Springer Series in
SOLID-STATE SCIENCES 156

Springer Series in

SOLID-STATE SCIENCES

Series Editors:

M. Cardona P. Fulde K. von Klitzing R. Merlin H.-J. Queisser H. Störmer

The Springer Series in Solid-State Sciences consists of fundamental scientific books prepared by leading researchers in the field. They strive to communicate, in a systematic and comprehensive way, the basic principles as well as new developments in theoretical and experimental solid-state physics.

138 **Phase Separation in Soft Matter Physics**
Micellar Solutions, Microemulsions, Critical Phenomena
By P.K. Khabibullaev and A.A. Saidov

139 **Optical Response of Nanostructures Microscopic**
Nonlocal Theory
By K. Cho

140 **Fractal Concepts in Condensed Matter Physics**
By T. Nakayama and K. Yakubo

141 **Excitons in Low-Dimensional Semiconductors**
Theory, Numerical Methods, Applications By S. Glutsch

142 **Two-Dimensional Coulomb Liquids and Solids**
By Y. Monarkha and K. Kono

143 **X-Ray Multiple-Wave Diffraction**
Theory and Application
By S.-L. Chang

144 **Physics of Transition Metal Oxides**
By S. Maekawa, T. Tohyama, S.E. Barnes, S. Ishihara, W. Koshibae, and G. Khaliullin

145 **Point-Contact Spectroscopy**
By Y.G. Naidyuk and I.K. Yanson

146 **Optics of Semiconductors and Their Nanostructures**
Editors: H. Kalt and M. Hetterich

147 **Electron Scattering in Solid Matter**
A Theoretical and Computational Treatise
By J. Zabloudil, R. Hammerling, L. Szunyogh, and P. Weinberger

148 **Physical Acoustics in the Solid State**
By B. Lüthi

149 **Solitary Waves in Complex Dispersive Media**
Theory · Simulation · Applications
By V.Yu. Belashov and S.V. Vladimirov

150 **Topology in Condensed Matter**
Editor: M.I. Monastyrsky

151 **Particle Penetration and Radiation Effects**
By P. Sigmund

152 **Magnetism**
From Fundamentals to Nanoscale Dynamics
By H.C. Siegmann and J. Stöhr

153 **Quantum Chemistry of Solids**
The LCAO First Principles Treatment of Crystals
By R.A. Evarestov

154 **Low-Dimensional Molecular Metals**
By N. Toyota, M. Lang and J. Müller

155 **Diffusion in Solids**
Fundamentals, Methods, Materials, Diffusion-Controlled Processes
By H. Mehrer

156 **Physics of Zero- and One-Dimensional Nanoscopic Systems**
By S.N. Karmakar, S.K. Maiti, J. Chowdhury

Volumes 91–137 are listed at the end of the book.

Sachindra Nath Karmakar · Santanu Kumar Maiti ·
Jayeeta Chowdhury (Eds.)

Physics of Zero- and One-Dimensional Nanoscopic Systems

With 141 Figures

Springer

Prof. Dr. Sachindra Nath Karmakar
Santanu Kumar Maiti
Jayeeta Chowdhury
Saha Institute of Nuclear Physics
Bidhannagar
700064 Kolkata
India

Library of Congress Control Number: 2007933209

ISSN 0171-1873

ISBN 978-3-540-72631-9 Springer Berlin Heidelberg New York

Typesetting: Digital data supplied by authors
Production: LE-TEX Jelonek, Schmidt & Vöckler GbR, Leipzig
Cover design: eStudio Calamar S.L., F. Steinen-Broo, Girona, Spain

SPIN 11768760 57/3180/YL – 5 4 3 2 1 0 Printed on acid-free paper

To our parents

Preface

The idea of this volume emerges from the "International Workshop on the Physics of Zero and One Dimensional Nanoscopic Systems," which was held on 1-9 February 2006 at Saha Institute of Nuclear Physics, India. The theme of the workshop was to understand physically the recent advances in nanoscale systems, like, quantum dots, quantum wires, 2D electron gases, etc. A limited number of distinguished physicists were invited to give pedagogical lectures and discuss core methods including the latest developments. This volume consists of self-contained review articles on recent theories of the evolution of Kondo effect in quantum dots, decoherence and relaxation in charged qubits, edge-state transport through nanographites and quantum Hall systems, transport through molecular bridges, coherence and interaction in diffusive mesoscopic systems, persistent current in mesoscopic rings, and, the thermoelectric phenomena of nanosystems. As these are rapidly growing subjects, we hope that this book with contributions from the leading experts will serve as a stimulus for new researchers and also become a landmark to the body of the knowledge in the field. We have presented the articles on quantum dots first, then on quantum wires and finally on 2D electron gases. A brief account of each chapter is given below:

The first chapter by Avraham Schiller starts with a brief historical note on the Kondo problem. The Anderson Hamiltonian for the ultra-small quantum dot is then mapped onto the Kondo Hamiltonian applying a suitable canonical transformation eliminating charge fluctuations. A detailed study of resistivity and conductance for tunneling through ultra-small quantum dots is given. The Toulouse limit, where the model can be solved exactly using standard techniques is studied here using Abelian bosonization. At $T = 0$ and $B = 0$, a Lorentzian zero-bias anomaly is observed in the differential conductance as a function of voltage bias. Nonzero temperature smears out the zero-bias anomaly and nonzero magnetic field splits the peak into two. In this article, a diagrammatic approach known as noncrossing approximation (NCA) to the Kondo problem is also introduced within slave boson representation. There is a sharp Abrikosov-Suhl resonance near the Fermi level in the equilibrium dot

density of states. This resonance splits as the voltage bias sufficiently exceeds the Kondo temperature which is also supported by experiments.

The second chapter by Yuval Oreg and David Goldhaber-Gordon reviews a theoretical analysis of a system consisting of a large electron droplet coupled to a small electron droplet. This system displays two-channel Kondo behavior at experimentally accessible temperatures. Special emphasis is put on the estimate of the two-channel Kondo energy scale using a perturbative renormalization group approach. Their predictions for the differential conductance in a scaling form is convenient for experimental analysis. They have also pointed out some open questions.

In the third chapter K. Kikoin and Y. Avishai show that a new ingredient in the study of the Kondo effect in quantum dots (also called artificial molecules) is the internal symmetry of the nano-object, which proves to play a crucial role in the construction of the effective exchange Hamiltonian. This internal symmetry combines continuous spin symmetry ($SU(2)$) and discrete point symmetry (such as mirror reflections for double dots or discrete C_{3v} rotation for equilateral triangular dots). When these artificial molecules are attached to metallic leads, the effective exchange Hamiltonian contains operators which couple states belonging to different irreducible representations of the internal symmetry group. In many cases, the set of dot operators appearing in the effective exchange Hamiltonian generate a group which is referred to as the dynamical symmetry group of the system dot-leads. These dynamical symmetry groups are mostly $SO(n)$ or $SU(n)$. One of the remarkable outcomes of their study is that the pertinent group parameters (such as the value of n) can be controlled by experimentalists. The reason for that is that the Kondo temperature turns out to be higher around the points of accidental degeneracy where the dynamical symmetry is "more exact" and these points can be tuned by experimental parameters such as gate voltages and tunneling strength. In this review the authors have clarified and expanded these concepts, and discussed some specific examples. They go from "light to heavy" starting from a simple quantum dot, moving on to discuss double quantum dot (where only permutation (reflection) symmetry can be considered as internal one) and finally elaborate on a triple quantum dot. In particular they concentrate on the difference between the chain geometry (where the three dots composing the triple dot are arranged in series) and the ring (triangular) geometry. When a perpendicular magnetic field is applied, the triple quantum dot in the ring geometry displays a remarkable combination of symmetries: $U(1)$ of the electromagnetic field, $SU(2)$ of the dot spin and C_{3v} of the dot orbital dynamics. The magnetic field controls the crossover between $SU(2)$ and $SU(4)$ dynamical symmetries and this feature shows up clearly in the conductance versus magnetic field curve.

The fourth chapter with contribution from Alex Grishin, Igor V. Yurkevich and Igor V. Lerner describes some essential features of loss of coherence by a qubit (controllable two-level system) coupled to the environment. They first presented the well-known semiclassical arguments that relate both de-

coherence and relaxation to the environmental noise. Then they show that models with pure decoherence (but no relaxation in qubit states) are exactly solvable. As an example, they have treated in detail the model of fluctuating background charges which is believed to describe one of the most important channels of decoherence for the charge Josephson junction qubit. They show that the decoherence rate is linear in T at low temperatures and saturates to a T-independent classical limit at 'high' temperatures, while depending in all the regimes non-monotonically on the coupling of the qubit to the fluctuating background charges. They have also considered, albeit only perturbatively, the qubit relaxation by the background charges and demonstrated that a quasi-linear behavior of the spectral density of noise deduced from the measurements of the relaxation rate can be qualitatively explained.

The contribution by Katsunori Wakabayashi in the fifth chapter elucidates the role of the edge states on the low-energy physical properties of nanographite systems. He first discussed the basics of the electronic properties of the nanographte ribbons and pointed out the existence of edge-localized states near the zigzag edge. He then presented the electronic properties of the nanographite systems in the presence of magnetic field and provides a simple picture for the origin of half-integer quantum Hall effect in graphene. The study of the orbital and Pauli magnetization shows that a nanographite system with zigzag edges exhibits strong paramagnetic response at low-temperature due to the edge states, and there exist a crossover from a weak diamagnetic response at room temperature to a strong paramagnetic response at low temperature. It is also observed that electron-electron interaction can produce a ferrimagnetic spin polarization along the zigzag edge. In this article author also describes the electron transport properties of nanographite ribbon junctions. A single edge state cannot contribute to electron conduction due to the non-bonding character of the edge states. However, in the zigzag ribbons edge states can provide a single-channel for electron conduction in the low-energy region due to the bonding and anti-bonding interaction between the edge states. The remarkable feature is the appearance of zero-conductance dips in the single-channel region where current vortex with Kekulé pattern is observed. Its relation with the asymmetric Aharonov-Bohm ring is also discussed.

The sixth chapter by K. A. Chao and Magnus Larsson is a review of the thermoelectric phenomena in nanosystems. Starting from the discovery of thermoelectric phenomenon in 1822 by Seebeck, the authors have divided the development of thermoelectricity into three stages. They pointed out that the thermodynamic theory was the driving force in the first stage, during which the Seebeck effect, the Peltier effect, the Thomson coefficient, the dual roles of thermoelectric power generation and refrigeration, and the efficiency of thermoelectric processes were extensively investigated and understood fairly well qualitatively. For a long time the practical use of thermoelectricity was measuring temperature with thermocouples. The beginning of the second stage was marked by the correct calculation of the efficiency of thermoelectric gen-

erator and refrigerator by Altenkirch in 1909. It was demonstrated that the efficiency depends mainly on a quantity which was later called the figure of merit. A higher value of this figure of merit indicates a better thermoelectric material. Using the free electron gas as a model system, Ioffe calculated the figure of merit and predicted doped semiconductors as favorable thermoelectric materials. Using the figure of merit as an indicator, and guided by the semi-classical transport theory, the search for better thermoelectric materials had lasted for a long time until around 1980s when the modern material technology enabled the fabrication of layer materials with nanometer thickness. This is the end of the second stage. In the second stage the search for new thermoelectric materials was based on the semi-classical Boltzmann transport equation, in which the dominating scattering process results in slow diffusive transport and so low value of the figure of merit. In layer materials it is possible to reduce the scattering and a new thermoelectric mechanism is found in the so-called thermionic transport. Thermionic emission of electrons from a hot surface is a well-studied physical process, and the emitted current density depends on the temperature and the work function of the emitting materials. In principle, large thermionic current can be achieved if one can reduce the work function to sufficiently low. With the advancement of material fabrication technology to produce high quality layer materials, there has been much progress in thermionics. The reduction of layer thickness in order to achieve efficient transport process also inevitably creates new fundamental problems, many of which are of quantum mechanical nature. Therefore, in the present third stage of thermoelectricity, we face the challenge of an entirely new field to which the macro-scale thermoelectric theory does not apply. This new field is the nano-scale thermoelectricity. The main theme of this chapter is to provide a smooth transition of thermoelectric phenomena from macro-scale systems to nano-scale systems.

The review article by Gilles Montambaux in the seventh chapter gives a nice introduction to coherent effects in disordered electronic systems. Avoiding technicalities as most as possible, he presented some personal points of view to describe well-known signatures of phase coherence like weak localization correction or universal conductance fluctuations. He showed that these physical properties of phase coherent conductors can be simply related to the classical return probability for a diffusive particle. The diffusion equation is then solved in various appropriate geometries and in the presence of a magnetic field. The important notion of quantum crossing is developed, which is at the origin of the quantum effects. The analogy with optics is exploited and the relation between universal conductance fluctuations and speckle fluctuations in optics is explained. The last part concerns the effect of electron-electron interactions. Using the same simple description, the author derived qualitatively the expressions of the Altshuler-Aronov anomaly of the density of states, and of the correction to the conductivity. The last part, slightly more technical, addresses the question of the lifetime of a quasi-particle in a disordered metal.

The eighth chapter by Georges Bouzerar is on the phenomenon of persistent current in mesoscopic normal metal rings. With a brief introductory note he first showed that the single particle picture can neither explain the magnitude nor the sign of the persistent current measured in diffusive metallic mesoscopic rings. This naturally lead him to the main part of the article – the interplay between electron-electron interaction and disorder. One important result is that electron-electron interaction can either enhance or suppress persistent current depending on the strength of the interaction. The underlying physics has been discussed in details.

The ninth chapter by Santanu K. Maiti and S. N. Karmakar focuses on electron transport through nanostructures. The authors first briefly introduce the Green's function technique in this study. Electron transmission through various molecular bridges are investigated in detail within the tight-binding framework. They show that the transport properties through such bridges are highly sensitive to relative position of the atoms in the molecule, coupling between molecule and electrodes, and also to the external magnetic or electric fields. The theoretical results are in qualitative agreement with the experimental observations. These model calculations provide better physical understanding of the transport problems through nanostructures. The authors have suggested some molecular devices in which electron transport can be tuned efficiently.

Finally, the tenth chapter by S. Sil, S. N. Karmakar and Efrat Shimshoni is on quantum Hall effect. This article provides an account of the exotic statistical nature of the quasi-particles in quantum Hall system. For instance, an electron in the presence of electron-electron interaction and strong magnetic field may undergo Bose condensation by charge-flux composite, and fractional charge excitations emerge as quasi-particles. These quasi-particles manifest lot of surprises in the studies of quantum Hall systems. This review is on both the integer and fractional quantum Hall effects within the field theoretic framework. In this review, the authors have also discussed the role of the edge states on integer and fractional quantum Hall effects to understand the experimental results.

It is our great pleasure to thank Prof. Yuval Gefen and Prof. Bikas K. Chakrabarti for their invaluable cooperation and support in organizing the international workshop without which this book might have not seen the light of the day. In this context, we also thank Prof. Hans Weidenmueller, Prof. Yoseph Imry, Prof. Markus Buttiker, Prof. Amnon Aharony and Prof. Yigal Meir for their advices and encouragements. We wish to thank all the invited speakers who made the workshop successful and all the authors who have contributed to this volume. We are very much grateful to Prof. Peter Fulde and Dr. Claus Ascheron for recommending the publication of this book. Thanks are also due to Dr. Angela Lahee and Dr. Elke Sauer from Springer-Verlag for friendly collaboration. Fine help and constant encouragement from our colleagues in this endeavor is also highly appreciated. Finally, we thank the

"Centre for Applied Mathematics and Computational Science (CAMCS)" of Saha Institute of Nuclear Physics in India for providing financial support.

