronics-still going strong after 40 years!” in *Simulation of Semiconductor Processes and Devices (SISPAD), 2015 International Conference on*. IEEE, 2015, pp. 1–3.
- [12] K. Bløtekjær, “Transport equations for electrons in two-valley semiconductors,” *Electron Devices, IEEE Transactions on*, vol. 17, no. 1, pp. 38–47, 1970.
- [13] T. Grasser, R. Kosik, C. Jungemann, H. Kosina, and S. Selberherr, “Non-parabolic macroscopic transport models for device simulation based on bulk Monte Carlo data,” *J. Appl. Phys.*, vol. 97, pp. 093710–1–12, 2005.
- [14] C. Jungemann, T. Grasser, B. Neinhüs, and B. Meinerzhagen, “Failure of moments-based transport models in nanoscale devices near equilibrium,” *Electron Devices, IEEE Transactions on*, vol. 52, no. 11, pp. 2404–2408, 2005.
- [15] T. Grasser, T.-W. Tang, H. Kosina, and S. Selberherr, “A review of hydrodynamic and energy-transport models for semiconductor device simulation,” *Proc. IEEE*, vol. 91, no. 2, pp. 251–274, 2003.
- [16] C. Jacoboni and P. Lugli, *The Monte Carlo Method for Semiconductor Device Simulation*. Wien: Springer, 1989.
- [17] N. Goldsman, L. Henrickson, and J. Frey, “A physics-based analytical/numerical solution to the Boltzmann transport equation for use in device simulation,” *Solid-State Electron.*, vol. 34, pp. 389–396, 1991.
- [18] A. Gnudi, D. Ventura, G. Baccarani, and F. Odeh, “Two-dimensional MOSFET simulation by means of a multidimensional spherical harmonics expansion of the Boltzmann transport equation,” *Solid-State Electron.*, vol. 36, no. 4, pp. 575 – 581, 1993.
- [19] J. A. Carrillo, I. M. Gamba, A. Majorana, and C.-W. Shu, “A WENO-solver for the transients of Boltzmann–Poisson system for semiconductor devices: performance and comparisons with Monte Carlo methods,” *Journal of Computational Physics*, vol. 184, no. 2, pp. 498–525, 2003.
- [20] —, “2d semiconductor device simulations by WENO-boltzmann schemes: efficiency, boundary conditions and comparison to monte carlo methods,” *Journal of Computational Physics*, vol. 214, no. 1, pp. 55–80, 2006.

- [21] M. Galler and F. Schürer, “A direct multigroup-WENO solver for the 2D non-stationary Boltzmann-Poisson system for GaAs devices: GaAs-MESFET,” *Journal of Computational Physics*, vol. 212, pp. 778–797, 2006.
- [22] C. Jungemann, B. Neinhüs, S. Decker, and B. Meinerzhagen, “Hierarchical 2-D DD and HD noise simulations of Si and SiGe devices: Part II—Results,” *Electron Devices, IEEE Transactions on*, vol. 49, no. 7, pp. 1258–1264, 2002.
- [23] B. Meinerzhagen, A. T. Pham, S. M. Hong, and C. Jungemann, “Solving Boltzmann transport equation without Monte-Carlo algorithms – new methods for industrial TCAD applications,” in *Proc. SISPAD*, 2010, pp. 293–296.
- [24] C. Jungemann, A.-T. Pham, B. Meinerzhagen, C. Ringhofer, and M. Bollhöfer, “Stable discretization of the Boltzmann equation based on spherical harmonics, box integration, and a maximum entropy dissipation principle,” *J. Appl. Phys.*, vol. 100, pp. 024 502–1–13, 2006.
