

Quantum Transport in the Transient Regime and Unconventional Geometries

By

Bozidar Novakovic

A dissertation submitted in partial fulfillment of  
the requirements for the degree of

Doctor of Philosophy

(Electrical and Computer Engineering)

at the

UNIVERSITY OF WISCONSIN-MADISON

2012

Date of final oral examination: 08/17/12

The dissertation is approved by the following members of the Final Oral Committee:

Irena Knezevic, Associate Professor, Electrical and Computer Engineering

John H. Booske, Professor, Electrical and Computer Engineering

Luke J. Mawst, Professor, Electrical and Computer Engineering

Max G. Lagally, Professor, Material Science and Engineering

Maxim G. Vavilov, Assistant Professor, Physics

Zhenqiang Ma, Professor, Electrical and Computer Engineering

## ACKNOWLEDGMENTS

I am grateful to my advisor Prof. Irena Knezevic. Her generous support made this work possible and her advising helped me to significantly improve my technical skills, as well as my presentation and writing skills. I would also like to thank her for always being very responsive and for allowing me to pursue my interest in physics.

My thanks also go to my committee members, Prof. John Booske, Prof. Luke Mawst, Prof. Max Lagally, Prof. Maxim Vavilov, and Prof. Zhenqiang Ma, for making my final examination interesting and for their useful comments on both the technical part of my work and on my future career. Special thanks go to Prof. John Booske, Prof. Luke Mawst, Prof. Mark Eriksson, Prof. Leon McCaughan, and Prof. Robert Joynt for providing very useful feedback at the time of my preliminary examination.

I would also like to thank my colleagues from the Nanoelectronics Theory Group with whom I had many useful discussions over the years. They are in alphabetical order: Dr. Zlatan Aksamija, Jie Chen, Amirhossein Davoody, Dr. Xujiao Gao, Dr. Edwin Ramayya, Yanbing Shi, Nishant Sule, and Dr. Han Zhou. Some of them were my good friends outside our laboratory setting as well. My thanks go to Dr. Minghuang Huang and Matt Kirley for their collaboration leading to interesting journal publications and to Kiritkumar Makwana from the Physics Department for his friendship.

Finally, I would like to thank Prof. Vitomir Milanovic, Prof. Jelena Radovanovic, and Prof. Dragan Indjin, for helping me make my first steps in science at the University of Belgrade, and for taking care of my subsequent career.

**DISCARD THIS PAGE**

# TABLE OF CONTENTS

	Page
<b>LIST OF FIGURES</b> . . . . .	iv
<b>ABSTRACT</b> . . . . .	xi
<b>1 Introduction</b> . . . . .	1
1.1 Theme 1: Quantum transport in the transient regime . . . . .	6
1.1.1 Device model . . . . .	6
1.1.2 Electronic Quantum Master Equations and the transient-regime transport . . . . .	7
1.2 Theme 2: Quantum transport in unconventional geometries . . . . .	10
1.3 Summary of completed tasks and contributions . . . . .	12
<b>2 Electronic quantum master equations</b> . . . . .	15
2.1 The Single-Particle Quantum Master Equation . . . . .	16
2.1.1 The Density Matrix and The Liouville-von Neumann Equation . . . . .	16
2.1.2 The BBGKY Chain and the Single-Particle QME . . . . .	17
2.1.3 The Pauli Master Equation . . . . .	20
2.1.4 A Single-Particle QME Beyond the Born-Markov Approximation . . . . .	23
2.1.5 Monte Carlo Solution to the QME . . . . .	26
2.2 Reduced Many-Particle QMEs . . . . .	30
2.2.1 The Nakajima-Zwanzig Equation . . . . .	30
2.2.2 The Born-Markov Approximation . . . . .	32
2.2.3 The Conventional Time-Convolutionless Equation of Motion . . . . .	34
2.2.4 The Eigenproblem of the Projection Operator . . . . .	37
2.2.5 A Partial-Trace-Free Equation of Motion . . . . .	41
2.2.6 Memory Dressing . . . . .	43
2.3 Coarse-Graining for the Steady State Distribution Function . . . . .	45
2.3.1 The Exact Dynamics and the Coarse-Graining Procedure . . . . .	46
2.3.2 The Short-Time Expansion of $\overline{\mathcal{F}}_\tau$ . . . . .	48
2.3.3 Steady State in a Two-Terminal Ballistic Nanostructure . . . . .	50

	Page
<b>3 Quantum transport in the transient regime</b> . . . . .	57
3.1 The time evolution of distribution functions . . . . .	58
3.2 The transient regime modeling in nanoscale devices . . . . .	61
3.2.1 The system of coupled equations . . . . .	61
3.2.2 The initial conditions . . . . .	63
3.2.3 Injection into the localized device states due to scattering . . . . .	65
3.3 Transient in an <i>nin</i> diode . . . . .	68
3.4 Conclusion . . . . .	73
<b>4 Quantum transport in unconventional geometries in a magnetic field</b> . . . . .	75
4.1 Theoretical and numerical framework . . . . .	76
4.1.1 Curvilinear Laplacian in the tight-binding form . . . . .	78
4.1.2 The tight-binding 2D Schrödinger equation . . . . .	82
4.1.3 The Peierls phase approximation . . . . .	85
4.1.4 Handling the magnetic field: the local Landau gauge . . . . .	87
4.2 Examples . . . . .	91
4.2.1 Cylindrical geometry . . . . .	91
4.2.2 Toroidal geometry . . . . .	93
4.3 Summary and Concluding Remarks . . . . .	101
<b>5 Summary and conclusions</b> . . . . .	103
<b>LIST OF REFERENCES</b> . . . . .	106
 <b>APPENDICES</b>	
Appendix A: The details of the algorithm for the transient-regime calculations . . . . .	116
Appendix B: The surface divergence of the magnetic vector potential . . . . .	121
Appendix C: Hermiticity of the exact and tight-binding Hamiltonians . . . . .	123
Appendix D: Stabilized Transfer Matrix Solution to Eq. (4.19) . . . . .	127
Appendix E: Local Landau Gauge . . . . .	133

**DISCARD THIS PAGE**

## LIST OF FIGURES

Figure	Page
1.1 Cut-off frequency $f_T=604$ GHz for an InP/InGaAs pseudomorphic heterojunction bipolar transistor (PHBT) that used to hold a record for speed in 2005. Reprinted with permission from [W. Hafez and M. Feng, Appl. Phys. Lett. <b>86</b> , 152101 (2005)]. ©[2005], American Institute of Physics. . . . .	2
1.2 Electromagnetic spectrum with the THz frequency window indicated with vertical red lines (300 MHz - 3 THz). Current optoelectronic devices, like quantum cascade lasers, are the most efficient at the frequencies above, while solid state devices at the frequencies below the THz window. Therefore, there is a need for increasing the performances of both types of devices, in order to be able to cover the THz window with efficient sources, detectors, mixers, and other types of devices. This work focuses on the lower-frequency-side devices, specifically modeling the transient response to a sudden turn on of bias in nanoscale ballistic devices. Source: <a href="http://www.lbl.gov/MicroWorlds/ALSTool/EMSpec/EMSpec2.html">http://www.lbl.gov/MicroWorlds/ALSTool/EMSpec/EMSpec2.html</a> . . . . .	3
1.3 A resonant tunneling diode (RTD) oscillator. The highest oscillating frequency achieved with this device is around 1.1 THz, as indicated by the blue short line in the right panel. The photograph of the physical layout of the device is shown in the left panel Reprinted with permission from [M. Feiginov, C. Sydlo, O. Cojocari, and P. Meissner, Appl. Phys. Lett. <b>99</b> , 233506 (2011)]. ©[2011], American Institute of Physics. . . . .	4
1.4 (a) Illustration of the curvature formation process based on selective underetching and exploitation of the lattice mismatch. (b) Various curved geometries fabricated. (c) Hall bar measurement setup. (d) Tube formed and (e) its scanning electron micrograph. (f) Two point measurement setup. Reprinted with permission from [N. Shaji, H. Qin, R. H. Blick, L. J. Klein, C. Deneke, and O. G. Schmidt, Appl. Phys. Lett. <b>90</b> , 042101 (2007)]. ©[2007], American Institute of Physics. . . . .	5
1.5 Two-terminal configuration displaying a quantum-mechanical, nanoscale active region connected to two large semiclassical reservoirs of charge. . . . .	6

Figure	Page
2.1 (a) The electron charge density and potential energy for an <i>nin</i> Si diode at 77 K, biased at 0.25 V, where the solid lines are results of using the master equation (2.13), while the dashed lines are results of using the Monte Carlo BTE. (b) Similar as (a), but with results for the average kinetic energy and drift velocity. Reprinted with permission from [M. V. Fischetti, Phys. Rev. B <b>59</b> , 4901 (1999)]. ©(1999) by the American Physical Society. . . . .	22
2.2 Drift velocity overshoot in silicon. The result of the quantum Monte Carlo technique is shown with the solid line, while the semiclassical result is shown with the dashed line. Reprinted with permission from [C. Jacoboni, Semicond. Sci. Technol. <b>7</b> , B6 (1992)]. ©(1992) IOP Publishing Ltd. . . . .	30
2.3 Decomposition of the total Liouville space $\mathcal{H}_{S+E}^2$ into the subspaces of the projection operator $\mathcal{P}$ and the isomorphism between the unit subspace $(\mathcal{H}_{S+E}^2)_{\mathcal{P}=1}$ and $\mathcal{H}_S^2$ for an operator $x$ acting on $\mathcal{H}_{S+E}$ . Reprinted with permission from [I. Knezevic and D. K. Ferry, Phys. Rev. A <b>69</b> , 012104 (2004)]. ©(2004) by the American Physical Society. 40	40
2.4 Schematic of the two-terminal ballistic nanostructure, negatively biased at the left contact, with the boundaries between the open system and contacts shown at $x_L$ and $x_R$ , and with the graphical representation of the wave function injected from and the hopping type interaction with the left contact. It is similar for the wave function and interaction for the right contact. . . . .	51
2.5 Schematic of the two-terminal ballistic nanostructure with non-ideal contacts, negatively biased at the left contact, with the boundaries between the open system and contacts shown at $x_L$ and $x_R$ , and with the graphical representation of the wave function injected from the left contact and the hopping type interaction between the left contact and the active region due to injection and between the active region and both contacts due to extraction. It is similar for the wave function and interaction for the right contact. Reprinted with permission from [I. Knezevic, Phys. Rev. B <b>77</b> , 125301 (2008)]. ©(2008) by the American Physical Society. . . . .	55
3.1 The difference between the right-propagating device and left contact distribution functions, $f_k(t) - f_k^L$ , normalized by $f_k(0) - f_k^L$ . The difference should decay to zero in the steady state, showing that a longer contact relaxation time $\tau$ leads to a shorter transient regime. It is similar for the left-propagating device states. The injection rate for this specific k-state is assumed to be $\Delta_k = 1$ ps. . . . .	61

Figure	Page	
3.2	A block diagram showing the overall coupling between different equations used to calculate the transient regime. Starting with an initial guess for $V$ , first the Schrödinger equation is solved to obtain the energies and wave functions for scattering states (lower left block). With this the lower right block is solved iteratively until the convergence of the drift wave vector $k_d$ is achieved. At that point we have the charge density $n$ , which serves as the input to the upper block, the Poisson equation. The solution to the Poisson equation yields a new guess for $V$ and the whole process is repeated, until the convergence of the outer loop is achieved. The outer loop is solved by the globally convergent Newton method designed to find the root of the functional $F[V]$ . The inner loop uses simple mixing of the previous and current $k_d$ solutions to find the next guess, with the use of a suitable damping coefficient. . . . .	64
3.3	Potential profile and the positions of the Fermi level throughout the contact-device-contact system for an <i>nin</i> diode with $10^{17} \text{ cm}^{-3} - 10^{13} \text{ cm}^{-3} - 10^{17} \text{ cm}^{-3}$ doping profile, room temperature, and 25 meV bias. The right contact is assumed to be grounded. The left figure is at equilibrium, before applying bias. The right figure is after applying bias, but before current flow at time $t = 0^+$ s. The left contact on the one hand and the device region along with the right contact on the other hand are each in a local equilibrium, and have different Fermi levels. Since there is still no current flow, $k_d(0^+) = 0$ . . . . .	66
3.4	Electron density at $t = 0^+$ s, corresponding to the initial conditions shown in Fig. 3.3. Electrons are depleted in the region near the left contact-device interface. These electrons are pushed into the right contact, when the perturbation due to the application of bias reaches the left contact-device interface, and will be replenished during the transient injection from the left contact. Consequently, there will be a difference between the contact and device currents, $J_{con}$ and $J_{dev}$ respectively, equal to $ dQ/dt $ . . . . .	67
3.5	Illustration of the scattering-enhanced injection into the states localized in the quantum well to the left of the device barrier. The structure used is the same as in Fig. 3.3. The two upper arrows are injection into the device current-carrying states due to the flow of electrons between the contacts and the device. The lower arrow is the enhanced injection into the localized quantum well states, due to the efficient scattering in the contacts. The thin densely spaced lines (black) designate current-carrying levels, while the widely spaced lines (blue) designate the localized states. As stated before, $Q$ is positive near the left contact, due to electron depletion. . . . .	68

Figure	Page
3.6 (a) Potential and (b) charge density in the <i>nin</i> diode as a function of time upon the application of -25 mV to the left contact. The <i>n</i> -type regions are doped to $10^{17} \text{ cm}^{-3}$ , while the middle region to $10^{13} \text{ cm}^{-3}$ . The contact momentum relaxation time is $\tau=120 \text{ fs}$ , as calculated from the textbook mobility value corresponding to the contact doping density. The temperature is $T = 300 \text{ K}$ and the device size is $100 \text{ nm} - 50 \text{ nm} - 100 \text{ nm}$ , for the two <i>n</i> -doped regions enclosing the middle one. The scattering-enhanced injection rate is chosen to be $\Delta_{scatt} = 1/\tau$ , and therefore the quantum well bottom rises quickly on the time scale comparable to the contact relaxation time. In the electron density plot, one can observe the initial depletion near the cathode side, and its subsequent population during the transient. . . . .	69
3.7 Current density versus time for the <i>nin</i> diode from Fig. 3.6 upon the application of -25 mV to the left contact. $J_{dev}$ plotted here refers to the device current from Eq. (3.5d). The transient time is relatively long, owing to small $\tau$ . . . . .	69
3.8 The influence of the injection into the localized device states due to scattering on the potential and device current transient evolution. The upper left panel, the upper right panel, and the lower left panel show the initial time-dependence of the device potential profile for three different scattering enhanced injection rates, $\Delta_{scatt} = 1/\tau$ , $\Delta_{scatt} = 10^{12} \text{ s}^{-1}$ , and $\Delta_{scatt} = 0$ , respectively. The <i>nin</i> diode used here has the same parameters as before. We see that with larger localized states injection rates the quantum well bottom is rising faster. However this initial difference in the quantum well bottom height does not affect the barrier height too much and therefore the transient current is very similar for different $\Delta_{scatt}$ as shown in the lower right panel. What does get affected is the initial difference between $J_{con}$ and $J_{dev}$ , equal to the change of electron charge over time $dQ/dt$ . . . . .	71
3.9 Comparison of current densities for the same <i>nin</i> diode used in the previous figures and for variable contact relaxation time $\tau$ . Each curve is shown just until the moment it saturates, to save the computational time. Having less scattering in the contacts helps to reach steady state faster. This phenomenon was discussed in connection to Fig. 3.1, where the same conclusion was reached using the low current limit and solving Eqs. (3.5c) analytically. If the current were limited by the scattering in the contacts, then one would expect that more scattering helps to establish steady state faster. However, in our model, the current is controlled by the ballistic device's potential profile. The device requests some current from the contacts, which act as charge reservoirs, but the contact ability to supply the charge of a given momentum will be limited due to momentum randomization in the presence of scattering. . . . .	72

Figure	Page
4.1 (a) The 4-point scheme for discretization of the first partial derivatives by using Eq. (4.9). (b) An example of a curved surface parametrized by curvilinear coordinates $u^1$ and $u^2$ . Leads are planar continuations of the curved surface in both directions along the longitudinal coordinate $u^1$ . In order to calculate the wave function value at the grid point denoted by the large solid circle (central point) from the 4-point scheme, we have to know the metric tensor values at all the surrounding points, denoted by the small solid circles, as well as at the central point. . . . .	80
4.2 An example of a curved nanoribbon for calculating the magnetic vector potential: the nanoribbon (dark blue) on top of a half-torus (light blue), with the leads extending downward at both ends. The Cartesian coordinate system in the left lead (shown in the lower left corner) is taken as a reference system, according to which the magnetic field components $B_x$ , $B_y$ , and $B_z$ are specified. Gauge transformations (GT), equivalent to coordinate system translations and rotations in the plane perpendicular to each component, are also shown. . . . .	87
4.3 (a) Normalized conductance vs. magnetic field for a nanoribbon with helicity in cylindrical geometry. The leads are assumed cylindrical to avoid the need for transitional regions and as a consequence have the Landau levels energies affected by the magnetic field flux in the leads (Landau level quantization), albeit weakly. The lead on the left-hand side is the injecting lead. Magnetic field direction with respect to the structure is shown in (b) and would be parallel to the leads if they were exactly planar. Even at zero field, not all of the three injected modes propagate through, due to the ribbon helicity. (b) Electron density at the Fermi level shown for $B_y=2.663$ T on the $\log_{10}$ scale, where there is resonant transmission, and for $B_y=4.5$ T, where the normalized conductance is zero (transmission suppressed). A quasibound state [right panel of (b)] is associated with the resonant transmission feature from (a). . . . .	96

Figure	Page
<p>4.4 (a) Normalized conductance vs. magnetic field obtained with the exact equation (4.19) (thick solid red curve) and within the Peierls phase approach (thin solid black curve) for a nanoribbon with helicity in cylindrical geometry. The leads are assumed cylindrical to avoid the need for transitional regions. The lead on the left-hand side is the injecting lead. The magnetic field direction with respect to the structure is shown in (b) and would be normal to the leads if they were exactly flat. The number of injected propagating modes is shown with a thin dashed black curve and is influenced by the magnetic field flux in the leads (Landau level quantization). There is a discrepancy in the conductance between the exact and Peierls phase approaches starting at around 2.3 T. Resonant reflection is present at <math>B = 4.582 T</math> in the exact (thick red) curve. (b) Electron density at the Fermi level, represented through color on the <math>\log_{10}</math> scale, corresponding to the resonant reflection at <math>B = 4.582 T</math>. Quasibound state is formed, primarily confined to the right hand side of the helical part, looking from the left, injecting lead. . . . .</p>	97
<p>4.5 (a) Normalized conductance vs. magnetic field for a nanoribbon in toroidal geometry. The leads are assumed cylindrical to avoid the need for transitional regions. The number of propagating modes, given by the dashed line, is influenced by the magnetic field in the leads (Landau level quantization). There are strong resonant reflections at high magnetic fields, where there is only one injected mode. (b) Electron density at the Fermi level, represented through color, for unit conductance (complete transmission) at <math>B_x=4.646 T</math> and zero conductance (resonant reflection) at <math>B_x=4.456 T</math>. For complete transmission at <math>B_x=4.646 T</math>, there is an edge channel formed that changes sides according to what can be expected from the classical Lorentz force. An extended quasibound state, responsible for the resonant reflection, is visible on the <math>\log_{10}</math> scale in the central region of the nanoribbon at <math>B_x=4.456 T</math>. . . . .</p>	98
<p>4.6 Sample electron density plot at the Fermi level for a toroidal nanoribbon subject to a magnetic field in the <math>y</math>-direction. Electron density is represented through color (light – high, dark – low). This situation illustrates the use of the gauge transformations, as explained in Sec. 4.1.4 and Appendix E.2, since the projection of the middle longitudinal grid line on the plane normal to the field is a half-circle, which requires the use of rotations in every slice of the toroidal part of the nanoribbon. The formed edge state reflects the symmetry of the problem and agrees qualitatively (the side at which it is formed) with the result obtained by applying the classical Lorentz force. . . . .</p>	99

Figure	Page
4.7 (a) Normalized conductance vs. Fermi level in the leads for a nanoribbon in toroidal geometry with and without helicity, in the absence of magnetic field. The leads are assumed cylindrical to avoid the need for transitional regions. The number of injected modes is changing due to the change in the Fermi energy. Comparison of the conductance for these two nanoribbons shows that helicity is the reason for the observed difference between the actual normalized conductance and the number of propagating modes. (b) Normalized electron density at the Fermi level, represented through color (white – high, black – low) for the nanoribbons with (right panel) and without helicity (left panel) when the Fermi level energy is equal to 13.5 meV. . . . .	100
A.1 Flowchart of the numerical algorithm for the calculation of the electronic response of a biased two-terminal nanostructure during a transient. . . . .	118

## ABSTRACT

This thesis covers aspects of the theory and simulation of small electronic semiconductor nanostructures. It addresses the transient regime and unconventional geometries, and the development of simulation algorithms and methodology to study them computationally. The thesis is divided into two themes: in Theme 1, we investigate the quantum electronic transport properties in the transient regime, while in Theme 2 the quantum electronic transport properties in nanostructures with unconventional geometries in a magnetic field. Both of these topics can lead to potential useful applications. Knowing the properties of the transient regime enables such device modeling that can lead to faster devices. Faster devices are needed not only in digital electronics, where the speed today is mostly limited by the dissipation, but also in high frequency signal generation, detection and manipulation in the area of sensing and telecommunications. On the other hand, nanostructures with unconventional, curved geometries are gaining popularity today, with various experimental groups producing one-dimensional and two-dimensional nonplanar nanostructures. It is reasonable to expect that curvature, potentially coupled with an external magnetic field, can have large effects on the coherent transport properties. Therefore, it may be possible to control and tune the conductance using mechanic or mechano-magnetic means.

Theme 1 starts by overviewing the most important quantum master equations, which give the time-dependence of the distribution functions  $f_k$  in the nanostructure. Since every nanostructure with current flow is coupled to and depend on its environment, it is treated by using an open system formalism. The major objective is to reduce the extremely complicated equations involving many particles/variables to a system which is analytically or numerically manageable. The reduction is

often done by using suitable approximations for the nanostructure (device, active region) and the environment (reservoirs of charge, contacts), while tracing out all unnecessary variables. Based on the previous work on the time coarse-graining and short-time expansion, we derive a quantum master equation for the device/active region and couple it to the Poisson, Schrödinger and current continuity equations, in order to calculate the charge density  $n(\mathbf{r}, t)$ , current density  $J(t)$ , and potential profile  $V(\mathbf{r}, t)$  of the nanostructure. We introduce suitable initial conditions at  $t = 0^+$  s, i.e. just after applying the excitation and before any current flow, and discuss the role of scattering during the transient using a simple model. Since the mentioned system of four equations is highly nonlinear, one must use suitable numerical algorithms to achieve satisfactory convergence rate and stability. The results show the following features. First, we find an interesting dependence of the device's transient time on  $\tau$ , the relaxation time in the contacts: the longer the  $\tau$ , the shorter the transient. Second, due to the initial conditions, there is a depletion of electrons in the device at the beginning of the transient, subsequently replenished by the injection from the higher potential contact. Furthermore, depending on the strength of the scattering enhanced injection into a set of device localized states, it is expected that the total current flowing through the contacts,  $J_{con}$ , can be appreciably different in magnitude from the device current,  $J_{dev}$ , at the beginning of the transient.

Theme 2 addresses the quantum transport in unconventional geometries in a magnetic field. We study the steady-state, coherent quantum conduction in curved nanomembranes/nanoribbons, with or without an external magnetic field, and calculate the electron density profile in the structure. The curvilinear Schrödinger equation with the electric scalar potential  $V$  and magnetic vector potential  $\mathbf{A}$  has a much more challenging form than the ordinary Schrödinger equation. We first cast this equation into a tight-binding form, using a curvilinear space grid on the curved surface. Unfortunately, once the curvilinear Schrödinger equation is discretized, the Hamiltonian matrix loses its exact Hermiticity, which cannot be fully restored. We use discretization schemes that specifically preserve the Hermiticity as much as possible and additionally introduce suitable matrix transformations that can transform our non-Hermitian Hamiltonian matrix into a Hermitian one. Having set up our tight-binding Schrödinger equation, we solve it by a specifically modified version of

the so called Usuki stabilized transfer matrix method, to obtain the conductance and electron density in the linear transport regime. This method of solving the tight-binding Schrödinger equation, which starts from one contact and integrates the equation through the curved nanoribbon to the other contact, introduces a preferable direction which affects the choice of gauge that gives reasonable physical results. We devise a type of a local Landau gauge in order to compute  $A$  from the given magnetic field  $B$ . By applying this method to several curved nanoribbons in cylindrical and toroidal geometries, with or without helicity, we observe several interesting features in the conductance, like conduction suppression, similar to the quantum point contact case when the gate bias is varied, and resonant reflections, previously observed in non-uniform, planar 2D nanostructures, such as various cavities.

# Chapter 1

## Introduction

This thesis covers two themes, each dealing with novel aspects of quantum electronic transport. Theme 1 is the study of the transient behavior in nanoscale electronic devices. Devices, considered here to be ballistic due to small dimensions, are connected to macroscopic contacts (reservoirs of charge) that have efficient scattering mechanisms. The novelty in this work is that the transient regime is studied using a quantum-mechanical open system formalism, which is necessary to address the mainly quantum-mechanical nature of transport in nanoscale electronic devices. The main results of this part of the thesis are the calculations of the transient behavior of charge density, current density, and potential, in response to a sudden change in bias applied to the contacts. Theme 2 is the study of the influence of curvature on the steady-state transport in nanoscale ballistic electronic devices in a magnetic field. In particular, we are interested in the charge density and conductance in the linear transport regime in curved devices, where the curvature can be cylindrical, toroidal, helical, or in principle arbitrary. The curvature can have strong effects on the coherent conduction properties of non-planar nanoribbons, especially in the presence of magnetic fields.

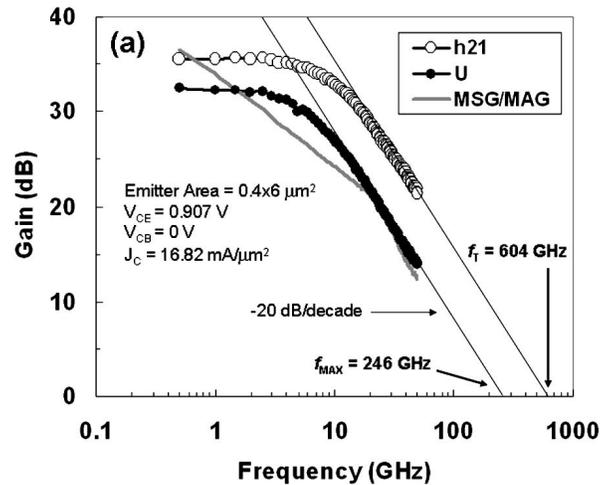


Figure 1.1 Cut-off frequency  $f_T=604$  GHz for an InP/InGaAs pseudomorphic heterojunction bipolar transistor (PHBT) that used to hold a record for speed in 2005. Reprinted with permission from [W. Hafez and M. Feng, Appl. Phys. Lett. **86**, 152101 (2005)]. ©[2005], American Institute of Physics.

Studying the transient response is important both in DC and AC operation regimes. In DC, the transient regime is responsible for establishing the highest operating frequency of switching in digital circuits/devices, while in AC operations for the cut-off frequency of amplification in analog devices (Fig. 1.1). The terahertz (THz) frequency window still lacks efficient electronic components for radiation, detection, and manipulation of THz signals. As shown in Fig. 1.2, there are two ways to deal with this problem: lowering the frequency of operation of optoelectronic devices, or increasing it in solid state devices. We focus on the latter, and as the first step to fully time-dependent transport, we model the transient regime due to a sudden application of bias. In particular, two-terminal devices, such as various diodes, are attracting attention in the area of solid-state high-frequency operations (Fig. 1.3). Two-terminal devices are much less affected by parasitic capacitances than multi-terminal devices, like transistors, owing to simpler geometry. For this reason, we focus on two-terminal ballistic devices, although it should be straightforward to expand the present method to multi-terminal devices.

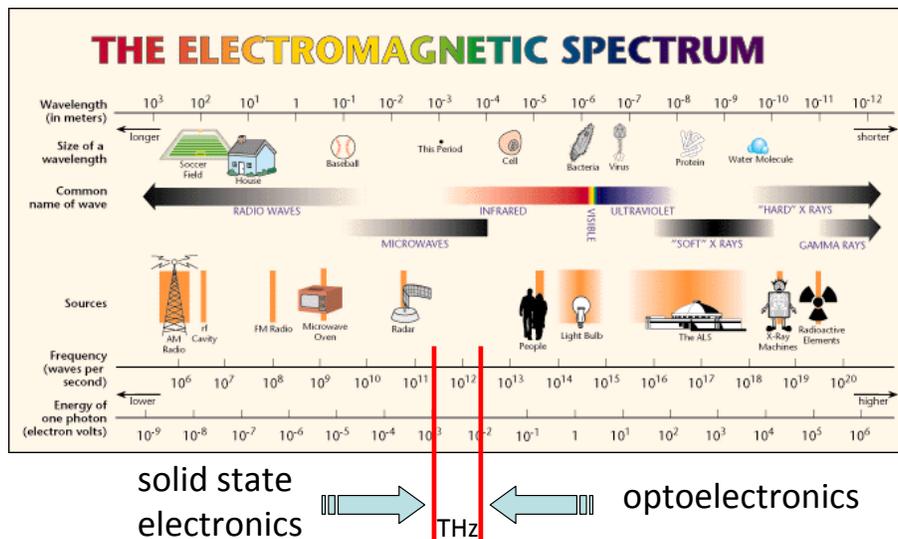


Figure 1.2 Electromagnetic spectrum with the THz frequency window indicated with vertical red lines (300 MHz - 3 THz). Current optoelectronic devices, like quantum cascade lasers, are the most efficient at the frequencies above, while solid state devices at the frequencies below the THz window. Therefore, there is a need for increasing the performances of both types of devices, in order to be able to cover the THz window with efficient sources, detectors, mixers, and other types of devices. This work focuses on the lower-frequency-side devices, specifically modeling the transient response to a sudden turn on of bias in nanoscale ballistic devices. Source: <http://www.lbl.gov/MicroWorlds/ALSTool/EMSpec/EMSpec2.html>.

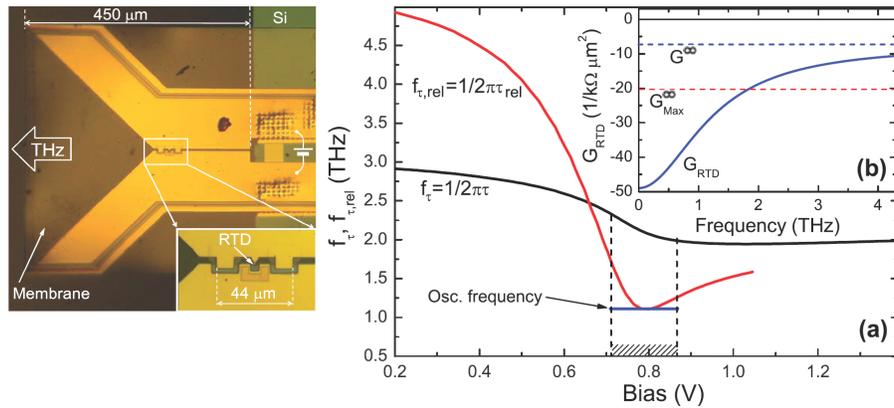


Figure 1.3 A resonant tunneling diode (RTD) oscillator. The highest oscillating frequency achieved with this device is around 1.1 THz, as indicated by the blue short line in the right panel. The photograph of the physical layout of the device is shown in the left panel Reprinted with permission from [M. Feiginov, C. Sydlo, O. Cojocari, and P. Meissner, *Appl. Phys. Lett.* **99**, 233506 (2011)]. ©[2011], American Institute of Physics.

The motivation to study curved devices is based on the fact that curvature alone, or coupled with a magnetic field, is a new and potentially useful mechanical degree of freedom in electronic device design. Such curved nanostructures can strongly affect the coherent electronic transport through them. In addition, various curved nanostructures are readily produced today, with one interesting example shown in Fig. 1.4.

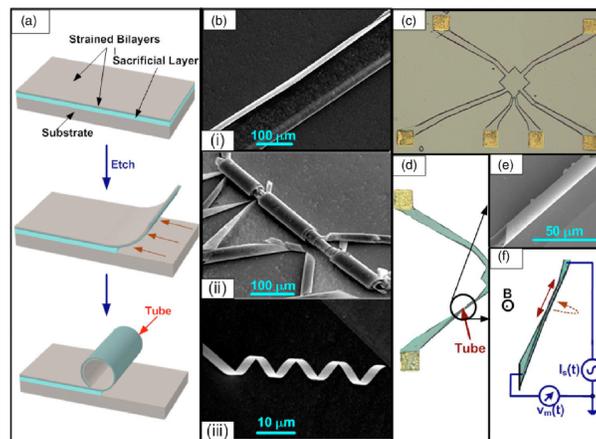


Figure 1.4 (a) Illustration of the curvature formation process based on selective underetching and exploitation of the lattice mismatch. (b) Various curved geometries fabricated. (c) Hall bar measurement setup. (d) Tube formed and (e) its scanning electron micrograph. (f) Two point measurement setup. Reprinted with permission from [N. Shaji, H. Qin, R. H. Blick, L. J. Klein, C. Deneke, and O. G. Schmidt, *Appl. Phys. Lett.* **90**, 042101 (2007)]. ©[2007], American Institute of Physics.

## 1.1 Theme 1: Quantum transport in the transient regime

### 1.1.1 Device model

Modern electronic devices are best described as physical systems consisting of a nanoscale active region (the device itself) attached to large reservoirs of charge (e.g. batteries, power supplies) via leads. A configuration of a two-terminal device is shown in Fig. 1.5. In this work, the active region is considered to be a quantum-mechanical, ballistic system (no scattering), thus exhibiting coherent electronic transport. The reservoirs are semiclassical in nature, meaning that electrons are spatially localized in them due to strong electron-electron and electron-phonon scattering. In the absence of scattering in the active region, scattering in the reservoirs is the only relaxation mechanism responsible for establishing the steady state in the reservoir-lead-active region system.

Localized electrons are naturally described as wave packets, but since the active region is on the scale of a few tens to a few hundred nanometers, the mean free path in the reservoirs, as determined by the scattering time and electron velocity, will be larger than the size of the active region, making electrons appear to be plane waves as they travel through the active region [1]. For example, the mean free path in heavily doped Si is on the order of 50 nm at room temperature. The scattering time in the reservoirs is much shorter than the transient time of the active region in the devices we consider: typical scattering times in the reservoirs are in the range of femtoseconds, while the

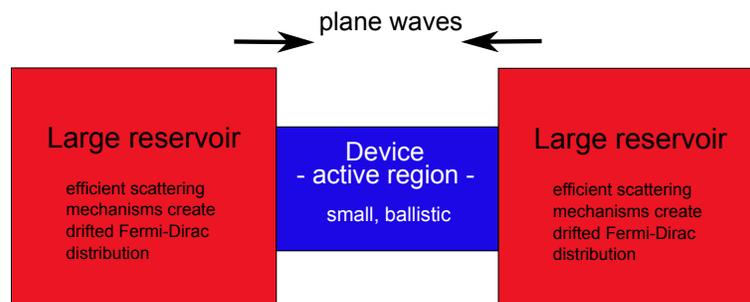


Figure 1.5 Two-terminal configuration displaying a quantum-mechanical, nanoscale active region connected to two large semiclassical reservoirs of charge.

transient times are in the range of picoseconds [2]. A quantity that is common to both the active region and the reservoirs is the drift wave vector  $k_d$ , ensuring continuity of current. On the one hand, this wave vector is proportional to the semiclassical current density in the reservoirs, while on the other hand it makes the Fermi-Dirac distribution function in them "drifted." The so-called drifted Fermi-Dirac distribution function is a good approximation for the current-carrying reservoirs in the presence of efficient electron-electron scattering, as was shown by detailed ensemble Monte Carlo–molecular dynamics simulations in bulk semiconductors [3–5]. The model approximations stated in this paragraph, along with the assumption of ballistic transport in the active region, will be used throughout this work.

The previously described model system has a large number of degrees of freedom, due to its intrinsic many-particle nature, and any viable theoretical formalism used to describe it has to reduce this number while keeping the physics sufficiently accurate [6,7]. We work within the open system formalism [8] by considering the active region to be open (exchanges particles with the reservoirs) and then reduce the total number of degrees of freedom to only the relevant ones, while performing an averaging over the rest. The open boundary conditions are employed both in the wave-vector space ( $k$ -space), to obtain the master equation that governs the time evolution of the density matrix and the electronic distribution function, and in real space, to obtain a numerically feasible scattering-state basis as a solution of the Schrödinger equation.

### **1.1.2 Electronic Quantum Master Equations and the transient-regime transport**

A quantum master equation is an equation for the time-evolution of the many-particle statistical operator. We solve this equation to obtain the time-evolution of the occupation of active region states due to their coupling with the contacts, which combined with the real space equations (here the Schrödinger and Poisson equations) gives a complete description of the time-evolution of important electronic quantities, like the electron density, potential, and current. As already mentioned, the systems we are dealing with are very complicated, since they consist of many particles that mutually interact. This means there are many degrees of freedom, either the position and

momentum for each particle in the semiclassical picture, or quantum numbers (position, linear, angular ...) in the quantum-mechanical description, that have to be included in the analysis. The fact that there are many degrees of freedom is connected to having many coupled equations that have to be solved, which leads to an intractable problem. Therefore, one must employ suitable approximations to make the problem tractable while keeping sufficient physical accuracy of the model.

Roughly speaking, each major approximation leads to a certain method or a class of methods that are used by device physicists and engineers to calculate the device transport properties. One possible classification of methods is done by approximating just how many particles/states in the many-particle problem are considered. We can speak of a one-body problem (single particle states), two-body problem, three-body problem and so on, all of which can be derived by truncating the BBGKY (Bogoliubov–Born–Green–Kirkwood–Yvon) hierarchy of equations [9,10], that is able to describe the many-particle problem exactly, with all the mutual interactions between many-particle subsets. Along with the assumption of how the many degrees of freedom per particle are treated exactly we arrive at the kinetic and hydrodynamic models, most commonly in use. Kinetic models are at the level of distribution functions defined on a single-particle phase space, therefore treating one-body problems with interactions exactly, while hydrodynamic models incorporate additional assumptions about the momentum, therefore not treating the momentum exactly [7]. Often [6, 11, 12], we account for interparticle interactions in the single-electron picture through the mean-field approximation (Hartree approximation), by self-consistently solving the Poisson equation along with any single-particle transport equation. Essentially, what we do in device physics is to solve the Poisson equation with the nonlinear charge density calculated by using the transport equation. When this system of equations converges, all other quantities of interest (e.g. current) can be calculated separately.

Another criterion we can use to distinguish between different models is whether they are quantum or semiclassical [13], classical being irrelevant in the context of small electronic devices. The simplest quantum model relies on particles populating the eigenstates of the single-particle Hamiltonian, obtained by solving the time-independent Schrödinger equation. This model can account

for quantum tunneling, interference effects, sharp potentials and other quantum mechanical features, but is unable to handle the time dynamics of far-from-equilibrium states in the presence of scattering and coupling to the contacts [7]. More advanced quantum models define mixed states allowing for spatial localization of particles due to their coupling to the surroundings. Among these methods we can mention the single-particle density matrix method where the central equation is the Liouville-von Neumann equation [14], the Wigner function method with the Wigner equation [15] and the non-equilibrium Green's function method with the Dyson equation [16, 17]. Usually, these are all quantum kinetic equations, with the Liouville-von Neumann equation being known as the quantum master equation (QME), since it is an equation of motion for the density matrix, either a single-particle (quantum kinetic level), or a full/reduced many-particle density matrix. In some situations one can use the single-particle Pauli master equation (PME) [18], which, by its ability to model dissipation of eigenstates, can be situated between the pure Schrödinger equation (eigenstates without dissipation) and the single-particle density matrix method (mixed states with dissipation). The Boltzmann transport equation (BTE) is semiclassical. Its solution is a distribution function in the phase space that, therefore, does not respect the uncertainty relations and represents electrons as pointlike particles for the purpose of drift and diffusion, making features like the tunneling, resonances, interference, etc. impossible. On the other hand, electrons are represented by plane waves during collisions, which makes the BTE unable to capture sharp potential changes (on the order of electron's wavelength). The BTE can be formally obtained by truncating the BBGKY chain [19]. Alternatively, it can be obtained from the NEGF method in the strong scattering limit [16].

Given the transport master equation that governs the time-evolution of the distribution functions for the device (active region) states, we can study the transient regime. In this theme, we develop a method to calculate the time-dependent charge density  $n(\mathbf{r}, t)$ , current density  $J(t)$ , and potential  $V(\mathbf{r}, t)$ , by coupling the time-dependent master equation to the Poisson, Schrödinger, and current continuity equations. The master equation is derived by introducing a model interaction Hamiltonian that gives Landauer-type distribution functions in the steady state, and a time-dependent drift wave vector  $k_d(t)$ , which is proportional to the semiclassical current in the contacts. A globally

convergent Newton method is used to solve the system of equations, given its robustness in the presence of high non-linearity. In order to obtain the proper transient behavior, we discuss the construction of initial conditions at  $t = 0^+$  s, the moment just after applying bias, but before any current flow through the device. Furthermore, we model the contact-to-device scattering, which enhances the initial electron injection into the device localized states and ensures reaching the proper steady state.

## 1.2 Theme 2: Quantum transport in unconventional geometries

Curved nanoribbons are narrow nonplanar strips fashioned out of semiconductor heterostructures that support a two-dimensional (2D) electron gas. [20, 21] They are among a growing variety of curved nanostructures that can be fabricated today. [22–25] Curved nanostructures can be produced from planar heterostructures by selective underetching in combination with strain mismatch. [22, 26–29] Other methods involve transfer from planar to curved substrates, [30] growth of curved layers using curved templates (nanowires, nanotubes, or nanorings) as substrates, [31–34] or thermal transformation of curved templates. [35] Among the most prominent curved structures are flexible electronic devices, [36, 37] used for applications ranging from flexible integrated circuits [38] to artificial retinas. [39] Techniques such as dry transfer printing [38, 40] and electrohydrodynamic jet printing [41] have been developed to enable deposition of electronic devices onto flexible substrates, such as polymers. [21, 42]

Curvature presents a novel, mechanical degree of freedom available for tailoring these nanostructures' physical properties. Theoretical and experimental investigations of optical properties of curved nanostructures for various resonators, [43–48] photoluminescence tuning, [49, 50] coherent emission and lasing [51, 52] and metamaterials [53] have been reported. There have also been a number of papers on magnetic properties of curved nanostructures: magnetization, [54–57] spin-orbit interaction, [58, 59] spin wave confinement and interference. [60, 61]

Theoretical work on the electronic properties of curved nanostructures has so far been largely focused on the effects of the curvature-induced geometric potential and the classification of states, [62–68] as well as transport [69–72] in simple structures with analytical shapes (e.g. cylindrical

or with periodic corrugation). Similar features have been addressed experimentally. [73–77] Considering that experimentally fabricated nanostructures typically have imperfections in shape, [22] there is a need for a theoretical approach, necessarily computationally based, that is capable of addressing transport in arbitrarily shaped curved nanostructures.