Kolkata,
January 2007

Sachindra Nath Karmakar
Santanu K. Maiti
Jayeeta Chowdhury

Contents

List of Contributors

Y. Avishai
Department of Physics
Ben-Gurion University
Beer, Sheva 84105, Israel
yshai@bgumail.bgu.ac.il

Georges Bouzerar
Laboratoire Louis Néel, CNRS
25 avenue des Martyrs BP 166
F-38042 Grenoble Cedex 9, France
bouzerar@ill.fr

K. A. Chao
Department of Physics
Lund University, Sölvegatan 14A
S-223 62 Lund, Sweden
chao@teorfys.lu.se

David Goldhaber-Gordon
Geballe Laboratory for Advanced
Materials and Department of Physics
Stanford University
Stanford, California 94305, USA
goldhaber-gordon@stanford.edu

Alex Grishin
School of Physics and Astronomy
University of Birmingham
UK
ag@th.ph.bham.ac.uk

S. N. Karmakar
Theoretical Condensed Matter
Physics Division
Saha Institute of Nuclear Physics
1/AF, Bidhannagar
Kolkata 700 064, India
sn.karmakar@saha.ac.in

K. Kikoin
Department of Physics
Ben-Gurion University
Beer, Sheva 84105, Israel
kikoin@bgu.ac.il

Magnus Larsson
Nanofreeze Technologies Lund AB
Ole Rmers Väg 12
S-223 70 Lund, Sweden
magnus.larsson@nanofreeze.se

Igor V. Lerner
School of Physics and Astronomy
University of Birmingham
UK
ivl@th.ph.bham.ac.uk

Santanu K. Maiti
Theoretical Condensed Matter
Physics Division
Saha Institute of Nuclear Physics
1/AF, Bidhannagar
Kolkata 700 064, India
santanu.maiti@saha.ac.in

Gilles Montambaux
Laboratoire de Physique des Solides
associé au CNRS
Université Paris–Sud
91405 Orsay, France
montambaux@lps.u-psud.fr

Yuval Oreg
Department of Condensed Matter
Physics
Weizmann Institute of Science
Rehovot 76100, Israel
yuval.oreg@weizmann.ac.il

Avraham Schiller
Racah Institute of Physics
The Hebrew University
Jerusalem 91904, Israel
avaraham@phys.huji.ac.il

Efrat Shimshoni
Department of Math-Physics
Oranim – University of Haifa
Tivon 36006, Israel
efrats@research.haifa.ac.il

S. Sil
Department of Physics
Visva-Bharati University
Santiniketan 731 235, India
sreekantha.sil@saha.ac.in

Katsunori Wakabayashi
Department of Quantum Matter
AdSM, Hiroshima University
Higashi-Hiroshima 739-8530, Japan
kwaka@hiroshima-u.ac.jp

Igor V. Yurkevich
School of Physics and Astronomy
University of Birmingham
UK
ivy@th.ph.bham.ac.uk

From Dilute Magnetic Alloys to Confined Nanostructures: Evolution of the Kondo Effect

Avraham Schiller

Racah Institute of Physics, The Hebrew University, Jerusalem 91904, Israel.
avaraham@phys.huji.ac.il

1 Introduction

The Kondo problem occupies a central chapter in condensed matter physics, with a long history in dilute magnetic alloys and valence-fluctuating systems. Originally observed some 70 years ago as a minimum in the resistivity of dilute magnetic alloys, the Kondo effect has evolved in time into a paradigmatic example for strong electronic correlations in condensed matter physics. It pertains to the many-body screening of an impurity spin by the surrounding conduction electrons, leading to the formation of a strong scattering center at low temperatures. Besides the dramatic effect on the resistivity of otherwise pure metals, the Kondo effect is manifested in anomalous enhancements of thermodynamic and dynamic properties such as the specific heat, magnetic susceptibility, and thermopower to name a few. Over the past 40 years, the Kondo effect has played a pivotal role in the development of the field of strongly correlated electron systems. Many of the basic concepts and notions of the field have either been conceived or significantly advanced in the Kondo arena. Notable examples are the renormalization-group ideas of Anderson [1, 2] and Wilson [3]. Nearly all techniques of modern many-body physics have been applied to the problem, which continues to serve as an important testing ground for new approaches.

The last decade has witnessed a dramatic resurgence of experimental interest in the Kondo effect following its discovery in lithographically defined quantum dots [4–6] and its measurement for isolated magnetic adatoms on metallic surfaces [7, 8]. In contrast to real magnetic impurities, quantum dots can be controlled in exquisite detail, and can be tuned at will from weak coupling to the Kondo regime. The precise control of the microscopic model parameters in combination with the advanced capabilities of detailed sample engineering have turned quantum-dot devices into a valuable testing ground

for our fundamental understanding of electronic correlations. Scanning tunneling microscopy of individual magnetic adatoms offers the complementary ability to spatially resolve the electronic structure around the impurity. Although lacking the enormous flexibility of quantum dots in terms of designing and tuning the microscopic parameters of an individual impurity, magnetic adatoms can be manipulated into forming small clusters [9] as well as novel resonators [10]. These settings offer an ideal setup for probing the interplay between interactions and quantum interference. Other nanostructures where Kondo physics has recently been observed include nanotube quantum dots [11, 12], single-atom transistors [13], and single-molecule [14] transistors. Each of these systems has its own distinct advantage toward sampling new aspects of Kondo physics. For example, nanotube quantum dots were deposited on superconducting electrodes in order to study the interplay of Kondo physics and superconductivity [15], whereas the effect of ferromagnetic leads was interrogated using C60 molecules as magnetic impurities [16].

Parallel to the flurry of activity in the mesoscopic realm, the renewed interest in Kondo physics has been amplified by important developments in the context of correlated electron systems. The first of these developments is of purely theoretical nature and goes under the name of dynamical mean-field theory [17] (DMFT). The DMFT has become one of the primary methods for studying strong electronic correlations. It is based on the mapping of a lattice problem onto that of a quantum impurity, self-consistently embedded in an effective medium. The main virtue of the method is that it captures all local time-dependent correlations, allowing for detailed studies of phenomena such as the Mott-Hubbard metal-insulator transition or the phase diagram of different Kondo lattices. In the last few years the method has matured into a highly advanced tool for studying real materials [18, 19]. However, its successful implementation relies on the availability of highly accurate, flexible, and efficient methods for solving the associated impurity problem in the presence of a structured density of states (DOS). This necessity has led to a vigorous quest for quantum-impurity solvers that can cope with the somewhat unconventional variants of the Kondo problem encountered in DMFT.

Another important development to be noted is the emergence of the concept of quantum criticality. There is increasing evidence that some of the deviations from conventional Fermi-liquid behavior observed in certain heavy fermion compounds and in the cuprates may be due to the proximity to a quantum critical point, where a transition temperature is suppressed to zero. In lattice systems, the nature of such quantum critical points is still not well-understood. While theoretical descriptions typically start from well-defined quasi-particle excitation modes, the non-Fermi-liquid behavior is interaction driven, and arises from persisting quantum-mechanical fluctuations between these modes. The multi-channel Kondo effect [20] provides one of the outstanding paradigms for a *local* quantum critical point, where the concept of quantum criticality can be studied in great detail. Suitably designed nanos-

tructures may again provide a valuable testing ground for confronting theory with well-controlled experiments on local quantum criticality.

As emphasized above, the Kondo effect is an old problem in condensed-matter physics. Numerous reviews have been devoted to this fundamental problem, the most comprehensive of which is the book by Hewson [21]. While earlier reviews of the problem were naturally focused on classical realizations of the Kondo effect, more recent reviews (see, e.g., [22]) mainly address its manifestations in quantum dots. It is the intention of the current article to provide somewhat of a bridge between these complementary points of view. The objective of this article is three-fold: (i) To briefly review the evolution of the Kondo problem from its classic realizations in dilute magnetic alloys and valence-fluctuation systems to its current manifestations in confined nanostructures. (ii) To provide an exposition of the basic physics of the Kondo effect in nanostructures. (iii) To present some of the theoretical techniques that are available for studying this fundamental effect. Since a comprehensive review of the many theoretical approaches that have been devised and applied over the years to the problem is beyond the scope of this article, we chose to highlight just a few leading methods. Our selection of methods is motivated in part by the particularly transparent physical picture they provide, and in part by their extensions to non-equilibrium conditions.

2 Brief Historical Notes

To put the renewed interest in the Kondo problem in historical perspective, we provide here a brief, albeit highly selective list of milestones in the history of this effect. The Kondo effect was first observed in the 1930s as a resistivity minimum in noble-metal samples containing small amounts of magnetic impurities [23]. This behavior marked a striking departure from Mathiessens' rule that prevailed at the time, and which states that the total resistivity of a crystalline metallic specimen is the sum of the resistivity due to electron-phonon scattering and the residual resistivity due to the presence of imperfections in the crystal. The cause of the resistivity minimum remained obscure for many years. It was only in 1964, thirty years after its discovery, that Kondo demonstrated [24] that it originated from scattering off individual magnetic impurities.

Kondo's theory was inspired by mounting experimental evidence in the late 1950s and early 1960s that correlated the occurrence of the resistivity minimum with a Curie-Weiss term in the impurity susceptibility. To this end, he considered the so-called antiferromagnetic *s-d* exchange Hamiltonian, which nowadays bears his name:

$$\mathcal{H} = \sum_{k,\sigma} \epsilon_k c^\dagger_{k\sigma} c_{k\sigma} + J \boldsymbol{S}_{\rm imp} \cdot \boldsymbol{s}(0) \; . \tag{1}$$

Here $\boldsymbol{S}_{\rm imp}$ is the impurity spin, $\boldsymbol{s}(0)$ is the conduction-electron spin density at the impurity site, and $J > 0$ is the spin-exchange coupling. Using third-order perturbation theory in J, Kondo predicted a logarithmic increase in the impurity contribution to the resistivity, which actually diverged for $T \to 0$. Combining the logarithmic impurity contribution to the resistivity with the T^5 contribution that stems from electron-phonon scattering, Kondo's calculations successfully explained the one-fifth power-law relation between the concentration of magnetic impurities, $c_{\rm imp}$, and the temperature $T_{\rm min}$ at which the resistivity develops it minimum: $T_{\rm min} \propto c_{\rm imp}^{1/5}$.

Kondo's discovery of unexpected logarithmic divergences in perturbation theory has generated considerable theoretical interest in the problem, aimed at finding a solution valid in the low-temperature regime. Numerous approaches were devised in the mid 1960s, but were met with only partial success. Using an infinite resummation procedure, the Suhl-Nagaoka theory [25,26] provided the first resistivity calculation respecting unitarity. However, it produced unphysical results for $T \to 0$. Yosida's variational wave function predicted a singlet ground state for antiferromagnetic exchange [27]. However, the approach was restricted to zero temperature and predicted an erroneous exponential form for the so-called Kondo temperature. The first flavor of scaling ideas had appeared in the work of Abrikosov [28], who introduced a convenient representation of the impurity spin in terms of auxiliary (slave) fermions. Carrying out an infinite summation of the leading logarithmic divergences for the Kondo Hamiltonian, Abrikosov showed that the bare coupling J can be replaced at temperature T with an effective coupling

$$J_{\rm eff}(T) = \frac{J}{1 + \rho J \ln(k_{\rm B}T/D)} \; . \tag{2}$$

Here ρ is the DOS at the Fermi energy and D is the half bandwidth. An important corollary of Abrikosov's calculation was the distinction between antiferromagnetic and ferromagnetic exchange. The impurity contribution to the resistivity was shown to be free of logarithmic diverges when the spin-exchange was ferromagnetic.

A major step forward in the understanding of the Kondo effect was taken in the late 1960s, in the ground-breaking work of Anderson and coworkers [1,2]. In these papers, the ideas of scaling and the renormalization group (RG) had been put forward. The basic idea underlying these works was the realization that lowering the temperature can be translated into a continuous evolution of the effective low-energy Hamiltonian, describing excitations on the scale of the temperature. This philosophy is particularly transparent in Anderson's "poor man's scaling" treatment of the Kondo Hamiltonian [2]. Upon lowering the temperature, high-energy electronic states are quenched. Integration of the frozen electronic degrees of freedom maps the Hamiltonian onto an effective low-energy Hamiltonian with renormalized parameters. By going continuously to smaller and smaller bandwidths, or temperature, a sequence of effective Hamiltonians is thus generated, with couplings that vary

smoothly as a function of the effective bandwidth. All Hamiltonians that flow along the same scaling trajectory share the same low-energy physics, with the Kondo temperature

$$k_{\mathrm{B}}T_{\mathrm{K}} = D\sqrt{\rho J}\ \exp\left[-\frac{1}{\rho J}\right] \tag{3}$$

playing the role of a scaling invariant.

The calculations of Anderson and coworkers were perturbative in nature, and consequently broke down as soon as the renormalized exchange coupling became large. Nonetheless, they suggested that the scaling procedure could be continued into the non-perturbative regime, and that an initial antiferromagnetic coupling would renormalize to infinity. This hypothesis was confirmed by Wilson [3], who devised a non-perturbative renormalization-group method — the numerical renormalization group (NRG) — for accurately solving the thermodynamics of the Kondo model for all parameter regimes. Wilson's approach provided the first complete solution of the Kondo problem. The fixed-point structure of the model was mapped out, and the universal scaling function for the susceptibility was obtained. The impurity spin was shown to be progressively screened by the conduction electrons, reaching complete screening for $T \to 0$. The low-temperature strong-coupling fixed point was found to be that of infinite coupling, corresponding to the formation of a local Fermi liquid. Thirty years after its development, Wilson's approach continues to be the leading non-perturbative method for solving quantum-impurity problems.

By the mid 1970s, the original Kondo problem was essentially solved. A first wave of renewed interest in the problem was sparked in the early 1980s by the exact Bethe *ansatz* solutions of the *s*-*d* [29,30] and Anderson [31] models, and by intense experimental investigations of alloys with rare earth elements such as Ce and Yb. The large orbital degeneracy in these ions ($N = 6$ for Ce^{3+} and $N = 8$ for Yb^{3+}) precluded the application of the NRG, which could not cope with the large computational effort involved. This has led to the pursuit of large-N approaches that treat $1/N$ as a small parameter, whether in the framework of mean-field theory [32, 33], diagrammatic calculations [34, 35], or variational wave functions [36]. Large-N approaches proved to be a very useful platform for comparison with experiment. They were extensively used to calculate dynamic quantities not accessible by the Bethe *ansatz* and NRG, such as the impurity DOS and the dynamic susceptibility.

Just as the theory of the Kondo effect appeared to be reaching its final plateau, another extremely powerful approach was added to the arsenal in the early 1990s: boundary conformal field theory (BCFT) [37]. The key idea of BCFT is that the impurity spin can be replaced at sufficiently low energies by a conformally invariant boundary condition for the linearly dispersed conduction electrons. The nature of the boundary condition must comply with the underlying symmetries of the problem, but cannot be generally determined based on symmetry considerations alone. A comparison with other solutions of the problem, e.g., the finite-size spectrum obtain by the NRG, is usually required in order to uniquely determine the boundary condition, or fixed point,

that applies. Once at hand, this enables a complete analytical characterization of the fixed point reached, including the leading irrelevant operators, their physical content, and the exact leading temperature and frequency dependences of thermodynamic and dynamic properties.

Although traditional investigations of the Kondo effect were mainly focused on bulk systems, the study of Kondo-assisted tunneling likewise dates back to the 1960s. The phenomenon was first discovered by accident [38, 39], when magnetic impurities were present in planner tunnel junctions between two normal metals. A zero-bias anomaly was seen, which enhanced the conductance at low voltages. Shortly after the original experiments, Appelbaum [40] and Anderson [41] developed a perturbative theory that captured the essential features of the experiment: a zero-bias conductance that increased logarithmically with decreasing temperature, and a zero-bias anomaly that split in the presence of a sufficiently large magnetic field. Although quite successful in explaining the qualitative and in some cases the quantitative results, the Appelbaum–Anderson theory was perturbative, and hence could not describe the strong-coupling regime of the Kondo effect. An adequate theory of nonequilibrium conditions remains an outstanding challenge today as well.

The early experiments on tunnel junctions probed the simultaneous tunneling through many impurities. It was not before the early 1990s that tunneling through a single impurity was first observed in two separate experiments: in a crossed-wire tunnel junction formed between tungsten wires [42], and in tunneling through individual charge traps formed in a point-contact tunnel barrier [43]. These two experiments marked an important step forward in the study of Kondo-assisted tunneling. However, in spite of the compelling evidence in favor of tunneling through a single magnetic impurity, neither experiment permitted a microscopic characterization of defects involved, let alone a microscopic control of their model parameters. This situation has changed dramatically in 1998 with the discovery of the Kondo effect in lithographically defined quantum dots [4, 5], and its measurement for isolated magnetic adatoms on metallic surfaces [7, 8].