- [25] K. Rupp, C. Jungemann, S.-M. Hong, M. Bina, T. Grasser, and A. Jüngel, “A review of recent advances in the spherical harmonics expansion method for semiconductor device simulation,” *Journal of Computational Electronics*, vol. 15, no. 3, pp. 939–958, 2016. [Online]. Available: <http://dx.doi.org/10.1007/s10825-016-0828-z>
- [26] D. Ventura, A. Gnudi, G. Baccarani, and F. Odeh, “Multidimensional spherical harmonics expansion of Boltzmann equation for transport in semiconductors,” *Appl. Math. Lett.*, vol. 5, p. 85, 1992.
- [27] W. Liang, N. Goldsman, I. Mayergoyz, and P. Oldiges, “2- D MOSFET modeling including surface effects and impact ionization by self-consistent solution of the Boltzmann, Poisson, and hole-continuity equations,” *Electron Devices, IEEE Transactions on*, vol. 44, no. 2, pp. 257–267, 1997.
- [28] S.-M. Hong, C. Jungemann, and M. Bollhöfer, “A deterministic Boltzmann equation solver for two-dimensional semiconductor devices,” in *Proc. SISPAD*, 2008, pp. 293–296.
- [29] S.-M. Hong, A. T. Pham, and C. Jungemann, *Deterministic solvers for the Boltzmann transport equation*, ser. Computational Microelectronics, S. Selberherr, Ed. Wien, New York: Springer, 2011.
- [30] S.-M. Hong and C. Jungemann, “Simulation of magnetotransport in nanoscale devices,” in *International Conference on Solid State and Integrated Circuits Technology*, 2008, pp. 377–380.

- [31] H. Kamrani, T. Kochubey, D. Jabs, and C. Jungemann, “Electrothermal simulation of sige hbt’s and investigation of experimental extraction methods for junction temperature,” in *2015 International Conference on Simulation of Semiconductor Processes and Devices (SISPAD)*, Sept 2015, pp. 108–111.
- [32] H. Kamrani, D. Jabs, V. d’Alessandro, N. Rinaldi, and C. Jungemann, “Physics-based hot-carrier degradation model for sige hbt’s,” in *2016 International Conference on Simulation of Semiconductor Processes and Devices (SISPAD)*, Sept 2016, pp. 341–344.
- [33] M. Ramonas and C. Jungemann, “Spherical harmonics solver for coupled hot-electron-hot-phonon system,” in *Simulation of Semiconductor Processes and Devices (SISPAD), 2013 International Conference on*, Sept 2013, pp. 360–363.
- [34] D. Jabs and C. Jungemann, “Avalanche breakdown of pn-junctions: Simulation by spherical harmonics expansion of the Boltzmann transport equation,” in *Simulation of Semiconductor Processes and Devices (SISPAD), 2014 International Conference on*, Sept 2014, pp. 173–176.
- [35] D. Jabs, C. Jungemann, and K. Bach, “A robust algorithm for microscopic simulation of avalanche breakdown in semiconductor devices,” *Electron Devices, IEEE Transactions on*, vol. 62, no. 8, pp. 2614–2619, Aug 2015.
- [36] W. R. Frensley, “Boundary conditions for open quantum systems driven far from equilibrium,” *Rev. Mod. Phys.*, vol. 62, pp. 745–791, Jul 1990. [Online]. Available: <http://link.aps.org/doi/10.1103/RevModPhys.62.745>
- [37] S. Datta, “Nanoscale device modeling: the Greens function method,” *Superlattices and microstructures*, vol. 28, no. 4, pp. 253–278, 2000.
- [38] G. Paasch and H. Übensee, “A modified local density approximation - electron density in inversion layers,” *Phys. Status Solidi B*, vol. 113, pp. 165–178, 1982.
- [39] L. Lucci, P. Palestri, D. Esseni, L. Bergagnini, and L. Selmi, “Multisubband Monte Carlo study of transport, quantization, and electron-gas degeneration in ultrathin SOI n-MOSFETs,” *Electron Devices, IEEE Transactions on*, vol. 54, pp. 1156–1164, 2007.