Within the envelope function and effective mass framework, electronic states in curved nanostructures can be obtained by solving the 2D Schrödinger equation (SE) in curvilinear coordinates, derived from the three-dimensional (3D) SE in the limit of zero thickness. [78–81] The problem of including arbitrary magnetic fields into the 2D SE through the covariant Peierls substitution has been solved recently. [82] It has been shown that the dynamics of a particle on the surface can be decoupled from the one along the normal to the surface by using a suitable gauge transformation in the derivation of the 2D SE. In the same work, analytic forms of the 2D SE with magnetic field have been given for spherical, cylindrical, and toroidal coordinate systems. As the low-field and nearly ballistic quantum transport in nanostructures at low temperatures is dictated by the transmission properties of the states in the vicinity of the Fermi level, [83–86] a theoretical account of electronic transport in curved nanostructures in this regime can therefore be described by solving the 2D SE in curvilinear coordinates with open boundary conditions.

In this theme, we study the linear and ballistic transport properties, specifically the electronic density and conductance, of curved nanoribbons subject to a static magnetic field. The presented theoretical and numerical framework is based on a tight-binding form of the time-independent two-dimensional Schrödinger equation in curvilinear coordinates, and is entirely general: it enables a computational description of ballistic transport in an arbitrarily shaped two-terminal curved nanostructure. A curved nanostructure is connected to large reservoirs via leads that are the source/drain for scattering states. We adopt a suitable orthogonal curvilinear coordinate system to parametrize a generic curved surface and discretize the 2D SE, due to Ferrari and Cuoghi, [82] by using a 2<sup>nd</sup> order finite-difference scheme. We explain the Hermiticity properties of the discretized curvilinear Hamiltonian and discuss the local Landau gauge used to include arbitrarily directed magnetic fields in the leads and in the curved nanoribbons. We also compare the Peierls phase approximation with the exact way of including the magnetic field in the curvilinear 2D SE. The obtained tight-binding

form of the 2D SE in curvilinear coordinates can be solved by well-established numerical methods for planar nanostructures, such as the transfer matrix, [87–89] scattering matrix, [90, 91] recursive Green’s function, [92] and nonequilibrium Green’s function techniques. [86] We use the stabilized transfer matrix approach due to Usuki *et al.*, [87] since it provides both the transmission/reflection coefficients and a convenient way to reconstruct the wave function. By using the Landauer formula [86] to calculate the conductance from the transmission coefficients, as well as by using the wave function to calculate the electron density, we show that coherent electron transport is affected by the curvature and magnetic field through the number of injected propagating modes (modulated by the magnetic field flux in the leads) and their transmission properties (determined by the interplay between the magnetic field and the nanoribbon’s curvature and helicity).

### 1.3 Summary of completed tasks and contributions

The project objective is to develop a theoretical and computational methods for calculating quantum electronic transport properties (mainly the electron density, the potential profile, and the current density) of nanoscale ballistic electronic devices, considering the following two novel phenomena: (1) transient regime established upon a sudden turn on of bias across the device/active region (Theme 1); (2) unconventional geometries with or without an external static magnetic field (Theme 2).

A list of completed tasks and contributions for Theme 1:

- Develop a self-consistent Poisson-Schrödinger solver with real-space open boundary conditions to obtain the potential profile, the electronic density and the current density in the steady state. This task is a necessary initial building block, and is implemented using the normal mode, or standing wave, boundary conditions introduced in [1, 93, 94].
- Derive the equations governing the evolution of the distribution functions in the ballistic nanoscale device coupled to macroscopic dissipative contacts. This derivation is performed using the time coarse-graining procedure and short-time expansion from [2] and introducing: (1) An interaction Hamiltonian which enables obtaining Landauer-type distribution

functions [86] in steady state, suitable for the electronic devices under consideration; (2) A time-dependent drift wave vector  $k_d(t)$ , which is proportional to the semiclassical current in the contacts and properly displaces the distribution functions in the contacts over time.

- Develop an algorithm to calculate the transient properties by solving the non-linear system of four equations – Schrödinger’s, Poisson’s, the current continuity, and the equation governing the time evolution of distribution functions. Special care must be taken to obtain a numerically stable algorithm with a satisfactory convergence rate. We use a globally convergent Newton’s method [95], with suitable Jacobians and a convergence method for  $k_d(t)$ . Initial conditions – which specify how the current density, potential profile, and charge density look at time  $t = 0^+$  s, or just after applying bias – are crucial for the subsequent time development.

The tasks of Theme 1 are treated in detail in Chapters 2 and 3. The contributions of Theme 1 resulted in one journal publication [96], one journal publication in preparation [B. Novakovic and I. Knezevic, “Transient quantum transport in ballistic nanostructures coupled to macroscopic contacts” (2012)], one book chapter [97], and six conference presentations [98–103], one winning the Best Student Paper Award [103].

A list of completed tasks and contributions for Theme 2:

- Develop a tight-binding form of the curvilinear 2D Schrödinger equation, by discretizing the curved surface, representing a thin semiconducting nanomembrane/nanoribbon with curvature. The Schrödinger equation in curvilinear coordinates, with the electric and magnetic fields included through the electric scalar  $V$  and magnetic vector potential  $\mathbf{A}$ , respectively, has a much more complicated form than the ordinary planar Schrödinger equation [78, 82]. Upon discretization, which allows for arbitrary real space modeling in principle, the curvilinear Hamiltonian can never be exactly Hermitian. Since it is desirable to have Hermitian matrices, we devise discretization schemes that allow for Hermitian matrices with additional matrix transformations. These matrix transformations are similar to the ones used earlier in non-uniform planar grid cases [104–106].

- Derive a suitable gauge to be used in calculating the magnetic vector potential,  $\mathbf{B} = \nabla \times \mathbf{A}$ . Although the physical problem should be gauge-independent, there is a preferred gauge in our case and other gauges may lead to wrong results. The calculation method that we use introduces a preferred direction in the simulation domain, the point described in the following item in this list. Subsequently, the gauge used must produce  $\mathbf{A}$  that conforms with that direction. We derive and use a gauge conveniently named a local Landau gauge. In connection with the presence of magnetic field, one also has to carefully calculate the velocities of electrons injected from planar contacts into a curved device, the velocities being important in calculating the transmission coefficient  $T$ .
- Develop a calculational method that uses the tight-binding form of the curvilinear Schrödinger equation, and calculates the electron density and conductance for a given curved nanoribbon geometry and magnetic field. The method used is a stabilized transfer matrix method, which is a modification of the method developed by Usuki and co-workers [87] for planar nanostructures. It starts at the injecting contact, by calculating the eigenmodes, and integrates the matrix equation representing a rearranged tight-binding curvilinear Schrödinger equation, until the second contact is reached, at which point the electron density and conductance can be calculated. This method in principle allows defining a wide range of geometries by assuming a simpler underlying coordinate system, like cylindrical or toroidal, and defining the actual shape of the nanoribbon/nanomembrane by specifying a potential on the surface that can constrain the particle motion.

The tasks of Theme 2 are treated in detail in Chapter 4. The contributions of Theme 2, resulted in one journal publication [107], and four conference presentations [101, 108–110].

The knowledge and experience gained while working on these main projects was used for two smaller projects, as a part of collaboration with the experimental groups of Prof. Lagally and Prof. Booske, both at the University of Wisconsin-Madison. This collaboration led to two papers published [111, 112] and four conference presentations [113–116].

## Chapter 2

### Electronic quantum master equations

Today, integrated circuits are made of many small electronic devices connected by leads to large reservoirs that supply them with charged particles (or other kind of matter/information). The natural framework in which modern electronic devices should be studied is the open system formalism, providing the necessary mathematical tools for handling a large number of variables and focusing on the most relevant ones [8, 117]. It requires the use of the reduced many-particle density matrix, that stores the information about the relevant variables after all the others have been traced out (a single-particle density matrix is generally insufficient). Most generally, we can refer to the electronic device in question as *the system* (i.e. the active region), which contains all the relevant variables, while everything else is *the environment* (e.g. reservoirs spatially separated from the system; other particles, like phonons, that share the same volume as the system). Therefore, the object of research is now a composite system, consisting of two, or more, physically coupled subsystems. The accuracy and the relevancy of our model will depend on what assumptions we apply to the environment. The environment in our problem is equivalent to reservoirs, since we do not consider other particles in the active region (*the system*), because it is ballistic (no scattering). In the rest of this document we will interchangeably use the term *contacts* to mean the same thing as *reservoirs*.

## 2.1 The Single-Particle Quantum Master Equation

The QME is an equation of motion for the density matrix. In the single-particle picture, with off-diagonal elements included, it is a kinetic equation, where diagonal elements provide information about the population of single-particle states, while off-diagonal elements represent coherences between different single-particle states, describing localized particles. The single-particle QME is approximate and can be formally derived by truncating the BBGKY chain of equations, similar to the BTE. It describes the time-irreversible, dissipative time evolution for the single-particle states. In this section, we will discuss the general form of the single-particle QME, as well as two particular equations, starting from the full many-particle density matrix and its equation of motion, the Liouville-von Neumann equation.

### 2.1.1 The Density Matrix and The Liouville-von Neumann Equation

The density matrix formalism was pioneered by John von Neumann in 1927 [118, 119] and is used to describe a mixed ensemble of states of a physical system, where by mixed we have in mind an ensemble that contain at least two, or more, different states of a physical system. Two extremes would be a pure ensemble, where all the states are the same, described by some state ket  $|\alpha\rangle$ , and a completely randomized ensemble, with each one of  $N$  states described by a different state ket  $|\alpha_i\rangle$ , where  $i = 1, \dots, N$ . Here, the state  $|\alpha\rangle$ , or  $|\alpha_i\rangle$ , is, in general, a linear combination of the eigenstates of the Hamiltonian. For a physical system with many particles the most exact density matrix is the one that describes a mixed ensemble of a full set of many-particle states, taking into account all the mutual interactions between the particles in the system. Such many-particle density matrix at some initial time 0 is defined as

$$\rho_{12\dots N}(0) = \sum_{i=0}^M W_{12\dots N}^{(i)} \left| \Psi_{12\dots N}^{(i)}(0) \right\rangle \left\langle \Psi_{12\dots N}^{(i)}(0) \right|, \quad (2.1)$$

where  $M$  is the maximum number of many-particle states in the ensemble and  $W_{12\dots N}^{(i)}$ 's are real positive numbers, representing the probability of occupation of the many-particle states  $|\Psi_{12\dots N}^{(i)}(0)\rangle$ , which are symmetrized or anti-symmetrized linear combinations of products of a complete set of

single-particle states [120]. The density matrix in (2.1) is normalized with the condition  $\text{Tr}(\rho_{12\dots N}(0)) = 1$ . From (2.1) follows that  $\rho$  is also hermitian,  $\rho_{12\dots N}^\dagger(0) = \rho_{12\dots N}(0)$ .

The time-evolution of the states  $|\Psi_{12\dots N}^{(i)}(0)\rangle$  is given by the many-particle time-dependent Schrödinger equation

$$i\hbar \frac{d}{dt} \left| \Psi_{12\dots N}^{(i)}(t) \right\rangle = H_{12\dots N} \left| \Psi_{12\dots N}^{(i)}(t) \right\rangle . \quad (2.2)$$

These states are not necessarily orthogonal. Since states  $|\Psi_{12\dots N}^{(i)}(0)\rangle$  in (2.1) evolve according to (2.2), we have that the many-particle density matrix at some later time  $t$  will be given by

$$\rho_{12\dots N}(t) = \sum_{i=0}^M W_{12\dots N}^{(i)} \left| \Psi_{12\dots N}^{(i)}(t) \right\rangle \left\langle \Psi_{12\dots N}^{(i)}(t) \right| . \quad (2.3)$$

By differentiating (2.3) with respect to time and making use of (2.2) we arrive at the most general form of the Liouville-von Neumann equation, describing the time evolution of the full many-particle density matrix for a closed system

$$i\hbar \frac{d}{dt} \rho_{12\dots N}(t) = [H_{12\dots N}, \rho_{12\dots N}(t)] \equiv \mathcal{L}_{12\dots N} \rho_{12\dots N} , \quad (2.4)$$

where  $\mathcal{L}_{12\dots N}$  is defined as a commutator superoperator generated by the many-particle Hamiltonian  $H_{12\dots N}$ . Because this equation was generated by the Schrödinger equation, it preserves the previously stated properties of the density matrix, namely the normalization and hermiticity. If we use a shorthand notation  $|\Psi_{12\dots N}^{(i)}(t)\rangle \equiv |\alpha_i\rangle$ , the expectation value of an observable  $A$  in a mixed ensemble described by the initial condition (2.1) and by (2.4), is given by

$$\begin{aligned} \langle A \rangle &= \sum_{i=1}^M w_i \langle \alpha_i | A | \alpha_i \rangle = \sum_{i=1}^M w_i \langle \alpha_i | \alpha_i \rangle \langle \alpha_i | A | \alpha_i \rangle \\ &= \sum_{i=1}^M \langle \alpha_i | \rho_{12\dots N} A | \alpha_i \rangle = \text{Tr}(\rho_{12\dots N} A) , \end{aligned} \quad (2.5)$$

where we use the fact the the many-particle states,  $|\alpha_i\rangle$  are properly normalized.

### 2.1.2 The BBGKY Chain and the Single-Particle QME

Instead of one exact many-particle Liouville-von Neumann equation (2.4), we can construct  $N$  coupled equations for the reduced density matrices,  $\rho_1, \rho_{12}, \dots, \rho_{12\dots N}$ , that form the BBGKY

chain of equations [10]. Similar to the way the BTE, as a single particle equation for the distribution function over a single-particle phase space  $(\mathbf{r}, \mathbf{p})$ , is derived by applying approximations to the BBGKY chain of equations [19], we can derive the single-particle QME for the time evolution of the single-particle density matrix. If we assume that the dissipation processes are sufficiently weak (the weak-coupling or Born approximation) and memoryless or Markovian (one collision is completed before the next one starts, so that collisions do not depend on their history or initial conditions), then we can consider that the transport consists of periods of “free flights” (generalized “free flights” generated by the single particle Hamiltonian) and temporally and spatially very localized collisions described by a linear collision operator. In this way we can obtain a Boltzmann like QME for the time evolution of the single-particle density matrix  $\rho(t)$  [7]

$$\frac{d\rho}{dt} = \frac{1}{i\hbar} \mathcal{L}\rho + \mathcal{C}\rho, \quad (2.6)$$

where  $\mathcal{C}$  is the collision superoperator, which is usually used to describe electron/phonon or electron/impurity interactions, and  $\mathcal{L}$  is a commutator superoperator (2.4) generated by the single-particle Hamiltonian  $H$ .  $H$ , for noninteracting particles of the same kind (usually we are interested in electrons), is a sum of the kinetic energy operator and the potential energy due to any external potential  $V_{ext}(\mathbf{r})$ , but if we couple the transport equation (2.6) with the Poisson equation it will also include the Hartree potential  $V_H(\mathbf{r})$  (mean-field approximation). So, we have in total

$$H = -\frac{\hbar^2}{2m} \nabla^2 + V_{ext}(\mathbf{r}) + V_H(\mathbf{r}). \quad (2.7)$$

Equation (2.6) is a limiting case of a density matrix completely reduced down to the single-particle states, with the additional assumptions about the nature of interactions in the system, stated above. The consequence of this derivation is the introduction of the time-irreversibility into the evolution of the single-particle density matrix  $\rho$  in (2.6), starting from the time-reversible equation (2.4).

So far we have considered a closed physical system for which  $\mathcal{L}$  in (2.6) is hermitian, i.e. with real eigenvalues. Therefore it will contribute with complex oscillatory solutions for  $\rho$  in (2.6). The collision operator  $\mathcal{C}$  will introduce negative real parts of eigenvalues which will cause an exponential decay of  $\rho$ . Therefore, this time-irreversible system is stable and behaves in an

expected way.  $\mathcal{L}$  is hermitian as a consequence of the hermiticity of the single-particle Hamiltonian for a closed system, where the hermiticity is defined through [7, 121]

$$\begin{aligned} \int_V [\psi^*(H\psi) - (H\psi)^*\psi] d^3r &= 0 \\ &= \int_S \left( \psi^* \frac{d\psi}{dn} - \frac{d\psi^*}{dn} \psi \right) d^2r = \int_S \mathbf{J} ds, \end{aligned} \quad (2.8)$$

where Green's identity was used,  $\psi$  is the wavefunction,  $H$  the single-particle Hamiltonian and  $\mathbf{J}$  the current density. We see that, when the number of particles is conserved in the volume  $V$  (closed system), the current density flux given by the last term in (2.8) is zero according to the current continuity equation and  $H$ , as well as  $\mathcal{L}$ , are hermitian.

If, on the other hand, the system is open, so that it exchanges particles with the environment, the number of particles is not conserved in general and both  $H$  and  $\mathcal{L}$  are non-hermitian. Therefore, the eigenvalues of  $\mathcal{L}$  will have imaginary parts and only non-positive imaginary parts are permissible in order to avoid having growing exponentials. To ensure this, it was shown by Frenslley [7] that the boundary conditions have to be carefully chosen. In particular it is necessary to use time-irreversible boundary conditions, which can be easily defined only in phase space. For example, if we have a 1D problem with two contacts and a region of interest (open system) in between we can choose different boundary conditions at  $(x_L, p_x)$  than at  $(x_R, -p_x)$ , where  $x_L$  and  $x_R$  are the left and right spatial boundaries of our open system. Now, under the time inversion those boundary conditions will apply to  $(x_L, -p_x)$  and  $(x_R, p_x)$ , respectively, and the problem will not be the same anymore. These BCs mean that the occupations of positive and negative propagating states are fixed by the left and right contacts, respectively. Even if we disregard the fact that the time-irreversible BCs are needed to achieve stability, they are a natural choice in the context of the following statement in [7] "if one's objective is to develop useful models of physical systems with many dynamical variables, rather than to construct a rigorously deductive mathematical system, it is clearly most profitable to adopt the view that irreversibility is a fundamental law of nature." The BCs of this form are naturally to be used with the Wigner function method. To include this kind of boundary conditions in (2.6) we can formally specify a contribution to the time evolution of the density matrix due to the injection/extraction through the contacts, a source term, the form

of which can be determined phenomenologically

$$\frac{d\rho}{dt} = \frac{1}{i\hbar}\mathcal{L}\rho + \mathcal{C}\rho + \left(\frac{\partial\rho}{\partial t}\right)_{\text{inj/extr}} . \quad (2.9)$$

### 2.1.3 The Pauli Master Equation

As already mentioned in Sect. 1, the PME describes the time evolution of the probabilities of occupation of the single-particle Hamiltonian's eigenstates. With  $p_n(t) \equiv \rho_{nn}(t)$  and for a closed system it is given by

$$\frac{d}{dt}p_n(t) = \sum_m [A_{nm}p_m(t) - A_{mn}p_n(t)] . \quad (2.10)$$

Equation (2.10) is easily justifiable at a phenomenological level, in situations when the exact Hamiltonian is not known, or when it is too complicated [117]. Then, we can always set up a master equation of the previous form, to describe the dissipative transport in the system. Coefficients  $A_{mn}$  represent transition rates between the levels and they can be found in a standard way, by using the quantum mechanical perturbation theory (Fermi's golden rule), or from experimental data. Alternatively, the PME follows from (2.6) by using Fermi's golden rule for the collision superoperator and a basis that diagonalizes the single-particle Hamiltonian that generates  $\mathcal{L}$ , since then the term  $\mathcal{L}\rho$  vanishes and there is only the collision operator, which corresponds to the right-hand side of (2.10). So, the PME is a closed equation for the diagonal elements of the single-particle density matrix in the eigenbasis of the single-particle Hamiltonian, obtained from (2.6) by using Fermi's golden rule to describe scattering. It will be a complete description of the problem in the case the off-diagonal elements in (2.6) can be neglected. We will say more on the conditions to satisfy that requirement in the following.

The simplicity of the PME (2.10) makes it attractive for applications to real problems of quantum transport in electronic devices. However, the major disadvantage of the PME is that it violates the current continuity, as shown by Frensky [7]. The reason for this is that open systems are inhomogeneous, making the eigenstates have different spatial distributions. Mathematically, if we combine the PME and the current continuity equation, with  $\rho(x, x; t)$  being the electron density, we can obtain for the rate of change of the electron density due to transitions between two eigenstates

$\psi_m$  and  $\psi_n$  [7]

$$\begin{aligned} \frac{\partial}{\partial t} \rho(x, x; t) &= \frac{\partial p_m}{\partial t} |\psi_m(x)|^2 + \frac{\partial p_n}{\partial t} |\psi_n(x)|^2 \\ &= [A_{nm} p_m(t) - A_{mn} p_n(t)] \times [|\psi_n(x)|^2 - |\psi_m(x)|^2] . \end{aligned} \quad (2.11)$$

The left-hand side of (3.5d) must be zero, because the divergence of an eigenstate's current density is zero. Since the second term on the right-hand side is non-zero, due to different spatial distributions of different eigenstates, we need the first term on the right-hand side to be zero, which is true only in equilibrium when detailed balance is satisfied. The conclusion is that the PME alone (i.e. without considering the off-diagonal terms) may be used at or very near equilibrium and in steady state, when  $\partial p_{m,n}/\partial t = 0$  and therefore  $\partial \rho(x, x; t)/\partial t = 0$ , as it should be because  $\nabla \cdot \mathbf{J}_m = \nabla \cdot \mathbf{J}_n = 0$ .

A good example of using the PME in modeling small electronic devices is the work done by Fischetti [1, 93]. There, the PME application to small devices was justified and the results of steady state simulations with [1] and without [93] the full band structure were compared with those obtained by using the BTE. Set-up is such that contacts to the device as well as phonons and other particles important for scattering belong to the environment, while the device region with electrons is the open system. The justification and conditions for using the PME go as follows:

- As shown by Van Hove [122] and Kohn and Luttinger [123], if one starts from a quasideagonal initial state and in the weak-scattering limit the off-diagonal terms remain negligible. Quasideagonal states satisfy the condition that the off-diagonal terms are nonvanishing only when mixing states with energy difference  $\delta E_{th} \ll \delta E_D$ , where  $\delta E_{th}$  is the thermal broadening of the states and  $\delta E_D$  is the energy scale over which the matrix elements of perturbing interactions are constant.
- If the size of the device is comparable or smaller than the dephasing length of the incoming electrons from the contacts,  $L \ll \lambda_\phi$  ( $\lambda_\phi \approx 30 - 50$  nm for Si at 300 K), then they appear as plane waves, i.e. the density matrix is diagonal in the momentum representation. Assuming the weak-scattering limit in the open system (device), we can say, with respect to the previous

statement, that neither are off-diagonal elements injected from the contacts nor do they form in the device region, so that the PME is applicable.

- The PME is unable to model the femtosecond time dynamics, because that is a genuinely off-diagonal problem on time-scales of the order of collision durations and strong-scattering effects beyond Fermi's golden rule. The PME's areas of applicability are steady state with the weak-scattering and long-time limits and "adiabatic" transients, when the number of particles in the system changes very slowly with time.

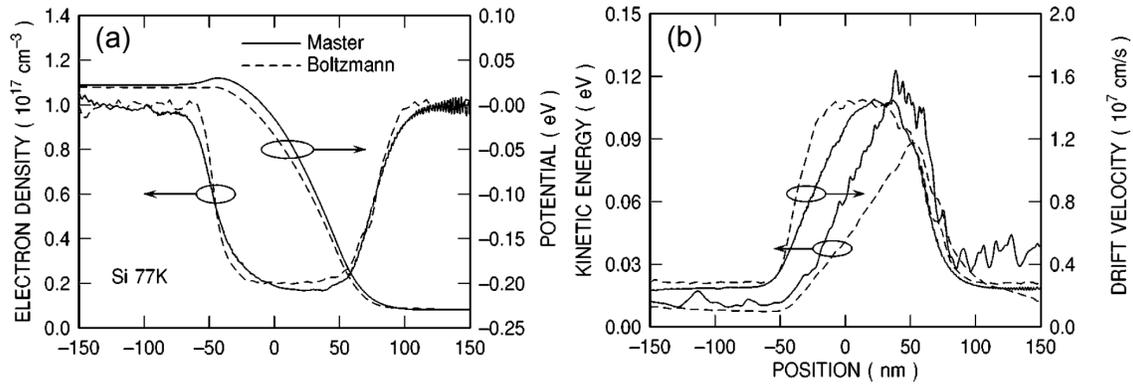


Figure 2.1 (a) The electron charge density and potential energy for an *nin* Si diode at 77 K, biased at 0.25 V, where the solid lines are results of using the master equation (2.13), while the dashed lines are results of using the Monte Carlo BTE. (b) Similar as (a), but with results for the average kinetic energy and drift velocity. Reprinted with permission from [M. V. Fischetti, Phys. Rev. B **59**, 4901 (1999)]. ©(1999) by the American Physical Society.

The PME with Fermi's golden rule can only be used to find occupation probabilities governed by scattering in the system, but not due to the coupling to the contacts. Following the work of Fischetti [1,93] this coupling can be introduced at a phenomenological level through a source term in the PME. The form of that source term for a general multiterminal configuration is given by [93]

$$\left( \frac{\partial \rho_{\mu}^{(s)}}{\partial t} \right)_{\text{res}} = |C_{\mu}^{(s)}|^2 v_{\perp}(\mathbf{k}_{\mu s}) [f^{(s)}(\mathbf{k}_{\mu s}) - \rho_{\mu}^{(s)}], \quad (2.12)$$

where  $s$  indicates the contact/terminal,  $v_{\perp}$  is the injecting velocity,  $f^{(s)}$  the  $s$ -th contact distribution function,  $\mu$  the full set of quantum numbers describing the eigenstates in the open system/device and  $C_{\mu}^{(s)}$  takes care of the proper normalization of the states. Additional assumption is that the injecting distributions are given by the drifted Fermi-Dirac distribution  $f^{(s)}(\mathbf{k}_{\mu}^{(s)} - \mathbf{k}_d^s)$ , where  $\mathbf{k}_d^s$  is calculated from the semiclassical current in the contact  $s$ . This takes into account the fast relaxation in the contacts and ensures the charge neutrality near the contacts/device boundaries as well as the current continuity. With this source term we can write the final steady state equation of motion for populations as

$$\begin{aligned} \sum_{\mu'r} \left[ A_{\mu s; \mu' r} \rho_{\mu'}^{(r)} - A_{\mu' r; \mu s} \rho_{\mu}^{(s)} \right] + |C_{\mu}^{(s)}|^2 v_{\perp}(\mathbf{k}_{\mu s}) \rho_{\mu}^{(s)} = \\ = |C_{\mu}^{(s)}|^2 v_{\perp}(\mathbf{k}_{\mu s}) f^{(s)}(\mathbf{k}_{\mu}^{(s)} - \mathbf{k}_d^s) . \end{aligned} \quad (2.13)$$

This is a set of equations over  $\mu$  that has to be solved self-consistently with  $\mathbf{k}_d$  by applying the condition of current continuity at the contact/device boundaries.

Some of the results of the full-band calculations with (2.13) are given in Fig. 2.1 for an *nin* Si diode at 77 K, biased at 0.25 V [93]. For comparison purposes, alongside them are the results of the simulation with the Monte Carlo BTE.

### 2.1.4 A Single-Particle QME Beyond the Born-Markov Approximation

A somewhat different QME to study semiconductors in a uniform electric field can be constructed using the perturbation expansion of the single-electron/many-phonon Liouville-von Neumann equation [124–126]. The difference with the previous one is that it was applied to homogeneous bulk problems (not devices), but on the other hand it makes no assumption about the electron-phonon coupling (it is beyond the Born-Markov or weak-scattering/long-time limit of the PME) and is able to simulate energy-nonconserving transitions, multiple collisions and intracollisional field effects [127, 128].

The perturbation expansion to the Liouville-von Neumann equation for bulk semiconductors in a uniform electric field can be constructed as follows [124]. The Hamiltonian of this system in the

effective mass approximation and with parabolic energy bands is a sum of several contributions

$$H = H_e + H_E + H_p + H_{e-ph} = H_0 + H_{e-ph} , \quad (2.14)$$

where

$$H_e = -\frac{\hbar^2}{2m^*} \nabla^2 , \quad H_E = e\mathbf{E}\mathbf{r} , \quad H_p = \sum_{\mathbf{q}} \hbar\omega_{\mathbf{q}} a_{\mathbf{q}}^\dagger a_{\mathbf{q}} \quad (2.15)$$

and  $H_{e-ph}$  is a standard Hamiltonian describing electron-phonon coupling and consisting of absorption and emission parts.  $H_0$ , describing the free and non-interacting electron gas, the equilibrium phonon distribution and the external homogeneous electric field is used to solve the time-dependent Schrödinger equation. Approximate solutions are the tensor products of the time-dependent accelerated plane waves (they would be accelerated Bloch waves beyond the effective mass approximation) normalized to 1 over the crystal volume  $V$  [129], and the many-body phonon states  $|n_{\mathbf{q}}, t\rangle$

$$|\mathbf{k}_0, n_{\mathbf{q}}, t\rangle = \frac{1}{\sqrt{V}} e^{i\mathbf{k}(t)\mathbf{r}} e^{-i\int_0^t ds \omega(\mathbf{k}(s))} |n_{\mathbf{q}}, t\rangle , \quad (2.16)$$

where  $\mathbf{k}(t) = \mathbf{k}_0 - e\mathbf{E}t/\hbar$  and  $\omega(\mathbf{k}(t)) = \hbar k^2/2m^*$ .

If we use this basis set (whose time evolution is generated by  $H_0$ ) for the density matrix, the Liouville-von Neumann equation contains only the interaction Hamiltonian

$$i\hbar \frac{\partial}{\partial t} \rho(\mu, \mu', t) = [H_{e-ph}(t), \rho(t)]_{\mu, \mu'} , \quad (2.17)$$

where  $\mu \equiv (\mathbf{k}_0, n_{\mathbf{q}})$ . Upon the formal integration and perturbation expansion we obtain the following Dyson series for the diagonal elements of the density matrix  $\rho(\mu, t) = \rho(\mu, \mu, t)$

$$\begin{aligned} \rho(\mu, t) &= \rho(\mu, 0) + \int_0^t dt_1 \left[ \tilde{H}_{e-ph}(t_1), \rho(0) \right]_{\mu, \mu} \\ &\quad + \int_0^t dt_1 \int_0^{t_1} dt_2 \left[ \tilde{H}_{e-ph}(t_1), \left[ \tilde{H}_{e-ph}(t_2), \rho(0) \right]_{\mu, \mu} \right] + \dots \\ &= \rho^{(0)}(\mu, t) + \rho^{(1)}(\mu, t) + \rho^{(2)}(\mu, t) + \dots , \end{aligned} \quad (2.18)$$

where  $\tilde{H}_{e-ph} = (1/i\hbar) H_{e-ph}$  and the initial condition is assumed to be diagonal and uncoupled,  $\rho(\mu, \mu', 0) = \rho(\mu, 0) = \rho^{(0)}(\mu, t) = f_0(\mathbf{k}_0) P_{eq}(n_{\mathbf{q}})$ , where  $f_0$  and  $P_{eq}$  are the initial distribution functions of electrons and phonons, respectively.

We are only interested in the diagonal elements, whose time-evolution is given by (2.18), since, first, we want to evaluate expectation values of electronic quantities only and, second, they are diagonal in the electronic part of the wave function. Furthermore, (2.18) is a closed equation for the diagonal elements of  $\rho(t)$ , which is a consequence of a diagonal initial condition and the fact that there are only initial values of  $\rho$  at the right hand side of the perturbation expansion. Remember that we have mentioned a similar effect in a somewhat different context in Sect. 2.1.3, i.e. that the closed equation for the diagonal elements of the PME can be obtained from the general form of the single-particle QME (2.6) by working in the basis of the single-particle Hamiltonian and by approximating the collision superoperator with Fermi's golden rule. The fact that each term in the perturbation expansion starts from a diagonal state and have to end up in some other (or the same) diagonal state means that only even order terms in the expansion will survive. This can be explained by the fact that each interaction Hamiltonian (being linear in creation/destruction operators) will either create or destroy a phonon in that state (left or right) of the initial diagonal outer product of states (since in general  $\rho = \sum |\alpha\rangle\langle\alpha|$ ) that is on the same side as that interaction Hamiltonian, after we expand the commutation relations. So to maintain the diagonalization we have to balance each absorption/emission at one of the sides by either the opposite process (emission/absorption) on the same side, or by the same process (absorption/emission) at the opposite side. This can only be achieved by having an even number of interaction Hamiltonians in a particular term in the perturbation expansion.

Equation (2.18) has several advantages over the steady state PME with Fermi's golden rule (of course within the limits of its applicability), beside the fact it can actually handle the transient regime. It is able to model quantum transitions of a finite duration and, because of the basis used, the acceleration of the plane waves during that time. The former ensures that the processes where the subsequent scattering effects begin before the previous ones have finished are accounted for (multiple collisions), while the latter ensures that the intracollisional field effect is not neglected. This approach also relaxes the constraint of the strict energy conservation during collisions, especially at short timescales. One of the disadvantages is that the trace over many-phonon degrees of freedom has to be taken in (2.18) [124].

## 2.1.5 Monte Carlo Solution to the QME

Using the Monte Carlo stochastic technique to solve the semiclassical BTE [130–133] is very common today, since it provides very accurate results (without using extensive approximations to make the problem numerically tractable), while the computational time is no more a bottleneck considering the availability of computing resources. The same idea of solving the semi-classical transport equation stochastically, instead of directly numerically, can be applied to the QME. In this section we will give a brief review of the ways this can be done in the case of a single-electron QME where we seek solutions (steady state and transient) to the diagonal elements of the density matrix. They will be algorithmically compared with the semiclassical Monte Carlo and shown to bear many similar characteristics, as far as the implementation is concerned.

### 2.1.5.1 The Steady-State PME for Inhomogeneous Devices

As has been shown in Sect. 2.1.3, the PME can be successfully applied to a certain class of problems which nowadays have high importance due to the down-scaling of electronic devices. The main equation of that section (2.13), which is a linear steady state equation for the occupations of levels with source terms modeling injection/extraction from the contacts, can be solved by using the Monte Carlo method [1]. For comparison purposes, let us write the standard BTE [133]

$$\frac{df(\mathbf{k}, \mathbf{r}, t)}{dt} + \frac{1}{\hbar} \nabla_{\mathbf{k}} E(\mathbf{k}) \nabla_{\mathbf{r}} f(\mathbf{k}, \mathbf{r}, t) + \frac{\mathbf{F}}{\hbar} \nabla_{\mathbf{k}} f(\mathbf{k}, \mathbf{r}, t) = \left. \frac{\partial f(\mathbf{k}, \mathbf{r}, t)}{\partial t} \right|_{\text{Coll}}. \quad (2.19)$$

Diagonal elements of the density matrix from the PME (2.10),  $p_n(t) = \rho_{n,n}(t)$  ( $n$  is a full set of basis quantum numbers), correspond to the distribution function  $f(\mathbf{k}, \mathbf{r}, t)$  in (2.19), while the right hand side of (2.10) corresponds to the right hand side of (2.19). The main difference is in the drift and diffusion terms (due to the external field and spatial inhomogeneity) present in (2.19). Their absence from (2.10) is a consequence of a specific basis chosen for the density matrix, which diagonalizes the total potential consisting of the Hartree potential and the potential due to the external field. Although the BTE is most often used in the form given by (2.19), it can also be cast in the form without those two terms by a change in coordinates, from the phase space variables  $(\mathbf{r}, \mathbf{k})$  into the collision-free trajectories (path variables) [134]. So, to solve the PME we can use

the conventional Monte Carlo procedure, used to solve the standard BTE (2.19), but without the free-flight part.

To better understand the relationship between (2.10) and (2.19) it can be shown that they are both limiting cases, but at the opposite ends of the domain [1]. As already mentioned in Sect. 2.1.3, the PME, being diagonal and therefore neglecting the off-diagonal elements, is justified for the quasideagonal initial state. As shown by Van Hove [122] it is the state obtained by mixing the eigenstates of the unperturbed Hamiltonian, but only in a very narrow energy range (amplitudes are non-zero only for a very narrow range of energies of the states being mixed). Therefore, those are the states highly delocalized. This physically corresponds to our assumption of devices much smaller than the dephasing length in the contacts, such that injecting electrons appear to them as spatially delocalized (but energetically very localized) wave packets, plane waves being the limiting case. There is one more group of states for which the diagonal form of the transport equation is justified and they are spatially very localized states, formed by linear combinations of eigenstates of the unperturbed Hamiltonian with amplitudes varying slowly with the energy. This opposite limit is satisfied by the BTE, which is therefore diagonal in the real space (the PME is diagonal in the wave vector space).

Finally, the implementation procedure would go as follows [1]:

- Electrons are initialized into the eigenstates  $|\mu\rangle$ , where  $\mu$  is a full set of quantum numbers for the open system considered, according to the thermal equilibrium occupations as determined by the solution to the ballistic problem (no scattering).
- The time step is chosen and all transition probabilities are calculated. Scattering probability  $P_{\text{scatter}}$  is proportional to the transition rates determined by Fermi's golden rule, while injection/extraction probabilities (the processes that can change the number of particles in the open system)  $P_{\text{in/out}}$  are proportional to the injection/extraction rates. Scattering or extraction events are selected according to the generated random number.
- If scattering is selected then the final state is chosen according to the final density of states and the matrix elements connecting the initial and final states, just like in the conventional

Monte Carlo procedure. If extraction (exit through a contact) is selected, the electron is simply removed. After all particles are processed, new particles are added to the states according to  $P_{in}$  and the drifted Fermi-Dirac distribution in the injecting contacts.

- After a few Monte Carlo steps the occupations of states, obtained from the Monte Carlo, are used to update the potential and wave functions with the Schrödinger/Poisson solver. The frequency of this update is determined by the plasma frequency of the whole device. The new potential is treated as a sudden perturbation which redistribute electrons from the old states  $|\mu^{(old)}\rangle$  to the new states  $|\mu^{(new)}\rangle$  according to the probability given by  $|\langle\mu^{(new)}|\mu^{(old)}\rangle|^2$ .

### 2.1.5.2 A Single-Electron QME in Homogeneous Bulk

The explanation of the similarity of (2.18) with the BTE can proceed by remembering what we said in Sect. 2.1.5.1, about the BTE written in the path variables, when it has the following form (after the drift-diffusion terms have disappeared)

$$f(t) = f(0) + P_i f - P_o f = f_0 + P_i f_0 - P_o f_0 + P_i P_i f_0 - P_i P_o f_0 - P_o P_i f_0 + P_o P_o f_0 + \dots, \quad (2.20)$$

where  $P_i$  and  $P_o$  are the integral operators for scattering “in” and “out”. This equation is of the same general form as (2.18) and so similar Monte Carlo procedures can again be used to solve both equations, as will be outlined below.

The Monte Carlo algorithm to solve (2.18) has several novelties comparing to the one explained in Sect. 2.1.5.1 [126]. Beside the initialization and the standard random selections of the type of the scattering process (in/out scattering and the type of scattering) like in the conventional Monte Carlo, here we have several new random selections due to the perturbation expansion. First, there is a selection of the perturbative order (just the even ones, as shown previously), second, the selection of  $n/2$  times where the first interaction Hamiltonians of each quantum process (a quantum process is defined as a pair of  $\tilde{H}_{e-ph}$ 's for a distinct  $\mathbf{q}$ ) are to be evaluated and, third, as already pointed out the average over the phonon variables  $\mathbf{q}$  have to be performed (equivalent of taking the trace over the phonon degrees of freedom), for which a separate random number is reserved. So far, this is

the same for both (2.18) and (2.20). The additional steps for the quantum case would be to select the side of  $\rho(0)$  where each process starts and the time for the second  $\tilde{H}_{e-ph}$  in the process.

The restoration of this quantum Monte Carlo algorithm to the standard one, consisting of periods of free flights interrupted by scattering events, can be achieved by introducing a quantum analog of the self-scattering in the standard Monte Carlo algorithm, that makes scattering rates constant [135, 136]. That can be achieved by the following transformation [125, 126]

$$\rho(t) \rightarrow \exp \left( \int_{t_0}^t \gamma(t_1) dt_1 \right) \rho(t) , \quad (2.21)$$

where  $\rho$  is understood to represent diagonal elements  $\rho_\mu$  as before. For constant  $\gamma = 1/\tau$  and  $t_0 = 0$  we have  $\rho \rightarrow e^{(t/\tau)}\rho$ , which gives the following equation instead of (2.18)

$$\begin{aligned} \rho(t) = e^{-(t/\tau)} & \left[ \rho_0 + \left( \tilde{H}\tilde{H} + \frac{1}{2\tau} \right) \rho_0 - \tilde{H}\rho_0\tilde{H} - \tilde{H}\rho_0\tilde{H} \right. \\ & \left. + \rho_0 \left( \tilde{H}\tilde{H} + \frac{1}{2\tau} \right) + \dots \right] . \end{aligned} \quad (2.22)$$

In this concise notation the integral signs as well as argument lists and subscripts are dropped, and the commutation relations are expanded. This equation is actually equal to (2.18), since the damping factor  $e^{-(t/\tau)}$  is going to cancel with all the factors  $1/2\tau$  when all the integration and summations are performed. Nevertheless, this form makes the quantum Monte Carlo very similar to the standard ensemble Monte Carlo, consisting of periods of free flights interrupted by scattering events. The change to the previously explained algorithm is that the times selected for the first  $\tilde{H}$  in each process is separated by a constant time  $\tau$ , the “free-flight” time, but only a few events will actually be quantum processes (scattering events) with a definite  $\mathbf{q}$ . Although this procedure does not really contribute to the physical side of the problem, the fact that it is made similar to the semiclassical approach makes comparison with it much more transparent.