3 Essentials of the Kondo Effect

3.1 The Anderson Impurity Model

The standard description of magnetic impurities in metals is by means of the Anderson model [44]. Introduced in 1961 in an effort to explain how local moments are formed in a metal, the model describes the hybridization of an interacting level, or ion, with the itinerant electrons of the metal. The Anderson model has three main ingredients: (i) a localized level with energy ϵ_d, corresponding to the outer-most atomic shell; (ii) an on-site repulsion U, representing the screened Coulomb repulsion between a pair of electrons in the outer-most shell; and (iii) an hybridization matrix element t between the

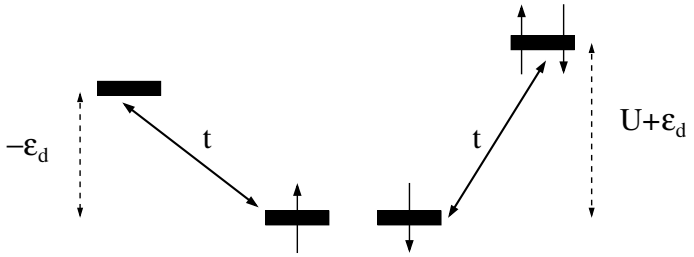

Fig. 1. Schematic description of the non-degenerate Anderson model

atomic electrons and the underlying conduction electrons of the metal. In its simplest non-degenerate form, the Anderson Hamiltonian reads

$$\mathcal{H} = \sum_{k,\sigma} \epsilon_k c^\dagger_{k\sigma} c_{k\sigma} + \sum_{k,\sigma} t \left\{ c^\dagger_{k\sigma} d_\sigma + \text{h.c.} \right\} + \epsilon_d \sum_\sigma d^\dagger_\sigma d_\sigma + U \hat{n}_\uparrow \hat{n}_\downarrow \,, \qquad (4)$$

where $d^\dagger_\sigma$ creates an atomic electron with spin projection σ, $c^\dagger_{k\sigma}$ creates a conduction electron with wave number k and spin projection σ in an s-wave state centered about the ion, and $\hat{n}_\sigma = d^\dagger_\sigma d_\sigma$ are the local number operators.

A realistic description of lanthanide and actinide ions requires inclusion of the orbital degeneracy of the f-shell electrons, which is lifted in turn by the spin-orbit and crystalline-electric-field terms. In practice this means replacing the spin index σ in (4) with a combined index m that runs over all relevant atomic states. The spin-orbit and crystalline-electric-field splittings of the level are accounted for by substituting $\epsilon_d \to \epsilon_m$. The orbitally degenerate model, commonly referred to as the degenerate Anderson model, has been extensively used for describing Ce- and Yb-based alloys. We shall focus, however, on the non-degenerate case, which usually suffices for modeling the Kondo effect in ultra-small quantum dots. A schematic illustration of the model is depicted in Fig. 1.

3.2 Anderson-Model Description of Ultra-Small Dots

Quantum dots behave in many respects as tunable "artificial" atoms. Due to quantum confinement, single-particle levels are discrete inside the dot with a finite level spacing Δ. When sufficiently "pinched off" from the leads, and for temperatures small as compared to Δ, the dot can be modeled by only one single-particle level ϵ_0. The electrostatic energy of the dot is well represented in this case by the classical charging term

$$\mathcal{H}_{\text{charging}} = E_C \left(\sum_\sigma d^\dagger_\sigma d_\sigma - N_g \right)^2 \,, \qquad (5)$$

where $E_C = e^2/2C$ is the charging energy, C is the total capacitance of the dot, and N_g is the dimensionless gate voltage. Here $d^\dagger_\sigma$ creates a spin-σ electron in the relevant dot level.

The charging energy E_C clearly plays the role of the on-site repulsion U for an Anderson impurity. The remaining ingredient required for an Anderson-model description of the dot is furnished by the tunneling to the leads. Denoting the creation of a conduction electron in the left and right lead by $c^\dagger_{\mathrm{L}k\sigma}$ and $c^\dagger_{\mathrm{R}k\sigma}$, respectively, the corresponding tunneling term reads

$$\mathcal{H}_{\mathrm{tunneling}} = \sum_{\alpha=\mathrm{L,R}} \sum_{k,\sigma} t_\alpha \left\{ c^\dagger_{\alpha k\sigma} d_\sigma + \mathrm{h.c.} \right\} , \tag{6}$$

where t_L and t_R are the tunneling matrix elements to the left and right lead. Although the coupling to two leads may appear different at first from the coupling to a single band in equation (4), this distinction turns out to be cosmetic. The physical reason is simple, as the dot couples only to the "bonding" combination

$$b^\dagger_{k\sigma} = \frac{1}{\sqrt{t_\mathrm{L}^2 + t_\mathrm{R}^2}} \left[t_\mathrm{L} c^\dagger_{\mathrm{L}k\sigma} + t_\mathrm{R} c^\dagger_{\mathrm{R}k\sigma} \right] . \tag{7}$$

The orthogonal "anti-bonding" combination

$$a^\dagger_{k\sigma} = \frac{1}{\sqrt{t_\mathrm{L}^2 + t_\mathrm{R}^2}} \left[t_\mathrm{R} c^\dagger_{\mathrm{L}k\sigma} - t_\mathrm{L} c^\dagger_{\mathrm{R}k\sigma} \right] \tag{8}$$

decouples from the impurity, much in the same way as all partial waves but one decouple from a magnetic impurity when placed in a metal. Thus, upon converting to the "bonding"–"anti-bonding" basis, one recovers the Hamiltonian of (4) with the following correspondence of parameters:

$$\epsilon_d = \epsilon_0 - 2E_C N_g , \quad U = E_C , \quad \text{and} \quad t = \sqrt{t_\mathrm{L}^2 + t_\mathrm{R}^2} . \tag{9}$$

Two comments should be made at this point about the hopping t quoted above. Firstly, in writing t of (9) we have implicitly assumed that the two leads share the same dispersion ϵ_k, or DOS ρ. If the conduction-electron DOS is different in both leads, i.e., $\rho_\mathrm{L} \neq \rho_\mathrm{R}$, then the correspondence of parameters is slightly modified. Specifically, the hybridization width associated with (4), $\Gamma = \pi\rho t^2$, must equal the sum of the hybridization widths to the two leads: $\Gamma = \Gamma_\mathrm{L} + \Gamma_\mathrm{R}$ with $\Gamma_\alpha = \pi\rho_\alpha t_\alpha^2$.

The second comment pertains to the role of a finite voltage bias. Although the "anti-bonding" modes remain decoupled from the impurity at the level of the Hamiltonian, they cannot be discarded for a finite voltage bias. This stems from the fact that the boundary condition imposed by the bias applies to the left- and right-lead electrons rather than the "bonding" and "anti-bonding" ones. Hence, the full two-lead Hamiltonian must be retained for a finite voltage bias.

3.3 From the Anderson Model to the Kondo Hamiltonian

The physics of the Anderson impurity model depends continuously on the interplay between the hybridization width Γ and the inter-configurational energies ϵ_d and $\epsilon_d + U$ (recall that $U > 0$). At low temperatures the model flows to a line of Fermi-liquid fixed points, parameterized by the scattering phase shift δ. The latter is given in the wide-band limit by $\delta = \pi n_d/2$, where $n_d = \langle \hat{n}_\uparrow + \hat{n}_\downarrow \rangle$ is the total impurity occupancy at zero temperature (see Sect. 4.1 below). It is customary to distinguish between several continuously connected regimes of the model.

- When $\epsilon_d \gg \Gamma$, the level is essentially depopulated for $k_{\rm B}T \ll \epsilon_d$. This regime, commonly referred to as the empty-orbital regime, is characterized by a small phase shift, $\delta \approx 0$, for $T = 0$.
- When $-\epsilon_d, U + \epsilon_d \gg \Gamma$, a local moment is formed on the level for $k_{\rm B}T \ll \min\{-\epsilon_d, U + \epsilon_d\}$. It is in this regime that the Kondo effect takes place. The local moment is initially coupled weakly to the band electrons, but is then progressively screened as the temperature T is lowered below the Kondo temperature $T_{\rm K}$. Eventually a Fermi-liquid fixed point is reached with a phase shift, δ, close to $\pi/2$.
- When $-(\epsilon_d + U) \gg \Gamma$, the level is doubly occupied for $k_{\rm B}T \ll -(\epsilon_d + U)$. This regime can be viewed as the inverted particle-hole image of the empty-orbital regime.
- If either $\Gamma \geq |\epsilon_d|$ or $\Gamma \geq |\epsilon_d + U|$, the impurity valence undergoes strong quantum fluctuations between two or more charge configurations. The impurity does not have stable valance in this case, and is said to be in the mixed-valent regime.

Let us focus on the local-moment regime, where the Kondo effect takes place. As the temperature is lowered below $T_L \sim k_{\rm B}^{-1} \min\{-\epsilon_d, U + \epsilon_d\}$, charge fluctuations are suppressed on the impurity and a stable local moment is formed. The role of the two excited charge configurations (the empty and doubly occupied ones) is demoted to mediating virtual transitions among the two spin states. The adequate way to formulate an effective low-energy Hamiltonian in this regime is by means of the so-called Schrieffer-Wolff transformation [45]. A suitable canonical transformation $\mathcal{H}' = e^S \mathcal{H} e^{-S}$ is applied to the Anderson Hamiltonian such that charge fluctuations are eliminated within $\mathcal{H}'$ up to second order in the hybridization matrix element t. To this end, the anti-hermitian operator S is taken to be

$$S = \sum_{k,\sigma} \left\{ \left[\frac{t}{\epsilon_k - \epsilon_d} (1 - \hat{n}_{-\sigma}) + \frac{t}{\epsilon_k - \epsilon_d - U} \hat{n}_{-\sigma} \right] c^\dagger_{k\sigma} d_\sigma - \text{h.c.} \right\} . \qquad (10)$$

Expanding $\mathcal{H}'$ up to second order in t and projecting the resulting Hamiltonian onto the singly occupied space, one recovers the Kondo Hamiltonian

$$\mathcal{H}_{\rm K} = \sum_{k,\sigma} \epsilon_k c^\dagger_{k\sigma} c_{k\sigma} + J \boldsymbol{S}_{\rm imp} \cdot \boldsymbol{s}(0) + V \sum_{k,k',\sigma} c^\dagger_{k\sigma} c_{k'\sigma} \; , \tag{11}$$

where

$$\boldsymbol{S}_{\rm imp} = \frac{1}{2} \sum_{\sigma,\sigma'} d^\dagger_\sigma d_{\sigma'} \boldsymbol{\tau}_{\sigma\sigma'} \tag{12}$$

is the impurity spin, and

$$\boldsymbol{s}(0) = \frac{1}{2} \sum_{k,k'} \sum_{\sigma,\sigma'} c^\dagger_{k\sigma} c_{k\sigma'} \boldsymbol{\tau}_{\sigma\sigma'} \tag{13}$$

is the conduction-electron spin density at the impurity site. Here $\boldsymbol{\tau}$ are the Pauli matrices. The exchange coupling J and potential scattering V are given in turn by

$$J = 2t^2 \left[\frac{1}{|\epsilon_d|} + \frac{1}{U + \epsilon_d} \right] \; , \qquad V = \frac{t^2}{2} \left[\frac{1}{|\epsilon_d|} - \frac{1}{U + \epsilon_d} \right] \; , \tag{14}$$

where we have omitted momentum dependence. The effective bandwidth pertaining to the Hamiltonian of (11) is $D_{\rm eff} = k_{\rm B} T_L \sim \min\{-\epsilon_d, U + \epsilon_d\}$.

In writing the couplings of (14), we have expressed them in terms of the bare parameters that appear in the parent Anderson Hamiltonian. As noted by Haldane [46], the energy level ϵ_d is actually slightly renormalized if $U + 2\epsilon_d \neq 0$. Specifically, ϵ_d is modified to

$$\epsilon_d + \frac{\Gamma}{\pi} \ln \left(\frac{U + \epsilon_d}{|\epsilon_d|} \right) \tag{15}$$

when $D \gg |\epsilon_d|, U$. Although the above correction to ϵ_d is typically small, it can substantially modify the pre-exponential factor that appears in the expression for the Kondo temperature, and therefore should generally be retained.

Once the mapping onto the Kondo Hamiltonian has been established, one can make use of known results for the latter model in order to unravel the low-energy physics. Specifically, the antiferromagnetic spin-exchange J flows to strong coupling, with $T_{\rm K}$ marking the departure from weak coupling. The most accurate expression for $T_{\rm K}$ is provided by the Bethe *ansatz* solution of the Anderson model [31]. Defining the Kondo temperature according to the zero-temperature susceptibility,

$$\chi(T = 0) = \frac{(\mu_{\rm B} g)^2}{2\pi k_{\rm B} T_{\rm K}} \; , \tag{16}$$

one finds [31]

$$k_{\rm B} T_{\rm K} = (\sqrt{2U\Gamma}/\pi) \, \exp\left[\pi \left(\Gamma^2 + \epsilon_d \, U + \epsilon_d^2\right) / (2 \, U \, \Gamma)\right] \tag{17}$$

in the wide-band limit.

4 Resistivity and Conductance

So far we mainly considered the low-energy fixed-point structure. The key quantity for computing transport properties, whether in dilute magnetic alloys or in a tunnel-junction geometry, is the conduction-electron $\mathcal{T}$-matrix. For the non-degenerate Anderson model, the latter is independent of the momentum of the scattered electron, and is related to the impurity Green function G_d through

$$\mathcal{T}(z) = t^2 G_d(z) \; . \tag{18}$$

In dilute magnetic alloys, scattering off the impurities hampers the ballistic motion of the electrons. For a small concentration of magnetic impurities, $c_{\rm imp} \ll 1$, the impurity contribution to the resistivity is given by [47]

$$\rho^{-1}(T) = \frac{ne^2}{m} \int_{-\infty}^{\infty} \tau(\epsilon) \left(-\frac{\partial f}{\partial \epsilon} \right) {\rm d}\epsilon \; , \tag{19}$$

where

$$\hbar \tau^{-1}(\epsilon) = c_{\rm imp} 2\pi t^2 A_d(\epsilon) \tag{20}$$

is the transport time of the electrons. Here n and m are the conduction-electron density and mass, while $A_d(\epsilon) = -\frac{1}{\pi}{\rm Im}\{G_d(\epsilon + {\rm i}\eta)\}$ is the impurity spectral function. The residual $T = 0$ resistivity is therefore equal to

$$\rho(T=0) = c_{\rm imp} \frac{2\pi m}{ne^2\hbar} t^2 A_d(0) \; . \tag{21}$$

For a tunnel geometry of the type realized for ultra-small quantum dots, transport is facilitated by scattering off the impurity. In this case the *conductance* is equal to

$$G(T) = \frac{e^2}{\hbar} \frac{4\Gamma_{\rm L}\Gamma_{\rm R}}{\Gamma_{\rm L} + \Gamma_{\rm R}} \int_{-\infty}^{\infty} A_d(\epsilon) \left(-\frac{\partial f}{\partial \epsilon} \right) {\rm d}\epsilon \; , \tag{22}$$

which reduces for $T = 0$ to

$$G(T=0) = \frac{e^2}{\hbar} \frac{4\Gamma_{\rm L}\Gamma_{\rm R}}{\Gamma_{\rm L} + \Gamma_{\rm R}} A_d(0) \; . \tag{23}$$

Evidently, there is an intimate relation between the resistivity in dilute magnetic alloys and the conductivity for tunneling through an ultra-small quantum dot. Indeed, both quantities increase logarithmically for $T > T_{\rm K}$, and are proportional to $A_d(0)$ at zero temperature. However, the two quantities do not exhibit identical temperature dependences. The resistivity in alloys is proportional to the reciprocal of the average of $A_d^{-1}(\epsilon)$ with respect to the derivative of the Fermi function, whereas the conductance for a quantum dot is directly proportional to the average of $A_d(\epsilon)$ with respect to $-\partial f(\epsilon)/\partial\epsilon$. Below we focus on zero temperature, where exact results can be obtained for the conductance $G(0)$.

4.1 Friedel-Langreth Sum Rule

One of the most important results of Fermi-liquid theory for the Anderson model is the Friedel-Langreth sum rule [48], relating the zero-temperature impurity Green function at the Fermi level to the number of electrons displaced by the impurity. The derivation of the sum rule is quite compact, and relies solely on the assumption that a Fermi liquid is formed. Below we briefly outline Langreth's derivation of the sum rule [48], allowing both for spin-dependent hybridization t_σ and for the presence of a finite magnetic field h_z. While the spin dependence of t_σ may appear superfluous in the context of dilute magnetic alloys, it occurs quite naturally in the case of ferromagnetic leads [49].