- [40] A.-T. Pham, B. Meinerzhagen, and C. Jungemann, “A fast  $\mathbf{k} \cdot \mathbf{p}$  solver for hole inversion layers with an efficient  $2d\mathbf{k}$  -space discretization,” *Journal of Computational Electronics*, vol. 7, no. 3, pp. 99–102, 2008. [Online]. Available: <http://dx.doi.org/10.1007/s10825-007-0155-5>

- [41] M. Noei and C. Jungemann, “Numerical investigation of junctionless nanowire transistors using a Boltzmann/Schrödinger/Poisson full Newton-Raphson solver,” in *Simulation of Semiconductor Processes and Devices (SISPAD), 2016 International Conference on*. IEEE, 2016, pp. 15–18.
- [42] M. Lenzi, P. Palestri, E. Gnani, S. Reggiani, A. Gnudi, D. Esseni, L. Selmi, and G. Baccarani, “Investigation of the Transport Properties of Silicon Nanowires Using Deterministic and Monte Carlo Approaches to the Solution of the Boltzmann Transport Equation,” *IEEE Transactions on Electron Devices*, vol. 55, no. 8, pp. 2086–2096, Aug 2008.
- [43] D. K. Shaeffer and T. H. Lee, “A 1.5-V, 1.5-GHz CMOS low noise amplifier,” *IEEE J. Solid-State Circuits*, vol. 32, no. 5, pp. 745–759, 1997.
- [44] T. H. L. R. Navid, C. Jungemann and R. Dutton, “High-frequency noise in nanoscale metal oxide semiconductor field effect transistors,” *J. Appl. Phys.*, vol. 101, pp. 124 501–1 – 124 501–8, 2007.
- [45] W. M. Leach, “Fundamentals of low-noise analog circuit design,” *Proceedings of the IEEE*, vol. 82, no. 10, pp. 1515–1538, Oct 1994.
- [46] S. E. Laux, “Techniques for small-signal analysis of semiconductor devices,” *Electron Devices, IEEE Transactions on*, vol. 32, no. 10, pp. 2028–2037, Oct. 1985.
- [47] F. Bonani, G. Ghione, M. R. Pinto, and R. K. Smith, “An efficient approach to noise analysis through multidimensional physics-based models,” *Electron Devices, IEEE Transactions on*, vol. 45, no. 1, pp. 261–269, 1998.
- [48] C. Jungemann, B. Neinhüs, and B. Meinerzhagen, “Hierarchical 2-D DD and HD noise simulations of Si and SiGe devices: Part I—Theory,” *Electron Devices, IEEE Transactions on*, vol. 49, no. 7, pp. 1250–1257, 2002.
- [49] J. C. Vaissiere, J. P. Nougier, L. Varani, P. Houlet, L. Hlou, E. Starikov, P. Shiktorov, and L. Reggiani, “Small-signal analysis of the Boltzmann equation from harmonic- and impulse-response methods,” *Phys. Rev. B*, vol. 49, pp. 11 144–11 152, Apr 1994. [Online]. Available: <http://link.aps.org/doi/10.1103/PhysRevB.49.11144>
- [50] S. M. Kogan and A. Y. Shul’man, “Theory of fluctuations in a nonequilibrium electron gas,” *Sov. Phys. JETP*, vol. 29, no. 467, p. 104, 1969.
- [51] S. Kogan, *Electronic Noise and Fluctuations in Solids*. Cambridge, New York, Melbourne: Cambridge University Press, 1996.

- [52] S. Smirnov, H. Kosina, M. Nedjalkov, and S. Selberherr, “Monte Carlo method for modeling of small signal response including the Pauli exclusion principle,” *J. Appl. Phys.*, vol. 94, pp. 5791–5799, 2003.
- [53] R. Hockney and J. Eastwood, *Computer Simulation Using Particles*. Bristol, Philadelphia: Institute of Physics Publishing, 1988.