A representative result of the application of this algorithm and a comparison with the semiclassical Monte Carlo is shown in Fig. 2.2 [126]. We see a clear discrepancy in the drift velocity overshoot between the two techniques, which is attributed to the intracollisional field effect favoring transitions oriented along the field direction, comparing with the standard isotropic cross section.

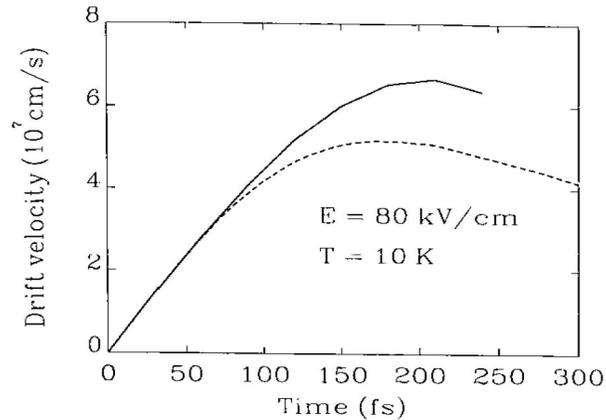


Figure 2.2 Drift velocity overshoot in silicon. The result of the quantum Monte Carlo technique is shown with the solid line, while the semiclassical result is shown with the dashed line.

Reprinted with permission from [C. Jacoboni, *Semicond. Sci. Technol.* **7**, B6 (1992)]. ©(1992) IOP Publishing Ltd.

## 2.2 Reduced Many-Particle QMEs

The reduced many-particle density matrix and the corresponding QME by its complexity fall between the single-particle and the full many-particle cases. This contributes to its flexibility, allowing us to find the optimal balance between the accuracy of modeling important physical processes in the open system and the computational complexity that results from including a large number of degrees of freedom. In this section we will first derive the formal, exact equation of motion for the reduced density matrix, the Nakajima-Zwanzig equation, and then introduce several approaches that make this equation more tractable for practical applications.

### 2.2.1 The Nakajima-Zwanzig Equation

Here, we will formally derive the Nakajima-Zwanzig equation for an exact reduced many-particle system. As already mentioned in Sect. 1 we are only interested in the time evolution of *the system*. Therefore, starting from (2.4) we need to trace out all *the environmental* degrees of

freedom. This can be formally done by introducing a projection superoperator pair  $\mathcal{P}$  and  $\mathcal{Q}$

$$\mathcal{P}\rho(t) = \rho_E \otimes Tr_E(\rho(t)) = \rho_E \otimes \rho_S(t), \quad \mathcal{Q}\rho(t) = \rho(t) - \mathcal{P}\rho(t), \quad (2.23)$$

where  $\rho(t)$  is the total density matrix,  $\rho_S(t)$  the density matrix of the system and  $\rho_E(t)$  represents the density matrix of the environment. Accordingly, we can split the Hamiltonian and the Liouvillian of the total system into three parts

$$H = H_S + H_E + H_I, \quad \mathcal{L} = \mathcal{L}_S + \mathcal{L}_E + \mathcal{L}_I, \quad (2.24)$$

where by index  $I$  we represent the interaction between the system and environment. Here, it is to be understood that each part acts in its corresponding Hilbert space (or Liouville space, for  $\mathcal{L}$ ), e.g.

$$H = I_E \otimes H_S + H_E \otimes I_S + H_I, \quad H_I = \sum_i A_i \otimes B_i, \quad (2.25)$$

where  $I_\alpha$  is the identity operator in the  $\alpha$ -subspace, and  $A$  and  $B$  are operators that act on the environment and system Hilbert spaces, respectively. The form of interaction in (2.25) is the most general one. By acting with projection operators (2.23) on (2.4) we get a system of two equations, one for  $\mathcal{P}\rho$  and one for  $\mathcal{Q}\rho$ . Upon formally solving it for the relevant part  $\mathcal{P}\rho$  we arrive at the formally exact equation of motion for the density matrix, the Nakajima-Zwanzig equation<sup>1</sup> [8, 137, 138]

$$\begin{aligned} \frac{d}{dt}\mathcal{P}\rho(t) = & - i\mathcal{P}\mathcal{L}(t)\mathcal{P}\rho(t) - \int_0^t ds\mathcal{K}(t,s)\mathcal{P}\rho(s) \\ & - i\mathcal{P}\mathcal{L}(t)\mathcal{G}(t,0)\mathcal{Q}\rho(0), \end{aligned} \quad (2.26)$$

where the convolution or memory kernel  $\mathcal{K}$  is

$$\mathcal{K}(t,s) = \mathcal{P}\mathcal{L}(t)\mathcal{G}(t,s)\mathcal{Q}\mathcal{L}(s)\mathcal{P}, \quad \mathcal{G}(t,s) = T_{\leftarrow} \exp \left[ -i \int_s^t ds' \mathcal{Q}\mathcal{L}(s') \right], \quad (2.27)$$

with  $T_{\leftarrow}$  being the time ordering operator which sorts the operators to the right of it according to increasing time argument from right to left.

Equation (2.26) is not very useful for practical applications in this form because it is very complex. It contains all orders of interaction  $H_I$  and some memory terms, which makes it an

---

<sup>1</sup>In the following we set  $\hbar = 1$ .

exact non-Markovian QME. Memory terms are incorporated through the non-local memory kernel, the integral over past times  $[0, t]$  and through the explicit dependence on the initial conditions in the second and third term. In the next section we will show some common approximations that are used to derive an approximate (to the second order in interaction) Markovian QME. Further modification to (2.26) that is commonly done is to choose the projection operator  $\mathcal{P}$  such that the third term is canceled in the situations when the initial state of the total system is uncoupled  $\rho(0) = \rho_E(0) \otimes \rho_S(0)$ . This is achieved if  $\mathcal{P}\rho$  is induced by  $\rho_E(0)$  in (2.23) because

$$\mathcal{Q}\rho(0) = \rho(0) - \mathcal{P}\rho(0) = \rho(0) - \rho_E(0) \otimes \rho_S(0) = 0 . \quad (2.28)$$

Now (2.26) is just

$$\frac{d}{dt}\mathcal{P}\rho(t) = -i\mathcal{P}\mathcal{L}(t)\mathcal{P}\rho(t) - \int_0^t ds\mathcal{K}(t, s)\mathcal{P}\rho(s) . \quad (2.29)$$

To finally obtain the reduced dynamics described by  $\rho_S(t)$  we have to take the trace over environmental variables  $Tr_E(\mathcal{P}\rho(t))$ .

### 2.2.2 The Born-Markov Approximation

Now, we will briefly sketch how to derive an approximate Markovian QME that ultimately lead to a QME whose time-evolution generator (equivalent to  $\mathcal{L}$  in (2.4)) satisfies the quantum dynamical semigroup property, meaning that if we define a dynamical map  $\mathcal{W}(t)$  as

$$\rho_S(t) = \mathcal{W}(t)\rho_S(0) , \quad (2.30)$$

its property is

$$\mathcal{W}(t_1)\mathcal{W}(t_2) = \mathcal{W}(t_1 + t_2) . \quad (2.31)$$

This defines a Markovian evolution and the necessary microscopic conditions for it will be stated in the following. The generator of this dynamical map can be defined as

$$\begin{aligned} \mathcal{W}(t) &= \exp(\mathcal{F}t) , \\ \frac{d}{dt}\rho_S(t) &= \mathcal{F}\rho_S(t) , \end{aligned} \quad (2.32)$$

from which it follows that the time evolution generator must be time-independent in order to have a Markovian QME.

The Born approximation is justified for weak coupling. This coupling is characterized by the interaction Hamiltonian  $H_I$ , which may refer to the coupling to reservoirs, phonons and everything else that can be encountered in real electronic devices. Since we assume that the coupling is weak we can keep only terms up to the second order in  $H_I$  in (2.29). Higher order interactions are contained in the memory term  $\mathcal{K}$  in the integral in (2.29) and in order to keep just the second order term we need to have  $\mathcal{L}_I$  in  $\mathcal{K}$  appearing twice at most. To achieve that we can approximate the propagator  $\mathcal{G}(t, s)$  with

$$\mathcal{G}(t, s) = T_{\leftarrow} \exp \left[ \int_s^t ds' \mathcal{Q}(\mathcal{L}_S(s') + \mathcal{L}_E(s')) \right], \quad (2.33)$$

which corresponds to leaving only zeroth order term in  $\mathcal{L}_I(t)$ . The Born approximation may be restated in several equivalent ways, depending on the way of derivation of final equations. The most obvious way, just mentioned, is to explicitly keep terms only up to the second order in interaction [117]. Equivalently, we can assume that, due to the weak-coupling, the density matrix of the system is always factorized during the evolution as [8]

$$\rho_S(t) = \rho_E \otimes \rho_S(t) \quad (2.34)$$

and that the density matrix of the reservoir is only negligibly affected by the interaction. The third way is somewhat less formal and is connected to the quantum mechanical scattering theory [119]. A variation of the Neumann series method, known as the Born series in this context, is used to approximate the form of the wave function after the scattering. This is also used in Fermi's golden rule, to calculate the transition rates which are valid in the weak-coupling and long-time limits.

The Markovian approximation would proceed by first replacing  $\mathcal{P}\rho(s)$  by  $\mathcal{P}\rho(t)$  in (2.29), thus removing any dependence at time  $t$  on the past states, for  $s < t$ ,

$$\frac{d}{dt} \mathcal{P}\rho(t) = -i\mathcal{P}\mathcal{L}(t)\mathcal{P}\rho(t) - \int_0^t ds \mathcal{K}(t, s)\mathcal{P}\rho(t). \quad (2.35)$$

This equation (in other forms and/or specific basis) is called the Redfield equation [8, 117, 139]. Second, there is an integral left which depends on the initial conditions, or in other words the

interval between the present and initial states. To get rid of this we make a simple substitution  $s \rightarrow t - s$  and let the upper limit of integration go to infinity, which gives us

$$\frac{d}{dt}\mathcal{P}\rho(t) = -i\mathcal{P}\mathcal{L}(t)\mathcal{P}\rho(t) - \int_0^\infty ds\mathcal{K}(t, t - s)\mathcal{P}\rho(t). \quad (2.36)$$

These two approximations, that make up the Markovian approximation, are possible provided  $\tau_E \ll \tau_S$ , where  $\tau_E$  is the environmental relaxation rate and  $\tau_S$  the open system relaxation rate. This means that the time evolution can be coarse-grained such that  $\rho_S(t)$  is almost constant during  $\tau_E$ , while the integral in (2.36) vanishes fast with decreasing  $t - s$  and, therefore, the Markovian approximation is justified.

Proceeding with some further less significant modifications to (2.36) we arrive at the most general form of the generator of the quantum dynamical semigroup [8, 117]. It constitutes the Lindblad form of the QME for an open system [140]

$$\frac{d}{dt}\rho_S(t) = -i[H, \rho_S(t)] + \sum_k \gamma_k \left( A_k \rho_S A_k^\dagger - \frac{1}{2} A_k^\dagger A_k \rho_S - \frac{1}{2} \rho_S A_k^\dagger A_k \right), \quad (2.37)$$

where  $H$  is the Hamiltonian that generates a unitary evolution, consisting of the system Hamiltonian and corrections due to the system–environment coupling, and  $A_k$ 's are the Lindblad operators that describe the interaction with the environment in the Born-Markov limit.

### 2.2.3 The Conventional Time–Convolutionless Equation of Motion

The Nakajima-Zwanzig equation (2.26), that relies upon the use of the projection-operator technique, has several shortcomings that are the motivation for the following sections. Various variants of the projection-operators have been used in the past to study a range of physical systems. Argyres and Kelley [141] applied it to a theory of linear response in spin-systems, Barker and Ferry [142] to quantum transport in very small devices, Kassner [143] to relaxation in systems with initial system-bath coupling, Sparpaglione and Mukamel [144] to electron transfer in polar media, followed by a study of condensed phase electron transfer by Hu and Mukamel [145], while Romero-Rochin and Oppenheim [146] studied relaxation of two-level systems weakly coupled to a bath. However, this approach is limited by two computationally intensive operations

needed to arrive at the final, reduced, density matrix of the open system: the time-convolution integral containing the memory kernel and the partial trace over environmental variables,  $\text{Tr}_E(\mathcal{P}\rho)$ . Specifically, these limits would be lifted by applying the Markov and Born approximations of Sect. 2.2.2, respectively, because then the time-convolution disappears and the trace is a trivial operation since the equation for  $\mathcal{P}\rho$  is already well factorized into the environmental and system parts.

Going beyond the Born-Markov approximation we have to think of different methods of leveraging the computational burden. In line with that, Tokuyama and Mory [147] proposed a time-convolutionless equation of motion in the Heisenberg picture. This was extended to the Schrödinger picture by Shibata *et al.* [148, 149] after which a stream of research appeared. Saeki analyzed the linear response of an externally driven systems coupled to a heat bath [150] and systems coupled to a stochastic reservoir [151, 152]. Ahn extended the latter to formulate the quantum kinetic equations for semiconductors [153], and a theory of optical gain in quantum-well lasers [154]. Later, he treated noisy quantum channels [155] and quantum information processing [156]. Chang and Skinner [157] applied the time-convolutionless approach to analyze relaxation of a two-level system strongly coupled to a harmonic bath, while Golosov and Reichmann [158] analyzed condensed-phase charge-transfer process. In the following, we will give a brief derivation of the time-convolutionless equation of motion and point out some of its shortcomings, resulting from the fact that it is still based on the projection-operator technique.

Let us choose some arbitrary, but proper and constant in time, environmental density matrix  $\tilde{\rho}_E$  as a generator for the time-independent projection operator (2.23). This means that  $\text{Tr}_E(\tilde{\rho}_E) = 1$  and therefore

$$\text{Tr}_E(\mathcal{P}\rho) = \text{Tr}_E(\tilde{\rho}_E) \cdot \text{Tr}_E(\rho) = \text{Tr}_E(\rho) = \rho_S . \quad (2.38)$$

The two equations for the projection operators  $\mathcal{P}$  and  $\mathcal{Q}$  are

$$\frac{d}{dt}(\mathcal{P}\rho(t)) = -i\mathcal{P}\mathcal{L}(t)\rho(t) = -i\mathcal{P}\mathcal{L}(t)\mathcal{P}\rho(t) - i\mathcal{P}\mathcal{L}(t)\mathcal{Q}\rho(t) , \quad (2.39)$$

$$\frac{d}{dt}(\mathcal{Q}\rho(t)) = -i\mathcal{Q}\mathcal{L}(t)\rho(t) = -i\mathcal{Q}\mathcal{L}(t)\mathcal{Q}\rho(t) - i\mathcal{Q}\mathcal{L}(t)\mathcal{P}\rho(t) . \quad (2.40)$$

A formal solution of (2.40) is

$$\mathcal{Q}\rho(t) = -i \int_0^t dt' \mathcal{G}(t, t') \mathcal{Q}\mathcal{L}(t') \mathcal{P}\mathcal{U}(t', t) \rho(t) + \mathcal{G}(t, 0) \mathcal{Q}\rho(0), \quad (2.41)$$

where for  $t > t'$

$$\begin{aligned} \mathcal{G}(t, t') &= T_{\leftarrow} \exp \left( -i \int_{t'}^t ds \mathcal{Q}\mathcal{L}(s) \mathcal{Q} \right), \\ \mathcal{U}(t', t) &= T_{\rightarrow} \exp \left( i \int_{t'}^t ds \mathcal{L}(s) \right). \end{aligned} \quad (2.42)$$

The superoperator  $\mathcal{U}(t, t')$  is defined by

$$\begin{aligned} \rho(t) &= \mathcal{U}(t, t_0) \rho(t_0), \\ \mathcal{U}(t, t') &= \Theta(t - t') T_{\leftarrow} \exp \left( -i \int_{t'}^t ds \mathcal{L}(s) \right) + \Theta(t' - t) T_{\rightarrow} \exp \left( i \int_t^{t'} ds \mathcal{L}(s) \right). \end{aligned} \quad (2.43)$$

By using it we make (2.41) time-local, which is the essence of this approach. Equation (2.41) can be rearranged in the following way

$$\mathcal{D}(t; 0) \mathcal{Q}\rho(t) = [1 - \mathcal{D}(t; 0)] \mathcal{P}\rho(t) + \mathcal{G}(t, 0) \mathcal{Q}\rho(0), \quad (2.44)$$

where  $\mathcal{D}(t; 0)$  is defined as

$$\mathcal{D}(t; 0) = 1 + i \int_0^t dt' \mathcal{G}(t, t') \mathcal{Q}\mathcal{L}(t') \mathcal{P}\mathcal{U}(t', t). \quad (2.45)$$

Assuming that  $\mathcal{D}(t; 0)$  is invertible, (2.41) finally becomes

$$\mathcal{Q}\rho(t) = [\mathcal{D}(t; 0)^{-1} - 1] \mathcal{P}\rho(t) + \mathcal{D}(t; 0)^{-1} \mathcal{G}(t, 0) \mathcal{Q}\rho(0). \quad (2.46)$$

Using the last equation in (2.39) we obtain

$$\frac{d}{dt} (\mathcal{P}\rho(t)) = -i \mathcal{P}\mathcal{L}(t) \mathcal{D}(t; 0)^{-1} \mathcal{P}\rho(t) - i \mathcal{P}\mathcal{L}(t) \mathcal{D}(t; 0)^{-1} \mathcal{G}(t, 0) \mathcal{Q}\rho(0). \quad (2.47)$$

The last step that is left to obtain the conventional time-convolutionless equation of motion is to take the trace over environmental variables of (2.47), which gives us

$$\begin{aligned}
\frac{d}{dt}\rho_S(t) &= -i\text{Tr}_E [\mathcal{P}\mathcal{L}(t)\mathcal{D}(t;0)^{-1}\mathcal{P}\rho(t)] - i\text{Tr}_E [\mathcal{P}\mathcal{L}(t)\mathcal{D}(t;0)^{-1}\mathcal{G}(t,0)\mathcal{Q}\rho(0)] \\
&= -i\text{Tr}_E [\mathcal{L}(t)\mathcal{D}(t;0)^{-1}\tilde{\rho}_E \otimes \rho_S(t)] - i\text{Tr}_E [\mathcal{L}(t)\mathcal{D}(t;0)^{-1}\mathcal{G}(t,0)\mathcal{Q}\rho(0)] \\
&= -i\text{Tr}_E [\mathcal{L}(t)\mathcal{D}(t;0)^{-1}\tilde{\rho}_E] \rho_S(t) - i\text{Tr}_E [\mathcal{L}(t)\mathcal{D}(t;0)^{-1}\mathcal{G}(t,0)\mathcal{Q}\rho(0)] .
\end{aligned} \tag{2.48}$$

This conventional form of the time-convolutionless equation of motion has three shortcomings. First, it explicitly depends on the choice of  $\tilde{\rho}_E$  that induces the projection operator, although the final result will not depend on it. Second, we have to evaluate complicated matrices  $\mathcal{U}$ ,  $\mathcal{G}$  and  $\mathcal{D}$  involving all the degrees of freedom in the system+environment, but at the end we will extract only those degrees belonging to the system, by taking the trace. Third, this approach depends on invertibility of  $\mathcal{D}$ , which might be difficult to fulfill. These issues will be addressed in the following sections.

## 2.2.4 The Eigenproblem of the Projection Operator

The projection operator, as defined in (2.38), is idempotent ( $\mathcal{P}^2 = \mathcal{P}$ ) because

$$\begin{aligned}
\mathcal{P}^2\rho = \mathcal{P}(\mathcal{P}\rho) &= \tilde{\rho}_E \otimes \text{Tr}_E [\tilde{\rho}_E \otimes \text{Tr}_E(\rho)] \\
&= \tilde{\rho}_E \otimes \text{Tr}_E(\tilde{\rho}_E) \text{Tr}_E[\text{Tr}_E(\rho)] = \tilde{\rho}_E \otimes \text{Tr}_E(\rho) = \mathcal{P}\rho .
\end{aligned} \tag{2.49}$$

Therefore, it has two eigenvalues, 0 and 1, and since they are both real we can conclude that  $\mathcal{P}$  is also hermitian,  $\mathcal{P} = \mathcal{P}^\dagger$ . In analogy with the notion that system states are members of the respective Hilbert space, while operators (like  $\rho$ ) act on it, we can introduce a Liouville space whose members are operators acting on the Hilbert space, while superoperators (like  $\mathcal{L}$ ) act on it. To complete the definition we have to define the inner product which is conveniently done as  $(A, B) = \text{Tr}(A^\dagger B)$ , where  $A$  and  $B$  are some operators belonging to the Liouville space. So, if the Hilbert spaces are  $\mathcal{H}_S$ ,  $\mathcal{H}_E$  and the composite space  $\mathcal{H}_{S+E} = \mathcal{H}_E \otimes \mathcal{H}_S$ , the respective Liouville spaces are  $\mathcal{H}_S^2$ ,  $\mathcal{H}_E^2$  and  $\mathcal{H}_{S+E}^2$ , where the dimensionality of Liouville spaces with respect

to that of the corresponding Hilbert spaces is obvious. It follows that  $\mathcal{P}$  is a superoperator acting on  $\mathcal{H}_{S+E}^2$ , which is  $d_E^2 d_S^2$ -dimensional. By construction (2.23) the image space of  $\mathcal{P}$  corresponds to  $\mathcal{H}_S^2$ , so that the subspace of  $\mathcal{P}$  spanned by the degenerate eigenvalue 1 is isomorphic to  $\mathcal{H}_S^2$ . We can write

$$\mathcal{H}_{S+E}^2 = (\mathcal{H}_{S+E}^2)_{\mathcal{P}=1} \oplus (\mathcal{H}_{S+E}^2)_{\mathcal{P}=0} , \quad (2.50)$$

where  $(\mathcal{H}_{S+E}^2)_{\mathcal{P}=1}$  is the  $d_S^2$ -dimensional unit subspace and  $(\mathcal{H}_{S+E}^2)_{\mathcal{P}=0}$  is the  $d_S^2 (d_E^2 - 1)$ -dimensional zero subspace of the eigenspace of  $\mathcal{H}_{S+E}^2$ .

We can always arrange the eigenbasis of  $\mathcal{P}$ ,  $\{|n\rangle | n = 1, \dots, d_E^2 d_S^2\}$ , such that the first  $d_S^2$  basis vectors span  $(\mathcal{H}_{S+E}^2)_{\mathcal{P}=1}$  and therefore

$$\mathcal{P} = \sum_{n=1}^{d_S^2} |n\rangle \langle n| . \quad (2.51)$$

The eigenstates of the composite space  $\mathcal{H}_{S+E}$  are constructed as  $|i\alpha\rangle = |i\rangle \otimes |\alpha\rangle$ , from which follows that the eigenstates  $|n\rangle$  of  $\mathcal{H}_{S+E}^2$  can be written by using four quantum numbers, i.e. as linear combinations of  $|i\alpha, j\beta\rangle$ . Here, states  $|i\rangle$  belong to the environment, while states  $|\alpha\rangle$  to the system. Furthermore, if we define  $\mathcal{P}$  by using a uniform density matrix

$$\bar{\rho}_E = d_E^{-1} \cdot \mathbf{1}_{d_E \times d_E} , \quad (2.52)$$

we can avoid mixing states with different  $\alpha$  and  $\beta$  to obtain a given  $|n\rangle$  [159]. One finds that the states defined as

$$|\overline{\alpha\beta}\rangle = \frac{1}{\sqrt{d_E}} \sum_{i=1}^{d_E} |i\alpha, i\beta\rangle \quad (2.53)$$

constitute an orthonormal basis within the unit subspace of  $\mathcal{P}$ , i.e.

$$\mathcal{P} |\overline{\alpha\beta}\rangle = |\overline{\alpha\beta}\rangle , \quad \langle \overline{\alpha\beta} | \overline{\sigma\gamma} \rangle = \delta_{\alpha\sigma} \delta_{\beta\gamma} . \quad (2.54)$$

Finally, we can write

$$\mathcal{P} = \sum_{\alpha, \beta=1}^{d_S} |\overline{\alpha\beta}\rangle \langle \overline{\alpha\beta}| = \frac{1}{d_E} \sum_{\alpha, \beta=1}^{d_S} \left( \sum_{i=1}^{d_E} |i\alpha, i\beta\rangle \right) \left( \sum_{j=1}^{d_E} \langle j\alpha, j\beta| \right) . \quad (2.55)$$

Since

$$\rho = \sum_{i,j=1}^{d_E} \sum_{\alpha,\beta=1}^{d_S} \rho_{j\beta}^{i\alpha} |i\alpha\rangle \langle j\beta| = \sum_{i,j=1}^{d_E} \sum_{\alpha,\beta=1}^{d_S} \rho^{i\alpha,j\beta} |i\alpha, j\beta\rangle , \quad (2.56)$$

we now have representations for both  $\mathcal{P}$  and  $\rho$ , which allows us to explicitly calculate  $\mathcal{P}\rho$  (with the help of  $\langle i\alpha, j\beta | p\sigma, q\nu \rangle = \delta_{ip}\delta_{jq}\delta_{\alpha\sigma}\delta_{\beta\nu}$ ) as

$$\mathcal{P}\rho = \frac{1}{\sqrt{d_E}} \sum_{\alpha,\beta=1}^{d_S} (\text{Tr}_E \rho)^{\alpha\beta} |\overline{\alpha\beta}\rangle = \sum_{\alpha,\beta=1}^{d_S} (\mathcal{P}\rho)^{\overline{\alpha\beta}} |\overline{\alpha\beta}\rangle , \quad (2.57)$$

where

$$(\mathcal{P}\rho)^{\overline{\alpha\beta}} = \frac{(\text{Tr}_E \rho)^{\alpha\beta}}{\sqrt{d_E}} . \quad (2.58)$$

Equation (2.58) defines an isomorphism between  $(\mathcal{H}_{S+E}^2)_{\mathcal{P}=1}$  and  $\mathcal{H}_S^2$  that allows us to calculate the trace over environmental variables by effectively doing the basis transformation (2.53).

The conclusion of the previous paragraph is that by working in the eigenbasis of  $\mathcal{P}$ , as one of the possible eigenbasis of  $\mathcal{H}_{S+E}^2$  (2.50), from the beginning we can avoid explicitly taking the trace over environmental variables at the end. In that eigenbasis, given by (2.53) and completed for  $(\mathcal{H}_{S+E}^2)_{\mathcal{P}=0}$  (details in [159]), the total density operator can be written as a  $d_S^2 d_E^2$ -dimensional column vector

$$\rho = \begin{bmatrix} \rho_1 \\ \rho_2 \end{bmatrix} , \quad (2.59)$$

where  $\rho_1$  is  $d_S^2$ -dimensional and  $\rho_2$  is  $d_S^2 (d_E^2 - 1)$ -dimensional, while the projection operators as  $d_S^2 d_E^2 \times d_S^2 d_E^2$  matrices

$$\begin{aligned} \mathcal{P} &= \begin{bmatrix} \mathbf{1}_{d_S^2 \times d_S^2} & \mathbf{0}_{d_S^2 \times d_S^2 (d_E^2 - 1)} \\ \mathbf{0}_{d_S^2 (d_E^2 - 1) \times d_S^2} & \mathbf{0}_{d_S^2 (d_E^2 - 1) \times d_S^2 (d_E^2 - 1)} \end{bmatrix} , \\ \mathcal{Q} &= \begin{bmatrix} \mathbf{0}_{d_S^2 \times d_S^2} & \mathbf{0}_{d_S^2 \times d_S^2 (d_E^2 - 1)} \\ \mathbf{0}_{d_S^2 (d_E^2 - 1) \times d_S^2} & \mathbf{1}_{d_S^2 (d_E^2 - 1) \times d_S^2 (d_E^2 - 1)} \end{bmatrix} . \end{aligned} \quad (2.60)$$

We see that  $\rho_S = \text{Tr}_E(\rho)$  is given just by (using (2.58))

$$\rho_S = \sqrt{d_E} \cdot \rho_1 . \quad (2.61)$$

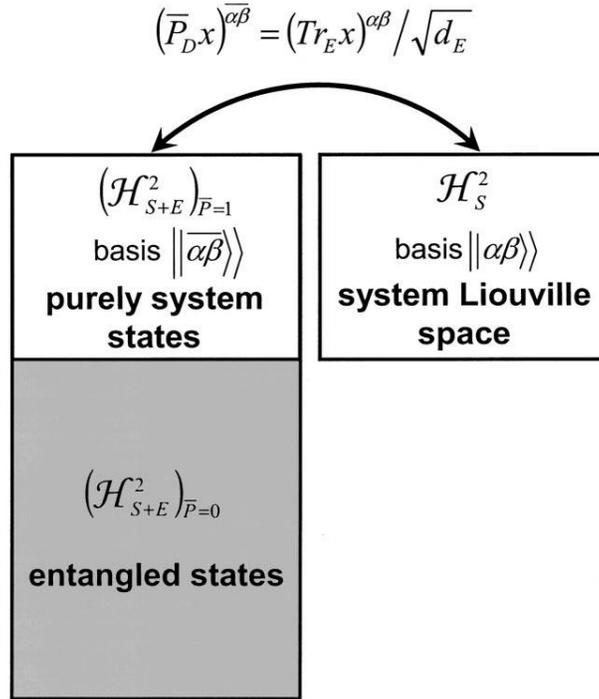


Figure 2.3 Decomposition of the total Liouville space  $\mathcal{H}_{S+E}^2$  into the subspaces of the projection operator  $\mathcal{P}$  and the isomorphism between the unit subspace  $(\mathcal{H}_{S+E}^2)_{\mathcal{P}=1}$  and  $\mathcal{H}_S^2$  for an operator  $x$  acting on  $\mathcal{H}_{S+E}$ . Reprinted with permission from [I. Knezevic and D. K. Ferry, Phys. Rev. A **69**, 012104 (2004)]. ©(2004) by the American Physical Society.

Similarly, any superoperator  $\mathcal{A}$  acting on  $\mathcal{H}_{S+E}^2$  is represented by

$$\mathcal{A} = \begin{bmatrix} \mathcal{A}_{11} & \mathcal{A}_{12} \\ \mathcal{A}_{21} & \mathcal{A}_{22} \end{bmatrix}. \quad (2.62)$$

Additionally, if an operator is a system operator, i.e.  $\mathcal{A}_{sys} = \mathbf{1}_E \otimes \mathcal{A}_S$ , then it commutes with  $\mathcal{P}$

$$\mathcal{P}\mathcal{A}_{sys}\rho = \bar{\rho}_E \otimes \text{Tr}_E[(\mathbf{1}_E \otimes \mathcal{A}_S)\rho] = \bar{\rho}_E \otimes \mathcal{A}_S \text{Tr}_E \rho \quad (2.63)$$

$$= (\mathbf{1}_E \otimes \mathcal{A}_S)(\bar{\rho}_E \otimes \text{Tr}_E \rho) = \mathcal{A}_{sys} \mathcal{P}\rho, \quad (2.64)$$

which means that it is block-diagonal in the eigenbasis of  $\mathcal{P}$ . Furthermore, it is easily shown that the upper left block matrix is just  $\mathcal{A}_S$  (see Appendix B of [160]), so that

$$\mathcal{A} = \begin{bmatrix} \mathcal{A}_S & \mathbf{0} \\ \mathbf{0} & \mathcal{A}_2 \end{bmatrix}. \quad (2.65)$$

The above mentioned isomorphism between  $(\mathcal{H}_{S+E}^2)_{\mathcal{P}=1}$  and  $\mathcal{H}_S^2$  and the decomposition of  $\mathcal{H}_{S+E}^2$  according to (2.50) are graphically shown in Fig. 2.3. Because of the isomorphism (2.58, 2.61) density matrices of the form

$$\rho = \begin{bmatrix} \rho_1 \\ \mathbf{0} \end{bmatrix} \quad (2.66)$$

are called “purely system states”, because they are completely determined by the state of the system  $S$  and depend on the environment only in an average sense (through the trace operation). On the other hand, density matrices for which  $\rho_1 = 0$  and  $\rho_2 \neq 0$  we call “entangled states” because they carry microscopic connections to the environmental states, beyond the point of easy separability like in the case of “purely system states”. This can be seen by explicitly deriving the part of the basis for  $(\mathcal{H}_{S+E}^2)_{\mathcal{P}=0}$ , with the help of the Gram-Schmidt procedure (see Appendix of [159] and, for more compact and explicit form, Appendix A of [2]).

### 2.2.5 A Partial–Trace–Free Equation of Motion

We proceed by writing the conventional time–convolutionless equation of motion from Sect. 2.2.3 in the basis of  $\mathcal{P}$  derived in Sect. 2.2.4. The Liouville operator and the time-evolution operator are given by the following block forms

$$\mathcal{L}(t) = \begin{bmatrix} \mathcal{L}_{11}(t) & \mathcal{L}_{12}(t) \\ \mathcal{L}_{21}(t) & \mathcal{L}_{22}(t) \end{bmatrix}, \quad \mathcal{U}(t, t') = \begin{bmatrix} \mathcal{U}_{11}(t, t') & \mathcal{U}_{12}(t, t') \\ \mathcal{U}_{21}(t, t') & \mathcal{U}_{22}(t, t') \end{bmatrix}. \quad (2.67)$$

The Liouville-von Neumann and equation for the time-evolution now have the following forms

$$\begin{aligned} \frac{d\rho_1}{dt} &= -i\mathcal{L}_{11}(t)\rho_1(t) - i\mathcal{L}_{12}(t)\rho_2(t), \\ \frac{d\rho_2}{dt} &= -i\mathcal{L}_{21}(t)\rho_1(t) - i\mathcal{L}_{22}(t)\rho_2(t) \end{aligned} \quad (2.68)$$

and

$$\begin{aligned}\rho_1(t) &= \mathcal{U}_{11}(t, t')\rho_1(t') + \mathcal{U}_{12}(t, t')\rho_2(t') , \\ \rho_2(t) &= \mathcal{U}_{21}(t, t')\rho_1(t') + \mathcal{U}_{22}(t, t')\rho_2(t') .\end{aligned}\quad (2.69)$$

The block matrix forms of  $\mathcal{G}$  and  $\mathcal{D}$  from (2.42) and (2.45) are

$$\mathcal{G}(t, t') = T_{\leftarrow} \exp \left( -i \int_{t'}^t ds \mathcal{Q} \mathcal{L}(s) \mathcal{Q} \right) = \begin{bmatrix} 1 & 0 \\ 0 & T_{\leftarrow} \exp \left( -i \int_{t'}^t ds \mathcal{L}_{22}(s) \right) \end{bmatrix}, \quad (2.70)$$

$$\begin{aligned}\mathcal{D}(t; 0) &= 1 + i \int_0^t dt' \begin{bmatrix} 1 & 0 \\ 0 & \mathcal{G}_{22}(t, t') \end{bmatrix} \begin{bmatrix} 0 & 0 \\ \mathcal{L}_{21}(t') & 0 \end{bmatrix} \begin{bmatrix} \mathcal{U}_{11}(t', t) & \mathcal{U}_{12}(t', t) \\ \mathcal{U}_{21}(t', t) & \mathcal{U}_{22}(t', t) \end{bmatrix} \\ &= \begin{bmatrix} 1 & 0 \\ i \int_0^t dt' \mathcal{G}_{22}(t, t') \mathcal{L}_{21}(t') \mathcal{U}_{11}(t', t) & 1 + i \int_0^t dt' \mathcal{G}_{22}(t, t') \mathcal{L}_{21}(t') \mathcal{U}_{12}(t', t) \end{bmatrix} .\end{aligned}\quad (2.71)$$

Since we need  $\mathcal{D}^{-1}(t; 0)$ , from (2.71) we obtain

$$\mathcal{D}^{-1}(t; 0) = \begin{bmatrix} 1 & 0 \\ -\mathcal{D}_{22}^{-1}(t; 0) \mathcal{D}_{21}(t; 0) & \mathcal{D}_{22}^{-1}(t; 0) \end{bmatrix} . \quad (2.72)$$

As a final step we use all previously defined block forms of necessary operators and superoperators, along with the equation of motion for  $\mathcal{P}\rho$  (2.47) and the isomorphism (2.58) to obtain

$$\begin{aligned}\frac{d\rho_S(t)}{dt} &= - i [\mathcal{L}_{11}(t) - \mathcal{L}_{12}(t) \mathcal{D}_{22}^{-1}(t; 0) \mathcal{D}_{21}(t; 0)] \rho_S(t) \\ &\quad - i \sqrt{d_E} \mathcal{L}_{12}(t) \mathcal{D}_{22}^{-1}(t; 0) \mathcal{G}_{22}(t, 0) \rho_2(0) .\end{aligned}\quad (2.73)$$

Equation (2.73) is a partial–trace–free time–convolutionless equation of motion for the reduced density matrix  $\rho_S(t)$ . It describes the evolution of the representation basis of  $\rho_S$ . Working with representation matrices is a necessary condition of this method and might help in the case when one is interested in numerical implementation. The increased transparency of working with representation forms may also help when introducing various approximations in the exact equation

of motion. Out of those three problems, mentioned at the end of Sect. 2.2.3, there is still one remaining. Namely, we still have the problem of evaluating the inverse of potentially large matrix  $\mathcal{D}_{22}^{-1}(t; 0)$  (if it exists at all). The solution to that problem will be discussed, among other things, in the next section.

## 2.2.6 Memory Dressing

Let us explicitly write the equations of motion for the density operator  $\rho$  in the eigenbasis of  $\mathcal{P}$  from the previous section, i.e. within the partial–trace–free approach. By using (2.47) and (2.46), or directly (2.73) for  $\rho_1$ , we obtain

$$\begin{aligned} \frac{d\rho_1(t)}{dt} &= -i [\mathcal{L}_{11}(t) - \mathcal{L}_{12}(t)\mathcal{D}_{22}^{-1}(t; 0)\mathcal{D}_{21}(t; 0)] \rho_1(t) \\ &\quad - i\mathcal{L}_{12}(t)\mathcal{D}_{22}^{-1}(t; 0)\mathcal{G}_{22}(t, 0)\rho_2(0) , \\ \rho_2(t) &= -\mathcal{D}_{22}^{-1}(t; 0)\mathcal{D}_{21}(t; 0)\rho_1(t) + \mathcal{D}_{22}^{-1}(t; 0)\mathcal{G}_{22}(t, 0)\rho_2(0) , \end{aligned} \quad (2.74)$$

where from (2.70) and (2.71) and by formally differentiating  $\mathcal{D}(t; 0)$ 's submatrices with respect to time we have

$$\begin{aligned} \mathcal{G}_{22}(t, 0) &= T_{\leftarrow} \exp \left( -i \int_0^t ds \mathcal{L}_{22}(s) \right) , \\ \frac{d\mathcal{D}_{21}(t; 0)}{dt} &= -i\mathcal{L}_{22}(t)\mathcal{D}_{21}(t; 0) + i\mathcal{D}_{21}(t; 0)\mathcal{L}_{11}(t) + i\mathcal{D}_{22}(t; 0)\mathcal{L}_{21}(t) , \\ \frac{d\mathcal{D}_{22}(t; 0)}{dt} &= -i\mathcal{L}_{22}(t)\mathcal{D}_{22}(t; 0) + i\mathcal{D}_{22}(t; 0)\mathcal{L}_{22}(t) + i\mathcal{D}_{21}(t; 0)\mathcal{L}_{12}(t) , \\ \mathcal{D}_{21}(0; 0) &= 0 , \quad \mathcal{D}_{22}(0; 0) = 1 , \end{aligned} \quad (2.75)$$

where in the last line the initial conditions are given. Taking the time derivative of the equation of motion for  $\rho_1(t)$  in (2.69) and comparing those two equations with (2.74) we obtain the following

relations for the representation of time evolution operator  $\mathcal{U}(t, 0)$

$$\begin{aligned}
\frac{d\mathcal{U}_{11}(t, 0)}{dt} &= -i [\mathcal{L}_{11}(t) - \mathcal{L}_{12}(t)\mathcal{D}_{22}^{-1}(t; 0)\mathcal{D}_{21}(t; 0)] \mathcal{U}_{11}(t, 0) , \\
\frac{d\mathcal{U}_{12}(t, 0)}{dt} &= -i [\mathcal{L}_{11}(t) - \mathcal{L}_{12}(t)\mathcal{D}_{22}^{-1}(t; 0)\mathcal{D}_{21}(t; 0)] \mathcal{U}_{12}(t, 0) \\
&\quad -i\mathcal{L}_{12}(t)\mathcal{D}_{22}^{-1}(t; 0)\mathcal{G}_{22}(t, 0) , \\
\mathcal{U}_{21}(t, 0) &= -\mathcal{D}_{22}^{-1}(t; 0)\mathcal{D}_{21}(t; 0)\mathcal{U}_{11}(t, 0) , \\
\mathcal{U}_{22}(t, 0) &= \mathcal{D}_{22}^{-1}(t; 0) [\mathcal{G}_{22}(t, 0) - \mathcal{D}_{21}(t; 0)\mathcal{U}_{12}(t, 0)] .
\end{aligned} \tag{2.76}$$

These are generic time–convolutionless equations of motions, the form of which results from using the specific basis within the partial–trace–free–approach. They have the general feature of time–convolutionless equations that  $\mathcal{U}_{21}$  and  $\mathcal{U}_{22}$  are expressed in terms of  $\mathcal{U}_{11}$  and  $\mathcal{U}_{12}$ . Formally, by solving Eqs. (2.76) (for which we first have to solve Eqs. (2.75)) we arrive at the final solution for the equation of motion of the reduced density operator  $\rho_S$ . However, this is a very difficult problem due to the sizes of the block matrices (the largest are at the position (2,2), being  $d_S^2(d_E^2 - 1) \times d_S^2(d_E^2 - 1)$ –dimensional) and because we need to evaluate the inverse of the matrix  $\mathcal{D}_{22}$  which is in turn the solution of coupled equations for  $\mathcal{D}_{21}$  and  $\mathcal{D}_{22}$ .