The number of displaced electrons in the spin channel σ is defined as

$$\Delta N_\sigma = \langle d_\sigma^\dagger d_\sigma \rangle + \sum_k \left[\langle c_{k\sigma}^\dagger c_{k\sigma} \rangle - \langle c_{k\sigma}^\dagger c_{k\sigma} \rangle_0 \right] , \tag{24}$$

where the zero subscript denotes averaging with respect to the unperturbed (i.e., $t = 0$) Hamiltonian. Setting $T = 0$, this quantity can be conveniently expressed in terms of the dressed impurity Green function $G_{d\sigma}(z)$ as

$$\Delta N_\sigma = -\int_{-\infty}^{0} \frac{\mathrm{d}\epsilon}{\pi} \mathrm{Im} \left\{ G_{d\sigma}(\epsilon + \mathrm{i}\eta) \left[1 + \sum_k \frac{t_\sigma^2}{(\epsilon - \epsilon_k + \mathrm{i}\eta)^2} \right] \right\} , \tag{25}$$

which follows from the fact that the conduction-electron $\mathcal{T}$-matrix equals $\mathcal{T}_\sigma(\epsilon + \mathrm{i}\eta) = t_\sigma^2 G_{d\sigma}(\epsilon + \mathrm{i}\eta)$. Writing the impurity self-energy in the form $\Sigma_\sigma(z) = \Sigma_\sigma^0(z) + \Sigma_\sigma^U(z)$, where

$$\Sigma_\sigma^0(z) = \sum_k \frac{t_\sigma^2}{\epsilon - \epsilon_k + \mathrm{i}\eta} \tag{26}$$

is the non-interacting ($U = 0$) self-energy, and using the Fermi-liquid relation

$$\int_{-\infty}^{0} \frac{\mathrm{d}\epsilon}{\pi} \mathrm{Im} \left\{ G_d(\epsilon + \mathrm{i}\eta) \frac{\partial \Sigma_\sigma^U(\epsilon + \mathrm{i}\eta)}{\partial \epsilon} \right\} = 0 , \tag{27}$$

(25) is rewritten as

$$\Delta N_\sigma = -\int_{-\infty}^{0} \frac{\mathrm{d}\epsilon}{\pi} \mathrm{Im} \left\{ G_{d\sigma}(\epsilon + \mathrm{i}\eta) \left[1 - \frac{\partial \Sigma_\sigma(\epsilon + \mathrm{i}\eta)}{\partial \epsilon} \right] \right\} . \tag{28}$$

Note that the expression in the curly brackets is nothing but the derivative of $\ln G_{d\sigma}^{-1}(\epsilon + \mathrm{i}\eta)$ with respect to ϵ. Hence the right-hand side of (28) can be readily integrated to obtain

$$\mathrm{Im} \left\{ \ln G_{d\sigma}(0 + \mathrm{i}\eta) \right\} = \pi \left(\Delta N_\sigma - 1 \right) . \tag{29}$$

Combined with the Fermi-liquid property $-\mathrm{Im}\{\Sigma(0+\mathrm{i}\eta, T=0)\} = \Gamma_\sigma$, this yields the exact result

$$G_{d\sigma}(0+\mathrm{i}\eta, T=0) = -\frac{\mathrm{e}^{\mathrm{i}\delta_\sigma}}{\Gamma_\sigma}\sin\delta_\sigma \ , \tag{30}$$

where $\delta_\sigma = \pi\Delta N_\sigma$ is the scattering phase shift in the spin channel σ.

Two comments should be made about this result. Firstly, ΔN_σ differs in principle from the corresponding impurity occupancy $n_{d\sigma}$. The difference between the two quantities is negligibly small in the conventional case where the conduction-electron DOS is a slowly varying function of energy, with a bandwidth far greater than all other energy scales in the problem (the so-called wide-band limit). However, this need not be the case in mesoscopic systems, where the DOS itself may vary on small energy scales. A case in point is that of a finite-length quantum wire, studied, e.g., in [50–52]. Second, from the expression above it is clear that the impurity DOS reaches its maximal value of $1/\pi\Gamma_\sigma$ when ΔN_σ equals one half. That is, when the corresponding scattering phase shift is $\pi/2$. In the wide-band limit and for SU(2) spin symmetry, this unitary limit is reached in two generic cases: (i) when the system is deep in the Kondo regime ($n_{d\sigma} \approx 1/2$), and (ii) for all values of U obeying $U + 2\epsilon_d = 0$. In the latter case $n_{d\sigma} = 1$ is guaranteed by symmetry, irrespective of whether the impurity is in the local-moment or mixed-valent regime.

4.2 Exact Zero-Temperature Conductance

From the discussion above it is clear that the zero-temperature conductance for tunneling through a quantum dot is exclusively determined in the wide-band limit by the occupancy of the dot and by the tunneling asymmetry to the left and right leads. Indeed, combining (23) and (30) one obtains

$$G(T=0) = G_0 \sin^2\left(\frac{\pi n_d}{2}\right) \ , \tag{31}$$

where n_d is the total occupancy of the dot and

$$G_0 = 2\frac{e^2}{h}\frac{4\Gamma_\mathrm{L}\Gamma_\mathrm{R}}{(\Gamma_\mathrm{L}+\Gamma_\mathrm{R})^2} \tag{32}$$

parameterizes the tunneling asymmetry. Here we have restored SU(2) spin symmetry, that is we have set again $t_\sigma = t$ and $h_z = 0$.

The occupancy n_d, and thus the conductance, can be computed exactly using the Bethe *ansataz* solution of the Anderson model [31]. The resulting conductance curves are depicted in Fig. 2 as a function of ϵ_d, for three different values of $\Gamma/U = 0.05$, 0.1, and 0.5. Here $\Gamma = \Gamma_\mathrm{L} + \Gamma_\mathrm{R}$ is the total hybridization width of the level. Experimentally ϵ_d can be tuned by varying a gate voltage, as described in (9). Upon going from the mixed-valent to the local-moment regime, $G(\epsilon_d)$ develops an extended plateaux at which the conductance is

pinned to the unitary value G_0. For $\Gamma \ll U$, the plateaux extends throughout most of the domain $-U < \epsilon_d < 0$. The shoulders on either side of the plateaux are smeared on the scale of Γ. For either $\epsilon_d \gg \Gamma$ or $-(U + \epsilon_d) \gg \Gamma$, the conducta

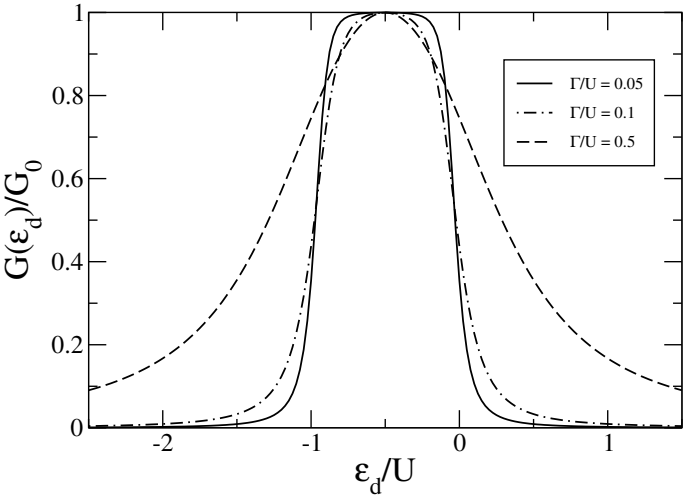

Fig. 2. Exact zero-temperature conductance vs ϵ_d, for tunneling through an ultrasmall quantum dot. The dot is modeled by an Anderson impurity model, with $\Gamma = \Gamma_{\rm L} + \Gamma_{\rm R}$ and $G_0 = (2e^2/h)(4\Gamma_{\rm L}\Gamma_{\rm R}/\Gamma^2)$. Here Γ/U equals 0.05, 0.1, and 0.5 for the solid, dotted-dashed, and dashed line, respectively

5 Toulouse Limit

One particularly revealing limit of the Kondo Hamiltonian is the so-called Toulouse point [53], where the model can be solved exactly using somewhat standard techniques. The Toulouse point corresponds to a particular value of the longitudinal exchange coupling given by

$$\rho J_z = \frac{4}{\pi} \tan\left[\frac{\pi}{2}\left(1 - \frac{1}{\sqrt{2}}\right)\right] . \tag{33}$$

Although well removed from realistic parameters in quantum dots and magnetic ions, the Toulouse limit can be used to extract universal low-temperature properties of the Kondo effect. For example, the Wilson ratio (see below). This follows from the fact that all Hamiltonians with antiferromagnetic spin exchange flow to the same strong-coupling fixed point as $T \to 0$ [54]. Thus, while physical properties away from the fixed point are model-dependent, any

solvable model can be used to extract the universal behavior near the common fixed point.

To expose the Toulouse point we resort to a continuum version of the anisotropic Kondo Hamiltonian, written in terms of one-dimensional fields. In the standard fashion [37], the conduction-electron modes that couple to the impurity are reduced to left-moving fields $\psi_{\alpha\sigma}(x)$, where $\sigma = \uparrow, \downarrow$ labels the spin projection. Here we have linearized the conduction-electron dispersion around the Fermi level: $\epsilon_k = \hbar v_{\mathrm{F}} k$, where ϵ_k and k are measured relative to the Fermi level and Fermi wave number, respectively. The coordinate x is a fictitious position variable conjugate to k. In terms of these fields, the Hamiltonian takes the form

$$\mathcal{H} = \mathrm{i}\hbar v_{\mathrm{F}} \sum_{\sigma} \int_{-\infty}^{\infty} \psi_{\sigma}^{\dagger}(x) \partial_x \psi_{\sigma}(x) \mathrm{d}x + \frac{J_{\perp} a}{2} \left[S^{+} \psi_{\downarrow}^{\dagger}(0) \psi_{\uparrow}(0) + \mathrm{h.c.} \right] \tag{34}$$

$$+ \frac{J_z a}{2} S_z \sum_{\sigma} \sigma \psi_{\sigma}^{\dagger}(0) \psi_{\sigma}(0) - \mu_{\mathrm{B}} g h_z \left[S_z + \sum_{\sigma} \frac{\sigma}{2} \int_{-\infty}^{\infty} \psi_{\sigma}^{\dagger}(x) \psi_{\sigma}(x) \mathrm{d}x \right] ,$$

where $\boldsymbol{S}$ is a spin-$\frac{1}{2}$ operator, a is a short-distance cutoff corresponding to a lattice spacing, μ_{B} is the magneton Bohr, and g is the Landé g-factor shared by the conduction electrons and the impurity spin. Here and throughout the section we use the labels $\sigma = \uparrow, \downarrow$ and $\sigma = \pm 1$ interchangeably to designate the spin projection.

To treat the Hamiltonian of (34), we resort to Abelian bosonization. According to standard prescriptions [55], two boson fields $\Phi_{\sigma}(x)$ are introduced, one boson field for each left-moving fermion field $\psi_{\sigma}(x)$. The fermion fields are written as

$$\psi_{\sigma}(x) = \frac{\hat{P}}{\sqrt{2\pi\alpha}} \mathrm{e}^{-\mathrm{i}\Phi_{\sigma}(x)} , \tag{35}$$

where the Φ_{σ}'s obey

$$[\Phi_{\sigma}(x), \Phi_{\sigma'}(y)] = -\mathrm{i}\pi \delta_{\sigma\sigma'} \operatorname{sign}(x - y) . \tag{36}$$

The ultraviolet momentum cutoff $\alpha^{-1} = \pi/a$ is related to the conduction-electron bandwidth D and the density of states per lattice site ρ through $D = \hbar v_{\mathrm{F}}/\alpha$ and $\rho = 1/(2D) = \alpha/(2\hbar v_{\mathrm{F}})$, respectively. The operator $\hat{P}$ in (35) is a phase-factor operator, which comes to ensure that the different fermion species anticommute. Our explicit choice for this operator is $\hat{P} = \mathrm{e}^{\mathrm{i}\pi N_{\uparrow}}$, where N_{σ} is the number operator for electrons with spin projection σ. When written in terms of the boson fields, the Hamiltonian reads

$$\mathcal{H} = \frac{\hbar v_{\mathrm{F}}}{4\pi} \sum_{\sigma} \int_{-\infty}^{\infty} \left[\partial_x \Phi_{\sigma}(x) \right]^2 \mathrm{d}x + \frac{J_{\perp}}{4} \left\{ S^{+} \mathrm{e}^{\mathrm{i}[\Phi_{\downarrow}(0) - \Phi_{\uparrow}(0)]} + \mathrm{h.c.} \right\} \tag{37}$$

$$+ \frac{a\delta_z}{\pi^2 \rho} S_z \sum_{\sigma} \sigma \partial_x \Phi_{\sigma}(0) - \mu_{\mathrm{B}} g h_z \left[S_z + \sum_{\sigma} \frac{\sigma}{4\pi} \int_{-\infty}^{\infty} \partial_x \Phi_{\sigma}(x) \mathrm{d}x \right] ,$$

where

$$\delta_z = \arctan\left(\frac{\pi\rho J_z}{4}\right) \tag{38}$$

is the phase shift associated with J_z (in the absence of $J_\perp$). Note that δ_z is bounded in magnitude by $\pi/2$, which stems from the cutoff scheme used in bosonization.

Exploiting the emergence of spin-charge separation in the Kondo problem, we proceed by converting to the spin and charge fields

$$\Phi_s(x) = \frac{1}{\sqrt{2}}\left[\Phi_\uparrow(x) - \Phi_\downarrow(x)\right] \ , \quad \Phi_c(x) = \frac{1}{\sqrt{2}}\left[\Phi_\uparrow(x) + \Phi_\downarrow(x)\right] \ , \tag{39}$$

which yields

$$\begin{aligned}\mathcal{H} = {} & \frac{\hbar v_\mathrm{F}}{4\pi}\sum_{\nu=c,s}\int_{-\infty}^{\infty}\left[\partial_x\Phi_\nu(x)\right]^2\mathrm{d}x + \frac{J_\perp}{4}\left\{S^+\mathrm{e}^{-\mathrm{i}\sqrt{2}\Phi_s(0)} + \mathrm{h.c.}\right\} \\ & + \sqrt{2}\frac{a\delta_z}{\pi^2\rho}S_z\partial_x\Phi_s(0) - \mu_\mathrm{B}gh_z\left[S_z + \frac{1}{\sqrt{8}\pi}\int_{-\infty}^{\infty}\partial_x\Phi_s(x)\mathrm{d}x\right] \ . \end{aligned} \tag{40}$$

To this Hamiltonian we apply the canonical transformation $\mathcal{H}' = U^\dagger\mathcal{H}U$ with $U = \exp\left[\mathrm{i}(1-\sqrt{2})S_z\Phi_s(0)\right]$ to obtain

$$\begin{aligned}\mathcal{H}' = {} & \frac{\hbar v_\mathrm{F}}{4\pi}\sum_{\nu=c,s}\int_{-\infty}^{\infty}\left[\partial_x\Phi_\nu(x)\right]^2\mathrm{d}x + \frac{J_\perp}{4}\left\{S^+\mathrm{e}^{-\mathrm{i}\Phi_s(0)} + \mathrm{h.c.}\right\} \\ & + \sqrt{2}\frac{a\delta_z'}{\pi^2\rho}S_z\partial_x\Phi_s(0) - \frac{1}{\sqrt{2}}\mu_\mathrm{B}gh_z\left[S_z + \frac{1}{2\pi}\int_{-\infty}^{\infty}\partial_x\Phi_s(x)\mathrm{d}x\right] \end{aligned} \tag{41}$$

with

$$\delta_z' = \delta_z - \frac{\pi}{2}(\sqrt{2}-1) \ . \tag{42}$$

At this point we transform to a new fermion representation. To this end we first express the impurity spin in terms of a fermion operator $d = S^-$. The boson fields are then "refermionized" according to

$$\psi_s(x) = \frac{\mathrm{e}^{\mathrm{i}\pi d^\dagger d}}{\sqrt{2\pi\alpha}}\mathrm{e}^{-\mathrm{i}\Phi_s(x)} \ , \tag{43}$$

$$\psi_c(x) = \frac{\mathrm{e}^{\mathrm{i}\pi(d^\dagger d + N_s)}}{\sqrt{2\pi\alpha}}\mathrm{e}^{-\mathrm{i}\Phi_c(x)} \ , \tag{44}$$

where N_s is the number operator for the ψ_s "spin" fermions. The end result of these manipulations is the following fermionic Hamiltonian

$$\begin{aligned}\mathcal{H}' = {} & \mathrm{i}\hbar v_\mathrm{F}\sum_{\nu=c,s}\int_{-\infty}^{\infty}\psi_\nu^\dagger(x)\partial_x\psi_\nu(x)\mathrm{d}x + \frac{J_\perp\sqrt{a}}{2\sqrt{2}}\left[d^+\psi_s(0) + \mathrm{h.c.}\right] \\ & + Va\left[d^\dagger d - 1/2\right]{:}\psi_s^\dagger(0)\psi_s(0){:} - \frac{\mu_\mathrm{B}gh_z}{\sqrt{2}}\left[S_z + \int_{-\infty}^{\infty}\psi_s^\dagger(x)\psi_s(x)\mathrm{d}x\right] \ , \end{aligned} \tag{45}$$