- [54] C. Jacoboni and L. Reggiani, “The Monte Carlo method for the solution of charge transport in semiconductors with applications to covalent materials,” *Rev. Mod. Phys.*, vol. 55, pp. 645–705, Jul 1983. [Online]. Available: <http://link.aps.org/doi/10.1103/RevModPhys.55.645>
- [55] L. Varani, L. Reggiani, T. Kuhn, T. González, and D. Pardo, “Microscopic simulation of electronic noise in semiconductor materials and devices,” *Electron Devices, IEEE Transactions on*, vol. 41, no. 11, pp. 1916–1925, 1994.
- [56] T. Gonzalez, J. Mateos, M. J. Martin-Martinez, S. Perez, R. Rengel, B. G. Vasallo, and D. Pardo, “Monte Carlo simulation of noise in electronic devices: limitations and perspectives,” *Proceedings of the 3rd International Conference on Unsolved Problems of Noise*, pp. 496–503, 2003.
- [57] H. K. Gummel, “A self-consistent iterative scheme for one-dimensional steady state transistor calculations,” *Electron Devices, IEEE Transactions on*, pp. 455–465, October 1964.
- [58] D. Ruić and C. Jungemann, “Numerical aspects of noise simulation in MOSFETs by a Langevin-Boltzmann solver,” *Journal of Computational Electronics*, vol. 14, no. 1, pp. 21–36, 2015. [Online]. Available: <http://dx.doi.org/10.1007/s10825-014-0642-4>
- [59] C.-K. Lin, N. Goldsman, Z. Han, I. Mayergoyz, S. Yu, M. Stettler, and S. Singh, “Frequency domain analysis of the distribution function by small signal solution of the Boltzmann and Poisson equations,” in *Proc. SISPAD*, 1999, pp. 39–42.
- [60] C. Jungemann, “A deterministic approach to RF noise in silicon devices based on the Langevin Boltzmann equation,” *Electron Devices, IEEE Transactions on*, vol. 54, no. 5, pp. 1185–1192, 2007.
- [61] —, “A deterministic solver for the Langevin Boltzmann equation including the Pauli principle,” in *SPIE: Fluctuations and Noise*, vol. 6600, 2007, pp. 660 007–1–660 007–12.

- [62] C. Jungemann, A.-T. Pham, S.-M. Hong, L. Smith, and B. Meinerzhagen, “Deterministic solvers for the Boltzmann transport equation of 3d and quasi-2d electron and hole systems in SiGe devices,” *Solid-State Electronics*, vol. 84, no. 0, pp. 112 – 119, 2013. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0038110113000956>
- [63] D. Ruić and C. Jungemann, “A self-consistent solution of the Poisson, Schrödinger and Boltzmann equations by a full Newton-Raphson approach for nanoscale semiconductor devices,” in *Simulation of Semiconductor Processes and Devices (SISPAD), 2013 International Conference on*, Sept 2013, pp. 356–359.
- [64] D. Ruić and C. Jungemann, “Microscopic noise simulation of long- and short-channel nMOSFETs by a deterministic approach,” *Journal of Computational Electronics*, vol. 15, no. 3, pp. 809–819, 2016. [Online]. Available: <http://dx.doi.org/10.1007/s10825-016-0840-3>
- [65] T. Ando, A. Fowler, and F. Stern, “Electronic properties of two-dimensional systems,” *Rev. Mod. Phys.*, vol. 54, pp. 437–672, 1982.
- [66] A. Spinelli, A. Benvenuti, and A. Pacelli, “Investigation of quantum effects in highly-doped MOSFET’s by means of a self-consistent 2d model,” pp. 399–402, 1996.
- [67] R. S. Varga, *Matrix Iterative Analysis*, ser. Series in Automatic Computation. Englewood Cliffs, New Jersey: Prentice-Hall, 1962.
- [68] M. Bollhöfer and Y. Saad, “ILUPACK — preconditioning software package,” release 2.2 available online at [www.icm.tu-bs.de/~bolle/ilupack/](http://www.icm.tu-bs.de/~bolle/ilupack/), Dept. of Mathematics and Computer Science, November 2008.