By inspection of (2.76) we see that we do not need all three large matrices  $\mathcal{G}_{22}$ ,  $\mathcal{D}_{21}$  and  $\mathcal{D}_{22}$  separately, but only the following combinations of them (we designate each of them with a new letter)

$$\begin{aligned}
\mathcal{R}(t) &= \mathcal{D}_{22}^{-1}(t; 0)\mathcal{D}_{21}(t; 0) , \\
\mathcal{S}(t; 0) &= \mathcal{D}_{22}^{-1}(t; 0)\mathcal{G}_{22}(t, 0) ,
\end{aligned} \tag{2.77}$$

where we left out the initial time in the argument list of  $\mathcal{R}(t; 0)$  for convenience. By using Eqs. (2.75) we can derive the equations of motion for the matrices  $\mathcal{R}$  and  $\mathcal{S}$

$$\begin{aligned}
\frac{d\mathcal{R}(t)}{dt} &= -i\mathcal{L}_{22}(t)\mathcal{R}(t) - i\mathcal{R}(t)\mathcal{L}_{12}(t)\mathcal{R}(t) + i\mathcal{R}(t)\mathcal{L}_{11}(t) + i\mathcal{L}_{21}(t) , \quad \mathcal{R}(0) = 0 ; \\
\frac{d\mathcal{S}(t; 0)}{dt} &= -i [\mathcal{L}_{22}(t) + i\mathcal{R}(t)\mathcal{L}_{12}(t)] \mathcal{S}(t; 0) , \quad \mathcal{S}(0; 0) = 1 .
\end{aligned} \tag{2.78}$$

Since we are really interested in the evolution of  $\rho_1$ , due to its direct connection with  $\rho_S$  via (2.61), we only need the time evolution matrices  $\mathcal{U}_{11}(t, 0)$  and  $\mathcal{U}_{12}(t, 0)$ . So, by starting from some initial

state  $\rho(0) = \begin{bmatrix} \rho_1(0) & \rho_2(0) \end{bmatrix}^T$ , we have a new system of equations completely describing the time evolution of the reduced density operator  $\rho_S$ , consisting of Eqs. (2.78) and

$$\begin{aligned} \frac{d\mathcal{U}_{11}(t, 0)}{dt} &= -i[\mathcal{L}_{11}(t) - \mathcal{L}_{12}(t)\mathcal{R}(t)]\mathcal{U}_{11}(t, 0), \quad \mathcal{U}_{11}(0, 0) = 1; \\ \frac{d\mathcal{U}_{12}(t, 0)}{dt} &= -i[\mathcal{L}_{11}(t) - \mathcal{L}_{12}(t)\mathcal{R}(t)]\mathcal{U}_{12}(t, 0) - i\mathcal{L}_{12}(t)\mathcal{S}(t; 0), \quad \mathcal{U}_{12}(0, 0) = 0. \end{aligned} \quad (2.79)$$

We see that by introducing  $\mathcal{R}(t)$  and  $\mathcal{S}(t; 0)$  there is no more problem with the cumbersome inverse matrix  $\mathcal{D}_{22}^{-1}(t; 0)$ . The equations for  $\mathcal{U}_{21}(t, 0)$  and  $\mathcal{U}_{22}(t, 0)$ , which we do not need here, but are sometimes important, for example in calculating two-time correlation functions in electronic transport where  $\mathcal{U}(t, t')$  for  $t' \neq 0$  are required [161–163], are

$$\mathcal{U}_{21}(t, 0) = -\mathcal{R}(t)\mathcal{U}_{11}(t, 0), \quad \mathcal{U}_{22}(t, 0) = \mathcal{S}(t; 0) - \mathcal{R}(t)\mathcal{U}_{12}(t, 0). \quad (2.80)$$

The concept of memory dressing from the title of this section refers to  $\mathcal{R}(t)$ . This is because  $\mathcal{R}(t)$  always goes along with  $\mathcal{L}_{12}(t)$ , which is the term representing physical interaction (as follows from the representation form (2.68)), in the “quasi-Liouvillian”  $\mathcal{L}_{11}(t) - \mathcal{L}_{12}(t)\mathcal{R}(t)$ . So, it is a memory dressing of the physical interaction. The self-contained non-linear equation of motion for the memory dressing  $\mathcal{R}(t)$  (first of Eqs. (2.78)) is a matrix Riccati equation, often encountered in control systems theory [164, 165]. It can be solved for  $\mathcal{R}$  to an arbitrary order by using the perturbation expansion, which also allows for a convenient diagrammatic representation [166].

### 2.3 Coarse-Graining for the Steady State Distribution Function

The purpose of this section is to derive the steady state distribution function for the open system, by solving for  $\rho_S(t)$  in a ballistic device (no scattering) that is attached to ideal contacts. We will show that, under these conditions, the distribution function is of Landauer-type. It says that the occupation of incoming states is fixed by the respective contact, while that of outgoing states by the open system alone. Furthermore, since there is no scattering in the open system, the occupation will remain the one determined by the contacts. We will use a coarse-graining procedure to approximate the exact non-Markovian time evolution towards the steady state. At the end,

an interaction Hamiltonian, suitable for ideal contacts, will be constructed and used to solve the approximate Markovian equation of motion.

### 2.3.1 The Exact Dynamics and the Coarse-Graining Procedure

By using Eqs. (2.61) and (2.77) in (2.74), we get the following form for the exact equation of motion for the reduced density matrix

$$\frac{d\rho_S(t)}{dt} = -i[\mathcal{L}_{11} - \mathcal{L}_{12}\mathcal{R}(t)]\rho_S(t) - i\sqrt{d_E}\mathcal{L}_{12}\mathcal{S}(t;0)\rho_2(0). \quad (2.81)$$

We will restrict our attention to the problems for which the initial density matrix is not correlated, i.e.

$$\rho(0) = \rho_E(0) \otimes \rho_S(0). \quad (2.82)$$

We see that when  $\rho_E(0) = \bar{\rho}_E$  then  $\rho_2(0) = 0$  and there exists a subdynamics ( $\rho_S$  does not depend on  $\rho_2(0)$ ). This is because  $\mathcal{P}$  is also generated by  $\bar{\rho}_E$ , so that  $\rho(0)$  is an eigenstate of  $\mathcal{P}$  and is of the form (2.66). Here, even though the environmental density matrix is not uniform, it can be proven that the following connecting relation holds

$$\rho_2(0) = \mathcal{M}\rho_1(0) = d_E^{-1/2}\mathcal{M}\rho_S(0), \quad (2.83)$$

where  $\mathcal{M}$  in the eigenbasis of  $\rho_E(0)$  is given by (see Appendix A of [2])

$$\mathcal{M}^i = \sqrt{\frac{d_E(d_E + 1 - i)}{d_E - 1}} \left( \rho_E^i(0) - \frac{1}{d_E + 1 - i} \sum_{j=1}^{d_E} \rho_E^j(0) \right). \quad (2.84)$$

So, in this more general case (for arbitrary  $\rho_E(0)$ ) there still exists the subdynamics in the following form

$$\rho_S(t) = [\mathcal{U}_{11}(t,0) + \mathcal{U}_{12}(t,0)\mathcal{M}]\rho_S(0) = \mathcal{W}(t,0)\rho_S(0), \quad (2.85)$$

which is in agreement with the statement made by Lindblad [167] that the subdynamics exists for an uncorrelated initial state. We can get a differential form of (2.85) by combining Eqs. (2.74) and (2.83)

$$\frac{d\rho_S(t)}{dt} = -i[\mathcal{L}_{11} - \mathcal{L}_{12}\mathcal{R}(t)]\rho_S(t) - i\mathcal{L}_{12}\mathcal{S}(t)\mathcal{M}\rho_S(0). \quad (2.86)$$

In general we can write

$$\mathcal{W}(t, 0) = T_{\leftarrow} \exp \left[ \int_0^t \mathcal{F}(s) ds \right] , \quad (2.87)$$

where  $\mathcal{F}(t)$  is the generator of  $\mathcal{W}(t, 0)$ .

It is very difficult to solve for the reduced system dynamics (2.85), because of the difficulties in obtaining  $\mathcal{W}(t, 0)$ . We can either be content with a Markovian approximation in the weak-coupling and van Hove limits [168], or by an expansion up to the second or fourth orders in the interaction if we need a non-Markovian approximation [8]. Although the weak-coupling limit has been used before to study tunneling structures in the Markovian approximation [169, 170], it is not generally applicable to nanostructures [169]. Here, we will apply an approximation beyond the weak-coupling limit, by approximating the exact reduced system dynamics using coarse-graining over the environmental relaxation time  $\tau$  [171, 172]. This limits the area of applicability to the open systems for which  $\tau \ll \tau_S$ , where  $\tau_S$  is the open system relaxation time, which is still a pretty wide area. For example, in typical small semiconductor devices (quasi-ballistic), with highly doped contacts at room temperature, the major energy relaxation mechanism is electron–electron scattering in the contacts (relaxation time for electron–electron scattering is about 10 fs for GaAs at  $10^{19} \text{ cm}^{-3}$  and room temperature [3], while about 150 fs for polar optical phonon scattering [136]). Electron–electron relaxation will drive the environmental distribution function to a drifted Fermi-Dirac distribution in a time interval  $\tau \approx 10 - 100$  fs, which is much shorter than the typical open system relaxation time for these devices  $\tau_S \approx 1 - 10$  ps.

The coarse-graining procedure proceeds by splitting the total evolution time interval  $[0, t]$  into segments of length  $\tau$ ,  $[1, 2, \dots, n] \times \tau$ , and defining the average of the generator of  $\mathcal{W}(t, 0)$  over each interval

$$\overline{\mathcal{F}}_j = \frac{1}{\tau} \int_{j\tau}^{(j+1)\tau} \mathcal{F}(s) ds . \quad (2.88)$$

This leads to the following connection between successive, discretized reduced density operators

$$\rho_{S,j+1} = \exp \left( \tau \overline{\mathcal{F}}_j \right) \rho_{S,j} , \quad (2.89)$$

which gives

$$\frac{\rho_{S,j+1} - \rho_{S,j}}{\tau} = \overline{\mathcal{F}}_j \rho_{S,j} \quad (2.90)$$

after expanding the exponent for small  $\tau$ . This is just a discretized version of the exact equation of motion.

There are three approximations applied in deriving (2.90). First, we do not have the information about the time evolution inside each  $\tau$ -interval, but only at its ends. Second, we cut the series after the first order in the expansion of  $\exp(\tau \overline{\mathcal{F}}_j)$  in order to get (2.90). Third, the time ordering from the exact equation (2.87) is violated at the  $\tau$  time scale, which can be shown explicitly by using the Dyson series to represent (2.87).

Finally, we will assume that the environmental state is nearly the same after every interval  $\tau$  during the transient, in other words  $\overline{\mathcal{F}}_0 = \overline{\mathcal{F}}_\tau \approx \overline{\mathcal{F}}_1 \approx \dots \approx \overline{\mathcal{F}}_n$ . This is also the most trivial way of ensuring that the coarse grained generators  $\overline{\mathcal{F}}_i$ 's commute (commute in an average sense). For this to be satisfied we have to ramp up the excitation (e.g. bias) to the system in small enough increments with sufficiently long time between two increments so that the open system is able to reach steady state, in the form of a drifted Fermi-Dirac distribution, after each small increment. This condition is more a thought experiment than a real constraint, because we are only interested in the steady state here. As the last step, we expand the discrete equation (2.90) to the continuum (since  $\tau$  is a small parameter) to obtain the final equation

$$\frac{d\rho_S}{dt} = \overline{\mathcal{F}}_\tau \rho_S(t) . \quad (2.91)$$

This equation is an approximate Markovian (because the generator  $\overline{\mathcal{F}}_\tau$  is constant in time) quantum master equation for the reduced density matrix in the limit of small-increments/long-pauses kind of ramping up the bias and we will use it to obtain the steady state distribution function for arbitrarily large bias.

### 2.3.2 The Short-Time Expansion of $\overline{\mathcal{F}}_\tau$

The practical value of (2.91) is in the fact that  $\overline{\mathcal{F}}_\tau$  can be calculated using the expansion of  $\mathcal{F}(t)$  in the small parameter  $\tau$  around zero. By introducing a definition  $\mathcal{F}(t) = -i\mathcal{L}_{\text{eff}} - \mathcal{G}(t)$ , where

$\mathcal{L}_{\text{eff}}$  is an effective system Liouvillian and  $\mathcal{G}$  a correction due to the system-environment coupling, expanding Eqs. (2.85) and (2.86) up to the second order in time and comparing coefficients it can be shown that (see Appendix B of [2])

$$\mathcal{F}(t) = -i\mathcal{L}_{\text{eff}} - 2\Lambda t + O(t^2) , \quad (2.92)$$

where  $\mathcal{L}_{\text{eff}}$  is a commutator superoperator generated by  $H_S + \langle H_{\text{int}} \rangle$ , while  $\Lambda$  in the basis  $\alpha\beta$  of the system's Liouville space is given by

$$\Lambda_{\alpha'\beta'}^{\alpha\beta} = \frac{1}{2} \left\{ \langle H_{\text{int}}^2 \rangle_{\alpha'}^{\alpha} \delta_{\beta}^{\beta'} + \langle H_{\text{int}}^2 \rangle_{\beta}^{\beta'} \delta_{\alpha'}^{\alpha} - 2 \sum_{j,j'} (H_{\text{int}})_{j\alpha'}^{j'\alpha} \rho_E^j (H_{\text{int}})_{j'\beta}^{j\beta'} - (\langle H_{\text{int}} \rangle^2)_{\alpha'}^{\alpha} \delta_{\beta}^{\beta'} + 2 \langle H_{\text{int}} \rangle_{\alpha'}^{\alpha} \langle H_{\text{int}} \rangle_{\beta}^{\beta'} - (\langle H_{\text{int}} \rangle^2)_{\beta}^{\beta'} \delta_{\alpha'}^{\alpha} \right\} , \quad (2.93)$$

where  $\rho_E^j$  are the eigenvalues of  $\rho_E(0)$  and  $\langle \dots \rangle \equiv \text{Tr}_E(\rho_E(0) \dots)$ .  $\Lambda$  contains important information on the directions of coherence loss. It has been implicitly defined previously [173], but only in the interaction (not Schrödinger) picture and for  $\langle H_{\text{int}} \rangle = 0$ .

If the following condition holds

$$\|\Lambda\|\tau \ll \|\mathcal{L}_{\text{eff}}\| , \quad (2.94)$$

then the short-time expansion of  $\mathcal{F}$  is valid and

$$\overline{\mathcal{F}}_{\tau} = -i\mathcal{L}_{\text{eff}} - \Lambda\tau , \quad (2.95)$$

which gives the final coarse-grained Markovian quantum master equation for the reduced density matrix

$$\frac{d\rho_S(t)}{dt} = (-i\mathcal{L}_{\text{eff}} - \Lambda\tau) \rho_S(t) . \quad (2.96)$$

We have already said that this coarse-grained Markovian approximation is valid if the environmental relaxation time  $\tau$  is much smaller than the system relaxation time (corresponding to  $1/\|\Lambda\|\tau$ ), or

$$\|\Lambda\|\tau^2 \ll 1 , \quad (2.97)$$

which, along with (2.94), gives in total

$$\|\Lambda\|\tau^2 \ll \min \{1, \|\mathcal{L}_{\text{eff}}\|\tau\} . \quad (2.98)$$

### 2.3.3 Steady State in a Two-Terminal Ballistic Nanostructure

In this section we will apply the main equation (2.96) to calculate the steady state distribution function for a two-terminal ballistic nanostructure attached to ideal and non-ideal contacts. By ideal contacts we mean contacts that behave like black bodies with respect to the emission/absorption of electrons. Therefore, they absorb all electrons coming from the open system. The consequence is that, as already mentioned, the occupation of states coming from the contacts is fixed by them, while the occupation of outgoing states is fixed by the open system. This gives a Landauer-type distribution function and specifically here, since the open system region is ballistic, the occupation of the incoming and outgoing states is the same and fixed by the injecting contact. In the case of non-ideal contacts the distribution function will be more complicated and will depend explicitly on the transmission and reflection coefficients.

#### 2.3.3.1 The Open System Model

Schematic of our two-terminal nanostructure is shown in Fig. 2.4. The device is biased negatively such that the negative polarity is at the left contact. All open system eigenenergies  $\epsilon_k$  above the bottom of the left contact have two eigenfunctions (double-degeneracy), one for the positive wave vector ( $\psi_k$ , injected from the left contact) and one for the negative wave vector ( $\psi_{-k}$ , injected from the right contact). The rest of the energy levels, made up of quasibound states that lay between the bottoms of the two contacts, have only one wave function for the states injected from the right contact and completely reflected. For doubly-degenerate scattering states we have the following asymptotic wave functions (assuming that the active region between  $x_L$  and  $x_R$  is wide enough)

$$\begin{aligned} \psi_k(x) &= \begin{cases} e^{ikx} + r_{-k,L}e^{-ikx}, & x < x_L \\ t_{k',L}e^{ik'x}, & x > x_R \end{cases}, \\ \psi_{-k}(x) &= \begin{cases} e^{-ik'x} + r_{k',L}e^{ik'x}, & x > x_R \\ t_{-k,L}e^{-ikx}, & x < x_L \end{cases}, \end{aligned} \quad (2.99)$$

where  $t$  and  $r$  are the transmission and reflection coefficients, respectively, while  $k$  and  $k'$  are the wave vectors for the same energy level  $\epsilon_k$  measured from the bottom of the left and right contacts, respectively.

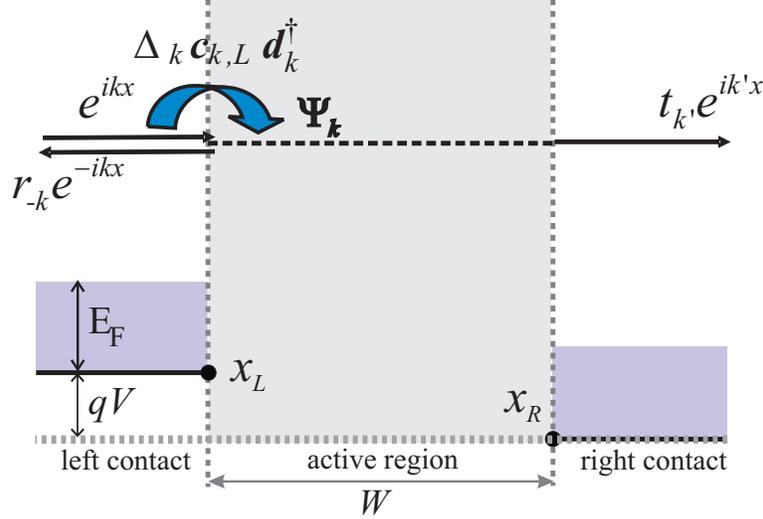


Figure 2.4 Schematic of the two-terminal ballistic nanostructure, negatively biased at the left contact, with the boundaries between the open system and contacts shown at  $x_L$  and  $x_R$ , and with the graphical representation of the wave function injected from and the hopping type interaction with the left contact. It is similar for the wave function and interaction for the right contact.

In the formalism of second quantization the non-interacting many-body Hamiltonian of the open system is given by (considering only scattering states in the following)

$$H_S = \sum_{k>0} \omega_k \left( d_k^\dagger d_k + d_{-k}^\dagger d_{-k} \right), \quad (2.100)$$

where  $\omega_k = \epsilon_k/\hbar$  and  $d_{\pm k}$  and  $d_{\pm k}^\dagger$  are the destruction and creation operators, respectively, for the open system states  $\psi_{\pm k}$ . The many-body effect that this Hamiltonian is able to model is the Pauli exclusion principle. Considering that the contacts are ideal, as explained above, the interaction Hamiltonian is modeled as a one-way coupling (for the particles that are injected only), while in the other way the electrons are free to propagate, because the contacts have ideal absorption characteristics. In other words, we do not have to enforce the Pauli exclusion principle (to “make

room”) by explicitly creating one electron in the contacts after destroying it in the open system region. So, the interaction Hamiltonians are given by

$$\begin{aligned} H_{\text{int}}^L &= \sum_{k>0} \Delta_k d_k^\dagger c_{k,L} + \text{h.c.} , \\ H_{\text{int}}^R &= \sum_{k>0} \Delta_{-k} d_{-k}^\dagger c_{-k,R} + \text{h.c.} , \end{aligned} \quad (2.101)$$

where  $c_{\pm k,L/R}$  and  $c_{\pm k,L/R}^\dagger$  are the destruction and creation operators, respectively, for the  $\pm k$  states in the left/right contact and the injection rates are given by

$$\Delta_k = \frac{\hbar k}{m \|\psi_k\|^2} , \quad \Delta_{-k} = \frac{\hbar k'}{m \|\psi_{-k}\|^2} , \quad (2.102)$$

where  $\|\psi_k\|^2 = \int_{x_L}^{x_R} |\psi_k(x)|^2 dx$ .

In Fig. 2.4 there are only  $H_{\text{int}}^L$  and  $\psi_k$  graphically represented, but the situation is similar for the right contact.

### 2.3.3.2 Steady State Distribution Functions

Since the interaction Hamiltonians (2.101) are linear in the contact creation and destruction operators, we conclude that  $\langle H_{\text{int}}^{L/R} \rangle = 0$ , which gives us the following equations

$$\begin{aligned} \mathcal{L}_{\text{eff}} &= \mathcal{L}_S , \\ (\Lambda^{L/R})_{\alpha'\beta'}^{\alpha\beta} &= \frac{1}{2} \left[ \langle (H_{\text{int}}^{L/R})^2 \rangle_{\alpha'}^\alpha \delta_{\beta'}^{\beta} + \langle (H_{\text{int}}^{L/R})^2 \rangle_{\beta'}^{\beta} \delta_{\alpha'}^\alpha \right] \\ &\quad - \sum_{j,j'} (H_{\text{int}}^{L/R})_{j\alpha'}^{j'\alpha} \rho_{L/R}^j (H_{\text{int}}^{L/R})_{j'\beta}^{j\beta'} . \end{aligned} \quad (2.103)$$

The quantities that we need to evaluate first are (for the left contact)

$$\langle (H_{\text{int}}^L)^2 \rangle = \sum_{k>0} \Delta_k^2 \left[ f_k^L d_k d_k^\dagger + (1 - f_k^L) d_k^\dagger d_k \right] , \quad (2.104)$$

which gives a contribution of the form  $\Lambda_{\alpha\beta}^{\alpha\beta}$ , and

$$\sum_{j,j'} (H_{\text{int}}^L)_{j\alpha'}^{j'\alpha} \rho_L^j (H_{\text{int}}^L)_{j'\beta}^{j\beta'} = \sum_{k>0} \Delta_k^2 \left[ (1 - f_k^L) (d_k^\dagger)_{\beta'}^{\beta} (d_k)_{\alpha'}^\alpha + f_k^L (d_k)_{\beta'}^{\beta} (d_k^\dagger)_{\alpha'}^\alpha \right] , \quad (2.105)$$

which gives a contribution of the form  $\Lambda_{\beta\beta}^{\alpha\alpha}$ . It is similar for the right contact.

Quantities  $f_{\pm k}^{L/R}$  define the drifted Fermi-Dirac distribution function in the contacts, as a consequence of the current flowing through the device. They take into account the feedback of the device under applied bias on the contacts and ensure that the charge neutrality and current continuity are satisfied near the device/contacts boundaries [1, 93]. The left contact distribution function is given by

$$f_{\pm k}^L = \frac{1}{\exp\left\{\frac{\hbar^2[(\pm k - k_d)^2 - k_F^2]}{2mk_B T}\right\} + 1}, \quad (2.106)$$

where  $k_d$  is the drift wave vector. Here, there is a common  $k_d$  for both contacts (since they carry the same current), but in a more general multi-terminal case there will be one drift wave vector for each contact, defined by the current density  $J_l$  through the  $l$ -th contact by  $k_d^l = mJ_l/n_lq\hbar$ , where  $n_l$  is the charge density of the  $l$ -th contact. This is an additional parameter that has to be determined self-consistently, by an additional equation  $J_l^{dev} = J_l^{contact}$  that ensures the current continuity across the device/contacts boundaries (but not on a state-by-state basis). Here,  $J_l^{dev}$  is the current density due to the injection from the  $l$ -th contact only. There is a similar equation to (2.106) for the right contact with the following changes:  $k \rightarrow k'$  in the denominator and  $L \rightarrow R$ . Detailed Monte Carlo–molecular dynamics simulations in bulk semiconductors show that when electron-electron scattering is the dominant relaxation mechanism the distribution function is very close to the one given by (2.106) [3–5].

Since  $\Lambda = \sum_k \Lambda_k$ , according to (2.104) and (2.105), it is just a sum of independent modes. Each mode can be represented with a two-state basis: one state for a particle being in the state  $\psi_k$  (“+” state) and another state for a particle being absent from it (“-” state). The reduced density matrix in this basis is a column vector with 4 elements,  $\rho_{S,k} = (\rho_{S,k}^{++}, \rho_{S,k}^{+-}, \rho_{S,k}^{-+}, \rho_{S,k}^{--})^T$ , and the equation of motion is

$$\frac{d\rho_{S,k}}{dt} = [-i\mathcal{L}_{S,k} - \Lambda_k\tau] \rho_{S,k}, \quad (2.107)$$

where

$$\mathcal{L}_{S,k} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 2\omega_k & 0 & 0 \\ 0 & 0 & -2\omega_k & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \quad (2.108)$$

$$\Lambda_k = \begin{bmatrix} A_k & 0 & 0 & -B_k \\ 0 & C_k & 0 & 0 \\ 0 & 0 & C_k & 0 \\ -A_k & 0 & 0 & B_k \end{bmatrix}. \quad (2.109)$$

Quantities  $A_k = \Delta_k^2 (1 - f_k^L)$ ,  $B_k = \Delta_k^2 f_k^L$  and  $C_k = (A_k + B_k) / 2 = \Delta_k^2 / 2$  are calculated using (2.104) and (2.105).

The elements  $\rho_{S,k}^{+-}$  and  $\rho_{S,k}^{-+}$  are zero in the steady state, because they decay as  $\exp(\mp i2\omega_k - \tau C_k) t$ . The two remaining elements,  $\rho_{S,k}^{++} = f_k(t)$  and  $\rho_{S,k}^{--} = 1 - f_k(t)$ , give the following equation

$$\frac{f_k(t)}{dt} = -\tau (A_k + B_k) f_k(t) + \tau B_k = -\tau \Delta_k^2 f_k(t) + \tau \Delta_k^2 f_k^L, \quad (2.110)$$

where  $f_k(t)$  is the distribution function of  $+k$  states in the open system region. In the steady state, this gives just

$$\begin{aligned} f_k^\infty &= f_k^L, \\ f_{-k}^\infty &= f_{-k'}^R, \end{aligned} \quad (2.111)$$

where  $f_{-k}^\infty$  is the steady state distribution function for  $-k$  states in the open system region, which can be derived in a similar way, starting by evaluating  $\langle (H_{\text{int}}^R)^2 \rangle$  and  $\sum_{j,j'} (H_{\text{int}}^R)_{j\alpha'}^{j'\alpha} \rho_R^j (H_{\text{int}}^R)_{j'\beta}^{j\beta}$ . We see that the result is the distribution function for the scattering states of the ballistic open system determined by the injecting contact only, which is what it should be considering the problem that we were solving.

### 2.3.3.3 Non-ideal Contacts

Contacts that are not ideal will not absorb electrons coming from the active region ideally. Therefore, we have to explicitly include interaction terms that model the extraction of outgoing

electrons by destroying one electron in the active region and recreating it in the contacts (for details of derivation see Ref. [2]). In this case the interaction Hamiltonian matrix elements are shown in Fig. 2.5.

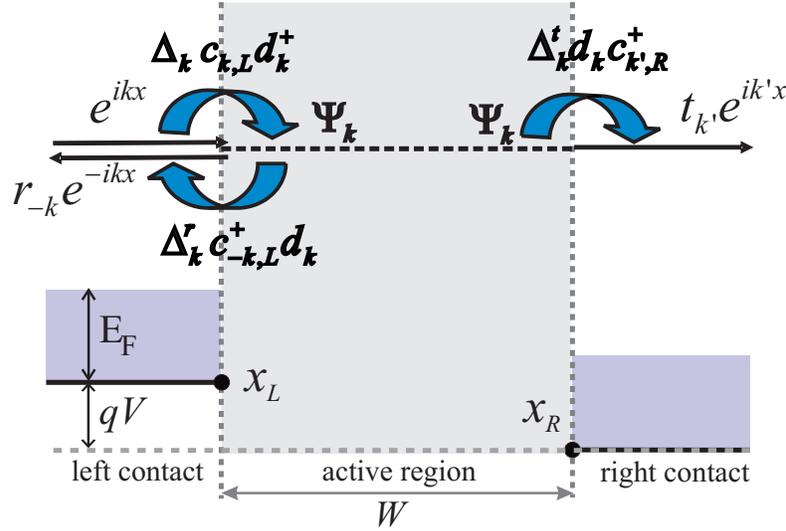


Figure 2.5 Schematic of the two-terminal ballistic nanostructure with non-ideal contacts, negatively biased at the left contact, with the boundaries between the open system and contacts shown at  $x_L$  and  $x_R$ , and with the graphical representation of the wave function injected from the left contact and the hopping type interaction between the left contact and the active region due to injection and between the active region and both contacts due to extraction. It is similar for the wave function and interaction for the right contact. Reprinted with permission from [I. Knezevic, Phys. Rev. B **77**, 125301 (2008)]. ©(2008) by the American Physical Society.

In this case the time evolution of the active region distribution function will be given by [2]

$$\frac{f_k(t)}{dt} = -\tau (A_k + B_k) f_k(t) + \tau B_k, \quad (2.112)$$

where  $A_k = \Delta_k^2 \{ (1 - f_k^L) + R_k^2 (1 - f_{-k}^L) + T_k^2 (1 - f_{k'}^R) \}$ ,  $B_k = \Delta_k^2 \{ f_k^L + R_k^2 f_{-k}^L + T_k^2 f_{k'}^R \}$  and  $R_k$  and  $T_k$  are reflection and transmission coefficients, respectively. It is similar for the injection from the right contact. In steady state Eq. (2.112) will give

$$f_k^\infty = \frac{f_k^L + R_k^2 f_{-k}^L + T_k^2 f_{k'}^R}{1 + R_k^2 + T_k^2}, \quad (2.113)$$

and similarly for the right contact injection

$$f_{-k}^{\infty} = \frac{f_{-k'}^L + R_k^2 f_{k'}^L + T_k^2 f_{-k}^R}{1 + R_k^2 + T_k^2}, \quad (2.114)$$

## Chapter 3

### Quantum transport in the transient regime

In nanoscale, quasiballistic electronic systems under bias, the process of relaxation towards a nonequilibrium steady state cannot be attributed to scattering, because these structures are small compared to the carrier mean free path [136, 174]. Rather, the active region of a nanostructure is an open quantum-mechanical system that exchanges particles and information with the dissipative reservoirs of charge, usually referred to as contacts [1, 93]. While qualitatively clear, a quantitative description of the irreversible evolution of the electronic system in this regime, where dissipation in the contacts coupled with the carrier exchange between the active region and contacts is the mechanism governing relaxation, is very challenging [175–177]. In this chapter, we treat the transient quantum transport in quasiballistic semiconductor nanostructures, which is based on the open system theory and valid on timescales much longer than the characteristic relaxation time in the contacts. The approach relies on the results of Ch. 2, specifically Sec. 2.3, where a model interaction between the current-limiting active region and ideal contacts is used to derive the Landauer-type steady state distribution function. Here, we derive a non-Markovian master equation for the irreversible time evolution of the active region’s distribution functions by introducing a time-dependent drift wave vector  $k_d(t)$ . In order to obtain the response quantities of a nanostructure under bias, such as the potential and the charge and current densities, the non-Markovian master equation must be solved numerically together with the Schrödinger, Poisson, and current continuity equations. We discuss how to numerically solve this coupled system of equations and illustrate the approach on the example of a silicon *nin* diode.

This chapter is organized as follows: In Sec. 3.1, using the results of Sec. 2.3, we derive the time evolution equations for the device, or active region, distribution functions and explain

their properties. The main point here is the introduction of the time-dependent drift wave vector  $k_d(t)$  that brings about the non-Markovian dynamical features. In Sec. 3.2 we couple the time dependent distribution functions with the Schrödinger, Poisson and current continuity equations and explain the algorithm used to find the solution to this system of equations, with the properly established initial conditions and scattering enhanced injection into the localized device states. The solution ultimately gives the time-dependent charge density, current density, and potential, which we illustrate using an *nin* diode in Sec. 3.3.

### 3.1 The time evolution of distribution functions

We focus here on the case of ideal contacts, explained in Sec. 2.3. Strictly speaking, the time-evolution of distribution functions given by Eq. (2.110) is valid only for a short time interval after applying the bias and not for the entire transient duration. This is because the generator  $\overline{\mathcal{F}}$  is derived approximately by using a short time expansion around  $t = 0$  s, which is the moment at which the excitation is applied or changed from the previous value (Sec. 2.3.2). However, to obtain the steady state distribution functions at some applied bias  $V = V_a$ , we can look at the time-evolution given by Eq. (2.110) upon applying a vanishing increment  $\Delta V \rightarrow 0$  to the applied bias. The small increment  $\Delta V$  requirement during the transient is necessary in order to ensure that the contact distribution function snaps back to approximately the same drifted Fermi-Dirac distribution it had at the beginning of the transient, when  $V = V_a$ . This corresponds to the small total change in the current flowing through the contacts during the transient, which is important for our assumption of memoryless contacts from Sec. 2.3.1 – the coarse-grained generators  $\overline{\mathcal{F}}_j$  are constant in each contact relaxation time  $\tau$  interval during a transient caused by the  $\Delta V$  potential increment.

Can we expand the constraint of using Eq. (2.110) with only small increments, to finite bias changes on the order of a couple of tens of *meV*? Knowing that the contact relaxation time  $\tau$  is much smaller than the transient duration, such that the time-dependent current is approximately constant during  $\tau$ , we can say that the reservoirs have enough time to relax in each  $\tau$  interval, but they relax to a new value of the drifted Fermi-Dirac distribution function as the current flowing

through them is changing. So, it is a good approximation to use the short time expansion for each  $\overline{\mathcal{F}}_j$ , where  $j$  indexes different  $\tau$  intervals, but with a different drifted Fermi-Dirac contact distributions, reflecting different values of current for different  $j$ . This means that we work beyond the approximation where all  $\overline{\mathcal{F}}_j$  are equal to some constant  $\overline{\mathcal{F}}_\tau$ , and by using Eq. (2.90) we get

$$\frac{d\rho_S}{dt} = \overline{\mathcal{F}}_\tau(t)\rho_S(t), \quad (3.1)$$

The time-evolution described by Eq. (3.1) is by definition non-Markovian (with memory) [8], because the generator of the time-evolution,  $\overline{\mathcal{F}}_\tau$ , is now time-dependent through the time dependent drift of the contact Fermi-Dirac distribution.

As we already mentioned in the introduction, the parameter that is proportional to the semi-classical current in the contacts, and at the same time is used to make the Fermi-Dirac distribution drifted, is the drift wavevector  $k_d$ . In our model, described in the previous paragraph,  $k_d$  will be time-dependent, so that the final time-evolution equation for the active region right-propagating states distribution function will be given by

$$\frac{df_k}{dt} = -\tau(A_k + B_k)f_k + \tau B_k = -\tau\Delta_k^2 f_k + \tau\Delta_k^2 f_k^L(t), \quad (3.2a)$$

An analogous relationship holds for the left-propagating states:

$$\frac{df_{-k}}{dt} = -\tau\Delta_{-k}^2 f_{-k} + \tau\Delta_{-k}^2 f_{-k'}^R(t), \quad (3.2b)$$

where  $k' = \sqrt{k^2 + 2mV_a/\hbar^2}$  for a single doubly degenerate device state, and  $V_a$  is the value of applied bias. The wave vector  $k$  is measured from the left contact conduction band bottom, while the wave vector  $k'$  from the right contact conduction band bottom.  $\Delta_k$  and  $\Delta_{-k}$  can also turn out to be time-dependent, depending on the numerical basis set used to solve the Schrödinger equation in the device region.

Equations (3.2) must now be self-consistently solved with the Schrödinger, Poisson, and current continuity equations, because they are coupled through  $k_d(t)$  and injection rates. To check certain properties of  $f_{\pm k}$ , we should first investigate a low-bias regime: (1) the potential and thus the scattering states, transmission coefficients, and the injection rates  $\Delta_{\pm k}$  are virtually constant

throughout the transient, and (2) the current density is low, so any changes to the contact distribution functions that result from the current flow can also be neglected, making Eqs. (3.2) Markovian. Therefore, we can solve the above equations analytically in the limit of low bias and low current densities. The solution to Eqs. (3.2) can be found as

$$f_k(t) = (f_k(0) - f_k^L) e^{-\tau \Delta_k^2 t} + f_k^L, \quad f_{-k}(t) = (f_{-k}(0) - f_{-k}^R) e^{-\tau \Delta_{-k}^2 t} + f_{-k}^R. \quad (3.3)$$

As expected, the steady-state values of the distribution functions are the contact distribution functions

$$f_k(\infty) = f_k^L, \quad f_{-k}(\infty) = f_{-k}^R. \quad (3.4)$$

Equations (3.3), can tell us about the dependence of the device transient time on the relaxation time in the contacts through  $\tau$ , as shown in Fig. 3.1. We see that the longer the  $\tau$ , the shorter the transient time, and vice versa. This means that with more scattering in the contacts one can expect a longer transient time. Although it seems counterintuitive, the explanation can be found in the following: the steady state current is not limited by the scattering in the contacts, but by the potential profile in the ballistic device region. The contacts must supply the current required by the device region, but they cannot do that fast enough if electrons in them undergo a lot of scattering.

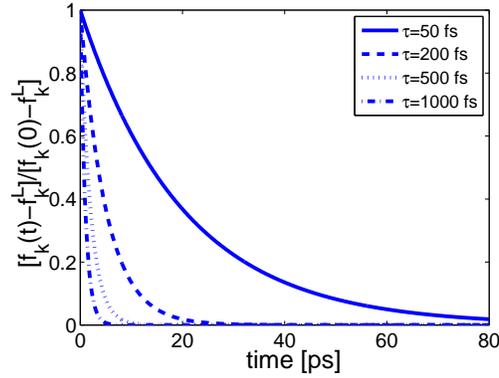


Figure 3.1 The difference between the right-propagating device and left contact distribution functions,  $f_k(t) - f_k^L$ , normalized by  $f_k(0) - f_k^L$ . The difference should decay to zero in the steady state, showing that a longer contact relaxation time  $\tau$  leads to a shorter transient regime. It is similar for the left-propagating device states. The injection rate for this specific  $k$ -state is assumed to be  $\Delta_k = 1$  ps.

## 3.2 The transient regime modeling in nanoscale devices

### 3.2.1 The system of coupled equations

As mentioned in the previous section, there is a system of four coupled equations that has to be solved in order to obtain the time-dependent quantities of interest: the charge density  $n(\mathbf{r}, t)$ , the current density  $J(t)$ , and the potential profile  $V(\mathbf{r}, t)$ . Below, these equations are given in their form independent of the problem dimensionality:

$$\nabla (\epsilon \nabla V) = e^2 [N_D - n] , \quad (3.5a)$$

$$-\frac{\hbar^2}{2} \nabla \left( \frac{1}{m^*} \nabla \psi_{\pm k} \right) + V \psi_{\pm k} = E_{|k|} \psi_{\pm k} , \quad (3.5b)$$

$$\frac{df_{\pm k}}{dt} = -\tau \Delta_{\pm k}^2 f_{\pm k} + \tau \Delta_{\pm k}^2 f_{k, -k'}^{L,R}(t) , \quad (3.5c)$$

$$J_{con} = J_{dev} - \frac{dQ}{dt} . \quad (3.5d)$$

Equation (3.5a) is the Poisson equation, where the device charge is represented by electron  $n$  and donor  $N_D$  concentrations only, the acceptor concentration being neglected.  $V$  is the conduction

band bottom potential energy in the device region, called here the potential or potential profile, while the permittivity  $\epsilon$  is in general a function of the coordinate. The absolute value of electron charge is represented by  $e$ . Eq. (3.5b) is the Schrödinger equation, with electron wave functions  $\psi_{\pm k}$ , total electron energies  $E_{|k|}$ , and an effective mass  $m^*$ , which is in general spatially dependent. The wave vectors for scattering states  $\psi_{\pm k}$  correspond to the wave vectors in the equation for the time-dependent distribution functions, Eq. (3.5c). Solving the Schrödinger and Poisson equations self-consistently is sometimes called the Hartree approximation, and means that even though the basis set is single-particle, we do include the mutual interaction between different particles through a mean field. Finally, the current continuity is given by Eq. (3.5d), and ensures current continuity across the contact-device interface. Any change in the total charge stored in the device during the transient is given by  $dQ/dt$ . The two terms on the right-hand side of Eq. (3.5d) do not depend on position, being governed by the current conserving Schrödinger equation. Therefore, this algebraic form of the continuity equation is sufficient, since the only place where the current can change is at the contact-device interface.

Let us now look at how Eqs. (3.5) are coupled. In order to solve the Poisson equation we need to know the electron density  $n$ . This electron density will depend on the distribution functions  $f_{\pm k}(t)$  and the wave functions  $\psi_{\pm k}$ . For example, in devices which allow 1D representation, the other two dimensions being integrated out due to spatial uniformity, the electron density is given by:

$$n(x, t) = \sum_k f_k(t) |\psi_k(x)|^2, \quad (3.6)$$

where the wave vector  $k$  can take both positive and negative values. First, we have to solve Eq. (3.5b) in order to obtain the wave functions. Next, to obtain  $f_{\pm k}$ , we need to solve Eq. (3.5c), where the contact distribution functions are given by:

$$f_k^L = \frac{m^* k_B T}{\pi \hbar^2} \ln \left( 1 + e^{\frac{E_F - E_k - k_d}{kT}} \right), \quad f_{-k'}^R = \frac{m^* k_B T}{\pi \hbar^2} \ln \left( 1 + e^{\frac{E_F - E_{-k'} - k_d}{kT}} \right), \quad (3.7)$$

where  $E_F$  is the right contact Fermi level, which is grounded, and  $E_k = \hbar^2 k^2 / 2m^*$ . The effective mass  $m^*$  is assumed to be equal in all directions in Eqs. (3.7). Furthermore, we need  $k_d$ , which is found from Eq. (3.5d), based on the necessity for the current continuity across the contact-device

interface. Equation (3.5d) is coupled to Eq. 3.5b through the device current density,  $J_{dev}$ , which is calculated from the wave functions using:

$$J_{dev}(t) = \frac{e\hbar}{m} \sum_k f_k(t) \text{Im} \left[ \psi_k^* \frac{\partial \psi_k}{\partial x} \right], \quad (3.8)$$

and through the changes in the total device charge, where the device charge is given by integrating the charge density over the device domain,  $Q = e \int_0^L [N_D - n(x)] dx$ ,  $L$  being the device length.