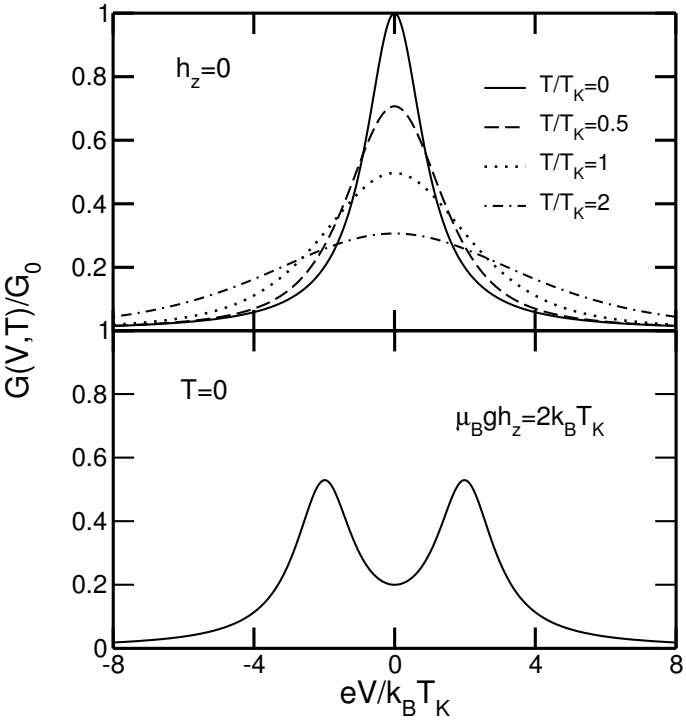

Fig. 3. The differential conductance as a function of voltage bias at the Toulouse limit. At $T = 0$ and $B = 0$ (upper panel), there is a Lorentzian zero-bias anomaly with half-width $k_{\rm B}T_{\rm K}$ and a peak height of G_0 that depends on the tunneling asymmetry to the left and right leads. For symmetric coupling, $G_0 = 2e^2/h$. The effect of a temperature is to smear the zero-bias anomaly (upper panel), while a non-zero magnetic field splits it into two peaks at $\pm\mu_{\rm B}gh_z$ (lower panel)

where

$$V = \frac{2}{\pi\rho}\tan\left(\sqrt{2}\delta_z'\right) \ . \tag{46}$$

Here $:\psi_s^\dagger(0)\psi_s(0):$ stands for normal ordering with respect to the unperturbed ψ_s Fermi sea. Note that in order to refermionize the Hamiltonian $\mathcal{H}'$ we had to assume that $|\delta_z'| < \frac{\pi}{\sqrt{8}}$. However, this condition is not very restrictive. It encompasses the entire antiferromagnetic regime, $J_z > 0$, and a fair part of the ferromagnetic regime, $J_z < 0$. Subject to this constraint, the mapping of (34) onto (45) is exact.

The Hamiltonian of (45) describes an interacting resonant level, where the strength of the interaction, V, depends on ρJ_z through (38), (42), and (46). Remarkably, V vanishes at the particular point described by (33). When tuned to this point, (45), and thus the original Kondo Hamiltonian, can be solved exactly using elementary methods. Any observable that can be expressed in

terms of simple combinations of the ψ_s and d fermions can be evaluated exactly. While not all physical quantities fall in this category, several do. For example, the spin component S_z equals $d^\dagger d - 1/2$. Hence, both the impurity susceptibility and the impurity contribution to the specific heat can be computed straightforwardly at the Toulouse point to obtain

$$\lim_{T\to 0} \chi_{\rm imp} = \frac{(\mu_{\rm B} g)^2}{2\pi \Gamma_{\rm K}} , \qquad \lim_{T\to 0} \frac{C_{\rm imp}}{T} = \frac{\pi}{3} \frac{(k_{\rm B} T)^2}{\Gamma_{\rm K}} . \tag{47}$$

Both quantities depend solely on $\Gamma_{\rm K} = \rho J_\perp^2/8$, which plays the role of a Kondo temperature at the Toulouse point. From these expressions one obtains the correct Wilson ratio

$$R_{\rm W} = \lim_{T\to 0} \frac{\chi_{\rm imp}}{C_{\rm imp}} \cdot \frac{C_{\rm bulk}}{\chi_{\rm bulk}} = 2 . \tag{48}$$

The discussion above was restricted to thermal equilibrium. An extension of the Toulouse limit to non-equilibrium tunneling through a Kondo impurity was devised in [56], using the Emery-Kivelson construction for the two-channel case [57]. Applications of the model to dc [56,58], ac [59], and pulsed-bias potentials [60] have shown all the qualitative features of Kondo-assisted tunneling: a zero-bias anomaly that splits in an applied magnetic field; Fermi-liquid characteristics in the low-T and low-V differential conductance; side peaks in the differential conductance at $eV = \pm n\hbar\omega$ for an ac drive of frequency ω; and a hierarchy of time scales for the rise, saturation and falloff of the current in response to a pulsed bias potential. The Toulouse limit was also recently applied to compute the full counting statistics for tunneling through a Kondo impurity [61]. Here we settle with a brief demonstration of the main point listed above, namely, the emergence of a zero-bias anomaly that splits in an applied magnetic field, see Fig. 3.

As emphasized above, the Toulouse point with all its success cannot be used to access all physical quantities. Observables that do not have a simple representation in terms of the new fermionic degrees of freedom cannot be computed based on this special point. One important quantity that falls in the latter category is the impurity spectral function. Hence, a different approach is required to track the evolution of the impurity DOS with an applied voltage bias. Such an approach will be presented in the following section.

6 Noncrossing Approximation

Among diagrammatic approaches to the Kondo problem, the most established is the so-called noncrossing approximation (NCA). The NCA was extensively used in the 1980s to study dilute magnetic alloys, especially Ce- and Yb-based alloys with large orbital degeneracy. Its usefulness for treating the multichannel Anderson model was later emphasized by Cox and Ruckenstein [62],

who noticed that the NCA pathology hampering the single-channel case [63] actually corresponds to the exact non-Fermi-liquid power laws and logarithms of the multi-channel Kondo effect. Compared to exact methods such as the Bethe ansatz and conformal field theory, the NCA has the crucial advantage that it is not restricted to idealized models, it could be used to compute dynamical properties over extended temperature and parameter regimes, and it has a natural extension to non-equilibrium. Indeed, the method has been successfully applied to numerous variants of the out-of-equilibrium Kondo effect, including dc [64], ac [65,66], pulsed-bias [67], and multi-lead [68] transport through a single-channel Anderson impurity, as well as Kondo-assisted tunneling through different multi-channel impurities [69–71]. In the following section we briefly review the approach, and discuss its application to the out-of-equilibrium impurity DOS.

6.1 Slave-Boson Representation

The NCA represents a self-consistent perturbative expansion about the atomic limit. Applied primarily to the infinite-U Anderson model, the approach is most conveniently formulated within the slave-boson representation of Barnes and Coleman [35, 72]. In order to facilitate usage of standard diagrammatic techniques, each of the three impurity states (double occupancy is forbidden for $U \to \infty$) is assigned a pseudo-particle according to the convention

$$b^\dagger|0\rangle \longleftrightarrow |n_d = 0\rangle \;, \tag{49}$$

$$f_\uparrow^\dagger|0\rangle \longleftrightarrow |n_d = 1, S_z = \uparrow\rangle \;, \tag{50}$$

$$f_\downarrow^\dagger|0\rangle \longleftrightarrow |n_d = 1, S_z = \downarrow\rangle \;. \tag{51}$$

Here $b^\dagger$ is a pseudo-boson, while $f_\uparrow^\dagger$ and $f_\downarrow^\dagger$ are pseudo-fermions analogous to ones used in Abrikosov's slave-fermion representation of the Kondo Hamiltonian [28]. The physical subspace corresponds to the constraint

$$\hat{N}_f \equiv b^\dagger b + \sum_\sigma f_\sigma^\dagger f_\sigma = 1 \;, \tag{52}$$

reflecting the fact that one is working within an enlarged Hilbert space. It is the latter constraint on the number of pseudo-particles that distinguishes $b^\dagger$ and $f_\sigma^\dagger$ from ordinary bosons and fermions. Subject to this constraint, the infinite-U Anderson model is written in the form

$$\mathcal{H} = \sum_{k,\sigma} \epsilon_k c_{k\sigma}^\dagger c_{k\sigma} + \sum_{k,\sigma} t \left\{ c_{k\sigma}^\dagger b^\dagger f_\sigma + \text{h.c.} \right\} + \epsilon_d \sum_\sigma f_\sigma^\dagger f_\sigma \;. \tag{53}$$

As noted above, the advantage of the slave-boson representation stems from the ability to apply standard diagrammatic many-body techniques to

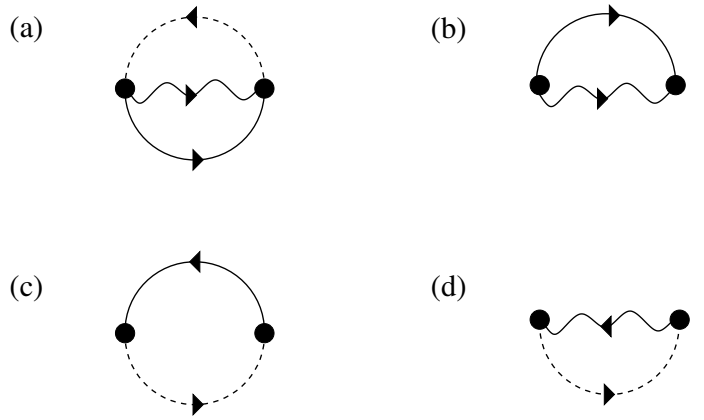

Fig. 4. Diagrammatic representation of the NCA. Figure **(a)** displays the NCA generating functional, $\Phi_{\rm NCA}$. Here dashed, wiggly, and solid lines represent a dressed pseudo-fermion, dressed pseudo-boson, and local conduction-electron Green function, respectively. Filled circles mark interaction vertices. Figures **(b)** and **(c)** depict the corresponding pseudo-fermion and pseudo-boson self-energies, obtained by a functional derivative of $\Phi_{\rm NCA}$. The impurity Green function is displayed in turn in figure **(d)**

compute physical observables. The difficulty lies in implementing the constraint, which necessitates the introduction of a fictitious "chemical potential" λ for the pseudo-particles. In order to project onto the $\hat{N}_f = 1$ physical subspace, the pseudo chemical potential is taken to minus infinity at the end of the calculation. Specifically, modifying the Hamiltonian of the system from $\mathcal{H}$ to $\mathcal{H} - \lambda\hat{N}_f$, physical averages are written as

$$\langle\hat{O}\rangle_{\rm phys} = \frac{1}{Z_{\rm imp}} \lim_{\lambda\to-\infty} {\rm e}^{-\beta\lambda}\langle\hat{O}\hat{N}_f\rangle_\lambda \;, \tag{54}$$

where

$$Z_{\rm imp} = \lim_{\lambda\to-\infty} {\rm e}^{-\beta\lambda}\langle\hat{N}_f\rangle_\lambda \tag{55}$$

is the "impurity contribution" to the partition function. Here λ subscripts denote averages within a grand-canonical ensemble for the pseudo-particles, and $\beta = 1/k_{\rm B}T$ is the reciprocal temperature. In practice, one can drop the $\hat{N}_f$ operator in (54) for those physical operators $\hat{O}$ that give zero when acting on the $\hat{N}_f = 0$ subspace, which greatly simplifies the calculations.

6.2 Noncrossing Diagrams

Equations (54) and (55) relate the physical average of any observable $\hat{O}$ to its average within the enlarged Hilbert space for the pseudo-particles. For

any given value of λ, one can compute the latter average using standard diagrammatic techniques, whose building blocks are the pseudo-fermion and pseudo-boson Green functions $G^{(\lambda)}_{f\sigma}(z)$ and $G^{(\lambda)}_{b}(z)$, respectively. The NCA consists of a particular set of diagrams for computing the pseudo-fermion and pseudo-boson self-energies, $\Sigma^{(\lambda)}_{f\sigma}(z)$ and $\Sigma^{(\lambda)}_{b}(z)$, respectively. The approximation scheme is defined by the generating function portrayed in Fig. 4(a), from which the self-energies are obtained by functional derivatives. The corresponding self-energy diagrams are depicted in Fig. 4(b) and Fig. 4(c).

The pseudo-particle Green functions enter the calculation of physical observables in their projected forms

$$G_{\gamma}(z) = \lim_{\lambda\to-\infty} G^{(\lambda)}_{\gamma}(z-\lambda)\,, \qquad \gamma = f\sigma, b\,. \tag{56}$$

They maintain the standard structure

$$G_{f\sigma}(z) = \frac{1}{z-\epsilon_d-\Sigma_{f\sigma}(z)}\,, \tag{57}$$

$$G_{b}(z) = \frac{1}{z-\Sigma_{b}(z)}\,, \tag{58}$$

where $\Sigma_{f\sigma}(z)$ and $\Sigma_{b}(z)$ are the projected self-energies. Evaluating the self-energy diagrams of Fig. 4(b) and Fig. 4(c) and implementing the projection rule of (56), one obtains the following pair of self-consistent equations for the pseudo-particle self-energies:

$$\Sigma_{f\sigma}(z) = \frac{\Gamma}{\pi}\int_{-\infty}^{\infty} G_{b}(\epsilon+z) f(\epsilon)\nu(-\epsilon)\mathrm{d}\epsilon\,, \tag{59}$$

$$\Sigma_{b}(z) = \sum_{\sigma}\frac{\Gamma}{\pi}\int_{-\infty}^{\infty} G_{f\sigma}(\epsilon+z) f(\epsilon)\nu(\epsilon)\mathrm{d}\epsilon\,. \tag{60}$$

Here $f(\epsilon) = 1/(\mathrm{e}^{\beta\epsilon}+1)$ is the Fermi-Dirac distribution function, and $\nu(\epsilon) = \rho(\epsilon)/\rho$ is the reduced conduction-electron density of states ($\nu = 1$ for a flat band).

In the slave-boson representation, the impurity annihilation operator d_σ is identified with $b^\dagger f_\sigma$. Therefore, the impurity Green function is given within the NCA by the diagram of Fig. 4(d). Upon evaluation and projection of the diagram onto the physical subspace we obtain

$$G_{d\sigma}(z) = \frac{1}{Z_{\mathrm{imp}}}\int_{-\infty}^{\infty}\mathrm{e}^{-\beta\epsilon}\left[G_{f\sigma}(\epsilon+z)A_{b}(\epsilon) - A_{f\sigma}(\epsilon)G_{b}(\epsilon-z)\right]\mathrm{d}\epsilon\,. \tag{61}$$

Here

$$A_{\gamma}(\epsilon) = -\frac{1}{\pi}\mathrm{Im}\left\{G_{\gamma}(\epsilon+\mathrm{i}\eta)\right\} \tag{62}$$

with $\gamma = f\sigma, b$ are the pseudo-particle spectral functions.

Figure 5 shows the impurity spectral function in thermal equilibrium, with the impurity tuned deep into the Kondo regime. Here the temperature is set equal to $T_{\rm K}$. In addition to a broad charge fluctuation peak slightly below the bare level ϵ_d, there is a sharp Abrikosov-Suhl resonance near the Fermi level. The width of the peak is roughly $T_{\rm K}$. Its height roughly equals 90% of the unitary limit, which is somewhat overestimated by the NCA. Indeed,

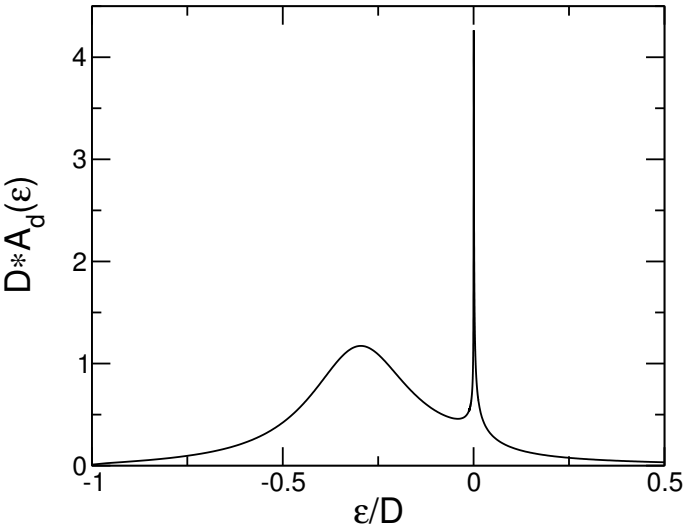

Fig. 5. Equilibrium dot density of states as obtained by the NCA for $U = \infty$. Here we have used a joint semi-circular conduction-electron DOS with half-width D for both leads. The remaining model parameters are $\epsilon_d/D = -0.278$, $\Gamma_{\rm L} + \Gamma_{\rm R} = D/15$ (symmetric junction), and $T = T_{\rm K}$. The corresponding Kondo temperature equals $k_{\rm B}T/D = 2.5 \times 10^{-4}$. In addition to a broad charge-fluctuation peak slightly below ϵ_d, there is a sharp Abrikosov-Suhl resonance near the Fermi level

the NCA overshoots the unitary limit as the temperature is further decreased below $T_{\rm K}$. The second charge fluctuation peak near $U + \epsilon_d$ cannot be seen as it was shifted to infinity along with U.