- [69] D. Esseni, P. Palestri, and L. Selmi, *Nanoscale MOS Transistors. Semi-Classical Transport and Applications*. Cambridge University Press, 2011.
- [70] C. Jungemann, A. Emunds, and W. Engl, “Simulation of linear and nonlinear electron transport in homogeneous silicon inversion layers,” *Solid-State Electronics*, vol. 36, no. 11, pp. 1529 – 1540, 1993. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/003811019390024K>
- [71] R. Brunetti, C. Jacoboni, F. Nava, L. Reggiani, G. Bosman, and R. J. J. Zijlstra, “Diffusion coefficient of electrons in silicon,” *J. Appl. Phys.*, vol. 52, pp. 6713–6722, 1981.
- [72] “Intel Math Kernel Library,” <https://software.intel.com/en-us/intel-mkl>.

- [73] R. Shankar, *Principles of quantum mechanics*. New York: Springer, 1994.
- [74] C. Kittel, *Quantum Theory of Solids*. New York: John Wiley & Sons, 1963.
- [75] G. Dresselhaus, A. F. Kip, and C. Kittel, “Cyclotron resonance of electrons and holes in silicon and germanium crystals,” *Phys. Rev.*, vol. 98, pp. 368–384, 1955.
- [76] B. N. Brockhouse, “Lattice vibrations in Silicon and Germanium,” *Phys. Rev. Lett.*, vol. 2, pp. 256–258, Mar 1959. [Online]. Available: <https://link.aps.org/doi/10.1103/PhysRevLett.2.256>
- [77] D. Long, “Scattering of conduction electrons by lattice vibrations in silicon,” *Phys. Rev.*, vol. 120, pp. 2024–2032, 1960.
- [78] T. Ohashi, T. Takahashi, N. Beppu, S. Oda, and K. Uchida, “Experimental evidence of increased deformation potential at MOS interface and its impact on characteristics of ETSOI FETs,” in *2011 International Electron Devices Meeting*, Dec 2011, pp. 16.4.1–16.4.4.
- [79] T. Ohashi, T. Tanaka, T. Takahashi, S. Oda, and K. Uchida, “Experimental Study on Deformation Potential (Dac) in MOSFETs: Demonstration of Increased (Dac) at MOS Interfaces and Its Impact on Electron Mobility,” *IEEE J. Electron Devices Soc*, vol. 4, no. 5, pp. 278–285, 2016.
- [80] M. V. Fischetti and S. E. Laux, “Band structure, deformation potentials, and carrier mobility in strained Si, Ge, and SiGe alloys,” *Journal of Applied Physics*, vol. 80, no. 4, pp. 2234–2252, 1996. [Online]. Available: <http://scitation.aip.org/content/aip/journal/jap/80/4/10.1063/1.363052>
- [81] S. Takagi, J. L. Hoyt, J. Welser, and J. Gibbons, “Comparative study of phonon-limited mobility of two-dimensional electrons in strained and unstrained Si metal-oxide-semiconductor field-effect transistors,” *J. Appl. Phys.*, vol. 80, pp. 1567–1577, 1996.
- [82] C. Jungemann, “Linear and nonlinear electron transport in homogeneous silicon inversion layers,” Vortrag am Physikalischen Institut der Università Modena, Nov. 1992.
- [83] W. A. Harrison, “Scattering of electrons by lattice vibrations in nonpolar crystals,” *Phys. Rev.*, vol. 104, pp. 1281–1290, 1956.
- [84] F. M. Bufler, “Full-Band Monte Carlo Simulation of Electrons and Holes in Strained Si and SiGe,” Dissertation, Universität Bremen, Bremen, 1997, (H. Utz Verlag Wissenschaft, München: 1998).

- [85] C. Jungemann, S. Keith, F. M. Bufler, and B. Meinerzhagen, “Effects of band structure and phonon models on hot electron transport in silicon,” *Electrical Engineering*, vol. 79, pp. 99–101, 1996.