In the following, we will discuss a numerical algorithm that can be used to solve Eqs. (3.5). Figure 3.2 shows a block diagram with relative positions and connections between different equations. The solution proceeds by iterations which should ideally have good convergence on the potential  $V$  and drift wave vector  $k_d$ . Iterations start by assuming some initial guess for  $V$  and solving the Schrödinger equation, to find energies and wave functions. Given this and equations for electron and current densities, Eqs. (3.5c) and (3.5d) are solved self-consistently to find  $k_d$ . At this point we have electron density at the output, which we use to solve the Poisson equation and obtain a new guess for potential  $V$ . The iterations are stopped when a good convergence is achieved on both  $V$  and  $k_d$ . As is clear at this point, there are two nested loops. The outer loop is used to solve for  $V$ , using the globally convergent Newton method, while the inner loop, used to solve for  $k_d$ , can be solved by a simpler mixing method, where the previous and new solutions are mixed with a suitable damping factor, to obtain the next guess. More details about the algorithm, discretization of each equation into a matrix form, and other important points are discussed in Appendix A.

### 3.2.2 The initial conditions

Before entering the iterative loop explained in the previous section, it is important to establish initial conditions on the potential  $V$  and distribution functions in the device region  $f_{\pm k}(0)$ . The initial condition is the one that exists at  $t = 0^+$  s, which is the moment just after applying bias and before any current flow in the active region. First, the size of the device region is found by requiring that the potential in steady state is sufficiently flat near the contact-device interface. Of course, one has to ensure that this size is comparable or smaller than the mean free path in

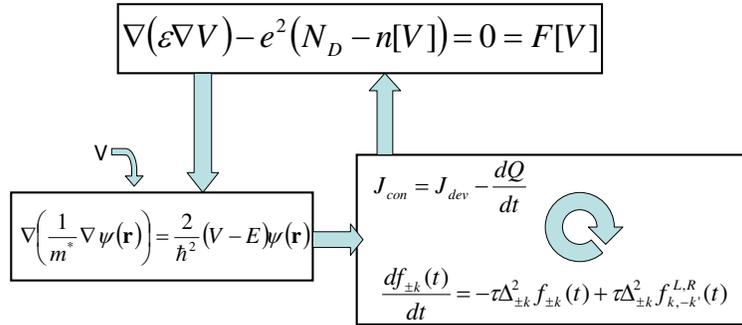


Figure 3.2 A block diagram showing the overall coupling between different equations used to calculate the transient regime. Starting with an initial guess for  $V$ , first the Schrödinger equation is solved to obtain the energies and wave functions for scattering states (lower left block). With this the lower right block is solved iteratively until the convergence of the drift wave vector  $k_d$  is achieved. At that point we have the charge density  $n$ , which serves as the input to the upper block, the Poisson equation. The solution to the Poisson equation yields a new guess for  $V$  and the whole process is repeated, until the convergence of the outer loop is achieved. The outer loop is solved by the globally convergent Newton method designed to find the root of the functional  $F[V]$ . The inner loop uses simple mixing of the previous and current  $k_d$  solutions to find the next guess, with the use of a suitable damping coefficient.

the contacts, granting the representation of injected particles as plane waves. Finding the size can be done by first performing a steady state calculation, which is much faster. In steady state Eq. (3.5c) becomes trivial and gives distribution functions of Landauer-type, as explained in Sec. 2.3. Once we know the position of the contact-device interfaces we assume that after applying the bias to the contacts, the perturbation will propagate through them until it reaches the device region. This happens at  $t = 0^+$  s, and it means that there is a potential difference between the two contact-device interfaces equal to the applied bias. However, since there is still no current flow, the device distribution functions at this moment are equal to the values they had in equilibrium, before

applying the bias. So we can write for the distribution functions:

$$f_k(0^+) = \frac{m^*k_B T}{\pi\hbar^2} \ln \left( 1 + e^{\frac{E_F - E_k|_{equil}}{kT}} \right), \quad f_{-k}(0^+) = \frac{m^*k_B T}{\pi\hbar^2} \ln \left( 1 + e^{\frac{E_F - E_{-k'}|_{equil}}{kT}} \right). \quad (3.9)$$

Note that  $k_d(0^+) = 0$ , since there is no current flow. In other words,  $f_k(0^+) = f_k^L|_{equil}$  and  $f_{-k}(0^+) = f_{-k'}^R|_{equil}$ . Here, we assume identical contacts, with aligned conduction band bottoms, and so  $f_k(0^+) = f_{-k}(0^+)$ , since then in equilibrium  $k = k'$ .

Figure 3.3 shows a plot of the potential profile distribution and the positions of the Fermi levels at  $t = 0^+$  s using an *nin* diode as an example. We see that when the perturbation due to the application of bias has reached the left contact-device interface, the moment we refer to as  $t = 0^+$  s, the Fermi level in the left contact is already raised according to the new local equilibrium in the presence of bias, while in the device the Fermi level correspond to the local equilibrium inherited from the prior equilibrium state. Once we know the initial condition on the distribution function, its subsequent evolution will be governed by Eq. (3.5c) and we do not need to know the Fermi level in the device region any more. Indeed, out of equilibrium the concept of the Fermi level has no use. One consequence of the previously described initial conditions is that  $Q(0^+)$  is initially positive, which is a result of some of the electrons being depleted near the left contact interface and pushed into the right contact. These charges will be replenished during the transient regime injection from the left contact, and are accounted for through the  $dQ/dt$  term in [Eq. (3.5d)]. The depletion of electrons near the left contact-device interface is shown in Fig. 3.4.

### 3.2.3 Injection into the localized device states due to scattering

As can be seen from Fig. 3.3, there is a quantum well formed near the left contact-device interface. There will be some states with wave functions localized in this well. The question is now where are these states populated from, since this will determine the injection rate and ultimately their distribution functions during the transient and in steady state. The injection into these states from the right contact, from which one would expect them to be populated, is very small, since they are on the other side of the barrier from where the right contact is situated. So, if the localized states were only populated from the right contact, their time-evolution would be very

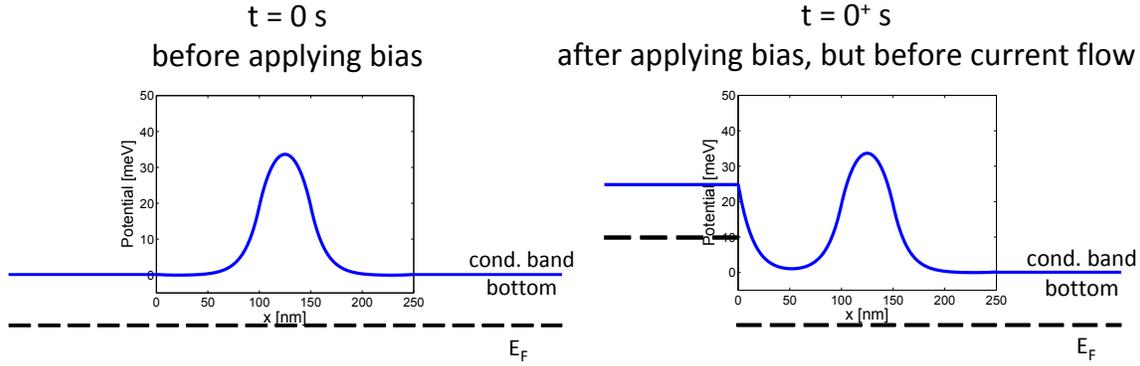


Figure 3.3 Potential profile and the positions of the Fermi level throughout the contact-device-contact system for an *nin* diode with  $10^{17} \text{ cm}^{-3} - 10^{13} \text{ cm}^{-3} - 10^{17} \text{ cm}^{-3}$  doping profile, room temperature, and 25 meV bias. The right contact is assumed to be grounded. The left figure is at equilibrium, before applying bias. The right figure is after applying bias, but before current flow at time  $t = 0^+$  s. The left contact on the one hand and the device region along with the right contact on the other hand are each in a local equilibrium, and have different Fermi levels. Since there is still no current flow,  $k_d(0^+) = 0$ .

slow to vanishing and their steady state distribution function would be determined by  $f_{-k'}^R$ . The right-traveling states immediately above the localized states are populated much faster and have a much larger steady state distribution function, given by  $f_k^L$ .

There will be an additional injection into those localized states from the left contact, as illustrated in Fig. 3.5, again using an *nin* diode as an example. This injection is due to the efficient scattering that is taking place in the contact and can produce much higher injection rates, comparable to those of the right-traveling waves immediately above the localized states, or even higher. For illustration, the injection rates due to the electron flow from the contact into the device are on the order of  $\Delta_k = 10^{12} \text{ s}^{-1}$  and usually less, while if we assume that the injection due to the scattering can be up to  $\Delta_{scatt} \approx 1/\tau$ ,  $\Delta_{scatt}$  can be many times larger than  $\Delta_k$ . This type of injection due to scattering is very important in some devices, like various heterostructure tunnel diodes, where there is a quantum well to the left of the barrier even in steady state. The absence

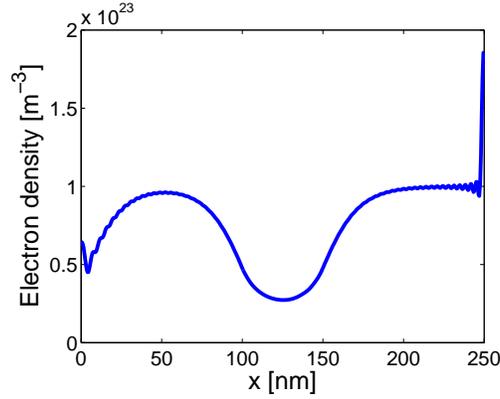


Figure 3.4 Electron density at  $t = 0^+$  s, corresponding to the initial conditions shown in Fig. 3.3. Electrons are depleted in the region near the left contact-device interface. These electrons are pushed into the right contact, when the perturbation due to the application of bias reaches the left contact-device interface, and will be replenished during the transient injection from the left contact. Consequently, there will be a difference between the contact and device currents,  $J_{con}$  and  $J_{dev}$  respectively, equal to  $|dQ/dt|$ .

of scattering-enhanced injection in such devices would fail to give the proper steady-state solution [1, 6, 7]. Since the localized states are predominantly populated from the left contact, we can model their time evolution in a similar fashion as for the current-carrying states, where the term analogous to  $f_{k,-k'}^{L,R}(t)$  in Eq. (3.5c) here will be related to the properties of the left contact, whose conduction band bottom is raised by the value of the applied bias. Taking this into consideration as well as the form of Eq. (3.5c), we can model the time-evolution of the localized states with the following equation:

$$\frac{df_{loc}}{dt} = -\tau \Delta_{scatt}^2 f_{loc} + \tau \Delta_{scatt}^2 f_{scatt} , \quad (3.10)$$

with

$$f_{scatt} = \frac{m^* k_B T}{\pi \hbar^2} \ln \left( 1 + e^{\frac{E_F + V_a - E_{loc}}{k_B T}} \right) . \quad (3.11)$$

Here,  $E_{loc}$  is the energy of localized states measured from the conduction band bottom in the anode, and we again assume a device which allows 1D modeling, and has an isotropic effective mass.

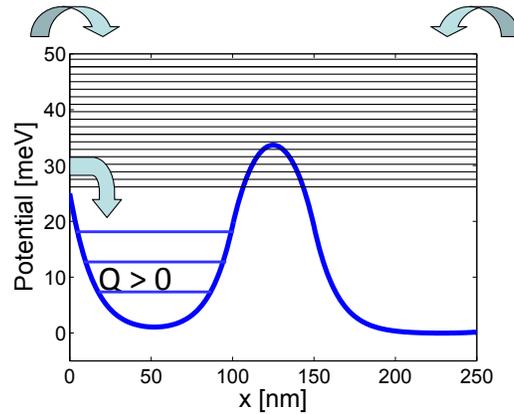


Figure 3.5 Illustration of the scattering-enhanced injection into the states localized in the quantum well to the left of the device barrier. The structure used is the same as in Fig. 3.3. The two upper arrows are injection into the device current-carrying states due to the flow of electrons between the contacts and the device. The lower arrow is the enhanced injection into the localized quantum well states, due to the efficient scattering in the contacts. The thin densely spaced lines (black) designate current-carrying levels, while the widely spaced lines (blue) designate the localized states. As stated before,  $Q$  is positive near the left contact, due to electron depletion.

### 3.3 Transient in an *nin* diode

Figures 3.6 and 3.7 depict the potential, charge density, and current density for a single ellipsoidal valley in an *nin* silicon diode at room temperature. The left and right contacts are doped to  $10^{17} \text{ cm}^{-3}$ , whereas the doping concentration in the middle region is assumed to be  $10^{13} \text{ cm}^{-3}$ . The momentum relaxation time in the contacts is taken to be  $\tau = 120 \text{ fs}$ , based on the textbook mobility values for the above doping density. Note that the characteristic response time of the current is of order hundreds of picoseconds, so three orders of magnitude greater than the contact relaxation time. The transient duration is long because of the relatively weak coupling between the active region and the contacts; the transient duration can be thought of as the inverse of a typical  $\Delta_k^2 \tau$  among the  $k$ 's participating in the current flow.

Figures 3.8 show the time development of the potential profile inside the device for the first few tens of picoseconds and three different scattering enhanced injection rates,  $\Delta_{scatt} = 1/\tau$ ,

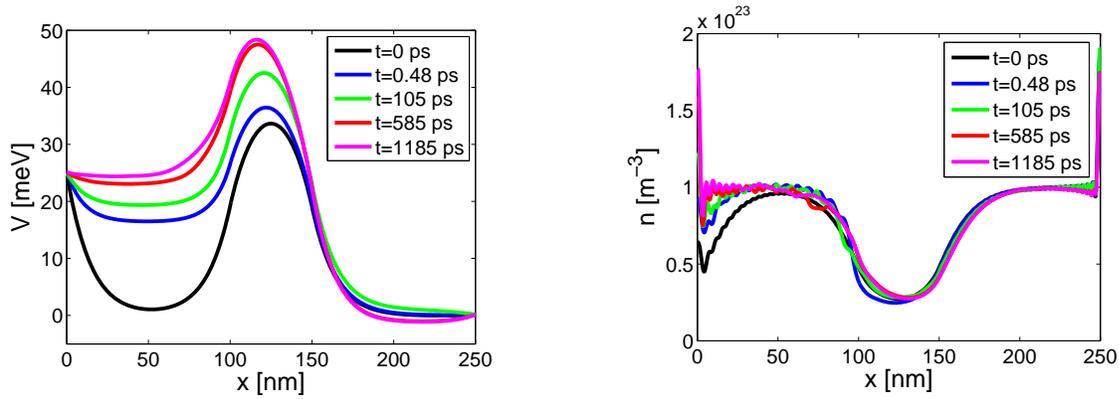


Figure 3.6 (a) Potential and (b) charge density in the *nin* diode as a function of time upon the application of  $-25$  mV to the left contact. The *n*-type regions are doped to  $10^{17}$   $\text{cm}^{-3}$ , while the middle region to  $10^{13}$   $\text{cm}^{-3}$ . The contact momentum relaxation time is  $\tau=120$  fs, as calculated from the textbook mobility value corresponding to the contact doping density. The temperature is  $T = 300$  K and the device size is  $100$  nm  $-$   $50$  nm  $-$   $100$  nm, for the two *n*-doped regions enclosing the middle one. The scattering-enhanced injection rate is chosen to be  $\Delta_{scatt} = 1/\tau$ , and therefore the quantum well bottom rises quickly on the time scale comparable to the contact relaxation time. In the electron density plot, one can observe the initial depletion near the cathode side, and its subsequent population during the transient.

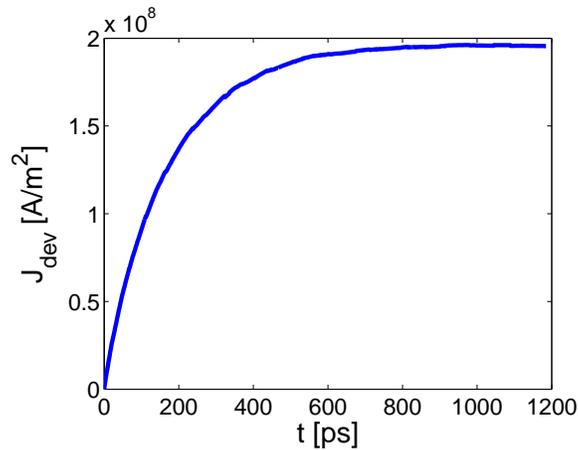


Figure 3.7 Current density versus time for the *nin* diode from Fig. 3.6 upon the application of  $-25$  mV to the left contact.  $J_{dev}$  plotted here refers to the device current from Eq. (3.5d). The transient time is relatively long, owing to small  $\tau$ .

$\Delta_{scatt} = 10^{12} \text{ s}^{-1}$ , and  $\Delta_{scatt} = 0$ . We see that the larger the  $\Delta_{scatt}$  the faster is the quantum well bottom rising, due to the more efficient filling of the localized states. If one compares the time-development of device currents only, then there are no big differences between these three cases. The reason for this is that the initial filling of localized states does not change the barrier height too much, and therefore does not significantly affect the current. However, what is different between these three cases is the contribution to the contact current  $J_{con} = J_{dev} - dQ/dt$  of the term  $dQ/dt$  that may be significantly affected by  $\Delta_{scatt}$  at the beginning of the transient regime.

Finally, we give a comparison of transient current densities for different contact relaxation times  $\tau$  in Fig. 3.9. The *nin* diode used is the same as before, but we allow  $\tau$  to be varied. In principle, when  $\tau$  changes something else usually changes as well, like  $T$ , or the contact doping density  $N_D$ , but here we keep everything constant and vary only  $\tau$ . The reason for this is that we want to achieve the same steady state currents for different relaxation times, and therefore make the influence of  $\tau$  more obvious. Looking at Fig. 3.9 the immediate conclusion is that for larger  $\tau$ , meaning less scattering in the contacts, the duration of the transient regime gets shorter. This was shown previously in Fig. 3.1 using the low current limit and solving Eqs. (3.5c) analytically. The same physical explanation of this phenomenon holds here as well.

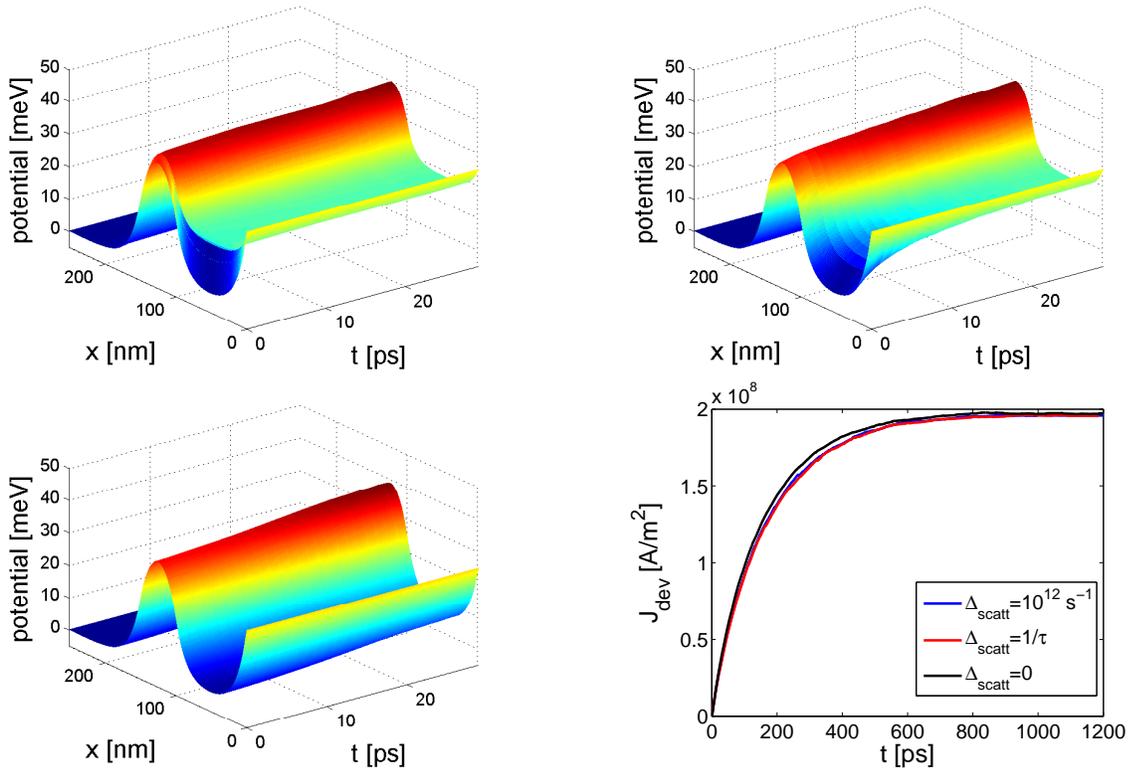


Figure 3.8 The influence of the injection into the localized device states due to scattering on the potential and device current transient evolution. The upper left panel, the upper right panel, and the lower left panel show the initial time-dependence of the device potential profile for three different scattering enhanced injection rates,  $\Delta_{scatt} = 1/\tau$ ,  $\Delta_{scatt} = 10^{12} s^{-1}$ , and  $\Delta_{scatt} = 0$ , respectively. The *nin* diode used here has the same parameters as before. We see that with larger localized states injection rates the quantum well bottom is rising faster. However this initial difference in the quantum well bottom height does not affect the barrier height too much and therefore the transient current is very similar for different  $\Delta_{scatt}$  as shown in the lower right panel. What does get affected is the initial difference between  $J_{con}$  and  $J_{dev}$ , equal to the change of electron charge over time  $dQ/dt$ .

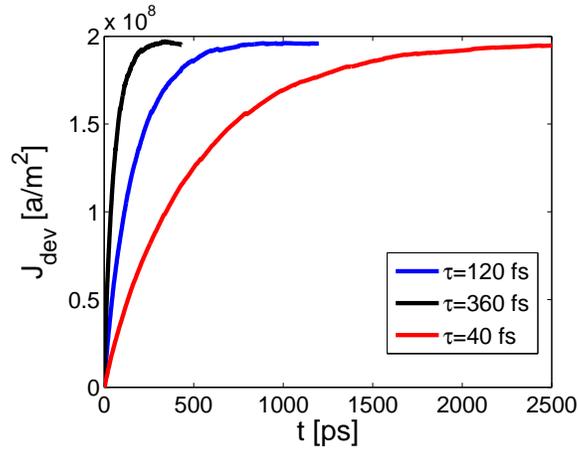


Figure 3.9 Comparison of current densities for the same *nin* diode used in the previous figures and for variable contact relaxation time  $\tau$ . Each curve is shown just until the moment it saturates, to save the computational time. Having less scattering in the contacts helps to reach steady state faster. This phenomenon was discussed in connection to Fig. 3.1, where the same conclusion was reached using the low current limit and solving Eqs. (3.5c) analytically. If the current were limited by the scattering in the contacts, then one would expect that more scattering helps to establish steady state faster. However, in our model, the current is controlled by the ballistic device's potential profile. The device requests some current from the contacts, which act as charge reservoirs, but the contact ability to supply the charge of a given momentum will be limited due to momentum randomization in the presence of scattering.

### 3.4 Conclusion

We presented a theoretical treatment of the transient-regime evolution of an electronic system in a two-terminal ballistic nanostructure coupled to dissipative contacts. The approach is rooted in the open system theory and is based on two key ingredients: (1) A model interaction Hamiltonian between the active region and the contacts, constructed specifically to conserve current during the process of carrier injection from/into the contacts, whose matrix elements are readily calculated from the single-particle transmission problem for structures with and without resonances alike. (2) In the absence of scattering in the active region, it is the rapid energy relaxation in the contacts (due to electron-phonon or, in good, highly-doped contacts, due to electron-electron scattering) that is the indirect source of irreversibility in the evolution of the current-limiting active region, owing to the contact-active region coupling. We account for the influence of the rapid relaxation in the contacts by coarse graining the exact active region evolution over the contact momentum relaxation time. The resulting equations of motion for the distribution functions of the forward and backward propagating states in the active region, Eqs. (3.2), have non-Markovian character as they incorporate the time-varying contact distribution functions through the time-dependent drift-wavevector that depends on the instantaneous current flowing. In order to obtain the response quantities of a nanostructure under bias, such as the potential and the charge and current densities, the non-Markovian master equations must be solved numerically together with the Schrödinger, Poisson, and continuity equations. In doing so, the initial conditions on the potential  $V$  and the device distribution functions  $f_{\pm k}(0)$  must be carefully constructed. Furthermore, the injection into the localized device states due to scattering is discussed and a simple model is devised to account for it. We presented an algorithm for the numerical solution of this coupled system of equations and illustrated the approach on the example of a silicon *nin* diode. The main conclusions are: for longer relaxation times in the contacts, the transient is shorter and vice versa; due to the initial conditions there is electron depletion in the device region at the beginning of the transient, leading to the subsequent population from the cathode and the necessity to introduce  $dQ/dt$  term representing the difference between the contact and device currents away from steady state; in the presence of

initial large injection into the localized states the device current is not affected, while in principle there can be a large difference between  $J_{con}$  and  $J_{dev}$  due to the existence of the  $dQ/dt$  term.

## Chapter 4

### Quantum transport in unconventional geometries in a magnetic field

In this chapter we study the low-field and ballistic electronic transport in curved nanoribbons subject to static and uniform magnetic fields, where the curved nanoribbon is placed between two leads that inject scattering states. The method we present is based on a tight-binding form of the time-independent two-dimensional Schrödinger equation in curvilinear coordinates, with electric scalar and magnetic vector potentials included, and it enables a numerical description of transport in arbitrarily shaped curved nanostructures. A description of transport in terms of the scattering-state basis in the presence of magnetic field, together with curvature and possibly misaligned contacts, requires the use of a local Landau gauge. Based on the use of a stabilized transfer matrix method, we compute the conductance and normalized electronic density at the Fermi level for several curved nanoribbons in cylindrical and toroidal geometry, with and without the magnetic field. The magnetic field determines the number of injected propagating modes by affecting the Landau-level energies in the contacts, while a complex interplay between the magnetic field and the nanoribbon shape (both its curvature and helicity) determines the transmission of injected modes.

This chapter is organized as follows: in Sec. 4.1, we lay out the theoretical concepts and the corresponding numerical model. We explain how the surface divergence, which appears in the curvilinear 2D SE, is related to the curvature properties and the normal component of the magnetic vector potential. In Sec. 4.1.1, we start from the 2D SE in curvilinear coordinates, with the electric scalar potential and the magnetic vector potential included, and focus on the curvilinear Laplacian by explaining its Hermitian properties and how are they affected by its finite-difference version. We

present several discretization schemes that may be used. In Sec. 4.1.2, we derive a tight-binding form of the 2D SE in curvilinear coordinates by using a 2<sup>nd</sup>-order finite-difference scheme. We analyze the consequences of the non-Hermitian nature of the tight-binding (discretized) curvilinear Hamiltonian matrix. In Sec. 4.1.3, we explore the conditions for the validity of the Peierls phase approximation. In Sec. 4.1.4, we explain how the magnetic field is included in the lead/curved nanoribbon/lead system. We discuss the necessity of using the Landau gauge in the leads to ensure proper mode injection, and discuss cases in which a local Landau gauge (amounting to a continuous gauge transformation) may need to be applied to the curved nanoribbon. In Sec. 4.2, we apply the described numerical technique to calculate the electron density at the Fermi level and conductance in several nanoribbons in cylindrical and toroidal geometries, with and without the magnetic field. A summary and concluding remarks are given in Sec. 4.3.

## 4.1 Theoretical and numerical framework

Conductance in the linear, ballistic, steady-state quantum transport regime in nanoribbons is determined by the transmission properties of the electronic states near the Fermi level. [84–86] In curved nanoribbons, we will calculate the transmission and electron density near the Fermi level from the time-independent 2D SE in curvilinear coordinates in the scattering-state basis. As shown previously, by starting from the 3D SE and using the covariant Peierls substitution and a layer thinning procedure, [82] one can arrive at its 2D version completely decoupled from the direction normal to the thin curved surface. We state this equation in its time-independent form

$$\begin{aligned} \frac{1}{2m} \left[ -\hbar^2 \nabla_S^2 \psi + ie\hbar \nabla_S \cdot \vec{A}_S \psi \right. \\ \left. + 2ie\hbar g^{ij} A_i \partial_j \psi + e^2 g^{ij} A_i A_j \psi \right] + V\psi = E\psi, \end{aligned} \quad (4.1)$$

where the 2D curvilinear Laplacian is given by

$$\nabla_S^2 \psi = \frac{1}{\sqrt{g}} \partial_i (\sqrt{g} g^{ij} \partial_j \psi), \quad (4.2)$$

and the 2D curvilinear divergence by

$$\nabla_S \cdot \vec{A}_S = \frac{1}{\sqrt{g}} \partial_i (\sqrt{g} g^{ij} A_j). \quad (4.3)$$

Here,  $\partial_i \equiv \partial/\partial u^i$ ,  $\vec{A}_S = (A_1, A_2, 0)$ ,  $u^i$  is a curvilinear coordinate on the surface  $S$  (here only  $i = 1, 2$  are included since the normal component  $i = 3$  has been decoupled),  $A_i$  are the covariant components of the magnetic vector potential,  $g^{ij}$  are the contravariant components of the metric tensor,  $g$  the determinant of the metric tensor with covariant components,  $V = e\Phi + V_s$ ,  $\Phi$  is the electric scalar potential, and  $V_s$  is the geometric potential given by

$$V_s = -\frac{\hbar^2}{8m} [k_1 - k_2]^2, \quad (4.4)$$

where  $k_1$  and  $k_2$  are the principal curvatures of the surface  $S$ . [178]

It can be shown (see Appendix B) that there is a connection between the surface divergence  $\nabla_S \cdot \vec{A}_S$ , the curvature properties, and the normal component of the vector potential  $A_3$ . Assuming an orthogonal coordinate system and  $\nabla \cdot \vec{A} = 0$ , the following relation holds

$$\nabla_S \cdot \vec{A}_S = \frac{1}{g} A_3|_{u^3=0} (g_{11}h_{22} + g_{22}h_{11}) - \frac{\partial A_3}{\partial u^3} \Big|_{u^3=0}, \quad (4.5)$$

where  $h_{ij}$  are the elements of the second fundamental form of the surface (covariant components of the metric tensor,  $g_{ij}$ , being the elements of the first fundamental form of the surface) [178] and  $A_3$  is the projection of  $\vec{A}$  onto the normal to the curved surface (it simply corresponds to the component of  $\vec{A}$  along the normal to the surface). We can further introduce effective potentials due to the magnetic vector potential

$$V_3 = \frac{1}{2m} e\hbar \nabla_S \cdot \vec{A}_S, \quad (4.6a)$$

$$V_S = \frac{1}{2m} e^2 g^{ij} A_i A_j, \quad (4.6b)$$

where  $V_3$  and  $V_S$  define the potentials due to the normal component  $A_3$  and the surface component  $\vec{A}_S$ , respectively. Now we can rewrite Eq. (4.1) in the form

$$\frac{1}{2m} [-\hbar^2 \nabla_S^2 \psi + 2ie\hbar g^{ij} A_i \partial_j \psi] + V_{\text{eff}} \psi = E\psi, \quad (4.7)$$

where  $V_{\text{eff}} = V + V_S + iV_3$ . Let us further rescale the previous equation, for convenience that will become apparent shortly, by multiplying it by  $\sqrt{g}$ . We get

$$\frac{1}{2m} [-\hbar^2 \overline{\nabla}_S^2 \psi + 2ie\hbar \sqrt{g} g^{ij} A_i \partial_j \psi] + \sqrt{g} V_{\text{eff}} \psi = \sqrt{g} E\psi, \quad (4.8)$$

where  $\bar{\nabla}_S^2 = \partial_i (\sqrt{g} g^{ij} \partial_j) \equiv \sqrt{g} \nabla_S^2$ . At the end, it should be mentioned that the Hamiltonian, as given by Eq. (4.1), is Hermitian (see Appendix C.1).

### 4.1.1 Curvilinear Laplacian in the tight-binding form

Unlike its Cartesian counterpart, the curvilinear Laplacian given by Eq. (4.2) can only be Hermitian up to the prefactor  $1/\sqrt{g}$  once it is cast into a finite-difference form. [179, 180] In other words, it is Hermitian only up to a truncation error of the Taylor expansion by which a particular finite-difference scheme is derived. A finite-difference approximation of the planar Hamiltonian gives a Hermitian Hamiltonian matrix, as well known, while the same approximation of the curvilinear Hamiltonian in the curved nanoribbon can yield a non-Hermitian Hamiltonian matrix. The latter will be derived in more detail in this section and Sec. 4.1.2. By rescaling the Laplacian, as in Eq. (4.8), we remove this Hermiticity problem from the rescaled Laplacian, and consequently from the rescaled Hamiltonian, by moving the prefactor that prevents achieving the Hermiticity to the main diagonal. The problem of finding appropriate discretization schemes for the rescaled Hamiltonian in Eq. (4.8) is easier than for the original Hamiltonian from Eq. (4.7). The non-Hermiticity can be minimized by choosing a sufficiently small grid size, such that  $\sqrt{g}$  does not vary too much across one grid cell. In the limit where curvature is constant (like in a cylinder), the curvilinear tight-binding Hamiltonian, obtained using the finite-differences, is exactly Hermitian. Sufficiently smooth curvature variation on the finite-difference grid is also important to be able to capture potentially sharp curvature features.

In order to determine which conditions should be met to have a small variation of  $\sqrt{g}$  over one grid cell, we can imagine an orthogonal coordinate system and refer to the definition of the metric tensor ( $g_{ii} = \partial \vec{r} / \partial u^i \partial \vec{r} / \partial u^i$ ), which is diagonal in that case. If we assume that the curvilinear coordinate  $u^i$  is angular, the metric tensor component  $g_{ii}$  at any point will be equal to the squared radius of the curvature at that point along  $u^i$ . First, we need enough points on the surface to capture all the curvature features. Therefore, the grid size (in radians) should satisfy  $\Delta u^i \ll 1$ , meaning that the radius of the curvature should be much larger than the grid size (in meters) along the geodesic in that point. So, a grid size of 1 nm will generate a fine enough mesh for curvature radii

not smaller than a few tens of nanometers. Second, the radii of curvature in both surface directions should vary smoothly and slowly over neighboring points, in other words the relative change in the grid size along the geodesic should be much smaller than 1. The second requirement can be visualized by first defining a uniform (constant grid size) rectangular grid on a planar surface and then introducing curvature, due to which the grid will become nonuniform (Fig. 4.1b). We can then assess the metric tensor variations by knowing the relative change in the grid size along the geodesic over neighboring points.

Even after the rescaling, Eq. (4.8), we still cannot use arbitrary finite-difference schemes, but only those that preserve the Hermiticity of the rescaled Laplacian  $\bar{\nabla}_S^2$ . The reason for possible non-Hermiticity in arbitrary finite-difference schemes are the off-diagonal terms of the metric tensor ( $g^{12} = g^{21}$  or  $g_{12} = g_{21}$ ), which are nonzero if the coordinate system is not orthogonal. Among possible finite-difference schemes to discretize  $\bar{\nabla}_S^2$  we will mention two: the 4-point formula (Fig. 4.1) given by

$$\begin{aligned} \frac{\partial f_{0,0}}{\partial u^1} &= \frac{1}{2\Delta_1} (f_{\frac{1}{2},\frac{1}{2}} - f_{-\frac{1}{2},\frac{1}{2}} + f_{\frac{1}{2},-\frac{1}{2}} - f_{-\frac{1}{2},-\frac{1}{2}}) \\ &\quad + \mathcal{O}(\Delta_1^2), \\ \frac{\partial f_{0,0}}{\partial u^2} &= \frac{1}{2\Delta_2} (f_{\frac{1}{2},\frac{1}{2}} + f_{-\frac{1}{2},\frac{1}{2}} - f_{\frac{1}{2},-\frac{1}{2}} - f_{-\frac{1}{2},-\frac{1}{2}}) \\ &\quad + \mathcal{O}(\Delta_2^2), \end{aligned} \tag{4.9}$$

where  $\Delta_1$  and  $\Delta_2$  are the grid sizes along  $u^1$  and  $u^2$  axes respectively, and the central difference scheme (by simultaneously using two different grid sizes) [179]

$$\frac{\partial f_{0,0}}{\partial u^{1(2)}} = \begin{cases} \frac{f_{\frac{1}{2},0(0,\frac{1}{2})} - f_{-\frac{1}{2},0(0,-\frac{1}{2})}}{\Delta_{1(2)}}, & i = j \text{ in } \bar{\nabla}_S^2 \\ \frac{f_{1,0(0,1)} - f_{-1,0(0,-1)}}{2\Delta_{1(2)}}, & i \neq j \text{ in } \bar{\nabla}_S^2 \end{cases}. \tag{4.10}$$

Here, we will work only with orthogonal coordinate systems, in which case there is no need for the more complicated finite-difference schemes given by Eqs. (4.9) and (4.10) and the ordinary central-difference scheme is enough. [Actually, in the limit  $g^{ij} = g_{ij} = 0$ ,  $i \neq j$ , the term in Eq. (4.10) for which  $i \neq j$  vanishes from  $\bar{\nabla}_S^2$  and we are left with the ordinary central-difference scheme]. In the case a non-orthogonal coordinate system is needed, we can go back and use (4.9) or

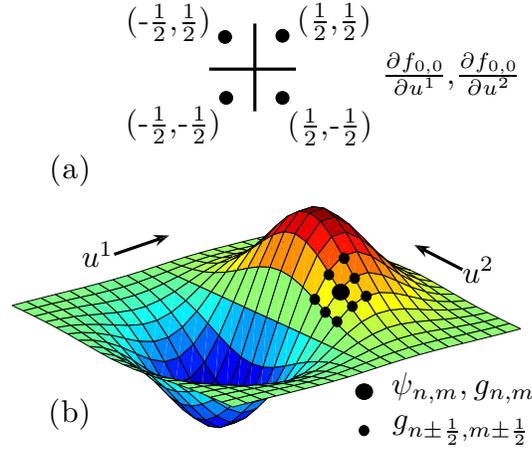


Figure 4.1 (a) The 4-point scheme for discretization of the first partial derivatives by using Eq. (4.9). (b) An example of a curved surface parametrized by curvilinear coordinates  $u^1$  and  $u^2$ . Leads are planar continuations of the curved surface in both directions along the longitudinal coordinate  $u^1$ . In order to calculate the wave function value at the grid point denoted by the large solid circle (central point) from the 4-point scheme, we have to know the metric tensor values at all the surrounding points, denoted by the small solid circles, as well as at the central point.

(4.10) and follow the procedure outlined in the rest of this section. So, considering an orthogonal coordinate system and applying the ordinary central-difference scheme we can come up with a tight-binding representation of  $\bar{\nabla}_S^2$ , discretized at some slice  $n$ , as

$$-\frac{\hbar^2}{2m} \left( \bar{\nabla}_S^2 \psi \right)_n = \mathbf{L}_n \Psi_n + \mathbf{L}_{n,n-1} \Psi_{n-1} + \mathbf{L}_{n,n+1} \Psi_{n+1}, \quad (4.11)$$

where  $\Psi_n$  is a column vector containing all wave function values on the  $n^{\text{th}}$  slice (slice is a set of all points with the longitudinal coordinate  $u^1$  constant) and the  $(M+1) \times (M+1)$  matrices  $\mathbf{L}$  are

given by

$$\mathbf{L}_n = \begin{bmatrix} c_{n,0} & -f_2^0 e_{n,0}^{+\frac{1}{2}} & 0 & \cdots & 0 \\ -f_2^1 e_{n,1}^{-\frac{1}{2}} & c_{n,1} & -f_2^1 e_{n,1}^{+\frac{1}{2}} & \cdots & 0 \\ 0 & -f_2^2 e_{n,2}^{-\frac{1}{2}} & c_{n,2} & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & 0 & -f_2^M e_{n,M}^{-\frac{1}{2}} & c_{n,M} \end{bmatrix}, \quad (4.12a)$$

$$\mathbf{L}_{n,n\pm 1} = \begin{bmatrix} -f_1^n d_{n,0}^{\pm\frac{1}{2}} & 0 & \cdots & 0 \\ 0 & -f_1^n d_{n,1}^{\pm\frac{1}{2}} & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & 0 & -f_1^n d_{n,M}^{\pm\frac{1}{2}} \end{bmatrix}, \quad (4.12b)$$

where

$$\begin{aligned} c_{n,m} &= f_1^n d_{n,m}^{+\frac{1}{2}} + f_1^n d_{n,m}^{-\frac{1}{2}} + f_2^m e_{n,m}^{+\frac{1}{2}} + f_2^m e_{n,m}^{-\frac{1}{2}}, \\ d_{n,m}^l &= t_1^{n+l} (\sqrt{g} g^{11})_{n+l,m}, \\ e_{n,m}^l &= t_2^{m+l} (\sqrt{g} g^{22})_{n,m+l}, \end{aligned} \quad (4.13)$$

with  $t_{1(2)}^{i+l} = \hbar^2/2m\Delta_{1(2)}^i$  for  $0 \leq l < 1$ ,  $f_{1(2)}^i = 2/(\Delta_{1(2)}^i + \Delta_{1(2)}^{i-1})$  and  $M + 1$  the number of points in a slice. The assumption is that  $M$  is constant for all slices. If the curved nanoribbon has varying width, we can always set a hardwall potential at the extra points around the narrow regions. We assume a 2D nonuniform curvilinear grid, with  $\Delta_1^n$  the longitudinal grid size between slices  $n$  and  $n + 1$  and  $\Delta_2^m$  the transversal grid size between points  $m$  and  $m + 1$  along a slice. Equation (4.11) reflects the lack of symmetry between the two curvilinear surface directions in our problem: there is the longitudinal direction with open boundaries, along which the current flows carried by complex traveling waves, and the transversal direction along which the potential defines the shape of real (confined) transverse modes. This form of Eq. (4.11) makes it suitable for use with the transfer matrix method, where the slice wave functions (containing all transverse points) are transferred along the longitudinal direction.