6.3 Non-Equilibrium Conditions

Equations (59), (60), and (61) were derived under the condition of thermal equilibrium. Several important modifications appear when a finite voltage bias is applied, mostly related to the need to adopt the Keldysh formulation [73]. In the Keldysh formulation, each pseudo-particle Green functions and self-energy is replaced with two independent components, a retarded function and a lesser function. In addition, $\Gamma f(\epsilon)$ terms are replaced within the NCA equations with terms of the form

$$\Gamma_{\rm L} f(\epsilon \mp \mu_{\rm L}) + \Gamma_{\rm R} f(\epsilon \mp \mu_{\rm R}) \ , \tag{63}$$

reflecting the different distribution functions on the two leads. Skipping the technical details (see, e.g., reference [64, 70] for details), we proceed directly to the results. Figure 6 shows the evolution of the Abrikosov-Suhl resonance with increasing voltage bias. As originally noted by Wingreen and Meir [64], the resonance splits as the voltage bias sufficiently exceeds the Kondo tem-

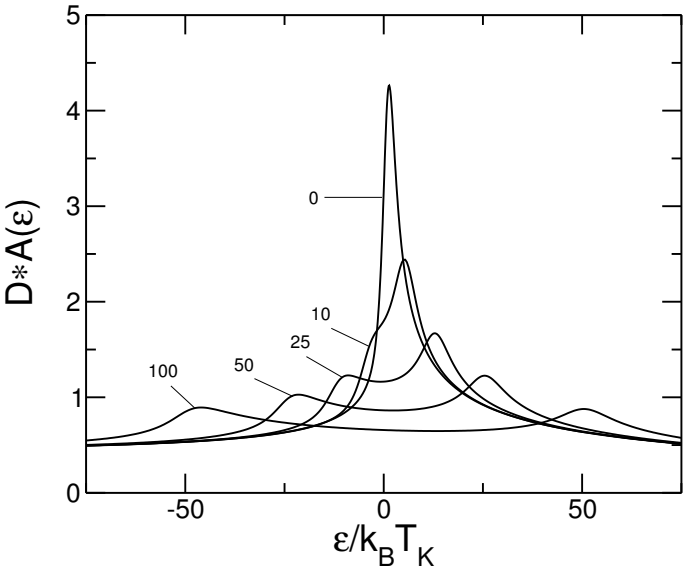

Fig. 6. Non-equilibrium splitting of the Abrikosov-Suhl resonance in the dot density of states for different values of the voltage bias, as obtained by the NCA for $U = \infty$. Here we have used a joint semi-circular conduction-electron DOS with half-width D for both leads. The remaining model parameters are $\epsilon_d/D = -0.278$, $\Gamma_{\rm L} = \Gamma_{\rm R} = D/30$ (symmetric junction), and $T = T_{\rm K}$. The corresponding Kondo temperature equals $k_{\rm B}T/D = 2.5 \times 10^{-4}$. The voltage bias eV takes the values $eV/k_{\rm B}T_{\rm K} = 0, 10, 25, 50$, and 100, according to the labels attached

perature. Physically this stems from the fact that there are now two sharp Fermi steps, one at the chemical potential of each lead. At the same time, the resonances are broadened by the dissipative lifetime induced by the bias [64]. Splitting of the Abrikosov-Suhl resonance was recently observed both in a three-terminal device [74], and in a setup where a ballistic quantum wire with a split distribution was side-coupled to the dot [75].

Acknowledgments

I wish to express my deepest gratitude to the organizers of the International Workshop on "The Physics of Zero and One Dimensional Nanoscopic Sys-

tems", particularly Prof. Sachindra Nath Karmakar, for the warm hospitality and the exceptionally stimulating environment at the Saha Institute of Nuclear Physics. I am grateful to Dr. Vyacheslavs Kashcheyevs for providing me with the Bethe *ansatz* occupancies used for the conductance curves of Fig. 2.

References

1. P. W. Anderson, G. Yuval, D. R. Hamman: Phys. Rev. B **1**, 4464 (1970)
2. P. W. Anderson: J. Phys. C **3**, 2436 (1970)
3. K. G. Wilson: Rev. Mod. Phys. **47**, 773 (1975)
4. D. Goldhaber-Gordon, H. Shtrikman, D. Mahalu, D. Abusch-Magder, U. Meirav, M. A. Kastner: Nature **391**, 156 (1998)
5. S. M. Cronenwett, T. H. Oosterkamp, L. P. Kouwenhoven: Science **281**, 540 (1998)
6. F. Simmel, R. H. Blick, J. P. Kotthaus, W. Wegscheider, M. Bichler: Phys. Rev. Lett. **83**, 804 (1999)
7. J. Li, W.-D. Schneider, R. Berndt, B. Delley: Phys. Rev. Lett. **80**, 2893 (1998)
8. V. Madhavan, W. Chen, T. Jamneala, M. F. Crommie, N. S. Wingreen: Science **280**, 567 (1998)
9. W. Chen, T. Jamneala, V. Madhavan, M. F. Crommie: Phys. Rev. B **60**, 8529 (1999)
10. H. C. Manoharan, C. P. Lutz, D. M. Eigler: Nature **403**, 512 (2000)
11. J. Nygard, D. H. Cobden, P. E. Lindelof: Nature **408**, 342 (2000)
12. P. Jarillo-Herrero, J. Kong, H. S. J. van der Zant, C. Dekker, L. P. Kouwenhoven, S. De Franceschi: Nature **434**, 484 (2005)
13. J. Park, A. N. Pasupathy, J. I. Goldsmith, C. Chang, Y. Yaish, J. R. Petta, M. Rinkoski, J. P. Sethna, H. D. Abruna, P. L. McEuen, D. C. Ralph: Nature **417**, 722 (2002)
14. W. J. Liang, M. P. Shores, M. Bockrath, J. R. Long, H. Park: Nature **417**, 725 (2002)
15. M. R. Buitelaar, T. Nussbaumer, C. Schönenberger: Phys. Rev. Lett. **89**, 256801 (2002)
16. A. N. Pasupathy, R. C. Bialczak, J. Martinek, J. E. Grose, L. A. K. Donev, P. L. McEuen, D. C. Ralph: Science **306**, 86 (2004)
17. For a comprehensive review, see, e.g., A. Georges, G. Kotliar, W. Krauth, M. Rozenberg: Rev. Mod. Phys. **68**, 13 (1996)
18. G. Kotliar, D. Vollhardt: Phys. Today **57**, 53 (2004)
19. G. Kotliar, S. Y. Savrasov, K. Haule, V. S. Oudovenko, O. Parcollet, C. A. Marianetti: Rev. Mod. Phys. **78**, 865 (2006)
20. For a comprehensive review of the multi-channel Kondo effect, see D. L. Cox, A. Zawadovski: Adv. Phys. **47**, 599 (1998)
21. A. Hewson: *The Kondo Problem to Heavy Fermions* (Cambridge Press, Cambridge 1993)
22. M. Pustilnik, L. I. Glazman: J. Phys.: Cond. Matt. **16**, R513 (2004)
23. W. J. de Haas, J. de Boer, G. J. van den Berg: Physica **1**, 115 (1934)
24. J. Kondo: Prog. Theor. Phys. **32**, 37 (1964)
25. H. Suhl: Phys. Rev. **138**, A515 (1965)

26. Y. Nagaoka: Phys. Rev. **138**, A1112 (1965)
27. K. Yosida: Phys. Rev. **147**, 223 (1966)
28. A. A. Abrikosov: Physics **2**, 5 (1965)
29. N. Andrei: Phys. Rev. Lett. **45**, 379 (1980)
30. P. B. Wiegmann: Sov. Phys. JETP Lett. **31**, 392 (1980)
31. P. B. Wiegmann, A. M. Tsvelick: J. Phys. C **16**, 2281 (1983)
32. N. Read, D. M. Newns: J. Phys. C **16**, L1055 (1983)
33. N. Read, D. M. Newns: J. Phys. C **16**, 3273 (1983)
34. N. Grewe: Zeit. Für Phys. B – Cond. Matter **53**, 271 (1983)
35. P. Coleman: Phys. Rev. B **29**, 3035 (1984)
36. O. Gunnarsson, K. Schönhammer: Phys. Rev. B **28**, 4315 (1983)
37. I. Affleck, A. W. W. Ludwig: Nucl. Phys. B **360**, 641 (1991)
38. A. F. G. Wyatt: Phys. Rev. Lett. **13**, 401 (1964)
39. R. A. Logan, J. M. Rowell: Phys. Rev. Lett. **13**, 404 (1964)
40. J. Appelbaum: Phys. Rev. Lett. **17**, 91 (1966)
41. P. W. Anderson: Phys. Rev. Lett. **17**, 95 (1966)
42. S. Gregory: Phys. Rev. Lett. **68**, 2070 (1992)
43. D. C. Ralph, R. A. Buhrman: Phys. Rev. Lett. **72**, 3401 (1994)
44. P. W. Anderson: Phys. Rev. **124**, 41 (1961)
45. J. R. Schrieffer, P. A. Wolff: Phys. Rev. **149**, 491 (1966)
46. F. D. M. Haldane: Phys. Rev. Lett. **40**, 416 (1978)
47. For a clear exposition, see, e.g., N. E. Bickers, D. L. Cox, J. W. Wilkins: Phys. Rev. B **36**, 2036 (1987), Appendix D
48. D. C. Langreth: Phys. Rev. **150**, 516 (1966)
49. J. Martinek, Y. Utsumi, H. Imamura, J. Barnaś, S. Maekawa, J. König, G. Schön: Phys. Rev. Lett. **91**, 127203 (2003)
50. P. Simon, I. Affleck: Phys. Rev. Lett. **89**, 206602 (2002)
51. P. S. Cornaglia, C. A. Balseiro: Phys. Rev. Lett. **90**, 216801 (2003)
52. T. Hand, J. Kroha, H. Monien: Phys. Rev. Lett. **97**, 136604 (2006)
53. G. Toulouse: Phys. Rev. B **2**, 270 (1970)
54. In the presence of potential scattering, the flow is to a line of strong-coupling fixed points with a phase shift that varies continuously with the strength of potential scattering
55. F. D. M. Haldane: J. Phys. C **14**, 2585 (1981)
56. A. Schiller, S. Hershfield: Phys. Rev. B **51**, 12896 (1995)
57. V. J. Emery, S. Kivelson: Phys. Rev. B **46**, 10812 (1992)
58. A. Schiller, S. Hershfield: Phys. Rev. B **58**, 14978 (1998)
59. A. Schiller, S. Hershfield: Phys. Rev. Lett. **77**, 1821 (1996)
60. A. Schiller, S. Hershfield: Phys. Rev. B **62**, 16271 (2000)
61. A. Komnik, A. O. Gogolin: Phys. Rev. Lett. **94**, 216601 (2005)
62. D. L. Cox, A. E. Ruckenstein: Phys. Rev. Lett. **71**, 1613 (1993)
63. N. E. Bickers: Rev. Mod. Phys. **59**, 845 (1987)
64. N. S. Wingreen, Y. Meir: Phys. Rev. B **49**, 11040 (1994)
65. M. H. Hettler, H. Schoeller: Phys. Rev. Lett. **74**, 4907 (1995)
66. P. Nordlander, N. S. Wingreen, Y. Meir, D. C. Langreth: Phys. Rev. B **61**, 2146 (2000)
67. M. Plihal, D. C. Langreth, P. Nordlander: Phys. Rev. B **61**, 13341 (2000)
68. E. Lebanon, A. Schiller: Phys. Rev. B **65**, 035308 (2002)
69. M. H. Hettler, J. Kroha, S. Hershfield: Phys. Rev. Lett. **73**, 1967 (1994)

70. M. H. Hettler, J. Kroha, S. Hershfield: Phys. Rev. B **58**, 5649 (1998)
71. A. Buxboim, A. Schiller: Phys. Rev. B **67**, 165320 (2003)
72. S. E. Barnes: J. Phys. F **6**, 1375 (1976)
73. L. V. Keldysh: Sov. Phys. JETP **20**, 1018 (1965)
74. R. Leturcq, L. Schmid, K. Ensslin, Y. Meir, D. C. Driscoll, A. C. Gossard: Phys. Rev. Lett. **95**, 126603 (2005)
75. S. De Franceschi, R. Hanson, W. G. van der Wiel, J. M. Elzerman, J. J. Wijpkema, T. Fujisawa, S. Tarucha, L. P. Kouwenhoven: Phys. Rev. Lett. **89**, 156801 (2002)

The Two Channel Kondo Effect in Quantum Dots

Yuval Oreg[1] and David Goldhaber-Gordon[2]

[1] Department of Condensed Matter Physics, Weizmann Institute of Science, Rehovot 76100, Israel. yuval.oreg@weizmann.ac.il
[2] Geballe Laboratory for Advanced Materials and Department of Physics, Stanford University, Stanford, California 94305, USA. goldhaber-gordon@stanford.edu

1 Introduction

The Kondo effect has become a hallmark of coherent electron transport in a variety of nanostructures ranging from lithographically-defined semiconductors [1] to carbon nanotubes [2] and molecules [3,4]. Kondo first introduced a phenomenological Hamiltonian [5] to describe how localized spin $\boldsymbol{S}$ couples antiferromagnetically with strength J to spins $\boldsymbol{s}$ of electrons in the surrounding reservoir.

The Kondo Hamiltonian was later found to be derivable from the more microscopic Anderson model, which consists of an electron bound to an impurity site in a metal host. The combined system of localized electron and metal host can lower its energy through tunneling: for example, an electron hops off the localized site into the Fermi sea and a new electron with opposite spin hops back from the Fermi level. Higher-order spin-flip tunneling processes coherently add at low temperature, screening the local spin.

Although the Kondo ground state is complex, its low-lying excitations look like individual weakly interacting quasi-particles. In contrast, some of the most intriguing problems in solid state physics arise when the motion of one electron dramatically affects the motion of other surrounding electrons, so that excitations become collective rather than single-particle-like. Examples of such highly-correlated systems include Tomonaga-Luttinger liquids [6], fractional quantum Hall Laughlin liquids [7], superconducting BCS condensates [8], and two-channel Kondo impurities [9].

In the two-channel Kondo (2CK) model, introduced 25 years ago by Nozières and Blandin [10], and independently by Zawadowski [11] , a localized spin $\boldsymbol{S}$ is antiferromagnetically coupled to two independent reservoirs of electrons according to the Hamiltonian

$$H_{\mathrm{2CK}} = J_1 \boldsymbol{s}_1 \cdot \boldsymbol{S} + J_2 \boldsymbol{s}_2 \cdot \boldsymbol{S} + H_{\mathrm{reservoirs}} \,. \tag{1}$$

The symmetric 2CK state is formed when the two independent channels (or reservoirs) are equally coupled to the magnetic impurity, i.e., $J_1 = J_2$. Each reservoir individually attempts to screen the local spin – if both succeeded, they would overcompensate the spin. This is an unstable situation, resulting in a new ground state in which a local "flavored spin" is only partially screened. Unlike the single-channel Kondo (1CK) the new 2CK state exhibits fascinating low energy behavior. In the 2CK state the quasi-particle concept of Fermi liquid theory does not apply: the decay rate for a low energy excitation ($\propto \sqrt{\epsilon}$) is greater than the energy ϵ of the excitation itself. Stable low-lying excitations must thus be collective [9, 12].

Any difference in channel coupling will force the system away from the non-Fermi liquid 2CK fixed point and toward the 1CK fixed point associated with the more strongly-coupled reservoir. Although the 2CK fixed point might therefore seem difficult to access, it has been invoked to explain remarkable low-energy properties of some heavy Fermion materials [9,13], and glassy metals [14–16]. However, the connections of these experimental observations to 2CK theory remain problematic [17], in part because the microscopic electronic structure of the various materials is unclear. Starting fifteen years ago there have been several theoretical suggestions for systems that should show the two-channel Kondo effect [18,19], but until recently none were realized in experiments.