- [86] S. Jin, M. V. Fischetti, and T. W. Tang, “Modeling of Surface-Roughness scattering in Ultrathin-Body SOI MOSFETs,” *IEEE Transactions on Electron Devices*, vol. 54, no. 9, pp. 2191–2203, Sept 2007.
- [87] R. E. Prange and T. W. Nee, “Quantum spectroscopy of the low-field oscillations in the surface impedance,” *Phys. Rev.*, vol. 168, pp. 779–785, 1968.
- [88] D. Esseni, A. Abramo, L. Selmi, and E. Sangiorgi, “Physically based modeling of low field electron mobility in ultrathin single- and double-gate SOI n-MOSFETs,” *IEEE Transactions on Electron Devices*, vol. 50, no. 12, pp. 2445–2455, Dec 2003.
- [89] S. M. Goodnick, D. K. Ferry, C. W. Wilmsen, Z. Liliental, D. Fathy, and O. L. Krivanek, “Surface roughness at the Si(100)-SiO<sub>2</sub> interface,” *Phys. Rev. B*, vol. 32, pp. 8171–8186, 1985.
- [90] C. Jungemann and B. Meinerzhagen, *Hierarchical Device Simulation: The Monte-Carlo Perspective*, ser. Computational Microelectronics. Wien, New York: Springer, 2003.
- [91] O. Madelung, *Introduction to Solid State Theory*. Berlin: Springer, 1978.
- [92] E. M. Lifshitz and L. P. Pitaevskii, *Physical Kinetics*, ser. Courses of Theoretical Physics. Oxford, UK: Butterworth–Heinemann, 1981, vol. 10.
- [93] K. Hess, *Advanced Theory of Semiconductor Devices*. Piscataway, NJ, USA: IEEE Press, 2000.
- [94] F. Gámiz, J. B. Roldán, J. A. López-Villanueva, P. Cartujo-Cassinello, and J. E. Carceller, “Surface roughness at the SiSiO<sub>2</sub> interfaces in fully depleted silicon-on-insulator inversion layers,” *Journal of Applied Physics*, vol. 86, no. 12, pp. 6854–6863, 1999. [Online]. Available: <http://dx.doi.org/10.1063/1.371763>
- [95] C. Herring and E. Vogt, “Transport and deformation-potential theory for many-valley semiconductors with anisotropic scattering,” *Phys. Rev.*, vol. 101, no. 3, pp. 944–962, 1956.
- [96] C. Ringhofer, “Space-time discretization of series expansion methods for the Boltzmann transport equation,” *SIAM J. Num. Ana.*, vol. 38, pp. 442–465, 2000.

- [97] C. Ringhofer, C. Schmeiser, and A. Zwirchmayer, “Moment methods for the semiconductor Boltzmann equation in bounded position domains,” *SIAM J. Num. Ana.*, vol. 39, pp. 1078–1095, 2001.
- [98] C. Ringhofer, “Numerical methods for the semiconductor Boltzmann equation based on spherical harmonics expansions and entropy discretizations,” *Transport Theory and Statistical Physics*, vol. 31, no. 4-6, pp. 431–452, 2002.
- [99] —, “A mixed spectral-difference method for the steady state boltzmann-poisson system,” *SIAMNumAna*, vol. 41, no. 1, pp. 64 – 89, 2003.
- [100] W. L. Engl, H. K. Dirks, and B. Meinerzhagen, “Device modeling,” *Proc. IEEE*, vol. 71, pp. 10–33, 1983.
- [101] B. Meinerzhagen, H. K. Dirks, and W. L. Engl, “Quasi-simultaneous solution method: A new highly efficient strategy for numerical MOST simulations,” *IEEE Trans. Computer-Aided Des.*, vol. 4, pp. 575–582, 1985.
- [102] W. Shockley, “Currents to conductors induced by a moving point charge,” *J. Appl. Phys.*, vol. 9, pp. 635–636, 1938.