The form of the matrix (4.12a) will depend on the boundary conditions in the  $u^2$  direction. While the system is always open in the longitudinal (i.e. transport) direction along the coordinate

$u^1$ , it might be either closed (hardwall potential) or periodic (e.g. closing onto itself, like in a cylinder) in the transverse direction  $u^2$ . In the former case we use the Dirichlet boundary conditions, which give the form (4.12a), while in the latter the periodic boundary conditions, where we have the following additional terms

$$\begin{aligned}\mathbf{L}_n(1, M + 1) &= -f_2^0 e_{n,0}^{-\frac{1}{2}}, \\ \mathbf{L}_n(M + 1, 1) &= -f_2^M e_{n,M}^{+\frac{1}{2}}.\end{aligned}\tag{4.14}$$

If the coordinate system is not orthogonal ( $g^{ij} \neq 0, i \neq j$ ) and we use either one of the schemes (4.9) or (4.10), then, besides having more complex matrix elements,  $\mathbf{L}_{n,n\pm 1}$  would also depend on the boundary condition in the transverse direction (being a three-diagonal matrix in that case).

### 4.1.2 The tight-binding 2D Schrödinger equation

Now we will discretize the rest of Eq. (4.8) to obtain the curvilinear 2D SE in the tight-binding form. To avoid having imaginary potential on the main diagonal of the Hamiltonian matrix, we will first rewrite the imaginary terms containing the magnetic vector potential in Eq. (4.8) as

$$\begin{aligned}2ie\hbar\sqrt{g}g^{ij}A_i\partial_j\psi + ie\hbar\nabla_S \cdot \vec{A}_S\psi = \\ ie\hbar\sqrt{g}g^{ij}A_i\partial_j\psi + ie\hbar\partial_i(\sqrt{g}g^{ij}A_j\psi).\end{aligned}\tag{4.15}$$

Since these terms contain the first derivative of the wave function, we will use the central-difference scheme like for the case  $i \neq j$  in Eq. (4.10), in order to avoid having wave function values at half-integer grid points. We obtain the following tight-binding form at some slice  $n$

$$\begin{aligned}\frac{ie\hbar}{2m} [\sqrt{g}g^{ij}A_i\partial_j\psi + \partial_i(\sqrt{g}g^{ij}A_j\psi)]_n = \\ = \mathbf{P}_n \Psi_n + \mathbf{P}_{n,n-1} \Psi_{n-1} + \mathbf{P}_{n,n+1} \Psi_{n+1},\end{aligned}\tag{4.16}$$

where

$$\begin{aligned}
\mathbf{P}_n(m, m+1) &= \frac{ie}{\hbar} f_2^{m-1} \frac{1}{2} [e_{n,m-1}^0 A_2(n, m-1) \Delta_2^{m-1} \\
&\quad + e_{n,m-1}^{+1} A_2(n, m) \Delta_2^m] , \\
\mathbf{P}_n(m, m-1) &= -\frac{ie}{\hbar} f_2^{m-1} \frac{1}{2} [e_{n,m-1}^0 A_2(n, m-1) \Delta_2^{m-1} \\
&\quad + e_{n,m-1}^{-1} A_2(n, m-2) \Delta_2^{m-2}] , \\
\mathbf{P}_{n,n\pm 1}(m, m) &= \pm \frac{ie}{\hbar} f_1^n \frac{1}{2} [d_{n,m-1}^0 A_1(n, m-1) \Delta_1^n \\
&\quad + d_{n,m-1}^{\pm 1} A_1(n \pm 1, m-1) \Delta_1^{n\pm 1}] .
\end{aligned} \tag{4.17}$$

(The matrix indices in the above equations are  $m \in [1, M]$ ,  $m \in [2, M+1]$ , and  $m \in [1, M+1]$ , respectively). Note that when  $m$  is used as a matrix index on the left-hand side of Eqs. (4.17) its range is  $[1, M+1]$ , while the corresponding transverse grid position on the right-hand side is in the range  $[0, M]$ , hence  $m-1$  is used on the right-hand side. The magnetic vector potential components in Eq. (4.17) are the covariant components, given by  $\vec{A} = A_i \vec{e}^i$ , where  $\vec{e}^i = \nabla u^i$ . The relations of the covariant components to the contravariant components  $\vec{A} = A^i \vec{e}_i$ , where  $\vec{e}_i = \partial \vec{r} / \partial u^i$ , are  $A_1 = g_{11} A^1$  and  $A_2 = g_{22} A^2$ , if the coordinate system is orthogonal. The relations to the magnetic vector components expressed with respect to the unit curvilinear vectors  $\vec{a}_i$ ,  $\vec{A} = \bar{A}^i \vec{a}_i$ , which is often done if one starts with calculating  $\vec{A}$  in the Cartesian coordinate system and then either performs a coordinate system transformation to the desired curvilinear system, or a vector projection in the case of arbitrary curvature, are  $A_1 = \sqrt{g_{11}} \bar{A}^1$  and  $A_2 = \sqrt{g_{22}} \bar{A}^2$ . In the case of using  $\bar{A}^{1(2)}$  we see that it should be multiplied by the actual grid size along the geodesic (in meters),  $(\sqrt{g_{11(22)}})_{n,m} \Delta_{1(2)}^{n(m)}$ , at the grid position  $(n, m)$ .

As with the  $\mathbf{L}$  matrices in Eqs. (4.12), if the transverse boundary condition is periodic, we have additional terms of the following form in Eqs. (4.17)

$$\begin{aligned}
\mathbf{P}_n(1, M+1) &= -\frac{ie}{\hbar} f_2^0 \frac{1}{2} [e_{n,0}^0 A_2(n, 0) \Delta_2^0 \\
&\quad + e_{n,0}^{-1} A_2(n, M) \Delta_2^M] , \\
\mathbf{P}_n(M+1, 1) &= \frac{ie}{\hbar} f_2^M \frac{1}{2} [e_{n,M}^0 A_2(n, M) \Delta_2^M \\
&\quad + e_{n,M}^{+1} A_2(n, 0) \Delta_2^0] .
\end{aligned} \tag{4.18}$$

Finally, the tight-binding 2D SE can be written as

$$\mathbf{H}_{n,n+1}\Psi_{n+1} + \mathbf{H}_{n,n-1}\Psi_{n-1} + \mathbf{H}_n\Psi_n = E\mathbf{G}_n\Psi_n, \quad (4.19)$$

with

$$\mathbf{G}_n = \begin{bmatrix} (\sqrt{g})_{n,0} & 0 & \cdots & 0 \\ 0 & (\sqrt{g})_{n,1} & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & 0 & (\sqrt{g})_{n,M} \end{bmatrix}, \quad (4.20)$$

and

$$\begin{aligned} \mathbf{H}_n &= \mathbf{L}_n + \mathbf{P}_n + \mathbf{G}_n \mathbf{V}_n, \\ \mathbf{H}_{n,n\pm 1} &= \mathbf{L}_{n,n\pm 1} + \mathbf{P}_{n,n\pm 1}, \end{aligned} \quad (4.21)$$

where  $\mathbf{V}_n$  is a column vector containing the potential  $V_{\text{eff}} - iV_3 = V + V_S$  of the  $n^{\text{th}}$  slice (since  $iV_3$  is already incorporated into the  $\mathbf{P}_n$  and  $\mathbf{P}_{n,n\pm 1}$  matrices).

Equation (4.19) is *the central equation of this paper*. Its form is similar to that of the planar case, [87, 92] but with more complicated matrix elements. In order to solve it, we make use of the stabilized transfer matrix technique due to Usuki *et al.*, [87] because it provides both the reflection/transmission coefficients and a convenient way to calculate the wave functions needed for electron density calculation. The modified stabilized transfer matrix technique, as used in this work, is described in detail in Appendix D.

Although Eq. (4.19) is derived from the exact curvilinear Schrödinger equation [Eq. (4.1)], the usual discretization error, associated with the second-order finite-difference scheme, here may lead to a non-Hermitian tight-binding Hamiltonian. The non-Hermiticity stems from the discretization of the curvilinear Laplacian in the case when curvature is spatially nonuniform, as explained in Sec. 4.1.1, and/or from the fact that we allow the curvilinear grid to be nonuniform in general. The Hermiticity of a discretized Hamiltonian matrix is a desirable property, because it makes the numerical solution of the eigenvalue problem easier. Even though our computational method requires the solution of the eigenvalue problem only in the leads – which are typically, but not necessarily, flat, so the Hamiltonian in them is Hermitian – we employed discretization schemes in Eq. (4.19) that enable the construction of a discretized non-Hermitian curvilinear Hamiltonian whose

eigenvalue problem can be mapped on that of a Hermitian matrix with the help of certain matrix transformations. In Appendix C, we give a detailed account of the Hermitian properties of both the exact and the tight-binding Hamiltonians, along with the aforementioned matrix transformations. After transforming the original non-Hermitian eigenvalue problem, we obtain a Hermitian eigenvalue problem with the same eigenvalues and transformed wave functions. From this we draw two conclusions: first, the eigenvalue problem becomes more manageable computationally, albeit with the additional reverse transformation necessary to reconstruct the original wave functions, and second, although the transformed wave functions do form a complete and orthonormal basis, being eigenvectors of a Hermitian matrix, our original wave functions may not form such a basis. However, the latter consequence is unavoidable, since it is a side effect of the discretization error of the finite-difference approximation in the case of spatially nonuniform curvature and/or nonuniform curvilinear grid. As explained in Sec. 4.1.1, this side effect can be minimized by choosing a sufficiently small grid size that prevents sharp curvature variations across one grid cell. If in addition the curvilinear grid is nonuniform, one should ensure sufficiently smooth grid size variations over neighboring grid cells.

### 4.1.3 The Peierls phase approximation

It is instructive to see how the matrix elements look in the limit of small curvature variations (smooth and slow transitions between different curvatures) and weak magnetic field. For the matrix element  $\mathbf{H}_n(m, m+1)$  the first condition allows us to write

$$\begin{aligned} e_{n,m}^{+\frac{1}{2}} &\approx e_{n,m}^0 \\ e_{n,m}^{+\frac{1}{2}} &\approx e_{n,m}^{+1}, \end{aligned} \tag{4.22}$$

since the metric tensor varies very slowly over one unit cell in this limit. Further taking a weak magnetic field limit such that terms of the form  $A_i \Delta_i$  are small, we obtain the following Peierls phase form

$$\begin{aligned} \mathbf{H}_n(m, m+1) &\approx -f_2^{m-1} e_{n,m-1}^{+\frac{1}{2}} \left\{ 1 - \frac{ie}{\hbar} \frac{1}{2} [A_2(n, m-1) \Delta_2^{m-1} + A_2(n, m) \Delta_2^m] \right\} \\ &\approx -f_2^{m-1} e_{n,m-1}^{+\frac{1}{2}} \exp \left\{ -\frac{ie}{\hbar} \frac{1}{2} [A_2(n, m-1) \Delta_2^{m-1} + A_2(n, m) \Delta_2^m] \right\}. \end{aligned} \tag{4.23}$$

Similarly

$$\begin{aligned} \mathbf{H}_n(m, m-1) &\approx -f_2^{m-1} e_{n,m-1}^{-\frac{1}{2}} \exp \left\{ \frac{ie}{\hbar} \frac{1}{2} [A_2(n, m-1)\Delta_2^{m-1} + A_2(n, m-2)\Delta_2^{m-2}] \right\}, \\ \mathbf{H}_{n,n\pm 1}(m, m) &\approx -f_1^n d_{n,m-1}^{\pm\frac{1}{2}} \exp \left\{ \mp \frac{ie}{\hbar} \frac{1}{2} [A_1(n, m-1)\Delta_1^n + A_1(n\pm 1, m-1)\Delta_1^{n\pm 1}] \right\}. \end{aligned} \quad (4.24)$$

Of course, there will be additional terms in Eq. (4.23) and the first of Eqs. (4.24) if we use the periodic boundary conditions in the  $u^2$ -direction, obtained by summing Eqs. (4.14) and (4.18) in a similar way. If we further neglect the remaining field-dependent term  $V_S$  from Eq. (4.6b), since it is second-order in magnetic vector components, we have represented the effect of the magnetic field in a curved nanoribbon through *the Peierls phase*, as is often done in planar nanostructures. [87,92]

In planar nanostructures subject to a perpendicular magnetic field, the Landau gauge with zero transverse component ( $A_2 = 0$ ) is typically chosen to ensure plane wave injection from the leads (more on this issue in the next section). Therefore, the phase in  $\mathbf{H}_n(m, m \pm 1)$  is typically zero in planar structures. In contrast, in curved nanostructures subject to a magnetic field, the curvature may make it impossible to eliminate  $A_2$  through gauge transformations, so there may be nontrivial phase factors in  $\mathbf{H}_n(m, m \pm 1)$ . We will discuss the issue of gauge transformations in the next section.

Let us estimate the magnitude of the magnetic field for which the Peierls phase approximation is justified. We can say  $|\vec{A}| \approx |\vec{B}| D$ , where  $D$  is the largest transverse dimension of the device. For the Peierls phase to hold we need

$$\frac{e}{\hbar} |\vec{A}| \Delta \ll 1, \quad |\vec{B}| \ll \frac{\hbar}{e\Delta D} = \frac{6.6 \times 10^{-16} \text{Tm}^2}{\Delta D}. \quad (4.25)$$

In a typical nanostructure,  $\Delta \approx 1$  nm and  $D \approx 100$  nm, so we have  $|\vec{B}| \ll 6.6$  T. To estimate  $V_S$ , which depends on both the curvature and the magnetic field, we introduce  $R$  as the radius of curvature in our device and notice that  $g_{ii} \sim R^2$ . By again using  $|\vec{A}| \approx |\vec{B}| D$  and Eq. (4.6b), we can easily show that  $V_S \ll 1$  eV for  $|\vec{B}| \ll \sqrt{m/|e|}/D$ . We see that for  $V_S$  the condition is inversely proportional to  $D$ , which is reasonable because  $V_S$  is of the second order in  $\vec{A}$  and therefore small for small device size. In GaAs and with  $D \approx 100$  nm we have  $V_S \ll 1$  eV for  $|\vec{B}| \ll 6$  T, which is a similar condition as the one given by Eq. (4.25).

The Peierls phase approach is often used for simulating the effect of magnetic field in planar nanostructures. [87, 92] Equation (4.19), modified such that matrix elements that contain the magnetic vector potential are given by Eqs. (4.23) and (4.24), along with setting  $V_S = 0$ , represent the generalization of that approach for curved nanostructures. Equations (4.23) and (4.24) reduce to the form used earlier for planar nanostructures in the limit of flat metric and a uniform grid, as do the other matrix elements in Eq. (4.19). Having in mind the approximations used to derive Eqs. (4.23) and (4.24) and to be able to neglect  $V_S$ , the Peierls phase approach for curved nanostructures has limited applicability for weak magnetic fields and small curvature variations. Since planar nanostructures are the limit of flat metric (unit metric tensor - no curvature variations), the only condition for the Peierls phase approximation validity in them is sufficiently weak magnetic field.

#### 4.1.4 Handling the magnetic field: the local Landau gauge

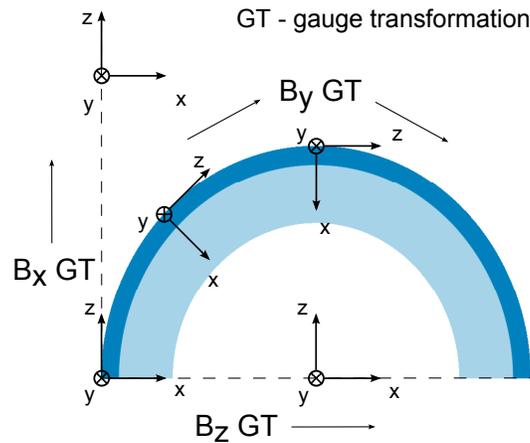


Figure 4.2 An example of a curved nanoribbon for calculating the magnetic vector potential: the nanoribbon (dark blue) on top of a half-torus (light blue), with the leads extending downward at both ends. The Cartesian coordinate system in the left lead (shown in the lower left corner) is taken as a reference system, according to which the magnetic field components  $B_x$ ,  $B_y$ , and  $B_z$  are specified. Gauge transformations (GT), equivalent to coordinate system translations and rotations in the plane perpendicular to each component, are also shown.

In order to properly inject transverse modes and compute their velocities (the latter being critical for the calculation of the transmission matrix, as described in Appendix D.1) we have to include the leads in the transport calculation. A curved nanoribbon is placed between two leads that connect it to macroscopic reservoirs of charge. The leads act as the source/sink for the scattering states, whose plane-wave form in the leads allows us to calculate the conductance using the Landauer-Büttiker formalism. [86] So, we will include a necessary portion of the leads in the simulation region; in the case of planar leads, we will in general have to connect a planar rectangular grid of the leads to a curvilinear grid of a nanoribbon via a transition grid. The magnetic field will be present in the entire simulation region, including the leads. The basis that describes the lead/curved nanoribbon/lead system consists of scattering states, whose form in the injecting lead is a sum of the incident and reflected waves and in the outgoing lead a single transmitted wave.

In order to keep the scattering-state basis for the purposes of transport calculation in the presence of magnetic field, one has to use the Landau gauge in the planar leads, such that there is no transverse component of the magnetic vector potential in the leads and no explicit dependence on the longitudinal coordinate. [86] The Landau gauge ensures that the Hamiltonian eigenstates are plane waves along the longitudinal direction. In Fig. 4.2, we have an example of a lead injecting scattering states into a nanoribbon in toroidal geometry. The lead connects to the structure at the bottom left corner. The lead's Cartesian coordinate system, shown at the bottom left ( $z$  is the longitudinal coordinate in the lead,  $y$  is the transverse, and  $x$  is the normal coordinate) serves as the reference. We can write an arbitrary magnetic field, present in the whole structure, in terms of the lead's Cartesian coordinate system,  $\vec{B} = (B_x, B_y, B_z)$  and define a Landau gauge for each component separately: namely, we can define  $\vec{A}^x = (0, 0, B_x y)$ ,  $\vec{A}^y = (0, 0, -B_y x)$ , and  $\vec{A}^z = (-B_z y, 0, 0)$ , so that for an arbitrarily directed magnetic field  $\vec{B}$ , we can employ the superposition principle and write  $\vec{A} = \vec{A}^x + \vec{A}^y + \vec{A}^z$ . I.e., the total magnetic vector potential will then be given by

$$\vec{A} = (-B_z y, 0, B_x y - B_y x) . \quad (4.26)$$

With  $\vec{A}$  in place for the injecting lead, we can calculate the injected scattering states by solving the eigenvalue problem of the transfer matrix between the first two slices that belong to the lead. [87]

Explanation of one technical point is needed here. In order to obtain the correct number of forward and backward propagating states by numerically solving the eigenvalue problem in the lead, as in Appendix D, it is important to properly choose the reference level for  $\vec{A}$ . This can be done by centering the reference coordinate system to follow the midline that “cuts” the lead in half in the transverse direction: in other words, we can choose the  $u^2 = 0$  grid line such that it extends into the lead and splits it in two halves symmetrically. Upon such a choice, the terms in  $\vec{A}$  in Eq. (4.26) due to the different  $\vec{B}$  components will always be either even or odd functions of the transverse position, which gives the correct number and distribution of forward and backward propagating states. We will refer to this  $u^2 = 0$  line simply as the midline or the bisector. More details on the calculation and sorting of injected modes can be found in Appendix D.

Now, our nonplanar structure is placed between two leads, which could in principle be misaligned, and we need to be able to inject scattering states from both of them. It was shown (see Appendix E in Baranger and Stone [181]) that, in planar structures with misaligned leads, in order to have the Landau gauge and properly inject plane waves from each lead one has to use suitable gauge transformations in order to properly account for the magnetic field in all leads. These gauge transformations ensure that the magnetic vector potential is calculated in the Landau gauge locally for each lead. The problem of including magnetic field in the presence of multiple misaligned leads is addressed in a similar way elsewhere. [182] In order to smoothly connect the magnetic vector potential in different leads, gauge transformations will be required in the device region as well. Since there is no unique analytical expression for gauge transformations for different geometries (like cylindrical or toroidal), we illustrate the approach with the aid of Fig. 4.2. We generalize the requirement to have a local Landau gauge in each lead to the requirement to have *a local Landau gauge in every slice of the lead/curved nanoribbon/lead system*.

The idea is the following: for a given magnetic field orientation, any coordinate system translation as well as coordinate system rotations in the plane perpendicular to the field can be shown to be equivalent to gauge transformations that make the gauge locally Landau. [181] In other words, if we define  $\vec{A}$  associated with a given  $\vec{B}$  in some coordinate system, then rotate and/or translate that system as specified above, after which we define  $\vec{A}'$  in the transformed coordinate system in

the same way as in the initial coordinate system (i.e. give it the same functional dependence on the new coordinates as we had in the old), then it can be shown that there exists a scalar function  $f$  such that (in the same set of coordinates – for instance, the old ones)  $\vec{A}' = \vec{A} + \nabla f$ . When employing the transfer matrix method, which uses slice-specific dynamical matrices as described in Sec. 4.1.2, the relevant matrices in each slice contain the magnetic vector potential according to the local Landau gauge (however, it is written in the curvilinear coordinates that parametrize the surface). Gauge-independent quantities, such as the electron density or conductance are, of course, insensitive to these gauge transformations. In Appendix E.1, we show that the multitude of gauge transformations, required to obtain the local Landau gauge slice-by-slice, can be represented as a single gauge transformation.

In Fig. 4.2, we apply this approach to the magnetic vector potential components ( $\vec{A}^x$ ,  $\vec{A}^y$ , or  $\vec{A}^z$ ) generated by each magnetic field component ( $B_x$ ,  $B_y$ , or  $B_z$ ) independently, starting from Eq. (4.26) and having in mind the superposition principle. As already mentioned, we can always identify a midline through our device on which the reference coordinate system origin is situated in the reference (injecting) contact. The gauge transformations are then performed for each magnetic vector potential component ( $\vec{A}^x$ ,  $\vec{A}^y$ , or  $\vec{A}^z$ ) independently such that the coordinate system origin moves along the bisector from one lead to the other, while the local  $z$ -axis is tangential to the projection of the bisector onto the plane normal to the respective magnetic field component. In the example of Fig. 4.2, nontrivial gauge transformations are only needed for the magnetic vector potential associated with the  $B_y$  field component, because it is the only field component requiring rotations – the projection of the bisector onto the plane normal to  $B_y$  is a semi-circle (more details can be found in Appendix E.2). Projections of the bisector onto the planes normal to  $B_x$  and  $B_z$  are straight lines, thus require only translations along the longitudinal axis (the  $z$ -axis of the reference system for  $B_x$  and the  $x$ -axis for  $B_z$ ), while the values of the components of  $\vec{A}$  in Eq. (4.26) related to these magnetic field components vary along the  $y$ -axis, which is normal to the respective longitudinal axes and therefore neither directions nor values of these magnetic vector components are changed going from slice to slice.

## 4.2 Examples

We apply the method of Sec. 4.1 to calculate the electron density and conductance for various nanoribbons. To define curved nanoribbons of various shapes, we use a virtual substrate specified by a curvilinear coordinate system, like cylindrical, spherical, and torodial – in other words, we mimic placing the ribbon on top (or wrapping it around) a cylinder, sphere, or a torus – and then we apply a hardwall potential to define the ribbon edges. This method allows us to use relatively simple coordinate systems to work with complicated shapes. For each curvilinear coordinate system there are several quantities, related to the non-zero curvature, that have to be precalculated: the first fundamental form (or the metric tensor)  $g$  and the two principal curvatures,  $k_1$  and  $k_2$ . Sufficient information for this task can be found in many textbooks and handbooks [178, 183] and online mathematical resources.

### 4.2.1 Cylindrical geometry

The cylindrical geometry is relatively simple and very useful for modeling various curved nanoribbons that are being produced today. [23, 26, 42, 184] We assume the ribbon will be placed on a cylinder of radius  $r$ , with  $(\phi, x)$  being the natural surface coordinates and  $x$  being the direction of cylinder axis. By applying a hardwall potential in  $(\phi, x)$  we can define the edges of the curved nanoribbon that is placed on top of or wrapped around the cylinder, and proceed to calculate its conductance and electronic density by the previously described numerical technique. The leads are assumed cylindrical to avoid the need for transitional regions, which is a good approximation if the nanoribbon is narrow compared to the circumference of the cylinder. The leads extend away from both ends of the nanoribbon and along the cylinder axis, as depicted in Fig. 4.3(b).

In Figs. 4.3 and 4.4, we present calculation results for a cylindrical nanoribbon with helicity (wrapped around the cylinder), for two different magnetic field directions. The radius of the cylinder is  $r = 40$  nm, its length is  $4\pi r$ , while the nanoribbon width corresponds to one third of the circumference. The helix occupies one half of the length. The Fermi level is at 8 meV, and the grid contains  $500 \times 250$  points. The curved nanoribbon does not have any scattering centers;

therefore, any difference between the actual normalized conductance and the number of injected propagating modes (formed in the injecting lead subject to the magnetic field), which can be seen in Figs. 4.3(a) and 4.4(a), is due to the curvature and helicity and their coupling with the magnetic field. For the direction of magnetic field as in Fig. 4.3(b), there is a very weak dependence of the Landau level energies in the injecting lead on the magnetic field, due to the fact that the field has a small flux through the narrow cylindrical lead, but the dependence is not strong enough to affect the number of injected propagating modes. (The flux would be zero if the leads were perfectly planar.) In Fig. 4.3(a), there are 3 modes injected throughout the magnetic field range, but even at  $B = 0$  only two are transmitted; the reason for the reflection is the ribbon's helicity (we will discuss helicity in more detail later). From Fig. 4.3(a), we can see that, for higher magnetic fields, the transmission gradually decreases, until it is completely suppressed at around 4 T. The suppression can also be observed in the plot of the electron density [Fig. 4.3(b)], where at 4.5 T the wave function is confined to the region close to the injecting lead. In contrast, at 2.663 T, a quasibound state is responsible for the resonant transmission, visible as a sharp peak in Fig. 4.3(a).

The variation in the Landau level energies in the leads with varying magnetic field is much more pronounced in Fig. 4.4, where the magnetic field is nearly perpendicular to the lead and produces maximal flux. Consequently, the number of injected propagating modes varies appreciably in the magnetic field range of interest. We use this example to illustrate the onset of discrepancies between a calculation based on the exact equation (4.19) and the Peierls phase approximation. The discrepancy between the curves calculated using the exact equation (4.19) and the Peierls phase approximation becomes apparent at fields higher than about 2.3 T, Fig. 4.4(a). The appearance of the resonant reflection (the sharp dip in transmission), often observed in 2D coherent electron systems of nonuniform width, [185–187] is present in the exact (thick red) curve at 4.582 T. In these nanostructures, although the nanoribbon has constant width (there are no cavities along the nanoribbon), the interplay between helicity and the magnetic field might lead to the formation of quasibound states responsible for resonant reflections. These reflections are very sensitive to the amplitude and direction of magnetic field. The interplay between the helicity and magnetic field is also responsible for the incomplete transmission from about 2 T to 3.3 T.

## 4.2.2 Toroidal geometry

Toroidal geometry is more complicated than the cylindrical one, and enables us to demonstrate the technique on nanoribbons with spatially nonuniform curvature. We use a standard parametrization of a torus of radii  $R$  and  $r$ , where  $R$  is the radius of the large circle going through the middle of the torus and  $r$  is the radius of the small circle, whose center remains on the larger circle as its rotation around the large circle's axis maps the torus surface. The surface coordinates are the two angles,  $(\theta, \phi)$ , where  $\theta$  is related to  $r$  and  $\phi$  to  $R$ . To avoid transitional regions, we adopt cylindrical leads here as well, because of their natural compatibility with a "straightened" torus. Having a curved nanoribbon that consists partially of a cylindrical shape and partially of a toroidal shape, causes problems in constructing the tight-binding Hamiltonian, since the adopted longitudinal curvilinear coordinates have different units: length in the cylindrical part and angle in the toroidal part. To avoid this we can rescale the longitudinal coordinate in the torus to get the units of length there as well:

$$\theta \rightarrow \tilde{\theta} = c \cdot \theta, \Delta_{1,tor} \rightarrow \tilde{\Delta}_{1,tor} = c \cdot \Delta_{1,tor}, \quad (4.27)$$

where  $\Delta_{1,tor}$  is the longitudinal grid size in the toroidal part. The constant  $c$  has units of length and if we choose  $c = \Delta_{1,cyl}/\Delta_{1,tor}$ , we can, in addition to equating the units, obtain a uniform grid along the cylindrical lead – toroidal nanoribbon – cylindrical lead system. The corresponding change in the metric tensor is

$$\tilde{g}_{11} = \frac{\partial \vec{r}}{\partial \tilde{\theta}} \frac{\partial \vec{r}}{\partial \tilde{\theta}} = \frac{1}{c^2} \frac{\partial \vec{r}}{\partial \theta} \frac{\partial \vec{r}}{\partial \theta} = \frac{g_{11}}{c^2}. \quad (4.28)$$

Similarly,  $\tilde{g}^{11} = c^2 g^{11}$  and  $\tilde{g} = g/c^2$ .

In Figs. 4.5-4.7 we present calculation results for several different nanoribbons in toroidal geometry, with and without magnetic field. The radii of the torus are  $R = 200$  nm and  $r = 40$  nm. The Fermi level is at 8 meV, while the grid size is  $575 \times 230$  for the toroidal part and  $50 \times 230$  for each of cylindrical leads. The leads extend downward from the nanoribbon as shown in the electron density plots. This particular geometry, in combination with the magnetic field in the  $x$ -direction [see the configuration in Fig. 4.5(b)] leads to one important point. Even though the

leads can be geometrically equivalent, due to the combination of curvature and magnetic field, *the calculated transverse mode profiles in the leads might not be the same*. This fact requires a slight modification of the transfer matrix method by Usuki *et al.* [87] [Eq. (2.17) there] to account for different mode-to-site conversion matrices in the incoming and outgoing leads. In Appendix D, we present the details of these necessary changes in case the two contacts cannot accommodate the same mode profile.

In Fig. 4.5, for the magnetic field along the  $x$ -axis (the largest flux through the leads), we see that the actual conductance follows the number of propagating modes for most of the magnetic field magnitudes applied (there is no nanoribbon helicity) and then undergoes oscillating behavior when the magnetic field is sufficiently strong so that there is only one propagating mode. This is a much more visible consequence of the resonant reflection, mentioned in connection to Fig. 4.4. In Fig. 4.5(b) electron density plots are shown for two characteristic points: zero conductance at  $B_z=4.456$  T (resonant reflection) and unit conductance (complete transmission) at  $B_z=4.646$  T. An extended quasibound state in the middle part of the nanoribbon is visible in the electron density plot in the  $\log_{10}$  scale at  $B_z=4.456$  T, while for the case of complete transmission there is an edge state which changes sides according to what can be expected from the classical Lorentz force.

Fig. 4.6 illustrates a situation where it is necessary to include gauge transformations, as explained in Sec. 4.1.4 and Appendix E.2. Since the magnetic field is in the  $y$ -direction (perpendicular to the plane of the torus and leads), the projection of the bisector (the middle longitudinal grid line of the nanoribbon) onto the plane normal to the magnetic field is a half-circle. In each slice perpendicular to the current flow, a local coordinate system (as depicted in Fig. 4.2) is rotated with respect to the reference system; each such rotation in the plane perpendicular to the magnetic field results in a gauge transformation. Therefore, we have a series of slice-dependent local Landau gauges as we move along the nanoribbon, which in this specific case means that  $A_2 = 0$ .

The edge state visible in the plot of electron density corresponds qualitatively (the side at which it is formed) to what can be expected from the classical Lorentz force and reflects well the symmetry of the problem. The details on how one performs the gauge transformations in this case in order to connect the local (slice-by-slice) Landau gauges are given in Appendix E.2.

In Fig. 4.7, we can see the results for a torodial nanoribbon, with and without helicity, in the absence of magnetic field. The normalized conductance for the nanoribbon without helicity agrees exactly with the number of injected propagating modes, meaning that there is complete transmission. Comparing the result of the conductance vs. Fermi level for the nanoribbons with and without helicity confirms that the observed difference between the normalized conductance and the number of propagating modes is due to helicity. As we also observed in Figs. 4.3 and 4.4, helicity plays a role similar to a quantum point contact in a flat wire (i.e. has the ability to quench propagation of certain modes). As shown earlier, [188] even without magnetic field one might expect resonant conductance features. The reason we do not see them here is because, due to the relatively large radius of the curvature of the nanoribbons considered here, the geometric potential is on the order of 0.1 meV, which is much smaller than the Fermi energy (8 meV). Another reason is that there are no sharp changes in curvature and, therefore, no sharp variations in the geometric vector potential, which can cause quantum interference effects.

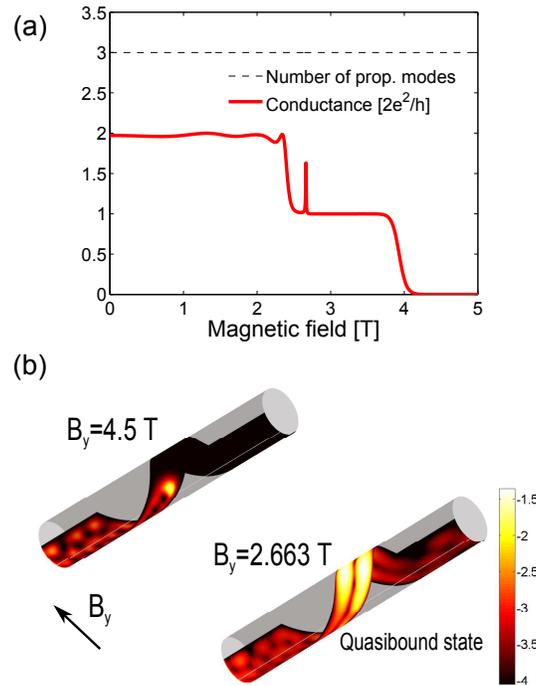


Figure 4.3 (a) Normalized conductance vs. magnetic field for a nanoribbon with helicity in cylindrical geometry. The leads are assumed cylindrical to avoid the need for transitional regions and as a consequence have the Landau levels energies affected by the magnetic field flux in the leads (Landau level quantization), albeit weakly. The lead on the left-hand side is the injecting lead. Magnetic field direction with respect to the structure is shown in (b) and would be parallel to the leads if they were exactly planar. Even at zero field, not all of the three injected modes propagate through, due to the ribbon helicity. (b) Electron density at the Fermi level shown for  $B_y = 2.663$  T on the  $\log_{10}$  scale, where there is resonant transmission, and for  $B_y = 4.5$  T, where the normalized conductance is zero (transmission suppressed). A quasibound state [right panel of (b)] is associated with the resonant transmission feature from (a).

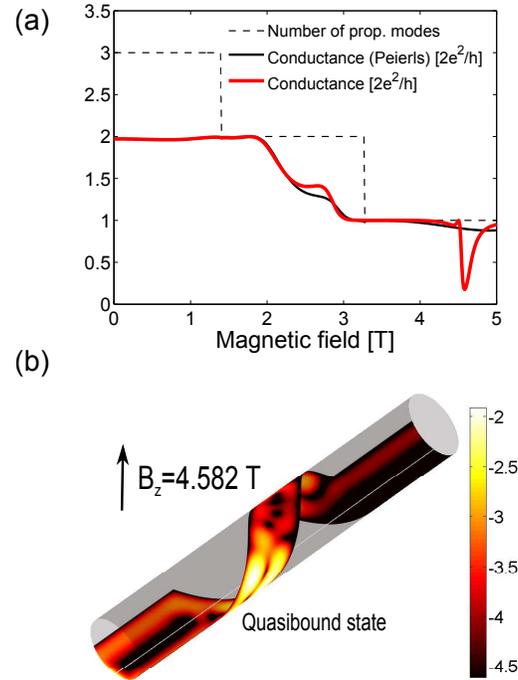


Figure 4.4 (a) Normalized conductance vs. magnetic field obtained with the exact equation (4.19) (thick solid red curve) and within the Peierls phase approach (thin solid black curve) for a nanoribbon with helicity in cylindrical geometry. The leads are assumed cylindrical to avoid the need for transitional regions. The lead on the left-hand side is the injecting lead. The magnetic field direction with respect to the structure is shown in (b) and would be normal to the leads if they were exactly flat. The number of injected propagating modes is shown with a thin dashed black curve and is influenced by the magnetic field flux in the leads (Landau level quantization). There is a discrepancy in the conductance between the exact and Peierls phase approaches starting at around 2.3 T. Resonant reflection is present at  $B = 4.582$  T in the exact (thick red) curve. (b) Electron density at the Fermi level, represented through color on the  $\log_{10}$  scale, corresponding to the resonant reflection at  $B = 4.582$  T. Quasibound state is formed, primarily confined to the right hand side of the helical part, looking from the left, injecting lead.

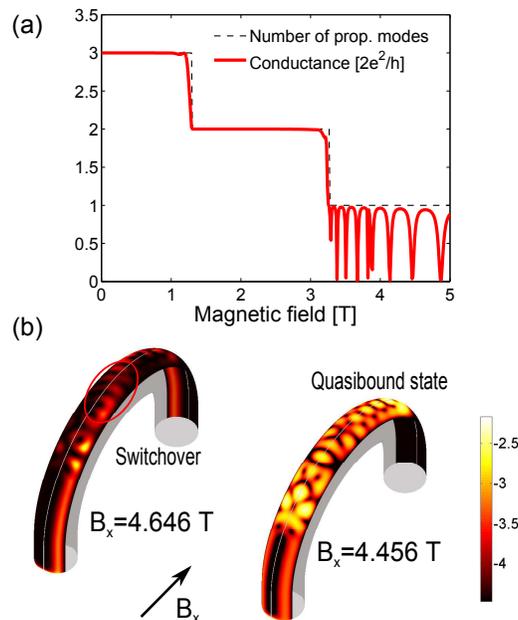


Figure 4.5 (a) Normalized conductance vs. magnetic field for a nanoribbon in toroidal geometry. The leads are assumed cylindrical to avoid the need for transitional regions. The number of propagating modes, given by the dashed line, is influenced by the magnetic field in the leads (Landau level quantization). There are strong resonant reflections at high magnetic fields, where there is only one injected mode. (b) Electron density at the Fermi level, represented through color, for unit conductance (complete transmission) at  $B_x=4.646$  T and zero conductance (resonant reflection) at  $B_x=4.456$  T. For complete transmission at  $B_x=4.646$  T, there is an edge channel formed that changes sides according to what can be expected from the classical Lorentz force. An extended quasibound state, responsible for the resonant reflection, is visible on the  $\log_{10}$  scale in the central region of the nanoribbon at  $B_x=4.456$  T.

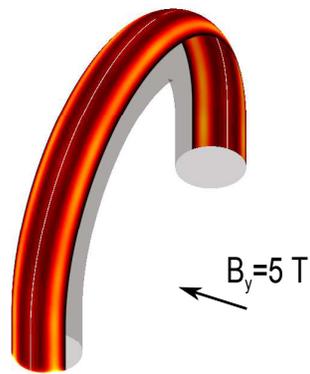


Figure 4.6 Sample electron density plot at the Fermi level for a toroidal nanoribbon subject to a magnetic field in the  $y$ -direction. Electron density is represented through color (light – high, dark – low). This situation illustrates the use of the gauge transformations, as explained in Sec. 4.1.4 and Appendix E.2, since the projection of the middle longitudinal grid line on the plane normal to the field is a half-circle, which requires the use of rotations in every slice of the toroidal part of the nanoribbon. The formed edge state reflects the symmetry of the problem and agrees qualitatively (the side at which it is formed) with the result obtained by applying the classical Lorentz force.

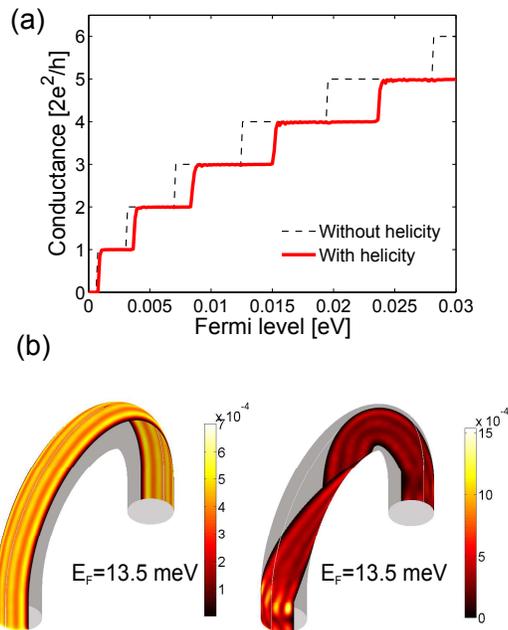


Figure 4.7 (a) Normalized conductance vs. Fermi level in the leads for a nanoribbon in toroidal geometry with and without helicity, in the absence of magnetic field. The leads are assumed cylindrical to avoid the need for transitional regions. The number of injected modes is changing due to the change in the Fermi energy. Comparison of the conductance for these two nanoribbons shows that helicity is the reason for the observed difference between the actual normalized conductance and the number of propagating modes. (b) Normalized electron density at the Fermi level, represented through color (white – high, black – low) for the nanoribbons with (right panel) and without helicity (left panel) when the Fermi level energy is equal to 13.5 meV.

### 4.3 Summary and Concluding Remarks

We presented a theoretical and numerical technique for investigating the linear, ballistic transport properties of curved nanoribbons in a static, uniform and arbitrarily directed magnetic field. Having started from the two-dimensional curvilinear Schrödinger equation with magnetic vector and electric scalar potentials included, after a rearrangement and rescaling of the Schrödinger equation we identified an effective potential, which is in general a complex quantity and depends on the magnetic field. The resulting equation was cast into a tight-binding form by using a second-order finite-difference scheme, which can be conveniently solved by a stabilized transfer matrix method. We analyzed the constraints on the Hermiticity of the discretized curvilinear Hamiltonian and proposed several suitable finite-difference schemes. While we adopted orthogonal coordinate systems throughout this paper, the framework is general enough that it can be extended to non-orthogonal coordinate systems if required.

Special attention was paid to the way in which the magnetic field should be included in the model. The requirement of working in the scattering-state basis, coupled with the structure's curvature, led us to a model for the inclusion of the magnetic field through *a local Landau gauge*. The use of the local Landau gauge is necessary to properly inject modes from the leads and avoid artificial numerical suppression of mode propagation, which can happen in finite-difference approaches to solving the Schrödinger equation in magnetic field, because an inadequate gauge in the effective complex potential *de facto* acts as a barrier to transmission. We have explained how slice-by-slice gauge transformations can be generated based on the structure's shape and the magnetic field direction.