Several years ago the present authors suggested [20] that a system consisting of a small single-level quantum dot coupled to two leads and a large but finite metallic dot (see Fig. 1) should exhibit the exotic two-channel Kondo physics at low temperature. The key observation is that due to the finite charging energy E_c of the large finite dot (c), electron transfer between leads and large dot is suppressed at energies smaller than E_c. Thus we have effectively two independent channels connected to the dot (one associated with the open leads, the other with the large dot.) By tuning gate voltages it is possible to control the values of the coupling constants and to drive the system to the isotropic two-channel point. Recently this system was realized experimentally and the two-channel Kondo effect was observed [21, 22].

The article is arranged in the following way. In Sect. 2 we introduce an Hamiltonian of the system which by assumption captures its most important characteristics. In Sect. 3 we analyzed the model using approximate Renormalization Group (RG) scaling methods and find the conditions for the stabilization of the two-channel Kondo effect. A theoretical estimate for the two-channel Kondo temperature [23] is given in (22). In Sect. 4 we give examples for the conductance [24] and the scaling curves that were used to analyze the experimental result of [21, 22]. A detailed derivation of the relation [24, 25] between the differential conductance and the $\mathcal{T}$ matrix is given in an Appendix (A) using the generalization of the Landauer formula for interacting systems [26] as a starting point.

2 Model of the System

Our system consists of a small quantum dot connected to two leads and to a large but finite dot with a continuous level spacing, as was suggested theoretically [20] and recently implemented experimentally [21, 22].

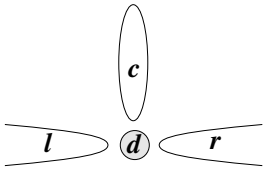

Fig. 1. Schematic view of the system

The Hamiltonian that describes the system is given by

$$\begin{aligned} H = & \sum_{k\sigma} \varepsilon_{lk\sigma} l^\dagger_{k\sigma} l_{k\sigma} + \sum_{k\sigma} \varepsilon_{rk\sigma} r^\dagger_{k\sigma} r_{k\sigma} \\ & + \sum_{k\sigma} \varepsilon_{ck\sigma} c^\dagger_{k\sigma} c_{k\sigma} + E_c (n - \mathcal{N})^2 \\ & + \varepsilon_{d\sigma} d^\dagger_\sigma d_\sigma + U n_{d\uparrow} n_{d\downarrow} \\ & + \sum_{k\sigma} \left(t_{kl} l^\dagger_{k\sigma} d_\sigma + t_{kr} r^\dagger_{k\sigma} d_\sigma + t_{kc} c^\dagger_{k\sigma} d_\sigma + \text{h.c.} \right) . \end{aligned} \tag{2}$$

In this Hamiltonian the first three terms describe the free electrons of the two leads and the large but finite dot. Let the index α run over the values l, r, c for the left lead, the right lead and the large dot respectively. Then the operator $\alpha_{\sigma k}(\alpha^\dagger_{\sigma k})$ are the annihilation (creation) operator of an electron at state k with spin σ in the one of the leads or the large dot.

The fourth term is the interaction energy in the large dot and its interaction with an external gate. The operator $n = \sum_{k\sigma} c^\dagger_{k\sigma} c_{k\sigma}$ is the number operator of the electrons in the large dot, and $\mathcal{N}$ denotes the effective interaction of the electrons in the large dot with an external gate. The fifth and sixth terms describe the Hamiltonian of the small dot, with $d_\sigma\,(d^\dagger_\sigma)$ the annihilation (creation) operator of an electron with spin σ on the small dot, and $n_{d\sigma} = d^\dagger_\sigma d_\sigma$. The last term is the hybridization between the small dot and its neighbors (the two leads and the large dot.) We will assume henceforth that the hopping matrix elements $t_{k\alpha}$ do not depend on k, and define $t_\alpha \equiv t_{k\alpha}$.

We have assumed that there is only one level in the small dot, so this Hamiltonian should be valid only at energy scale D smaller than the level spacing of the small dot Δ_d, which serves as the ultraviolet cutoff of our theory, D_0. For simplicity of presentation we have ignored in this Hamiltonian the mutual electrostatic coupling between dots d and c. By changing external gates one may control the values of t_α, $\mathcal{N}$ and ε_d.

3 Model Analysis

In this section we analyze the model Hamiltonian (2) by progressively integrating out the "fast" degrees of freedom in the renormalization group sense. Hamiltonian (2) has a few energy scales: U the charging energy of the small dot (d), $\varepsilon_{d\sigma}$ the position of the single-particle (spin full) level on the small dot, the charging energy[3] E_c of the large dot, and the tunneling matrix elements t_α, $\alpha = l, r, c$.

The parameters' phase space is large, so we limit ourself to the case where $U \gg E_c$ and obtain the proper effective Hamiltonian at the different running scales D. The different scales and the procedure used to derive the RG flow are depicted schematically in Fig. 2.

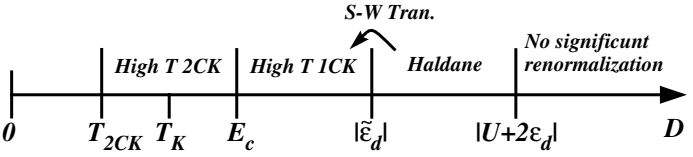

Fig. 2. We assume $|U + 2\epsilon_d| \gg |\epsilon_d| \gg E_c \gg T_{K1} > T_{K2}$. For each energy scale we use the appropriate RG method, see details in the text

3.1 Effective Hamiltonian at $D > E_c$ and its Scaling Analysis

At energy scales much bigger than E_c we can safely mix operators in the free leads and in the big dot, such that we have one linear combination of the l, r and c operators that is coupled to the small dot (while the other two combinations are decoupled). We will find these rotated operators in two steps:

[3] In fact there are two energy scales associated with the large dot, determined by the distance of the working position from the adjacent Coulomb blockade peaks. The two-channel Kondo effect may emerge when electron transfer between the leads and the large dot is not possible. Therefore, we tune the effective gate potential $\mathcal{N}$ so that the system is between Coulomb blockade peaks of the large dot. While the charging energy E_c sets the distance between Coulomb blockade peaks, $\mathcal{N}$ sets where the system sits relative to those peaks. We have two energy scales because the distances from the nearest Coulomb peaks are not necessarily identical, and in principle they may be parametrically different. Since typically both are fractions of E_c we will ignore that effect in the following analysis.

l-*r* Rotation: *l*-*r* ⇒ ψ, ϕ

First write the Hamiltonian in terms of the rotated operators:

$$\begin{pmatrix} \psi_k^\dagger \\ \phi_k^\dagger \end{pmatrix} = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} \mathrm{e}^{\mathrm{i}\varphi_l} & 0 \\ 0 & \mathrm{e}^{\mathrm{i}\varphi_r} \end{pmatrix} \begin{pmatrix} l_k^\dagger \\ r_k^\dagger \end{pmatrix} \tag{3}$$

with $\cos\theta = |t_l|/t_\psi$, $t_\psi^2 = |t_r|^2 + |t_l|^2$ and $\varphi_{r(l)} = \arg t_{r(l)}$ [see also discussion before (47)]. The Hamiltonian in terms of the rotated basis is:

$$\begin{aligned} H_{c\psi\phi} = & \sum_{k\sigma} \varepsilon_{\phi k\sigma} \phi_{k\sigma}^\dagger \phi_{k\sigma} + \sum_{k\sigma} \varepsilon_{\psi k\sigma} \psi_{k\sigma}^\dagger \psi_{k\sigma} \\ & + \sum_{k\sigma} \varepsilon_{ck\sigma} c_{k\sigma}^\dagger c_{k\sigma} + E_c (n - \mathcal{N})^2 \\ & + \varepsilon_{d\sigma} d_\sigma^\dagger d_\sigma + U n_{d\uparrow} n_{d\downarrow} \\ & + \sum_{k\sigma} \left(t_\psi \psi_{k\sigma}^\dagger d_\sigma + t_c c_{k\sigma}^\dagger d_\sigma + \text{h.c.} \right) . \end{aligned} \tag{4}$$

Thus we have two channels that are coupled to the small dot (ψ and c) and one free channel (ϕ).

ψ-c Rotation: ψ-c ⇒ ξ, ζ

At scales above E_c we can ignore the term $\propto E_c$ in (4) and mix the ψ and c operators using the rotation:

$$\begin{pmatrix} \xi_k^\dagger \\ \zeta_k^\dagger \end{pmatrix} = \begin{pmatrix} \cos\theta' & \sin\theta' \\ -\sin\theta' & \cos\theta' \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & \mathrm{e}^{\mathrm{i}\varphi_c} \end{pmatrix} \begin{pmatrix} \psi_k^\dagger \\ c_k^\dagger \end{pmatrix} , \tag{5}$$

with $\cos\theta' = t_\psi/t_\xi$, $t_\xi^2 = t_\psi^2 + |t_c|^2$ and $\varphi_c = \arg t_c$.

The Hamiltonian in terms of the channel ξ and the two free channels ζ and ϕ is given by:

$$\begin{aligned} H_{\xi\zeta\phi} = & \sum_{k\sigma} \varepsilon_{\phi k\sigma} \phi_{k\sigma}^\dagger \phi_{k\sigma} + \sum_{k\sigma} \varepsilon_{\zeta k\sigma} \zeta_{k\sigma}^\dagger \zeta_{k\sigma} + \sum_{k\sigma} \varepsilon_{\xi k\sigma} \xi_{k\sigma}^\dagger \xi_{k\sigma} \\ & + \varepsilon_{d\sigma} d_\sigma^\dagger d_\sigma + U n_{d\uparrow} n_{d\downarrow} + \sum_{k\sigma} \left(t_\xi \xi_{k\sigma}^\dagger d_\sigma + \text{h.c.} \right) . \end{aligned} \tag{6}$$

Hamiltonian (6) is valid up to a scale $D \gg D_1 = aE_c$ where $a \gtrsim 1$. For $D > D_1$ we can follow the standard RG poor man's scaling procedure for the Anderson model [27]. For completeness we reproduce the scaling equations here.

3.2 Haldane Flow: $D > \min\{|U+\varepsilon_d|, |\varepsilon_d|\}$

To perform the renormalization group transformation in this regime, we look at the many-body energies of the small dot with zero ($E_0 = 0$), one ($E_1 = \epsilon_d$) and two ($E_2 = 2\epsilon_d + U$) electrons in second order perturbation theory and integrate out the fast degree of freedom. Following [28] and the pedagogical presentation in [27], we find

$$\begin{aligned}\frac{\delta\varepsilon_d}{\delta D} &= \Gamma_\xi \left[\frac{2}{D+\varepsilon_d} - \frac{1}{D-\varepsilon_d} - \frac{1}{D+\varepsilon_d+U}\right] , \\ \frac{\delta U}{\delta D} &= 2\Gamma_\xi \left[\frac{1}{D-\varepsilon_d} - \frac{1}{D+\varepsilon_d} \right. \\ &\quad \left. + \frac{1}{D+\varepsilon_d+U} - \frac{1}{D-\varepsilon_d-U}\right] ,\end{aligned} \tag{7}$$

with $\Gamma_\xi = \nu|t_\xi|^2$. There are number of parameter regimes to study. There is no significant renormalization if $|\varepsilon_d|, |\varepsilon_d + U| \ll D$. At these scales the impurity is practically non-interacting. In the regime $|\varepsilon_d + U| \gg D \gg |\varepsilon_d|$, U does not flow and the scaling equation for ε_d become:

$$\frac{\mathrm{d}\log\varepsilon_d}{\mathrm{d}\log D} = -\Gamma_\xi \ . \tag{8}$$

It can be shown that Γ_ξ does not flow in this limit so that integration of (8) gives:

$$\varepsilon_d(D) = \varepsilon_d^* - \Gamma_\xi \log\frac{D}{\Gamma_\xi} \tag{9}$$

with $\varepsilon_d^* = \epsilon_d(D_0) + \Gamma_\xi \log\frac{D_0}{\Gamma_\xi}$ is a scaling invariant constant expressed in term of the 'bare' parameters at the initial scale $D_0 \sim \Delta_d$.

The scaling in this form can be continued until D reaches a scale $\tilde{D}$ so that $|\tilde{D}| \gtrsim |\varepsilon(\tilde{D})|$. For $\varepsilon(D), \varepsilon(D) + U \gg D$ real charge fluctuations cannot occur and we may explicitly eliminate them by a Schrieffer-Wolff transformation. The scale $\tilde{\varepsilon}_d \sim \tilde{D}$ at which the Haldane flow (8) applies is set by:

$$\tilde{\varepsilon}_d + \Gamma_\xi \log\left(\tilde{a}\frac{|\tilde{\varepsilon}_d|}{\Gamma_\xi}\right) = \epsilon_d^* \ , \tag{10}$$

with $\tilde{a}$ a positive constant of $\mathcal{O}(1)$. At this stage we may perform the Schriffer-Wolff transformation [29] with the effective parameters $\tilde{\varepsilon}_d, U, \Gamma_\xi$.

3.3 Schrieffer-Wolff Transformation $D \sim \tilde{\epsilon}_d$

The Schrieffer-Wolff transformation [29] lead us now to the Hamiltonian:

$$H_{1\mathrm{CK}} = \sum_{k\sigma} \varepsilon_{\phi k\sigma} \phi^\dagger_{k\sigma} \phi_{k\sigma} + \sum_{k\sigma} \varepsilon_{\zeta k\sigma} \zeta^\dagger_{k\sigma} \zeta_{k\sigma} + \sum_{k\sigma} \varepsilon_{\xi k\sigma} \xi^\dagger_{k\sigma} \xi_{k\sigma} + 2\tilde{J}_\xi \boldsymbol{S} \cdot \boldsymbol{s}_\xi(0) \tag{11}$$

with

$$\tilde{J}_\xi = J_\xi(\tilde{D}) = t_\xi^2 \left[\frac{1}{|\tilde{\varepsilon}_d|} + \frac{1}{|\tilde{\varepsilon}_d + U|} \right] . \tag{12}$$

Here, $\boldsymbol{S} = (S_x, S_y, S_z)$; $S^\pm = S_x \pm \mathrm{i}S_y$, $S^\pm = d^\dagger_{\uparrow(\downarrow)} d_{\downarrow(\uparrow)}$, $2S^z = d^\dagger_\uparrow d_\uparrow - d^\dagger_\downarrow d_\downarrow$. The ξ channel spin operators are: $\boldsymbol{s}_\xi(0) = \sum_{kq} \boldsymbol{s}^{kq}_\xi$ with $s^{\pm kq}_\xi = \xi^\dagger_{k\uparrow(\downarrow)} \xi_{q\downarrow(\uparrow)}$ and $2s^{zkq}_\xi = \xi^\dagger_{k\uparrow} \xi_{q\uparrow} - \xi^\dagger_{k\downarrow} \xi_{q\downarrow}$.

Now we may continue the RG flow utilizing the single-channel Hamiltonian of (11) up to a scale of order but greater than the charging energy E_c of the large dot.

3.4 Flow for $\tilde{\varepsilon}_d > D \gtrsim E_c$

To perform the scaling transformation on the Hamiltonian (11), we follow the poor man's scaling transformation of the Kondo model [27, 30]. Formally this is done by progressively integrating out the fast degrees contribution to the $\mathcal{T}$-matrix $\mathcal{T}_\xi(E)$ at energy $E \ll D$. This procedure leads to the following equation for the spin isotropic Kondo model:

$$\delta J_\xi = -J_\xi^2 \nu \, |\delta D| \left[\frac{1}{E - D + \varepsilon_{\xi k}} + \frac{1}{E - D - \varepsilon_{\xi k'}} \right] . \tag{13}$$

The first (second) term correspond to an intermediate electron (hole) process. As long as $D \gg E_c, E$ we ignore the dependence of the denominator on E and $\varepsilon_{\xi k}$. Therefore for these scales E can be above or below E_c.

Inclusion of the next order terms in the RG equation leads to the scaling equation:

$$\frac{\delta J_\xi}{\mathrm{d} \log D} = -2\nu J_\xi^2 + 2\nu^2 J_\xi^3, \text{ for } U > D > aE_c, \; a \gtrsim 1 . \tag{14}$$

The solution of this equation can be written in terms of the single-channel Kondo temperature T_K (that means the Kondo temperature of the single-channel ξ we would have if E_c was zero) as:

$$2\nu J_\xi(D) = \frac{1}{\log(D/T_K)} ,$$

$$T_K = \tilde{D} \sqrt{2J_\xi(\tilde{D})\nu} \mathrm{e}^{-1/\left[2\nu J_\xi(\tilde{D})\right]} \approx \sqrt{\frac{\Gamma_\xi \tilde{\varepsilon}_d U}{|\tilde{\varepsilon}_d + U|}} \mathrm{e}^{-\frac{|\tilde{\varepsilon}_d||\tilde{\varepsilon}_d + U|}{2\Gamma_\xi U}} , \tag{15}$$

where in the last approximate equality we have used (10) and (12) for $\tilde{\varepsilon}_d$ and $J_\xi(\tilde{D})$.