- [103] S. Ramo, “Currents induced by electron motion,” *Proc. IRE*, vol. 27, pp. 584–585, 1939.
- [104] H. Kim, H. S. Min, T. W. Tang, and Y. J. Park, “An extended proof of the Ramo-Shockley theorem,” *Solid-State Electron.*, vol. 34, pp. 1251–1253, 1991.
- [105] F. H. Branin, “Network sensitivity and noise analysis simplified,” *IEEE Transactions on circuit theory*, vol. 20, pp. 285–288, 1973.
- [106] N. Wiener, “Generalized harmonic analysis,” *Acta Mathematica*, vol. 55, no. 1, pp. 117–258, 1930. [Online]. Available: <http://dx.doi.org/10.1007/BF02546511>
- [107] A. Khintchine, “Korrelationstheorie der stationären stochastischen Prozesse,” *Mathematische Annalen*, vol. 109, no. 1, pp. 604–615, 1934. [Online]. Available: <http://dx.doi.org/10.1007/BF01449156>
- [108] H. Nyquist, “Thermal agitation of electric charge in conductors,” *Phys. Rev.*, vol. 32, no. 1, pp. 110–113, Jul 1928.
- [109] J. B. Johnson, “Thermal agitation of electricity in conductors,” *Phys. Rev.*, vol. 32, no. 1, p. 97, Jul 1928.

- [110] W. Schottky, "Über spontane Stromschwankungen in verschiedenen Elektrizitätsleitern," *Annalen der Physik*, vol. 362, no. 23, pp. 541–567, 1918. [Online]. Available: <http://dx.doi.org/10.1002/andp.19183622304>
- [111] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, 4th ed. McGraw–Hill, 2001.
- [112] S. Takagi, A. Toriumi, M. Iwase, and H. Tango, "On the universality of inversion layer mobility in Si MOSFET's: Part I—Effects of substrate impurity concentration," *Electron Devices, IEEE Transactions on*, vol. 41, pp. 2357–2362, 1994.
- [113] K. Uchida, J. Koga, and S.-i. Takagi, "Experimental study on electron mobility in ultrathin-body silicon-on-insulator metal-oxide-semiconductor field-effect transistors," *Journal of Applied Physics*, vol. 102, no. 7, 2007. [Online]. Available: <http://scitation.aip.org/content/aip/journal/jap/102/7/10.1063/1.2785957>
- [114] G. Masetti, M. Severi, and S. Solmi, "Modeling of carrier mobility against carrier concentration in arsenic- and phosphorus- and boron-doped silicon," *Electron Devices, IEEE Transactions on*, vol. 30, no. 7, pp. 764–769, 1983.
- [115] M. Bollhöfer and Y. Saad, "Multilevel preconditioners constructed from inverse-based ILUs," *SIAM Journal on Scientific Computing*, vol. 27, no. 5, pp. 1627–1650, 2006.
- [116] Y. Tsvividis and C. McAndrew, *The MOS transistor*. Oxford University Press, 2012.
- [117] Z. Kargar, D. Ruić, and C. Jungemann, "A Self-consistent Solution of the Poisson, Schrödinger and Boltzmann Equations for GaAs Devices by a Deterministic Solver," in *Simulation of Semiconductor Processes and Devices (SISPAD), 2015 International Conference on*. IEEE, 2015, pp. 361–364.
- [118] J. Rollett, "Stability and Power-Gain Invariants of Linear Twoports," *IRE Transactions on Circuit Theory*, vol. 9, no. 1, pp. 29–32, March 1962.
- [119] S. Mason, "Power Gain in Feedback Amplifier," *Transactions of the IRE Professional Group on Circuit Theory*, vol. 1, no. 2, pp. 20–25, June 1952.
- [120] H. B. Callen and T. A. Welton, "Irreversibility and generalized noise," *Phys. Rev.*, vol. 83, no. 1, pp. 34–40, Jul 1951.