Since the Peierls phase is often used to account for the magnetic field in planar nanostructures, we have also investigated the requirements for the validity of the Peierls phase approximation generalized to curved structures, where the magnetic field influence is captured through complex exponential terms. We find that there exists an upper limit on the magnetic field strength and curvature variations, connected to the structure's dimensions as well as the grid size, beyond which the Peierls phase approach is no longer valid.

We illustrated the method by numerically calculating the electron density at the Fermi level and conductance for several cylindrical and toroidal nanoribbons, with and without helicity, and subject to magnetic fields of different orientations. Interesting features that demonstrate the interplay between geometry and the magnetic field can be observed in these structures: for example, the prominent resonant reflection conductance features that occur due to a formation of bound states can be manipulated by curvature, helicity, and magnetic field.

## Chapter 5

### Summary and conclusions

The aim of this work is the development of a theoretical and computational framework for studying quantum transport in the transient regime and unconventional geometries. Conveniently, the work is split into two themes, Theme 1 dealing with the transient regime, and Theme 2 with the unconventional geometries. The transient regime modeling in solid state devices is important from the aspect of increasing the operating frequencies in both DC and AC regimes, especially having in mind the THz frequency window, currently lacking efficient electronic device for radiation, detection, signal manipulation and other functions. As the first step toward the goal of modeling the fully time-dependent quantum transport we developed a method to model the transient regime as a response to the sudden turn on of bias. On the other hand, curvature, being a novel mechanical aspect of device design, can have profound effects on device steady state coherent conduction, especially when the curved device is additionally placed in an external magnetic field.

The theory and results of Theme 1 are discussed in detail in Chapters 2 and 3. The physical systems in question are modeled in the following way: large, bulk reservoirs of charge with efficient scattering mechanisms are connected to the nanoscale electronic devices, which have the role of active regions. This configuration corresponds well to modern electronic devices that are the building blocks of integrated circuits. In that situation, the main steady state driving mechanism are efficient scattering mechanisms in the reservoirs (contacts). Our aim is to look at the coarse-grained time dynamics of the current limiting active regions attached to large reservoirs, through calculating the time-evolution of the current density, potential profile and electron density as the main electronic properties in device modeling. We develop the time-dependent quantum master equation and couple it to the Schrödinger, Poisson and current continuity equations. Subsequently,

we develop a numerical method to solve this non-linear system of equations efficiently. In addition, we construct proper initial conditions in time that result in the initial depletion of electrons in the device region. These electrons are replenished during the transient regime transport and cause the difference between the contact and device currents away from steady state. To that end we develop a simple model to account for the contact-to-device scattering that efficiently fills the localized device states in the depletion region.

In Ch. 4 we report the study of Theme 2 topics, quantum transport in curved devices in a magnetic field. We give a detailed account of the coherent, linear, steady state transport in curved nanoribbons in a magnetic field, along with the results for the conductance and electronic densities in several curved nanoribbons in cylindrical and toroidal geometries. First, a suitable tight-binding form of the curvilinear Hamiltonian is devised. This form is constrained by our wish to retain as much as possible the Hermitian properties of the exact curvilinear Hamiltonian. Since the numerical method to solve the Schrödinger equation for conductance and electron density introduces a preferred direction, we derive a special gauge, a local Landau gauge, that is compatible with the method. The numerical method itself is a modification of the method due to Usuki and coworkers [87], devised originally for planar nanostructures. Our results for electron density and conductance show that curvature, coupled with an external magnetic field, has a large effect on the device's conduction properties. Based on examples involving cylindrical and toroidal geometries, with and without helicity, we observe conduction suppression, reminding of the quantum point contact properties, and resonant reflections observed before in various planar quantum cavities and similar non-uniform 2D nanostructures. The effects on the conduction are visible even without the magnetic field, being then controlled only by the curvature and its effects on the coherent quantum transport.

This work can be expanded in several directions, depending on the importance and impact one assigns to them. As far as Theme 1 is concerned, possible topics of interest, ordered approximately from less to more time consuming, that build on the current work include: (1) microscopic modeling of the contact to device scattering, either by supplementing the current quantum master

equation in the spirit of rate equations, or deriving a new equation from scratch; (2) carefully modeling the contact scattering mechanisms and relaxation rates, by possibly coupling a Monte Carlo semiclassical contact simulation with the quantum mechanical treatment of the device, based on the quantum master equation; (3) Modeling effects that arise beyond the current coarse-grained time scale, going to even shorter time scales. This would require new quantum master equation, which would be non-Markovian from the start. As far as Theme 2 is concerned possible future topics, again ordered in a similar fashion as the topics for Theme 1, include: (1) establishing collaboration with experimental groups with access to high magnetic field facilities and attempting to model and match experimental results; (2) lifting the current constraint of the linear regime quantum transport, valid at low temperature and bias, and coupling the curvilinear Schrödinger equation with the Poisson equation, to obtain self-consistency.

## LIST OF REFERENCES

- [1] M. V. Fischetti, J. Appl. Phys. **83**, 270 (1998).
- [2] I. Knezevic, Phys. Rev. B **77**, 125301 (2008).
- [3] A. M. Kriman, M. J. Kann, D. K. Ferry, and R. Joshi, Phys. Rev. Lett. **65**, 1619 (1990).
- [4] P. Lugli and D. K. Ferry, IEEE Trans. Electron Devices **32**, 2431 (1985).
- [5] P. Lugli and D. K. Ferry, Phys. Rev. Lett. **56**, 1295 (1986).
- [6] W. Pötz, J. Appl. Phys. **66**, 2458 (1989).
- [7] W. R. Frensley, Rev. Mod. Phys. **62**, 745 (1990).
- [8] H.-P. Breuer and F. Petruccione, *The Theory of Open Quantum Systems* (Oxford University Press, Oxford, 2002).
- [9] L. E. Reichl, *A Modern Course in Statistical Physics* (WILEY-VCH, Weinheim, 1980).
- [10] M. Bonitz, *Quantum Kinetic Theory* (Teubner, Stuttgart; Leipzig, 1998).
- [11] N. C. Kluksdahl, A. M. Kriman, D. K. Ferry, and C. Ringhofer, Phys. Rev. B **39**, 7720 (1989).
- [12] S. E. Laux, A. Kumar, and M. V. Fischetti, IEEE Trans. Nanotechnol. **1**, 255 (2002).
- [13] A. Gehring and S. Selberherr, J. Comput. Electron. **3**, 149 (2004).
- [14] J. von Neumann, *Mathematical Foundations of Quantum Mechanics* (Princeton University Press, Princeton, 1955).
- [15] E. Wigner, Phys. Rev. **40**, 749 (1932).
- [16] L. P. Kadanoff and G. Baym, *Quantum Statistical Mechanics* (Benjamin, New York, 1962).
- [17] L. V. Keldysh, Sov. Phys. JETP **20**, 1018 (1965).
- [18] W. Pauli, *Festschrift zum 60. Geburtstage A. Sommerfeld* (Hirzel, Leipzig, 1928), p. 30.

- [19] K. Huang, *Statistical Mechanics* (John Wiley & Sons, New York, 1987).
- [20] O. G. Schmidt and K. Eberl, *Nature* **410**(6825), 168 (March 2001).
- [21] A. J. Baca, J. H. Ahn, Y. G. Sun, M. A. Meitl, E. Menard, H. S. Kim, W. M. Choi, D. H. Kim, Y. Huang, and J. A. Rogers, *Angew. Chem. Int. Ed.* **47**(30), 5524 (2008).
- [22] N. Shaji, H. Qin, R. H. Blick, L. J. Klein, C. Deneke, and O. G. Schmidt, *Appl. Phys. Lett.* **90**, 042101 (2007).
- [23] F. Cavallo, R. Songmuang, and O. G. Schmidt, *Appl. Phys. Lett.* **93**, 143113 (2008).
- [24] S. A. Scott and M. G. Lagally, *J. Phys. D: Appl. Phys.* **40**, R75 (2007).
- [25] X. Li, *J. Phys. D: Appl. Phys.* **41**, 193001 (2008).
- [26] V. Y. Prinz, V. A. Seleznev, A. K. Gutakovskiy, A. V. Chehovskiy, V. V. Preobrazhenskii, M. A. Putyato, and T. A. Gavrilova, *Physica E* **6**(1-4), 828 (February 2000).
- [27] S. V. Golod, V. Y. Prinz, V. I. Mashanov, and A. K. Gutakovskiy, *Semicond. Sci. Technol.* **16**, 181 (2001).
- [28] M. H. Huang, C. Boone, M. Roberts, D. E. Savage, M. G. Lagally, N. Shaji, H. Qin, R. Blick, J. A. Nairn, and F. Liu, *Adv. Mater.* **17**, 2860 (2005).
- [29] O. G. Schmidt, C. D. Y. M. Manz, and C. Muller, *Physica E* **13**, 969 (2002).
- [30] A. Lorke, S. Bohm, and W. Wegscheider, *Superlattice. Microst.* **33**, 347 (2003).
- [31] A. L. Ivanovskii, *USP. KHIM.* **71**, 203 (2002).
- [32] J. Goldberger, R. R. He, Y. F. Zhang, S. W. Lee, H. Q. Yan, H. J. Choi, and P. D. Yang, *Nature* **422**, 599 (2003).
- [33] A. K. Roy, S. Knohl, and W. A. Goedel, *J. Mater. Sci.* **46**, 4812 (2011).
- [34] H. W. Kim, H. G. Na, and J. C. Yang, *Chem. Eng. J.* **170**, 326 (2011).
- [35] H. M. Fan, J. B. Yi, Y. Yang, K. W. Kho, H. R. Tan, Z. X. Shen, J. Ding, X. W. Sun, M. C. Olivo, and Y. P. Feng, *ACS NANO* **3**, 2798 (2009).
- [36] H. Yuan, Z. Ma, M. M. Roberts, D. E. Savage, and M. G. Lagally, *J. Appl. Phys.* **100**, 013708 (2006).
- [37] H. Yuan, G. K. Celler, and Z. Ma, *J. Appl. Phys.* **102**, 034501 (2007).
- [38] D. Kim, J. Ahn, W. M. Choi, H. Kim, T. Kim, J. Song, Y. Y. Huang, Z. Liu, C. Lu, and J. A. Rogers, *Science* **320**, 507 (2008).

- [39] H. C. Ko, M. P. Stoykovich, J. Song, V. Malyarchuk, W. M. Choi, C.-J. Yu, J. G. III, J. L. Xiao, S. Wang, Y. G. Huang, *et al.*, *Nature* **454**, 748 (2008).
- [40] E. Menard, R. G. Nuzzo, and J. A. Rogers, *Appl. Phys. Lett.* **86**, 093507 (2005).
- [41] J.-U. Park, M. Hardy, S. J. Kang, K. Barton, K. Adair, D. K. Mukhopadhyay, C. Y. Lee, M. S. Strano, A. G. Alleyne, J. G. Georgiadis, *et al.*, *Nature Mater.* **6**, 782 (2007).
- [42] Y. Mei, G. S. Huang, A. A. Solovev, E. B. Urena, I. Moench, F. Ding, T. Reindl, R. K. Y. Fu, P. K. Chu, and O. G. Schmidt, *Adv. Materials* **20**(21), 4085 (November 2008).
- [43] T. Kipp, H. Welsch, C. Strelow, C. Heyn, and D. Heitmann, *Phys. Rev. Lett.* **96**, 077403 (2006).
- [44] S. Mendach, R. Songmuang, S. Kiravittaya, A. Rastelli, M. Benyoucef, and O. G. Schmidt, *Appl. Phys. Lett.* **88**, 111120 (2006).
- [45] S. Vicknesh, F. Li, and Z. Mi, *Appl. Phys. Lett.* **94**, 081101 (2009).
- [46] G. S. Huang, S. Kiravittaya, V. A. B. Quinones, F. Ding, M. Benyoucef, A. Rastelli, Y. F. Mei, and O. G. Schmidt, *Appl. Phys. Lett.* **94**, 141901 (2009).
- [47] M. Hosoda and T. Shigaki, *Appl. Phys. Lett.* **90**, 181107 (2007).
- [48] F. Li, S. Vicknesh, and Z. Mi, *Electron Lett.* **45**, 645 (2009).
- [49] Y. J. Xing, Z. H. Xi, Z. Q. Xue, X. D. Zhang, J. H. Song, R. M. Wang, J. Xu, Y. Song, S. L. Zhang, and D. P. Yu, *Appl. Phys. Lett.* **83**, 1689 (2003).
- [50] I. S. Chun, K. Bassett, A. Challa, and X. L. Li, *Appl. Phys. Lett.* **96**, 251106 (2010).
- [51] F. Li, Z. T. Mi, and S. Vicknesh, *Opt. Lett.* **34**, 2915 (2009).
- [52] F. Li and Z. T. Mi, *Opt. Express* **17**, 19933 (2009).
- [53] S. Schwaiger, M. Broll, A. Krohn, A. Stemmann, C. Heyn, Y. Stark, D. Stickler, D. Heitmann, and S. Mendach, *Phys. Rev. Lett.* **102**, 163903 (2009).
- [54] D. V. Bulaev, V. A. Geyler, and V. A. Margulis, *Phys. Rev. B* **69**, 195313 (2004).
- [55] C. Deneke, J. Schumann, R. Engelhard, J. Thomas, C. Muller, M. S. Khatri, A. Malachias, M. Weisser, T. H. Metzger, and O. G. Schmidt, *Nanotechnology* **20**, 045703 (2009).
- [56] C. Muller, M. S. Khatri, C. Deneke, S. Fahler, Y. F. Mei, E. B. Urena, and O. G. Schmidt, *APL* **94**, 102510 (2009).
- [57] T. R. Ger, C. C. Huang, H. T. Huang, and Z. H. Wei, *J. Appl. Phys.* **109**, 07E534 (2011).

- [58] L. I. Magarill, D. A. Romanov, and A. V. Chaplik, *J. Exp. Theor. Phys.* **86**, 771 (1998).
- [59] L. I. Magarill and A. V. Chaplik, *J. Exp. Theor. Phys.* **88**, 815 (1999).
- [60] S. Mendach, J. Podbielski, J. Topp, W. Hansen, and D. Heitmann, *Appl. Phys. Lett.* **93**, 262501 (2008).
- [61] F. Balhorn, S. Mansfeld, A. Krohn, J. Topp, W. Hansen, D. Heitmann, and S. Mendach, *Phys. Rev. Lett.* **104**, 037205 (2010).
- [62] S. Ono and H. Shima, *Phys. Rev. B* **79**, 235407 (2009).
- [63] S. Ono and H. Shima, *Physica E* **42**, 1224 (2010).
- [64] H. Taira and H. Shima, *Surf. Sci.* **601**, 5270 (2007).
- [65] G. Ferrari, A. Bertoni, G. Goldoni, and E. Molinari, *Phys. Rev. B* **78**, 115326 (2008).
- [66] V. Atanasov, R. Dandoloff, and A. Saxena, *Phys. Rev. B* **79**, 033404 (2009).
- [67] C. Ortix and J. van den Brink, *Phys. Rev. B* **81**, 165419 (2010).
- [68] G. D. Valle and S. Longhi, *J. Phys. B-At. Mol. Opt.* **43**, 051002 (2010).
- [69] C. L. Foden, M. L. Leadbeater, and M. Pepper, *Phys. Rev. B* **52**, R8646 (1995).
- [70] V. A. Margulis and M. A. Pyataev, *Phys. Rev. B* **72**, 075312 (2005).
- [71] G. J. Meyer, N. L. Dias, R. H. Blick, and I. Knezevic, *IEEE Trans. Nanotechnol.* **6**(4), 446 (2007).
- [72] G. Papp and F. M. Peeters, *Solid State Commun.* **149**, 778 (2009).
- [73] M. L. Leadbeater, C. L. Foden, J. H. Burroughes, M. Pepper, T. M. Burke, L. L. Wang, M. P. Grimshaw, and D. A. Ritchie, *Phys. Rev. B* **52**, R8629 (1995).
- [74] Y. Meir and N. S. Wingreen, *Physica E* **23**, 274 (2004).
- [75] A. V. Chaplik and R. H. Blick, *New J. Phys.* **6**, 33 (2004).
- [76] A. B. Vorob'ev, K. J. Friedland, H. Kostial, R. Hey, U. Jahn, E. Wiebicke, J. S. Yukecheva, and V. Y. Prinz, *Phys. Rev. B* **75**, 205309 (2007).
- [77] K. J. Friedland, R. Hey, H. Kostial, A. Riedel, and K. H. Ploog, *Phys. Rev. B* **75**, 045347 (2007).
- [78] R. C. T. da Costa, *Phys. Rev. A* **23**, 1982 (1981).
- [79] M. Encinosa, *Phys. Rev. A* **73**, 012102 (2006).

- [80] H. Jensen and H. Koppe, *Ann. Phys.* **63**, 586 (1971).
- [81] M. Ikegami, Y. Nagaoka, S. Takagi, and T. Tanzawa, *Prog. Theor. Phys.* **88**, 229 (1992).
- [82] G. Ferrari and G. Cuoghi, *Phys. Rev. Lett.* **100**, 230403 (2008).
- [83] R. Landauer, *IBM J. Res. Develop.* **1**, 233 (1957).
- [84] M. Büttiker, *Phys. Rev. Lett.* **57**, 1761 (1986).
- [85] R. Landauer, *Z. Phys. B* **68**, 217 (1987).
- [86] S. Datta, *Electronic Transport in Mesoscopic Systems* (Cambridge University Press, 1995).
- [87] T. Usuki, M. Saito, M. Takatsu, R. A. Kiehl, and N. Yokoyama, *Phys. Rev. B* **52**, 8244 (1995).
- [88] S. Brand and D. T. Hughes, *Semicond. Sci. Technol.* **2**, 607 (1987).
- [89] C. Mailhot and D. L. Smith, *Phys. Rev. B* **33**, 8360 (1986).
- [90] H. Tamura and T. Ando, *Phys. Rev. B* **44**, 1792 (1991).
- [91] David Yuk Kei Ko and J. C. Inkson, *Phys. Rev. B* **38**, 9945 (1988).
- [92] T. Ando, *Phys. Rev. B* **44**, 8017 (1991).
- [93] M. V. Fischetti, *Phys. Rev. B* **59**, 4901 (1999).
- [94] S. E. Laux, A. Kumar, and M. V. Fischetti, *J. Appl. Phys.* **95**, 5545 (2004).
- [95] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery.
- [96] B. Novakovic and I. Knezevic, *Fortschr. Phys.* (2012), doi: 10.1002/prop.201200071.
- [97] B. Novakovic and I. Knezevic, in D. Vasileska and S. M. Goodnick, eds., *Nano-Electronic Devices: Semiclassical and Quantum Transport Modeling* (New York: Springer, 2011), pp. 249–287.
- [98] B. Novakovic and I. Knezevic, in *Proc. 8th IEEE Conf. Nanotechnology (NANO '08), Arlington, TX* (2008).
- [99] B. Novakovic and I. Knezevic, *Transient-regime quantum transport in two-terminal nanostructures* (2010), poster presentation at the 14th Int. Workshop Computational Electronics, Pisa, Italy.
- [100] B. Novakovic and I. Knezevic, *Transient response of a quantum point contact due to the coupling with reservoirs* (2011), oral presentation at the 2011 APS March Meeting, Dallas, TX.

- [101] B. Novakovic and I. Knezevic, *Coherent electron transport in planar and curved nanostructures and decoherence due to the coupling with contacts* (2011), oral presentation at the 17th Int. Conf. Electron Dynamics in Semiconductors, Optoelectronics and Nanostructures (EDISON 17), Santa Barbara, CA.
- [102] B. Novakovic and I. Knezevic, *Transient-regime transport in nanostructures* (2011), oral presentation at the Frontiers of Quantum and Mesoscopic Thermodynamics (FQMT '11), Prague, Czech Republic.
- [103] B. Novakovic and I. Knezevic, *Modeling transients in nanostructures* (2012), oral presentation at the 15th International Workshop on Computational Electronics (IWCE), Madison, WI.
- [104] I.-H. Tan, G. L. Snider, L. D. Chang, and E. L. Hu, *J. Appl. Phys.* **68**, 4071 (1990).
- [105] M. Luisier, A. Schenk, and W. Fichtner, *J. Appl. Phys.* **100**, 043713 (2006).
- [106] M. Luisier and A. Schenk, *IEEE Trans. Electron Devices* **55**, 1494 (2008).
- [107] B. Novakovic, R. Akis, and I. Knezevic, *Phys. Rev. B* **84**, 195419 (2011).
- [108] B. Novakovic, K. J. Messer, and I. Knezevic, *Magnetotransport in curved 2degs using a real-space numerical technique* (2010), oral presentation at the 2010 APS March Meeting, Portland, OR.
- [109] B. Novakovic, K. J. Messer, R. Akis, and I. Knezevic, in *Proc. ICCES'10 (Int. Conf. Computational and Experimental Engineering and Sciences)*, Las Vegas, NV (2010).
- [110] B. Novakovic and I. Knezevic, *Transport properties of curved nanostructures in magnetic field* (2010), poster presentation at the 14th Int. Workshop Computational Electronics, Pisa, Italy.
- [111] M.-H. Huang, C. S. Ritz, B. Novakovic, D. Yu, Y. Zhang, F. Flack, D. E. Savage, P. G. Evans, I. Knezevic, F. Liu, *et al.*, *ACS Nano* **3**, 721 (2009).
- [112] M. P. Kirley, B. Novakovic, N. Sule, M. J. Weber, I. Knezevic, and J. H. Booske, *J. Appl. Phys.* **111**, 063717 (2012).
- [113] M. Huang, C. S. Ritz, B. Novakovic, F. Flack, D. E. Savage, P. G. Evans, D. Y. I. Knezevic, Y. Zhang, F. Liu, and M. G. Lagally, *Mechano-electronic superlattices in silicon nanoribbons* (2009), 2009 APS March Meeting, Pittsburgh, PA.
- [114] C. Ritz, M.-H. Huang, B. Novakovic, H.-J. Kim-Lee, N. Appathurai, R. Metzler, P. Gilbert, K. Turner, I. Knezevic, and M. G. Lagally, *Controlling mechano-lectronic superlattices in hybrid nanowire ribbon/quantum dot structures* (2009), 2009 MRS Fall Meeting, Boston, MA.

- [115] N. Sule, M. Kirley, B. Novakovic, J. Scharer, I. Knezevic, and J. H. Booske, *Examination of field emission from copper knife edge cathodes with low-work function coatings* (2010), 2010 IEEE Int. Vacuum Electronics Conf. (IVEC 2010), Monterey, CA.
- [116] M. Kirley, B. Novakovic, M. Weber, N. Sule, J. Scharer, I. Knezevic, and J. H. Booske, *Examination of field emission from lanthanum hexaboride coated knife edge cathodes* (2011), 2011 IEEE Int. Vacuum Electronics Conf. (IVEC 2011), Bangalore, India.
- [117] U. Weiss, *Quantum Dissipative Systems* (World Scientific, Singapore, 1999).
- [118] J. von Neumann, *Nachr. Ges. Wiss.* p. 245 (1927).
- [119] J. J. Sakurai, *Modern Quantum Mechanics* (Addison–Wesley, USA, 1994).
- [120] A. L. Fetter and J. D. Walecka, eds., *Quantum theory of many-particle systems* (McGraw-Hill, Inc., 1971).
- [121] A. Messiah, *Quantum Mechanics* (Dover, Mineola, 1999).
- [122] L. van Hove, *Rev. Mod. Phys.* **21**, 517 (1955).
- [123] W. Kohn and J. M. Luttinger, *Phys. Rev.* **108**, 590 (1957).
- [124] R. Brunetti, C. Jacoboni, and F. Rossi, *Phys. Rev. B* **39**, 10781 (1989).
- [125] F. Rossi and C. Jacoboni, *Europhys. Lett.* **18**, 169 (1992).
- [126] C. Jacoboni, *Semicond. Sci. Technol.* **7**, B6 (1992).
- [127] J. R. Barker and D. K. Ferry, *Phys. Rev. Lett.* **42**, 1779 (1979).
- [128] I. B. Levinson and Y. Yasevichyute, *Sov. Phys. JETP* **35**, 991 (1972).
- [129] W. V. Houston, *Phys. Rev.* **57**, 184 (1940).
- [130] P. J. Price, *Semicond. Semimet.* **14**, 249 (1979).
- [131] C. Jacoboni and L. Reggiani, *Rev. Mod. Phys.* **55**, 645 (1983).
- [132] M. V. Fischetti, S. E. Laux, P. M. Solomon, and A. Kumar, *J. Comput. Electron.* **3**, 287 (2004).
- [133] D. Vasileska and S. M. Goodnick, *Mater. Sci. Eng.* **R 38**, 181 (2002).
- [134] H. Budd, *Phys. Rev.* **158**, 798 (1967).
- [135] R. Chambers, *Proc. R. Soc. London* **65**, 458 (1952).

- [136] M. Lundstrom, *Fundamentals of Carrier Transport* (Cambridge University Press, Cambridge, 2000).
- [137] S. Nakajima, *Prog. Theor. Phys.* **20**, 948 (1958).
- [138] R. Zwanzig, *J. Chem. Phys.* **33**, 1338 (1960).
- [139] A. G. Redfield, *IBM J. Res. Dev.* **1**, 19 (1957).
- [140] G. Lindblad, *Commun. Math. Phys.* **48**, 199 (1976).
- [141] P. Argyres and P. Kelley, *Phys. Rev.* **134**, A98 (1964).
- [142] J. R. Barker and D. K. Ferry, *Solid-State Electron.* **23**, 531 (1980).
- [143] K. Kassner, *Phys. Rev. A* **36**, 5381 (1987).
- [144] M. Sparpaglione and S. Mukamel, *J. Chem. Phys.* **88**, 3263 (1988).
- [145] Y. Hu and S. Mukamel, *J. Chem. Phys.* **91**, 6973 (1989).
- [146] V. Romero-Rochin and I. Oppenheim, *Physica A* **155**, 52 (1989).
- [147] M. Tokuyama and H. Mori, *Prog. Theor. Phys.* **55**, 411 (1975).
- [148] F. Shibata, Y. Takahashi, and N. Hashitsume, *J. Stat. Phys.* **17**, 171 (1977).
- [149] N. Hashitsume, F. Shibata, and M. Shingu, *J. Stat. Phys.* **17**, 155 (1977).
- [150] M. Saeki, *Prog. Theor. Phys.* **67**, 1313 (1982).
- [151] M. Saeki, *Prog. Theor. Phys.* **79**, 396 (1988).
- [152] M. Saeki, *Prog. Theor. Phys.* **89**, 607 (1993).
- [153] D. Ahn, *Phys. Rev. B* **51**, 2159 (1995).
- [154] D. Ahn, *Prog. Quantum Electron.* **21**, 249 (1997).
- [155] D. Ahn, J. H. Oh, K. Kimm, and S. Hwang, *Phys. Rev. A* **61**, 052310 (2000).
- [156] D. Ahn, J. Lee, M. S. Kim, and S. W. Hwang, *Phys. Rev. A* **66**, 012302 (2002).
- [157] T. M. Chang and J. L. Skinner, *Physica A* **193**, 483 (1993).
- [158] A. A. Golosov and D. R. Reichmann, *J. Chem. Phys.* **115**, 9849 (2001).
- [159] I. Knezevic and D. K. Ferry, *Phys. Rev. E* **66**, 016131 (2002).
- [160] I. Knezevic and D. K. Ferry, *Phys. Rev. E* **67**, 066122 (2003).

- [161] D. Semkat, D. Kremp, and M. Bonitz, *Phys. Rev. E* **59**, 1557 (1999).
- [162] D. Semkat, D. Kremp, and M. Bonitz, *J. Math. Phys.* **41**, 7458 (2000).
- [163] K. M. et al., *Phys. Rev. E* **63**, 020102(R) (2001).
- [164] W. T. Reid, *Riccati Differential Equations* (Academic Press, New York, 1972).
- [165] S. Bittanti, A. J. Laub, and J. C. Willems, eds., *The Riccati Equation* (Springer-Verlag, Berlin, 1991).
- [166] I. Knezevic and D. K. Ferry, *Phys. Rev. A* **69**, 012104 (2004).
- [167] G. Lindblad, *J. Phys. A* **29**, 4197 (1996).
- [168] E. B. Davies, *Commun. Math. Phys.* **39**, 91 (1974).
- [169] X. Q. Li, J. Y. Luo, Y. G. Yang, P. Cui, and Y. J. Yan, *Phys. Rev. B* **71**, 205304 (2005).
- [170] J. N. Pedersen and A. Wacker, *Phys. Rev. B* **72**, 195330 (2005).
- [171] D. Bacon, D. A. Lidar, and K. B. Whaley, *Phys. Rev. A* **60**, 1944 (1999).
- [172] D. A. Lidar, Z. Bihary, and K. B. Whaley, *Chem. Phys.* **268**, 35 (2001).
- [173] R. Alicki, *Phys. Rev. A* **40**, 4077 (1989).
- [174] D. K. Ferry and S. M. Goodnick, *Transport in Nanostructures* (Cambridge University Press, Cambridge, UK, 1997).
- [175] D. K. Ferry, R. Akis, J. P. Bird, M. Elhassan, I. Knezevic, C. Prasad, and A. Shailos, *J. Vac. Sci. Technol. B* **21**, 1891 (2003).
- [176] G. Ferrari, N. Giacobbi, P. Bordone, A. Bertoni, and C. Jacoboni, *Semicond. Sci. Technol.* **19**, S254 (2004).
- [177] R. Gebauer and R. Car, *Phys. Rev. Lett.* **93**, 160404 (2004).
- [178] G. A. Korn and T. M. Korn, *Mathematical Handbook for Scientists and Engineers; Definitions, Theorems, and Formulas for Reference and Review* (McGraw-Hill, New York, 1968).
- [179] N. A. Modine, G. Zumbach, and E. Kaxiras, *Phys. Rev. B* **55**, 10289 (1997).
- [180] F. Gygi and G. Galli, *Phys. Rev. B* **52**, R2229 (1995).
- [181] H. U. Baranger and A. D. Stone, *Phys. Rev. B* **40**, 8169 (1989).
- [182] D. Guan, U. Ravaioli, R. W. Giannetta, M. Hannan, I. Adesida, and M. R. Melloch, *Phys. Rev. B* **67**, 205328 (2003).

- [183] K. F. Riley, M. P. Hobson, and S. J. Bence, *Mathematical methods for physics and engineering* (Cambridge University Press, Cambridge, 1997), p. 722.
- [184] V. Y. Prinz and S. V. Golod, *J. Appl. Mech. Tech. Phys.* **47**, 867 (2006).
- [185] W. Porod, Z. Shao, and C. S. Lent, *Appl. Phys. Lett.* **61**, 1350 (1992).
- [186] H. Wu and D. W. L. Sprung, *Phys. Rev. B* **47**, 1500 (1993).
- [187] R. Akis, P. Vasilopoulos, and P. Debray, *Phys. Rev. B* **52**, 2805 (1995).
- [188] A. Marchi, S. Reggiani, M. Rudan, and A. Bertoni, *Phys. Rev. B* **72**, 035403 (2005).
- [189] (1972).
- [190] V. I. Tselyaev, *J. Comput. Appl. Math.* **170**, 103 (2004).
- [191] L. D. Landau and E. M. Lifshitz, *Quantum Mechanics - Non-relativistic Theory*, vol. 3 (Pergamon Press, 1977).
- [192] C. Kittel, *Introduction to Solid State Physics* (John Wiley and Sons, Hoboken, 2005).
- [193] R. Lake, G. Klimeck, R. C. Bowen, and D. Jovanovic, *J. Appl. Phys.* **81**, 7845 (1997).
- [194] R. Akis, R. Barnes, and G. Kirczenow, *Can. J. Phys.* **73**, 147 (1995).
- [195] D. K. Ferry, *Quantum Mechanics: An Introduction for Device Physicists and Electrical Engineers* (IOP Publishing, Bristol, UK, 2001), p. 192.

## Appendix A: The details of the algorithm for the transient-regime calculations

Before attempting a numerical solution to Eqs. (3.5), we discretize each equation by using a finite-difference method and transform them into a matrix form. The Poisson equation is discretized using a standard second-order central difference scheme. Eq. (3.5c) describing the time evolution of distribution functions is discretized in time using a backward difference scheme and solving the resulting algebraic equation for the distribution function at the new time step. The discretized form of Eq. (3.5c) at time  $i\Delta t$  is given by:

$$f_k(i) = \frac{f_k(i-1) + \Delta t \tau \Delta_k^2 f_k^L(i)}{1 + \Delta t \tau \Delta_k^2}, f_{-k}(i) = \frac{f_{-k}(i-1) + \Delta t \tau \Delta_{-k}^2 f_{-k}^R(i)}{1 + \Delta t \tau \Delta_{-k}^2}, \quad (\text{A.1})$$

where  $\Delta t$  is the chosen time step. In this equation, the time-dependent contact distribution functions are calculated according to the instantaneous value of  $k_d(i\Delta t)$ , which is to be found self-consistently as explained below. The Schrödinger equation is discretized using a 6th order, three-point Numerov method. This higher order scheme is chosen in order to ensure sufficient accuracy in calculating the wave functions with small wave lengths and the dispersion relation. This method has slightly different forms depending on whether the effective mass is constant [189], or spatially dependent [190], like in heterostructures.

The solution to the Schrödinger equation requires a more in depth explanation. In open systems, such as devices that carry current, the energy spectrum is continuous. In principle, since all energies are allowed, one can solve the Schrödinger equation as a differential equation with known energy  $E$  and not as an eigenvalue problem. However, one must know how to properly normalize the wave functions in order to restore a proper density of states. If the energy spectrum is continuous and the wave functions are represented as plane wave scattering states, as is here, the textbook normalization method would be to normalize the wave functions to a delta function in the wave vector space. In other words, plane waves cannot be properly normalized, since they are completely delocalized in real space, and one would have problems in attempting to find a self-consistent solution to the Poisson and Schrödinger equations. For this reason we employ a normal mode, or standing wave, energy spectrum discretization method, as introduced in [1, 93, 94]. We

will give a brief overview of the method in the following few sentences. The boundary conditions are chosen such that the wave function has some fixed phase at the contact-device interface. For each phase there will be one set of normal modes, represented by a set of discrete energies and real wave functions, upon solving the eigenvalue problem. This means that the wave function is essentially localized to the device region,  $\int_0^L |\psi|^2 dx = 1$ . This is similar to a closed system, and therefore the energy spectrum discretization, giving the density of levels, is now consistent with the wave function normalization and a proper electron density can be obtained. In order to increase the accuracy, one can choose  $N$  phases, or  $N$  sets of discrete energy levels and associated standing waves, in which case the normalization becomes  $\int_0^L |\psi|^2 dx = 1/N$ . Since real standing waves are obtained, instead of complex traveling waves necessary for current flow, one has to additionally decompose each standing wave into traveling waves injected from different contacts.

Having the numerical solutions of each equation, it is necessary to couple them and obtain the self-consistent solution. A detailed algorithm, incorporating the initial conditions, is shown in the flowchart in Fig. A.1. The algorithm starts just after applying bias, which we refer to as  $t = 0^+$  s moment. First, we need some initial guess for the potential. A satisfactory guess can be obtained using the so called Thomas-Fermi approximation from physics (e.g. [191, 192]), which in device physics amounts to solving the Poisson equation with the charge density represented by the formula for its bulk value, assuming a spatially dependent Fermi level which is in a local quasi-equilibrium. This initial guess is a crude approximation whenever the current density is non-negligible and the quantum effects are pronounced, but it serves the purpose as a guess that will lead to the proper solution. With this guess the algorithm proceeds to the next block, which solves for  $n(\mathbf{r}, 0^+)$  and  $V(\mathbf{r}, 0^+)$  according to the initial conditions discussed in Sec. 3.2.2. Only the Poisson and Schrödinger equations are solved here using the distribution functions in the device region given by Eqs. (3.9), since  $J_{con}(0^+) = J_{dev}(0^+) = 0$  and  $k_d(0^+) = 0$ . The quantities  $V(\mathbf{r}, 0^+)$ ,  $n(\mathbf{r}, 0^+)$ , and  $k_d(0^+)$ , are subsequently used as an initial guess for the next time step  $\Delta t$  ( $i = 1$ ), where the full system of Eqs. (3.5) must be solved. This part of the algorithm consists of the outer loop, which solves for  $V$ , and the inner loop, which solves for  $k_d$ . The outer loop is solved by the globally convergent Newton method, while the inner loop with the simpler mixing method,

the details of which will be given in the following paragraph. After the outer loop converges for  $i = 1$ , we increase the counter by one,  $i = i + 1$ , and repeat the same procedure, where  $V(\mathbf{r}, \Delta t)$ ,  $n(\mathbf{r}, \Delta t)$ , and  $k_d(\Delta t)$  are used as an initial guess to calculate  $V(\mathbf{r}, 2\Delta t)$ ,  $n(\mathbf{r}, 2\Delta t)$ , and  $k_d(2\Delta t)$ . The counter is being increased until the steady state is reached.

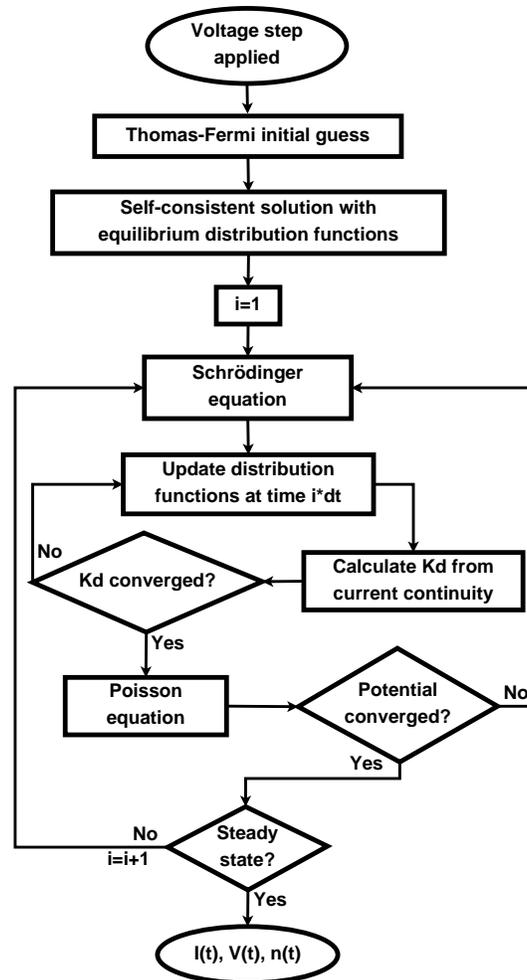


Figure A.1 Flowchart of the numerical algorithm for the calculation of the electronic response of a biased two-terminal nanostructure during a transient.

The globally convergent Newton method [95] is a stable and robust method to find the root of non-linear equations. As in the ordinary Newton method we need to know the direction in which to look for the new solution guess, given the old one, and the step size, which says how far we go

along that direction. It is easy to imagine that, even if we know the direction very well, if the step size is too big we can run into stability problems. The globally convergent method chooses the step size in each iteration such that we are always getting closer to the root. A potential drawback of this method is that one has to calculate the Jacobian that gives the direction in which to search for the solution in the next step. Depending on the equations that we are solving, the analytical Jacobian may not be accessible and we may be forced to use numerical methods, which are usually very time consuming. However, in some cases it is possible to find an approximate analytical Jacobian which is sufficiently good. It is such a case that we have here. The equation that we want to solve in the outer loop is:

$$\nabla (\epsilon \nabla V) - e^2 [N_D - n] = 0 = F[V] . \quad (\text{A.2})$$

By using the globally convergent Newton method we want to find the root of the functional  $F[V]$ , given in the column vector form upon the discretization of the Poisson equation. The electron density  $n$  also depends on  $V$  in a complicated way, through the solution of Eqs. (3.5b), (3.5d), and (3.5c). The Jacobian  $\partial F/\partial V$ , given as a matrix upon the discretization, is inaccessible in an exact analytical form. However, we find that the approximate Jacobian given in [193] leads to a satisfactory convergence. To construct this Jacobian we first calculate a local quasi-equilibrium Fermi level from the electron density, using the bulk formula:

$$n(x) = 2 \left( \frac{m^* kT}{2\pi \hbar^2} \right)^{\frac{3}{2}} \frac{2}{\sqrt{\pi}} F_{\frac{1}{2}}^D \left( \frac{E_F^{loc}(x) - V(x)}{kT} \right) , \quad (\text{A.3})$$

where it is assumed that  $n$  is calculated at some time  $i\Delta t$ ,  $F^D$  is the Fermi-Dirac integral and again we assume a 1D device with an isotropic and homogeneous effective mass. Having  $E_F^{loc}(x)$ , the Jacobian is now obtained by taking the derivative of the discrete form of Eq. A.2 with respect to a column vector  $V$  representing the device potential. To find  $\partial n/\partial V$ , we use Eq. A.3:

$$\frac{\partial n}{\partial V} = -2 \left( \frac{m^* kT}{2\pi \hbar^2} \right)^{\frac{3}{2}} \frac{1}{\sqrt{\pi} kT} F_{-\frac{1}{2}}^D \left( \frac{E_F^{loc}(x) - V(x)}{kT} \right) , \quad (\text{A.4})$$

which is a diagonal matrix. One has to note that the globally convergent Newton method [95] may need to be slightly modified, to account for the fact that the Jacobian here is not exact and may not

always get us closer to the solution regardless of how small the step size is. In such cases, usually choosing a slightly different initial guess solves the problem.

## Appendix B: The surface divergence of the magnetic vector potential

By using the fact that  $\nabla \cdot \vec{A} = 0$  for the Landau gauge in the Cartesian coordinate system and that the curvilinear coordinate system we are using is orthogonal, we derive a relationship between the surface curvilinear divergence  $\nabla_S \cdot \vec{A}_S$  and the normal component  $A_3$  and curvature properties, given by Eq. (4.5).