- [121] H. B. Callen and R. F. Greene, "On a theorem of irreversible thermodynamics," *Physical Review*, vol. 86, no. 5, p. 702, 1952.
- [122] R. F. Greene and H. B. Callen, "On a theorem of irreversible thermodynamics. II," *Physical Review*, vol. 88, no. 6, p. 1387, 1952.
- [123] H. B. Callen, M. L. Barasch, and J. L. Jackson, "Statistical mechanics of irreversibility," *Physical Review*, vol. 88, no. 6, p. 1382, 1952.
- [124] F. M. Klaassen and J. Prins, "Thermal noise of MOS transistors," *Philips Res. Repts*, pp. 505–514, 1967.
- [125] A. van der Ziel, *Noise in Solid State Devices and Circuits*. Canada: John Wiley and Sons, 1986.
- [126] G. Smit, A. Scholten, R. Pijper, L. Tiemeijer, R. van der Toorn, and D. Klaassen, "RF-Noise Modeling in Advanced CMOS Technologies," *Electron Devices, IEEE Transactions on*, vol. 61, no. 2, pp. 245–254, Feb 2014.
- [127] A. J. Scholten, R. V. Langevelde, L. F. Tiemeijer, and D. B. M. Klaassen, "Compact modeling of noise in CMOS," in *IEEE Custom Integrated Circuits Conference 2006*, Sept 2006, pp. 711–716.
- [128] C. Jungemann, B. Neinhüs, C. D. Nguyen, A. J. Scholten, L. F. Tiemeijer, and B. Meinerzhagen, "Numerical modeling of RF noise in scaled MOS devices," *Solid-State Electron.*, vol. 50, pp. 10–17, 2006.
- [129] M. Shoji, "Analysis of high-frequency thermal noise of enhancement mode MOS field-effect transistors," *Electron Devices, IEEE Transactions on*, vol. ED-13, no. 6, pp. 520–524, June 1966.
- [130] H. E. Halladay and A. Van der Ziel, "On the high frequency excess noise and equivalent circuit representation of the MOSFET with n-type channel," *Solid-State Electronics*, vol. 12, no. 3, pp. 161–176, 1969.
- [131] A. van der Ziel, "Gate noise in field effect transistors at moderately high frequencies," *Proc. IEEE*, vol. 51, no. 3, pp. 461–467, 1963.
- [132] T.-Y. Oh, C. Jungemann, and R. W. Dutton, "Hydrodynamic simulation of RF noise in deep-submicron MOSFETs," in *SISPAD'03*, Boston (USA), 2003, pp. 87–90.
- [133] G. Iannaccone, "Analytical and numerical investigation of noise in nanoscale ballistic field effect transistors," *J. Comput. Electron.*, no. 3, pp. 199–202, 2004.

- [134] A. Betti, G. Fiori, and G. Iannaccone, “Shot Noise Suppression in Quasi-One-Dimensional Field-Effect Transistors,” *Electron Devices, IEEE Transactions on*, vol. 56, no. 9, pp. 2137–2143, 2009.
- [135] D. Ruić and C. Jungemann, “Small Signal and Microscopic Noise Simulation of an nMOSFET by a Self-Consistent, Semi-Classical and Deterministic Approach,” in *Simulation of Semiconductor Processes and Devices (SISPAD), 2015 International Conference on*. IEEE, Sept 2015, pp. 20–23.
- [136] O. Dieball, Z. Kargar, D. Ruić, and C. Jungemann, “A self-consistent solution of the Poisson and Boltzmann equations for electrons in graphene with a deterministic approach,” in *2016 International Conference on Simulation of Semiconductor Processes and Devices (SISPAD)*, Sept 2016, pp. 205–208.
- [137] Z. Kargar, T. Linn, D. Ruić, and C. Jungemann, “Investigation of transport modeling for plasma waves in THz devices,” *IEEE Transactions on Electron Devices*, vol. 63, no. 11, pp. 4402–4408, Nov 2016.