Since  $\nabla \cdot \vec{A} = 0$  in the rectangular coordinate system, it will remain zero in any other coordinate system. So if we assume a general curvilinear coordinate system suitable for defining a curved surface in the limit where the thickness goes to zero, [78] we get

$$\begin{aligned} \nabla \cdot \vec{A} &= \frac{1}{\sqrt{G}} \partial_i \left( \sqrt{G} G^{ij} A_j \right) = \\ &\frac{1}{\sqrt{G}} \partial_a \left( \sqrt{G} G^{ab} A_b \right) + \frac{1}{\sqrt{G}} \partial_3 \left( \sqrt{G} A_3 \right) = 0 . \end{aligned} \quad (\text{B.1})$$

Here,  $a$  and  $b$  are the surface curvilinear coordinates and

$$\mathbf{G} = \begin{bmatrix} G_{11} & G_{12} & 0 \\ G_{21} & G_{22} & 0 \\ 0 & 0 & 1 \end{bmatrix} , \quad (\text{B.2})$$

$G$  is the determinant of the matrix (B.2) and  $G^{ij}$  the contravariant components. If we evaluate Eq. (B.1) at  $u^3 = 0$ , we get

$$\frac{1}{\sqrt{g}} \partial_a \left( \sqrt{g} g^{ab} A_b \Big|_{u^3=0} \right) = - \frac{1}{\sqrt{G}} \partial_3 \left( \sqrt{G} A_3 \right) \Big|_{u^3=0} , \quad (\text{B.3})$$

since

$$G_{ab} = g_{ab} + \left[ \tilde{\alpha} \mathbf{g} + (\tilde{\alpha} \mathbf{g})^T \right]_{ab} u^3 + (\tilde{\alpha} \mathbf{g} \tilde{\alpha}^T)_{ab} (u^3)^2 \quad (\text{B.4})$$

and  $\tilde{\alpha}$  is a matrix with elements satisfying Weingarten equations [78]

$$\alpha_{11} = \frac{1}{g} (g_{12} h_{21} - g_{22} h_{11}) , \quad \alpha_{12} = \frac{1}{g} (h_{11} g_{21} - h_{21} g_{11}) , \quad (\text{B.5})$$

$$\alpha_{21} = \frac{1}{g} (h_{22} g_{12} - h_{12} g_{22}) , \quad \alpha_{22} = \frac{1}{g} (h_{21} g_{12} - h_{22} g_{11}) .$$

Using the fact that the coordinate system is orthogonal ( $g_{12} = g_{21} = 0$ ) and that in our notation  $\vec{A}_S = (A_1|_{u^3=0}, A_2|_{u^3=0}, 0)$ , where  $A_{1,2,3}$  are the covariant components of the magnetic vector potential, after some algebra we arrive at the final equation (4.5) from (B.3).

## Appendix C: Hermiticity of the exact and tight-binding Hamiltonians

### C.1 Exact Hamiltonian

Here, we will investigate the Hermiticity property of the exact Hamiltonian, given by Eq. (4.1). The first term on the left-hand side of Eq. (4.1) is the curvilinear Laplacian and it is well known that it is Hermitian. Therefore, we will not give the formal proof, but the method is similar to the one we will use for other terms. The fourth term on the left-hand side is real and does not contain any derivatives, so it is Hermitian too. The terms that are not obviously Hermitian are the imaginary terms containing the magnetic vector potential (the second and third terms on the left hand side). If we introduce  $h = (ie\hbar/m)g^{ij}A_i\partial_j + iV_3$ , we need to prove

$$\int \phi^* (h\psi) dS = \int (h\phi)^* \psi dS, \quad (\text{C.1})$$

where  $dS$  is an infinitesimal surface element on the curved surface  $S$ . It is known that the infinitesimal volume element can be expressed in a curvilinear coordinate system as [183]  $dV = \sqrt{G}du^1du^2du^3$ , where the three-dimensional metric tensor  $G_{ij}$  is defined in Appendix B. Now we can write

$$(dV)_{u^3=0} = \sqrt{g}du^1du^2du^3 \Rightarrow dS = \sqrt{g}du^1du^2. \quad (\text{C.2})$$

The first term in  $h$  consists of a sum of four terms (two if the curvilinear coordinate system is orthogonal). For example, for  $(i, j) = (1, 1)$ , we can use the integration by parts and the fact that the wave function vanishes at infinity:

$$\begin{aligned} \frac{ie\hbar}{m} \int du^2 \int du^1 \phi^* g^{11} A_1 \sqrt{g} \partial_1 \psi &= \left\{ \begin{array}{l} u = \phi^* g^{11} A_1 \sqrt{g} \\ dv = \partial_1 \psi du^1 \end{array} \right\} = \\ &= \frac{ie\hbar}{m} \int du^2 \left[ \psi \phi^* g^{11} A_1 \sqrt{g} \Big|_{u^1=-\infty}^{u^1=+\infty} - \int \psi \partial_1 (\phi^* g^{11} A_1 \sqrt{g}) du^1 \right] = \\ &= -\frac{ie\hbar}{m} \int \psi \partial_1 (\phi^* g^{11} A_1 \sqrt{g}) du^1 du^2 = \\ &= -\frac{ie\hbar}{m} \left[ \int \psi g^{11} A_1 \partial_1 \phi^* dS - \int \psi \phi^* \frac{1}{\sqrt{g}} \partial_1 (\sqrt{g} g^{11} A_1) dS \right]. \end{aligned} \quad (\text{C.3})$$

Similar derivations can be carried out for the other three terms, so that we have in total for the first term in  $h$

$$\frac{ie\hbar}{m} \int \phi^* g^{ij} A_i \partial_j \psi dS = -\frac{ie\hbar}{m} \left[ \int \psi g^{ij} A_i \partial_j \phi^* dS + \int \psi \phi^* \frac{1}{\sqrt{g}} \partial_i (\sqrt{g} g^{ij} A_j) dS \right]. \quad (\text{C.4})$$

The second term on the right-hand side in Eq. (C.4) is the matrix element of  $-2iV_3$ . Then, after the partial cancelation with  $iV_3$  in  $h$  we have

$$\int \phi^* (h\psi) dS = -\frac{ie\hbar}{2m} \int \psi \phi^* \frac{1}{\sqrt{g}} \partial_i (\sqrt{g} g^{ij} A_j) dS - \frac{ie\hbar}{m} \int \psi g^{ij} A_i \partial_j \phi^* dS = \int (h\phi)^* \psi dS, \quad (\text{C.5})$$

which proves the Hermiticity of the exact Hamiltonian.

## C.2 Tight-binding Hamiltonian

The tight-binding Hamiltonian derived by using finite-difference schemes does not have to be Hermitian, regardless of the fact the exact Hamiltonian is Hermitian. However, it is computationally advantageous to have Hermitian matrices when one wants to solve the eigenvalue problem. Here, we will show that the discretization schemes employed in Secs. 4.1.1 and 4.1.2, along with some matrix transformations, allow for the definition of Hermitian matrices. To obtain the tight-binding Hamiltonian, we can rewrite Eq. (4.19) in the form of an eigenvalue problem. We have

$$\begin{aligned} & \begin{bmatrix} \mathbf{H}_0 & \mathbf{H}_{0,1} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{H}_{1,0} & \mathbf{H}_1 & \mathbf{H}_{1,2} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{H}_{2,1} & \mathbf{H}_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \mathbf{0} & \cdots & \mathbf{0} & \mathbf{H}_{N,N-1} & \mathbf{H}_N \end{bmatrix} \begin{bmatrix} \Psi_0 \\ \Psi_1 \\ \Psi_2 \\ \vdots \\ \Psi_N \end{bmatrix} = \\ & = E \begin{bmatrix} \mathbf{G}_0 & \mathbf{0} & \cdots & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{G}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{G}_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \mathbf{0} & \cdots & \cdots & \mathbf{0} & \mathbf{G}_N \end{bmatrix} \begin{bmatrix} \Psi_0 \\ \Psi_1 \\ \Psi_2 \\ \vdots \\ \Psi_N \end{bmatrix} - \begin{bmatrix} \mathbf{A} \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \\ \mathbf{B} \end{bmatrix}, \end{aligned} \quad (\text{C.6})$$

where  $\mathbf{A} = \mathbf{H}_{0,-1} \Psi_{-1}$  and  $\mathbf{B} = \mathbf{H}_{N,N+1} \Psi_{N+1}$  are the open boundary conditions. If we extend all the matrices to include slices sufficiently far away from the curved region, we will eventually be able to use the Dirichlet boundary conditions (analytically wave function vanishes at infinity), so that  $\mathbf{A} = \mathbf{0}$  and  $\mathbf{B} = \mathbf{0}$ . Then, we can write the eigenvalue problem as

$$\mathbf{H}\Psi = E\mathbf{G}\Psi, \quad (\text{C.7})$$

where the matrix definitions follow from Eq. (C.6) with  $\mathbf{A} = \mathbf{0}$  and  $\mathbf{B} = \mathbf{0}$ .

As mentioned in Sec. (4.1.1), one of the consequences of using the second-order discretization scheme is that the final tight-binding Hamiltonian matrix  $\mathbf{G}^{-1}\mathbf{H}$  will not be Hermitian due to the existence of the curvilinear Laplacian. By rescaling the Schrödinger equation [Eq. (4.8)] we eventually arrived at a generalized eigenvalue problem given by Eq. (C.7), where  $\mathbf{H}$  is the rescaled tight-binding Hamiltonian matrix. Using a suitable matrix transformation we can obtain the eigenvalue problem in a standard form

$$\begin{aligned} \mathbf{H}\tilde{\mathbf{G}}^{-1}\tilde{\mathbf{G}}\Psi &= E\tilde{\mathbf{G}}\tilde{\mathbf{G}}\Psi, \\ (\tilde{\mathbf{G}}^{-1}\mathbf{H}\tilde{\mathbf{G}}^{-1})\tilde{\mathbf{G}}\Psi &= E\tilde{\mathbf{G}}\Psi, \end{aligned} \quad (\text{C.8})$$

where the definition of  $\tilde{\mathbf{G}}$  is obvious, since  $\mathbf{G} = \tilde{\mathbf{G}}\tilde{\mathbf{G}}$  and  $\mathbf{G}$  is diagonal. In this transformed eigenvalue problem, the Hamiltonian is given by  $\tilde{\mathbf{G}}^{-1}\mathbf{H}\tilde{\mathbf{G}}^{-1}$  and the wave function by  $\tilde{\mathbf{G}}\Psi$ . Since the matrix  $\tilde{\mathbf{G}}$  is diagonal, the condition for Hermiticity of this transformed Hamiltonian reduces to  $\mathbf{H} = \mathbf{H}^\dagger$ , where  $\dagger$  represents the conjugate transpose. Furthermore, this condition reduces to

$$\left\{ \begin{array}{l} \mathbf{H}_n = \mathbf{H}_n^\dagger \\ \mathbf{H}_{n,n\pm 1} = \mathbf{H}_{n\pm 1,n}^\dagger \end{array} \right\} \Rightarrow \left\{ \begin{array}{l} \mathbf{L}_n = \mathbf{L}_n^\dagger \\ \mathbf{L}_{n,n\pm 1} = \mathbf{L}_{n\pm 1,n}^\dagger \\ \mathbf{P}_n = \mathbf{P}_n^\dagger \\ \mathbf{P}_{n,n\pm 1} = \mathbf{P}_{n\pm 1,n}^\dagger \end{array} \right\}, \quad (\text{C.9})$$

where we used definitions from Eq. (4.21).

If our curvilinear grid is uniform, such that  $\Delta_1^n = \text{const.} = \Delta_1$  and  $\Delta_2^m = \text{const.} = \Delta_2$ , the conditions in Eq. (C.9) are satisfied, as can be seen from Eqs. (4.12a)-(4.14) and Eqs. (4.17)-(4.18). Therefore, the eigenvalue problem in Eq. (C.8) contains a Hermitian matrix. Since, in

general, we prefer the possibility of using a non-uniform curvilinear grid, with  $\Delta_1^n$  varying in the longitudinal direction and  $\Delta_2^m$  in the transversal direction, the rescaled Hamiltonian matrix  $\mathbf{H}$  will not be Hermitian. In that case we can introduce further matrix transformations [104–106], defined by a diagonal block matrix  $\mathbf{F}$ , with diagonal matrices  $\mathbf{F}_n$  on its main diagonal given by

$$\mathbf{F}_n(m, m) = \frac{2}{\sqrt{(\Delta_1^n + \Delta_1^{n-1})(\Delta_2^{m-1} + \Delta_2^{m-2})}}, \quad (\text{C.10})$$

where, as before, the index  $m$  on the left-hand side designates matrix elements with limits  $m \in [1, M + 1]$ , while on the right-hand side all the quantities are defined on the grid having limits  $[0, M]$  and, therefore, we use  $m - 1$  instead of  $m$ . We can now construct a Hermitian eigenvalue problem by the following matrix transformation in the second of Eqs. (C.8)

$$\begin{aligned} \left( \mathbf{F}^{-1} \tilde{\mathbf{G}}^{-1} \mathbf{H} \mathbf{F} \mathbf{F}^{-1} \tilde{\mathbf{G}}^{-1} \right) \tilde{\mathbf{G}} \Psi &= E \mathbf{F}^{-1} \tilde{\mathbf{G}} \Psi, \\ \left( \tilde{\mathbf{G}}^{-1} \mathbf{F}^{-1} \mathbf{H} \mathbf{F} \tilde{\mathbf{G}}^{-1} \right) \tilde{\mathbf{G}} \mathbf{F}^{-1} \Psi &= E \tilde{\mathbf{G}} \mathbf{F}^{-1} \Psi, \end{aligned}$$

which can be rewritten as

$$\bar{\mathbf{H}} \Phi = E \Phi, \quad (\text{C.11})$$

where  $\bar{\mathbf{H}} = \tilde{\mathbf{G}}^{-1} \mathbf{F}^{-1} \mathbf{H} \mathbf{F} \tilde{\mathbf{G}}^{-1}$  is a Hermitian matrix and  $\Phi = \tilde{\mathbf{G}} \mathbf{F}^{-1} \Psi$ .

## Appendix D: Stabilized Transfer Matrix Solution to Eq. (4.19)

To solve Eq. (4.19) for the transmission/reflection coefficients and electron density, we first rearrange it in the transfer matrix form:

$$\begin{bmatrix} \Psi_n \\ \Psi_{n+1} \end{bmatrix} = \mathbf{T}_n \begin{bmatrix} \Psi_{n-1} \\ \Psi_n \end{bmatrix}, \quad (\text{D.1})$$

where the transfer matrix  $\mathbf{T}_n$  is given by

$$\mathbf{T}_n = \begin{bmatrix} 0 & \mathbf{1} \\ -\mathbf{H}_{n,n+1}^{-1} \mathbf{H}_{n,n-1} & \mathbf{H}_{n,n+1}^{-1} (E \cdot \mathbf{G}_n - \mathbf{H}_n) \end{bmatrix}. \quad (\text{D.2})$$

For our case of linear (small temperature and bias) and ballistic transport we solve this equation only for  $E = E_F$ , where  $E_F$  is the Fermi level energy. Writing equations of the form given by Eq. D.1 for every slice in our curved nanoribbon we can connect the two slices at the interface with the left lead to the two slices at the interface with the right lead

$$\begin{bmatrix} \Psi_N \\ \Psi_{N+1} \end{bmatrix} = \mathbf{T}_N \cdots \mathbf{T}_0 \begin{bmatrix} \Psi_{-1} \\ \Psi_0 \end{bmatrix}, \quad (\text{D.3})$$

where the domain bounded by slices 0 and  $N$  belongs to the curved nanoribbon and the rest belongs to the left (smaller than 0) and right (greater than  $N$ ) leads.

We now express the wave functions on the left and right hand sides of Eq. (D.3) in terms of the reflection and transmission coefficients multiplying forward and backward propagating eigenfunctions of the leads. In the Landau gauge, these eigenfunctions are proportional (up to the normalization constant) to the product of plane waves along the lead and transverse modes [86]. The decomposition of the wave functions in Eq. (D.3) is possible because each  $\Psi_n$  ("slice wave function") is a superposition of forward and backward traveling waves, with amplitudes determined by specific injection conditions as well as any reflection and transmission that the wave undergoes.

For the wave incident at the left lead (with unit amplitude) we can write

$$\begin{aligned} \begin{bmatrix} \Psi_{-1} \\ \Psi_0 \end{bmatrix} &= \mathbf{T}_L \begin{bmatrix} \mathbf{1} \\ \mathbf{r} \end{bmatrix}, \\ \begin{bmatrix} \Psi_N \\ \Psi_{N+1} \end{bmatrix} &= \mathbf{T}_R \begin{bmatrix} \mathbf{t} \\ \mathbf{0} \end{bmatrix}, \end{aligned} \quad (\text{D.4})$$

where

$$\mathbf{T}_{L/R} = \begin{bmatrix} \Psi_{L/R}^+ & \Psi_{L/R}^- \\ \Psi_{L/R}^+ \lambda_{L/R}^+ & \Psi_{L/R}^- \lambda_{L/R}^- \end{bmatrix}, \quad (\text{D.5})$$

with

$$\begin{aligned} \Psi_{L/R}^\pm &= [\psi_{0,L/R}^\pm, \dots, \psi_{M,L/R}^\pm], \\ \lambda_{L/R}^\pm &= \text{diag} [\lambda_{0,L/R}^\pm, \dots, \lambda_{M,L/R}^\pm]. \end{aligned} \quad (\text{D.6})$$

$\psi_{l,L/R}^\pm$  is the eigenfunction of the  $l$ -th transverse mode at the  $(-1)$ -th (for L) and  $N$ -th (for R) slice, where  $+$  stands for forward and  $-$  for backward propagation, while  $\lambda_{l,L/R}^\pm$  (similar notation as for eigenfunctions) is a phase factor which represents the difference in phase between eigenfunction values at two neighboring slices (eigenfunctions are plane waves along the lead). It should be noted that  $\Psi_{L/R}^\pm$  is a matrix of dimension  $(M+1) \times (M+1)$ , while  $\Psi_n$  is a column vector of dimension  $(M+1) \times 1$ .

We can construct the matrices  $\mathbf{T}_{L/R}$  by using the transfer matrix form [Eq. D.1] within the leads (planar geometry) and knowing that wave function values at two neighboring slices in the leads differ only by a phase factor. We obtain the following eigenvalue problems for the left and right leads

$$\begin{aligned} \begin{bmatrix} 0 & \mathbf{1} \\ -\mathbf{H}_{0,1}^{-1} \mathbf{H}_{0,-1} & \mathbf{H}_{0,1}^{-1} (E_F \cdot \mathbf{1} - \mathbf{H}_0) \end{bmatrix} \begin{bmatrix} \psi_{l,L}^\pm \\ \lambda_{l,L}^\pm \psi_{l,L}^\pm \end{bmatrix} \\ = \lambda_{l,L}^\pm \begin{bmatrix} \psi_{l,L}^\pm \\ \lambda_{l,L}^\pm \psi_{l,L}^\pm \end{bmatrix}, \end{aligned} \quad (\text{D.7})$$

$$\begin{aligned}
& \begin{bmatrix} 0 & \mathbf{1} \\ -\mathbf{H}_{N+1,N+2}^{-1} \mathbf{H}_{N+1,N} & \mathbf{H}_{N+1,N+2}^{-1} (E_F \cdot \mathbf{1} - \mathbf{H}_{N+1}) \end{bmatrix} \times \\
& \times \begin{bmatrix} \psi_{l,R}^{\pm} \\ \lambda_{l,R}^{\pm} \psi_{l,R}^{\pm} \end{bmatrix} = \lambda_{l,R}^{\pm} \begin{bmatrix} \psi_{l,R}^{\pm} \\ \lambda_{l,R}^{\pm} \psi_{l,R}^{\pm} \end{bmatrix}.
\end{aligned} \tag{D.8}$$

Here, for the purpose of keeping the same notation as in Eq. (D.1), slices 0 and 1 appear, but it should be considered that they belong to the leads. In fact, since the leads are planar and uniform and the magnetic field is also uniform, we could have chosen any other combination of three consecutive slices that belong to the same lead in Eqs. (D.7) and (D.8). Matrices  $\mathbf{H}_0$ ,  $\mathbf{H}_{0,\pm 1}$  and  $\mathbf{H}_{N+1}$ ,  $\mathbf{H}_{N+1,N+1\pm 1}$  follow from the matrices in Eq. (4.21) in the zero-curvature limit and are given in earlier papers on coherent transport in planar nanostructures [87, 92]. Finally, we obtain the following matrix equation to be solved

$$\begin{bmatrix} \mathbf{t} \\ \mathbf{0} \end{bmatrix} = \mathbf{T}_R^{-1} \mathbf{T}_N \cdots \mathbf{T}_0 \mathbf{T}_L \begin{bmatrix} \mathbf{1} \\ \mathbf{r} \end{bmatrix}. \tag{D.9}$$

Straightforward solution of Eq. (D.9) is not stable numerically because of the existence of evanescent waves (exponentially growing or decaying), whose numerical value can become extremely small or large after only a few matrix multiplications. We employ a stabilized iterative technique due to Usuki *et al.* [87] with two minor modifications: in the wave function calculation and in allowing the left and right lead Hamiltonians to be different (either through different geometry or magnetic field flux). The fact that the lead Hamiltonians can be different is made explicit in the above equations. The stabilized iterative version of Eq. (D.9) for the calculation of transmission coefficients in the column vector  $\mathbf{t}$  is

$$\begin{aligned}
& \begin{bmatrix} \mathbf{C}_1^{(j+1)} & \mathbf{C}_2^{(j+1)} \\ \mathbf{0} & \mathbf{1} \end{bmatrix} = \mathbf{T}_j \begin{bmatrix} \mathbf{C}_1^{(j)} & \mathbf{C}_2^{(j)} \\ \mathbf{0} & \mathbf{1} \end{bmatrix} \mathbf{P}^{(j)}, \\
& -1 \leq j \leq N+1,
\end{aligned} \tag{D.10}$$

with  $\mathbf{T}_{-1} = \mathbf{T}_L$  and  $\mathbf{T}_{N+1} = \mathbf{T}_R^{-1}$  and

$$\mathbf{P}^{(j)} = \begin{bmatrix} \mathbf{1} & \mathbf{0} \\ \mathbf{P}_1^{(j)} & \mathbf{P}_2^{(j)} \end{bmatrix}. \tag{D.11}$$

By using the block matrix inversion to calculate  $\mathbf{T}_R^{-1}$  and setting  $\mathbf{T}_R^{12} = \mathbf{1}$ ,  $\mathbf{T}_R^{22} = \mathbf{0}$ , where the position of the block matrix in  $\mathbf{T}_R$  is denoted by the superscript, (their value is not important because they are multiplied by  $\mathbf{0}$  in Eq. D.4) we get

$$\mathbf{T}_{N+1} = \begin{bmatrix} \mathbf{0} & [\Psi_R^+ \lambda_R^+]^{-1} \\ \mathbf{1} & -\Psi_R^+ [\Psi_R^+ \lambda_R^+]^{-1} \end{bmatrix}, \quad (\text{D.12})$$

The form of  $\mathbf{P}^{(j)}$  and its block matrices is determined from the condition that the form of the matrix on the left hand side of Eq. (D.10) is satisfied. We obtain

$$\begin{aligned} \mathbf{P}_1^{(j)} &= -\mathbf{P}_2^{(j)} \mathbf{T}_j^{21} \mathbf{C}_1^{(j)}, \\ \mathbf{P}_2^{(j)} &= \left[ \mathbf{T}_j^{21} \mathbf{C}_2^{(j)} + \mathbf{T}_j^{22} \right]^{-1}, \end{aligned} \quad (\text{D.13})$$

where different block matrices in  $\mathbf{T}_j$  are specified with the row and column numbers in the superscript. Starting from the initial conditions  $\mathbf{C}_1^{(0)} = \mathbf{1}$  and  $\mathbf{C}_2^{(0)} = \mathbf{0}$  and solving Eq. (D.10) iteratively we obtain the transmission coefficients from

$$\mathbf{t} = \mathbf{C}_1^{(N+2)}. \quad (\text{D.14})$$

In a similar way for the reflection coefficients we have

$$\left[ \mathbf{D}_1^{(j+1)} \mathbf{D}_2^{(j+1)} \right] = \left[ \mathbf{D}_1^{(j)} \mathbf{D}_2^{(j)} \right] \mathbf{P}^{(j)}, \quad \mathbf{r} = \mathbf{D}_1^{(N+2)}, \quad (\text{D.15})$$

with the initial conditions  $\mathbf{D}_1^{(0)} = \mathbf{0}$  and  $\mathbf{D}_2^{(0)} = \mathbf{1}$ .

For the calculation of electron density we define a new operator  $\Phi^{(j)}$ , using the already calculated operator  $\mathbf{P}^{(j)}$ ,

$$\Phi^{(j-1)} = \mathbf{P}_1^{(j-1)} + \mathbf{P}_2^{(j-1)} \Phi^{(j)}, \quad N+1 \geq j \geq 1, \quad (\text{D.16})$$

with the initial condition  $\Phi^{(N+1)} = \mathbf{P}_1^{(N+1)}$ . Now the electron density originating from a single transverse mode  $l$  in the injecting lead will be

$$n_l(n, m) = \left| \Phi^{(n)}(m, l) \right|^2, \quad n \in [0, N], \quad (\text{D.17})$$

where the indices  $(n, m)$  represent longitudinal and transverse grid positions, respectively.

## D.1 Velocities of propagating states in the presence of magnetic field

With  $\vec{A}$  in place as in Eq. (4.26), we can calculate the injected (evanescent and propagating) scattering states from each lead by solving the eigenvalue problem of the transfer matrix between the first two slices ( $n=-1,0$ ) that belong to the leads [here we make use of the fact that the scattering states in the leads are plane waves in the longitudinal (transport) direction] [87]

$$\exp(ik_1\Delta_1) \begin{bmatrix} \Psi_{-1} \\ \Psi_0 \end{bmatrix} = T_0 \begin{bmatrix} \Psi_{-1} \\ \Psi_0 \end{bmatrix}, \quad (\text{D.18})$$

$$T_0 = \begin{bmatrix} 0 & \mathbf{1} \\ -\mathbf{H}_{0,1}^{-1}\mathbf{H}_{0,-1} & \mathbf{H}_{0,1}^{-1}(\mathbf{E} \cdot \mathbf{1} - \mathbf{H}_0) \end{bmatrix},$$

where  $k_1$  denotes the longitudinal component of the wavevector. Leads are assumed planar, so the  $\mathbf{H}$  matrices satisfy the tight-binding 2D SE for a planar grid, [87] into which Eq. (4.19) reduces in the limit of zero curvature ( $g_{ii} = g^{ii} = 1$ ). Explanation of one technical point is needed here. In order to obtain the correct forward and backward propagating states from a numerical solution to Eq. (D.18), it is important to properly choose the reference level for  $\vec{A}$ . This can be done by centering the reference coordinate system to follow the midline that ‘‘cuts’’ the lead in half in the transverse direction: in other words, we can choose the  $u^2 = 0$  grid line such that it extends into the lead and splits it in two halves symmetrically.) Upon such a choice, the terms in  $\vec{A}$  in Eq. (4.26) due to the different  $\vec{B}$  components will always be odd (linear in  $y$ ) or even (independent of  $y$ ) functions of the transverse position, which gives correct number and distribution of forward and backward propagating states. [92]

A correct evaluation of velocities is important since one of our goals is to calculate the transmission through a curved nanoribbon,

$$T_{ij} = |t_{ij}|^2 v^{(i)}/v^{(j)}, \quad (\text{D.19})$$

where  $i$  is the outgoing mode index,  $j$  is the incoming mode index, with  $v_i$  and  $v_j$  being the corresponding velocities, while  $t_{ij}$  is the matrix of transmission amplitudes, Eq. (D.14).

It turns out that the relation  $v^{(l)} = \hbar k_1^{(l)}/m^*$ , where  $k_1^{(l)}$  for a propagating mode of index  $l$  is calculated as explained above, does not hold in the presence of a magnetic field. [181, 194] To

calculate the appropriate relationship we can use a tight-binding Hamiltonian on a planar grid in a magnetic field. By working in the  $|n, m\rangle$  basis ( $n$  - slice index,  $m$  - transverse index in a slice)

$$\hat{x}|n, m\rangle = n\Delta_1|n, m\rangle, |\psi\rangle = \sum_{n,m} \psi_{n,m}|n, m\rangle \quad (\text{D.20})$$

and defining the velocity operator in the  $x$ -direction (which is assumed longitudinal in the leads) as usual

$$\hat{v} = \frac{1}{i\hbar} [\hat{x}, \hat{H}] , \quad (\text{D.21})$$

we can define the velocity at the slice  $n$  as

$$\begin{aligned} v_n^{(l)} &= [\langle\psi|\hat{v}|\psi\rangle]_n^{(l)} = \frac{2t_1\Delta_1}{\hbar} \sum_m |\psi_{n,m}|^2 \times \\ &\times \left[ \sin\left(k_1^{(l)}\Delta_1\right) - \frac{e}{\hbar} \cos\left(k_1^{(l)}\Delta_1\right) A_1(n, m)\Delta_1 \right] . \end{aligned} \quad (\text{D.22})$$

Here,  $\hat{H}$  is the Hamiltonian with the matrix representation given by Eq. (C.6) in flat metric and with a uniform grid assumed in the leads, so that  $t_1$  is the longitudinal hopping energy. We also make use of the fact that for a sufficiently small grid size one can employ the translational operator of the form  $\psi_{n\pm 1, m} = \psi_{n, m} \exp(\pm ik_1\Delta_1)$ . [195] In the Peierls phase approximation (weak magnetic field so a small  $\vec{A}$ ) Eq. (D.22) equals

$$v_n^{(l)} \approx \frac{2t_1\Delta_1}{\hbar} \sum_m |\psi_{n,m}|^2 \sin \left\{ \left[ k_1^{(l)} - \frac{e}{\hbar} A_1(n, m) \right] \Delta_1 \right\} . \quad (\text{D.23})$$

As  $A_1$  depends only on the transverse coordinate [a consequence of the Landau gauge, see (4.26)], the only dependence of the current on the slice index  $n$  would occur through the term  $|\psi_{n,m}|^2$ . However, since a mode's transversal profile is constant inside a lead of constant width, and is modulated by a plane wave in the longitudinal direction, its square amplitude does not depend on  $n$  for any  $m$  and velocity (and current) per mode is conserved between slices in the lead. Therefore, it is sufficient to evaluate the mode velocity at ( $n = -1$ ) in Eqs. (D.22) and (D.23).

## Appendix E: Local Landau Gauge

### E.1 Local Landau gauge. Gauge transformations

The local Landau gauge transformation can be defined in a formal mathematical language. In analogy with a misaligned multi-terminal structure, [181] here we can derive very similar equations with a slightly different interpretation: a local Landau gauge in every lead becomes a local Landau gauge in every slice, while the planar device region between the leads, used to smoothly connect local Landau gauges in the leads, becomes the curved regions between slices used to smoothly connect local Landau gauges in slices. The slice at the position of the reference coordinate system is numbered 0 (slice  $-1$  belongs to the left lead), so we can write for the gauge transformation between slices 0 and  $n$

$$\vec{A}_n(\vec{x}_0) = \vec{A}_0(\vec{x}_0) + \nabla f_n(\vec{x}_0) , \quad (\text{E.1})$$

where  $\vec{x}_0$  is a point in the reference coordinate system, which is positioned at the slice 0. It is straightforward to derive the function  $f_n$  for an arbitrary rotation, [181] and translation needed for the gauge transformation between the magnetic vector potential at 0-th slice,  $\vec{A}_0$ , and the magnetic vector potential at  $n$ -th slice,  $\vec{A}_n$ . Then, the gauge transformation generated by  $f_n$  in Eq. (E.1) allows us to express  $\vec{A}_n$  in the same form as  $\vec{A}_0$ , but in a different coordinate system, translated and rotated with respect to the reference coordinate system in which  $\vec{A}_0$  is calculated (see Fig. 4.2). For example, if  $\vec{A}_0 = -By_0\hat{x}_0$ , then  $\vec{A}_n = -By_n\hat{x}_n$ .

The gauge transformation given by  $f_n$  should be turned on only at the position of  $n$ -th slice and turned off at the positions of all other slices, while in the rest of the domain (curved nanoribbon's surface) it should be a smooth function. This can be achieved by multiplying  $f_n$  by the following auxiliary function:

$$\xi_n = \begin{cases} 1 , & \text{on slice } n \\ \text{smooth} , & \text{between slices } n-1 \text{ and } n+1 . \\ 0 , & \text{everywhere else} \end{cases} \quad (\text{E.2})$$

Now Eq. (E.1) becomes

$$\vec{A}_n = \vec{A}_0 + \nabla (\xi_n f_n) = \vec{A}_0 + \nabla \xi_n f_n + \xi_n \nabla f_n . \quad (\text{E.3})$$

At the slice  $n$ ,  $\nabla \xi_n = 0$  and  $\xi_n = 1$  and we have a correct gauge transformation, given by Eq. (E.1) evaluated at the slice  $n$ . In the domain bounded by slices 0 to  $n-1$  and  $n+1$  to  $N$  both the auxiliary function and its gradient are zero, so there is no gauge transformation of the form generated by  $f_n$ . In the domain bounded by slices  $n-1$  to  $n+1$  both the auxiliary function and its gradient are nonzero and we have some unavoidable gauge transformation, but we are not interested in its exact form since we work on the discrete grid only (space between discrete slices is not included in our calculation). To generate the total gauge transformation we calculate functions  $f_i$  and  $\xi_i$  for each slice  $i$  and sum them up

$$f = \sum_{i=1}^N \xi_i f_i . \quad (\text{E.4})$$

For a magnetic field of general direction we can write the total magnetic vector potential in the following form

$$\vec{A}_0 = \vec{A}_0^x + \vec{A}_0^y + \vec{A}_0^z , \quad (\text{E.5})$$

where the magnetic vector potential components are each due to a magnetic field component denoted by the superscript. Therefore, for Eq. (4.26) we have that  $\vec{A}_0^x = B_{x_0} y_0 \hat{z}_0$ ,  $\vec{A}_0^y = -B_{y_0} x_0 \hat{z}_0$  and  $\vec{A}_0^z = -B_{z_0} y_0 \hat{x}_0$ . Since the curl operator is linear we can use the principle of superposition and apply the necessary gauge transformations independently to these magnetic vector potential components. By applying the semi-graphical and formal mathematical methods explained above, we can decide which component needs a gauge transformation and calculate it.

In deriving Eq. (4.5) we used the fact that  $\nabla \cdot \vec{A} = 0$  in the Landau gauge. If we use the separation into components explained in the previous paragraph, we see that the components of  $\vec{A}$  not requiring gauge transformations will not alter the fact that  $\nabla \cdot \vec{A} = 0$ . However, in general we might have a field component that requires a gauge transformation generated by the function in

Eq. (E.4). If we designate the field component requiring a gauge transformation by  $q$  we have

$$\begin{aligned}
\vec{A}^q &= \vec{A}_0^q + \nabla f \\
&= \vec{A}_0^q + \sum_{i=0}^N \nabla (\xi_i f_i) \\
&= \vec{A}_0^q + \sum_{i=0}^N [f_i \nabla \xi_i + \xi_i \nabla f_i] .
\end{aligned} \tag{E.6}$$

Therefore

$$\begin{aligned}
\nabla \cdot \vec{A}^q &= \nabla \cdot \vec{A}_0^q + \sum_{i=0}^N [2\nabla f_i \cdot \nabla \xi_i + f_i \nabla^2 \xi_i + \xi_i \nabla^2 f_i] \\
&= \sum_{i=0}^N [2\nabla f_i \cdot \nabla \xi_i + f_i \nabla^2 \xi_i + \xi_i \nabla^2 f_i] ,
\end{aligned} \tag{E.7}$$

since  $\nabla \cdot \vec{A}_0^q = 0$ ,  $\vec{A}_0^q$  (as well as any other  $\vec{A}_i^q$ ) being in the Landau gauge. This is clearly nonzero for arbitrary point on the surface of the curved nanoribbon, but since we work on a discrete grid let us evaluate Eq. (E.7) at the points belonging to some slice  $n$ :

$$\begin{aligned}
\left( \nabla \cdot \vec{A}^q \right)_n &= \left( \nabla^2 f_n \right)_n = \left( \nabla \cdot \nabla f_n \right)_n = \left[ \nabla \cdot \left( \vec{A}_n^q - \vec{A}_0^q \right) \right]_n \\
&= \left( \nabla \cdot \vec{A}_n^q - \nabla \cdot \vec{A}_0^q \right)_n \\
&= 0 ,
\end{aligned} \tag{E.8}$$

where we used the properties of the function  $\xi_i$ , and the fact that all  $\vec{A}_i^q$ 's are in the Landau gauge. So, at the points on the grid we still have zero divergence, meaning that we can still use Eq. (4.5) to calculate  $V_3$  on the grid.

## E.2 Gauge Transformation for Fig. 4.6

From Eq. (4.26) we see that

$$\vec{A}_0 = (0, 0, -B_{y_0} x_0) . \tag{E.9}$$

From our discussion in Sec. E.1 we know that the slice 0 does not need a gauge transformation: we can directly use Eq. E.9 evaluated at the slice 0, because the longitudinal axis (the  $z_0$ -axis for the

$B_{y_0}$  component) is already tangential to the bisector at the point of its intersection with the slice 0, and the coordinate system is positioned at that point. When we transform the coordinates in Eq. E.9 to our toroidal coordinates  $(\theta, \phi, r)$  and evaluate the result at the position of the slice 0 ( $\phi_0 = 0$ ) we get

$$\vec{A}_0(\vec{x}_0^{(0)}) = \hat{\phi} B_{y_0} r (\cos\theta - 1) , \quad (\text{E.10})$$

where  $\vec{x}_0^{(0)}$  are points for which  $\vec{x}_0 \in \{\text{slice } 0\}$ .

Angle  $\phi$  is measured from the slice 0, which is at the interface with the left lead in Fig. 4.2, while angle  $\theta$  is measured from the bisector toward the negative  $y_0$ -axis (where  $y_0$ -axis belongs to the reference coordinate system in Fig. 4.2). If the Cartesian coordinate system is like the reference coordinate system in Fig. 4.2, but positioned in the center of the torus (from where  $R$  is measured), then the formulas relating the Cartesian coordinates of the torus surface  $(x, y, z)$  to the curvilinear coordinates  $(\theta, \phi)$  are given by

$$\begin{aligned} x &= -(R + r\cos\theta) \cos\phi, \\ y &= -r\sin\theta, \\ z &= (R + r\cos\theta) \sin\phi, \end{aligned} \quad (\text{E.11})$$

where  $\phi$  is measured from the negative  $x$ -axis and  $\theta$  in the clockwise direction looking along the current flow (from left to right in Fig. 4.2).

The transformation to the toroidal coordinates can be done by using Eq. (E.11) and the following relations:

$$\begin{aligned} x_0 &= x + R + r , \\ y_0 &= y , \\ z_0 &= z , \end{aligned} \quad (\text{E.12})$$

$$\begin{aligned} \hat{\mathbf{x}}_0 &= \hat{\phi} \sin\phi - \hat{\mathbf{r}} \cos\theta \cos\phi + \hat{\boldsymbol{\theta}} \sin\theta \cos\phi , \\ \hat{\mathbf{y}}_0 &= -\hat{\boldsymbol{\theta}} \cos\theta - \hat{\mathbf{r}} \sin\theta , \\ \hat{\mathbf{z}}_0 &= \hat{\phi} \cos\phi + \hat{\mathbf{r}} \cos\theta \sin\phi - \hat{\boldsymbol{\theta}} \sin\theta \sin\phi . \end{aligned} \quad (\text{E.13})$$

Since we know that after the gauge transformation the form of  $\vec{A}$  is the same, but expressed in a new coordinate system, which is translated and rotated according to the slice position (see Fig. 4.2

for the  $B_y$  component), we conclude that

$$\vec{A}_i(\vec{x}_i^{(i)}) = \hat{\phi} B_{y_0} r (\cos\theta - 1) , \quad 0 \leq i \leq N . \quad (\text{E.14})$$

Eq. (E.14) can be derived in a formal way, by using Eq. (E.1). If we assume that for some slice  $n$  the coordinate system (as shown in Fig. 4.2) is rotated by the angle  $\alpha_n$  (clockwise here) and translated by the lengths  $a$  along the  $x_0$ -axis and  $b$  along the  $z_0$ -axis we obtain for  $f_n$

$$f_n(\vec{x}_0) = B_{y_0} \left[ (x_0 z_0 - b x_0) \sin^2 \alpha_n + a z_0 \cos^2 \alpha_n + \frac{1}{2} \left( \frac{1}{2} z_0^2 - \frac{1}{2} x_0^2 - b z_0 + a x_0 \right) \sin 2 \alpha_n \right] . \quad (\text{E.15})$$

Now, since  $\vec{A}_0 = -B_{y_0} x_0 \hat{z}_0$ , by using Eq. (E.1) we obtain

$$\vec{A}_n(\vec{x}_0) = -B_{y_0} \left\{ \hat{z}_0 \left[ (x_0 - a) \cos^2 \alpha_n - \frac{1}{2} (z_0 - b) \sin 2 \alpha_n \right] + \hat{x}_0 \left[ \frac{1}{2} (x_0 - a) \sin 2 \alpha_n - (z_0 - b) \sin^2 \alpha_n \right] \right\} . \quad (\text{E.16})$$

Using the transformations between coordinates  $(x_0, y_0, z_0)$  and  $(x_n, y_n, z_n)$

$$x_0 = a + z_n \sin \alpha_n + x_n \cos \alpha_n , \quad (\text{E.17})$$

$$z_0 = b + z_n \cos \alpha_n - x_n \sin \alpha_n ,$$

$$\hat{x}_0 = \hat{z}_n \sin \alpha_n + \hat{x}_n \cos \alpha_n , \quad (\text{E.18})$$

$$\hat{z}_0 = \hat{z}_n \cos \alpha_n - \hat{x}_n \sin \alpha_n ,$$

it can be shown that from Eq. (E.16) follows  $\vec{A}_n = -B_{y_0} x_n \hat{z}_n$ . On the other hand, by using the transformations between coordinates  $(x_0, y_0, z_0)$  and  $(\theta, \phi, r)$  given by Eqs. (E.11) and (E.12), along with the transformations between unit vectors in Eq. (E.13) we obtain

$$\vec{A}_i(\vec{x}_0^{(i)}) = \hat{\phi} B_{y_0} r (\cos\theta - 1) , \quad 0 \leq i \leq N , \quad (\text{E.19})$$

where we used the facts that  $\alpha_n = \phi_n$ ,  $\sin \phi_n = b / (R + r)$  and  $\cos \phi_n = (R + r - a) / (R + r)$ . Eq. (E.19) is the same as Eq. (E.14), previously obtained in an easier way, with the formal distinction in the argument of  $\vec{A}_i$ : the transformation to the toroidal coordinates in Eq. (E.14) is done from the rotated and translated coordinate system  $(x_i, y_i, z_i)$ , since we know  $\vec{A}_i(\vec{x}_i)$  and  $\vec{A}_0(\vec{x}_0)$  have the same form, while in Eq. (E.19) from the reference coordinate system by formally deriving the gauge transformation according to Eq. (E.1).