3.5 Elimination of the c - ψ Couplings at $D \sim E_c$

At the scale $D = aE_c$ we have to stop the RG procedure that led to (14) as at this scale real charge excitations on the large dot c are not energetically favorable. These charge fluctuations were implicitly included until this scale as the rotation (5) mixes channel c with channel ψ and therefore allows charge fluctuations on the large dot.

Thus now we rotate back Hamiltonian (11) (with the scaled parameter) with the help of the transformation (5) to the basis ψ, c and ϕ. During the flow from D_0 to D_1 the channels ζ and ϕ remain free, so the angle θ', which does not appear in Hamiltonian (11), does not change during that flow. The rotation gives:

$$\begin{aligned} H^{\mathrm{2CK}}_{c\psi\phi} &= \sum_{k\sigma} \varepsilon_{\phi k\sigma} \phi^\dagger_{k\sigma} \phi_{k\sigma} + \sum_{k\sigma} \varepsilon_{\psi k\sigma} \psi^\dagger_{k\sigma} \psi_{k\sigma} \\ &+ \sum_{k\sigma} \varepsilon_{ck\sigma} c^\dagger_{k\sigma} c_{k\sigma} + E_c (n - \mathcal{N})^2 \\ &+ 2J_{\psi\psi} \boldsymbol{S} \cdot \boldsymbol{s}_{\psi\psi}(0) + 2J_{c\psi} \boldsymbol{S} \cdot \boldsymbol{s}_{c\psi}(0) \\ &+ 2J_{\psi c} \boldsymbol{S} \cdot \boldsymbol{s}_{\psi c}(0) + 2J_{cc} \boldsymbol{S} \cdot \boldsymbol{s}_{cc}(0) \end{aligned} \tag{16}$$

with $J_{\psi\psi} = \cos^2\theta' J_\xi(D_1)$, $J^*_{c\psi} = J_{\psi c} = \cos\theta' \sin\theta' \mathrm{e}^{-\mathrm{i}\varphi_c} J_\xi(D_1)$, $J_{\psi\psi} = \cos^2\theta' J_\xi(D_1)$. The $a, b = \psi, c$ channel spin operators are: $\boldsymbol{s}_{ab}(0) = \sum_{kq} \boldsymbol{s}^{kq}_{ab}$ with $s^{\pm kq}_{ab} = a^\dagger_{k\uparrow(\downarrow)} b_{q\downarrow(\uparrow)}$ and $2s^{zkq}_a = a^\dagger_{k\uparrow} b_{q\uparrow} - a^\dagger_{k\downarrow} b_{q\downarrow}$.

Elimination of Charge Fluctuations on Large Dot at Scales $< E_c$

We now continue and integrate out energy scales in the vicinity of E_c. This can be done in a single RG step where we integrate out all degrees of freedom from the band width $D_1 = aE_c$, $a \gtrsim 1$ which is wider than E_c to a band width $D_2 = a'E_c$, $a' \lesssim 1$ which is narrower than E_c . At the scale D_2 charge fluctuation on the small dot are not possible, thus the off-diagonal terms $J_{\psi c}$ and $J_{c\psi}$ are strongly suppressed and we set them to zero.

At the same time we should take into consideration the flow of the diagonal terms $J_{\psi\psi}$ and J_{cc} which do not cause charge fluctuations on the large dot. However, if $E_c \gg T_K$ the flow of these parameters is not significant. We therefore may substitute $J_\xi(D_1)\nu$ by $1/\log(E_c/T_K)$ knowing that we are making a mistake by a factor of order a/a' inside the logarithm. This type of mistake is within the "logarithmic" accuracy of the RG scaling procedure. We arrive therefore at scales below E_c to the Hamiltonian:

$$\begin{aligned} H^{\mathrm{2CK}}_{c\psi\phi} &= \sum_{k\sigma} \varepsilon_{\phi k\sigma} \phi^\dagger_{k\sigma} \phi_{k\sigma} + \sum_{k\sigma} \varepsilon_{\psi k\sigma} \psi^\dagger_{k\sigma} \psi_{k\sigma} \\ &+ \sum_{k\sigma} \varepsilon_{ck\sigma} c^\dagger_{k\sigma} c_{k\sigma} + E_c (n - \mathcal{N})^2 \\ &+ 2\tilde{J}_\psi \boldsymbol{S} \cdot \boldsymbol{s}_{\psi\psi}(0) + 2\tilde{J}_c \boldsymbol{S} \cdot \boldsymbol{s}_{cc}(0) \end{aligned} \tag{17}$$

with $2\tilde{J}_\psi\nu \equiv 2\tilde{J}_{\psi\psi}\nu = \cos^2\theta'/\Lambda$ and $2\tilde{J}_c\nu = 2\tilde{J}_{cc}\nu = \sin^2\theta'/\Lambda$, where $\Lambda = \log(E_c/T_K)$.

Hamiltonian (17) is identical to the (possibly anisotropic) 2CK Hamiltonian, discussed extensively in the literature [9]. To stabilize the two-channel fixed point one needs to tune

$$\cos^2\theta' = \sin^2\theta' = 1/2 \Rightarrow |t_\psi| = |t_c| \,. \tag{18}$$

This condition for the stabilization of the two-channel Kondo effect is obtained with the assumption $E_c \ll U$.

3.6 Scaling Below E_c

Below E_c we should examine the two-channel Kondo Hamiltonian (17). Nozieres and Blandin [10] analyzed it and found the scaling equations:

$$\begin{aligned}\frac{\mathrm{d}J_\psi}{\mathrm{d}\log D} &= -2\nu J_\psi^2 + 2\nu^2 J_\psi(J_\psi^2 + J_c^2) \,,\\ \frac{\mathrm{d}J_c}{\mathrm{d}\log D} &= -2\nu J_c^2 + 2\nu^2 J_c(J_\psi^2 + J_c^2) \,.\end{aligned} \tag{19}$$

When $\cos^2\theta' = 1/2$ we have $J_\psi = J_c \equiv J$, so the system should flow to the two-channel fixed point. The scaling equation then becomes:

$$\frac{\mathrm{d}J}{\mathrm{d}\log D} = -2\nu J^2 + 4\nu^2 J^3 \,. \tag{20}$$

The extra factor 2 in front of the third order term (compared to the single-channel case (14)) appears because there are now two channels.

Integrating out these equations we find that J diverges at the scaling-invariant temperature:

$$T_{2\mathrm{CK}} \approx E_c 2J(E_c)\nu \mathrm{e}^{-1/2\nu J(E_c)} \,. \tag{21}$$

Using now (15) and the symmetric condition $\cos^2\theta' = 1/2$ we have $2\nu J(E_c) = 1/\left(2\log\left(E_c/T_K\right)\right)$ and we find, as long as $T_K \ll E_c$

$$T_{2\mathrm{CK}} \approx \frac{1}{2\log\left(E_c/T_K\right)}\frac{T_K^2}{E_c} \,. \tag{22}$$

Our treatment ceases to be valid when $T_K \sim E_c$.

Experimentally one would like to make the level spacing in the large dot small, maintaining the validity of the assumption of a continuum of levels in the big dot. This will necessarily make E_c small as well. At the same time one would like to make $T_{2\mathrm{CK}}$ (21) as large as possible, since two-channel Kondo phenomena will be most dramatic well below $T_{2\mathrm{CK}}$. We conclude that one should tune the system so that $E_c \sim T_K$. These were the guidelines for the choice of dot sizes and coupling strengths in [21, 22].

The case $E_c \ll T_K$ was analyzed in [23] and it was found that the highest $T_{2\mathrm{CK}}$ is obtained for $E_c \sim T_K$.

4 Examples of Conductance Curves

After establishing the conditions for the observation of the 2CK, we will now derive the conductance scaling curves appropriate for comparison with experiments. In a typical experiment one fixes the temperature T, applies a bias voltage V between the left and the right leads, and measures the current I through them. The differential conductance of the system is defined as:

$$G(V,T) = \frac{\mathrm{d}I}{\mathrm{d}V} \,. \tag{23}$$

Starting from a general relation between the conductance and the dot propagators [26] we give in the Appendix A a detailed derivation of the general relation between the conductance [24,25] and the scattering matrix. Applying this relation to the specific expressions for the scattering matrix $\mathcal{T}$ of [12] we can calculate the scaling curves.

4.1 2CK Case $G(V,T)$

Zero Bias

The conditions for the realization of the two-channel Kondo case are discussed in Sect. 3. When they are fulfilled we can utilize (for $V, T \ll T_{K2}$) the results for the $\mathcal{T}$-matrix obtained by conformal field theory [12]. We repeat them here for convenience:

$$\mathrm{Im}\mathcal{T}(\varepsilon) = -\frac{1}{2\pi\nu}\left[1 - 3\lambda\sqrt{\pi T} I_1\left(\frac{\varepsilon}{2\pi T}\right)\right] \tag{24}$$

with $\lambda \propto 1/\sqrt{T_{k2}}$, (the proportionality coefficient cannot be obtained by conformal field theory) and $I_1(x)$ is a function defined in (53) of Appendix B.1.

Substituting $\mathrm{Im}\mathcal{T}$ from (24) in the expression (50) for $G(0,T)$ we find:

$$G(0,T) = \frac{e^2}{2\pi\hbar}\sin^2(2\theta)\left(1 - 3\lambda\sqrt{\pi T} I\right) \,. \tag{25}$$

Using the results: $\pi/2 \int \mathrm{d}x \cos(x \log u) / \cosh^2(\pi x) = \sqrt{u}|\log u|/(1-u)$ for $0 < u < 1$ we have

$$I = \frac{\pi}{2}\int \mathrm{d}x \frac{1}{\cosh^2 x\pi} I_1(x) \approx -4/3 \,. \tag{26}$$

The last integral was estimated numerically [12].

Substituting (26) in (25) we find:

$$G(0,T) = \frac{e^2}{2\pi\hbar}\sin^2(2\theta)\left(1 + 4\lambda\sqrt{\pi T}\right) \,. \tag{27}$$

Since by assumption the initial coupling constants are small we may assume (see the discussion of [12] after their Eq. (3.64)) that λ is negative, and we choose the definition

$$\lambda = -\frac{1}{4\sqrt{T_{K2}}} , \tag{28}$$

this choice[4] yields:

$$G(0,T) = \frac{e^2}{2\pi\hbar}\sin^2(2\theta)\left(1-\sqrt{\frac{\pi T}{T_{K2}}}\right) . \tag{29}$$

Defining

$$G_0 = 2\tfrac{e^2}{2\pi\hbar}\sin^2(2\theta) = 8\frac{e^2}{2\pi\hbar}\frac{|t_l|^2|t_r|^2}{|t_l|^2+|t_r|^2} \tag{30}$$

$$\xrightarrow{|t_l|\ll|t_r|} \qquad 8\frac{e^2}{2\pi\hbar}\frac{|t_l|^2}{|t_r|^2} , \tag{31}$$

we find[5]:

$$G(0,T) = \frac{1}{2}G_0\left(1-\sqrt{\frac{\pi T}{T_{K2}}}\right) . \tag{32}$$

Finite Bias

Using (52) and manipulating $\mathrm{Im}\mathcal{T}$ as done in the previous section and in [12], we find for the extremely left-right asymmetric case ($|t_l| \ll |t_r|$):

$$G(V,T) = \frac{e^2}{2\pi\hbar}\left(\frac{2|t_l|}{|t_r|}\right)^2\left[1-3\lambda\sqrt{\pi T}I_2\left(\frac{eV}{2\pi T}\right)\right] . \tag{33}$$

With the definition (30) and the choice (28) for λ we have:

$$G(V,T) = G_0\frac{1}{2}\left[1-\sqrt{\frac{\pi T}{T_{K2}}}F_{2\mathrm{CK}}\left(\frac{eV}{\pi T}\right)\right] \tag{34}$$

with

$$F_{2\mathrm{CK}}(x) = -3/4 I_2(x/2) . \tag{35}$$

The function $I_2(x)$ is defined in (54) of Appendix B.2.

[4] We emphasize that the coefficient 1/4 as the proportionality coefficient is a bit arbitrary. A different choice would yield a different value for the crossover energy scale T_{K2}.

[5] Notice that in the symmetric case $\theta = \pi/4$ the conductance G_0 is twice $e^2/2\pi\hbar$.

Scaling Curves

Examining (34) we notice that

$$\frac{2}{G_0}\frac{G(0,T)-G(V,T)}{\sqrt{\pi T/T_{K2}}} = Y\left(\frac{|eV|}{\pi T}\right) \tag{36}$$

with the scaling function $Y(x) = F_{2\mathrm{CK}}(x) - 1$, and $F_{2\mathrm{CK}}$ given by (35). The scaling curve $Y(x)$ was used in [21, 22].

The asymptotes of $Y(x)$ are:

$$Y(x) \approxeq \begin{cases} c\,x^2 & \text{for} \quad x \ll 0 \\ \frac{3}{\sqrt{\pi}}\sqrt{x} - 1 & \text{for} \quad x \gg 1 \end{cases} \tag{37}$$

and we found numerically that $c = 0.748336$.

4.2 1CK Case $G(V,T)$

In case of the asymmetric two-channel Kondo Problem, one of the channels (the finite reservoir or the free reservoirs) "wins", so that at temperature smaller then the crossover temperature T_Δ [24] we should have a regular single-channel behavior. Here Δ is the asymmetry parameter – we define Δ to be positive when the free reservoirs "wins".

Zero Bias

At T, $eV \ll T_\Delta$ the scattering matrix is given by [12]

$$\mathrm{Im}\mathcal{T}(\varepsilon) = -\frac{1}{2\pi\nu}\left[2\theta(\Delta) - \mathrm{sign}(\Delta)\frac{(3\pi\lambda)^2}{4}\left(3\omega^2 + \pi^2T^2\right)\right] . \tag{38}$$

Choosing now $(T_\Delta^2)^{-1} = (3\pi\lambda)^2/4$ and substituting into (50) we find:

$$G(0,T) = G_0\left\{\theta(\Delta) - \mathrm{sign}(\Delta)\left(\frac{\pi T}{T_\Delta}\right)^2\right\} . \tag{39}$$

Finite Bias

At finite bias we can derive a closed formula only for the case $|t_l| \ll |t_r|$

$$G(V,T) = G_0\left\{\theta(\Delta) - \mathrm{sign}(\Delta)\left(\frac{\pi T}{T_\Delta}\right)^2\left[1 + \frac{3}{2}\left(\frac{eV}{\pi T}\right)^2\right]\right\} . \tag{40}$$

Scaling Curves

Examining (40) we note that

$$\frac{1}{G_0}\frac{G(0,T)-G(V,T)}{\left(\pi T/T_\Delta\right)^2}=\mathrm{sign}\left(\Delta\right)\frac{3}{2}\left(\frac{eV}{\pi T}\right)^2 . \tag{41}$$

This scaling curve was used in the analysis of the experimental data [21, 22].

5 Summary and Conclusions

To summarize we reviewed the temperature and voltage dependence of the differential conductance of a small quantum dot coupled to both two free leads and a large dot. The relation between the single-channel Kondo temperature and the two-channel Kondo temperature (for $E_c \gg T_K$) is given in (21). The scaling functions of the conductance for the two-channel Kondo and the single-channel Kondo cases are given in (36) and (41). As explained in the text the system crosses from the single-channel Kondo behavior to a two-channel Kondo behavior at a scale T_Δ determined by the asymmetry between the couplings to the two channels. Finding the scaling functions that describe the crossover between the single- and the two-channel Kondo behavior remains a challenge for future studies.

Acknowledgments

We would like to thank Ron Potok, Ileana Rau, and Eran Sela for many enlightening discussions. We also acknowledge support of the BSF, ISF and DIP grants.

Appendixes

A Conductance Through the Dot $G(V,T)$

In a typical experimental set up a bias voltage V is applied between the left and the right leads and the current I is measured between them. In this section we express the differential conductance defined as

$$G(V,T)=\frac{\mathrm{d}I}{\mathrm{d}V} , \tag{42}$$

in terms of other properties of the system. In addition to the temperature T and the bias potential V, the differential conductance depends upon other parameters that control the state of the system, such as $E_c, \mathcal{N}, U, \varepsilon_d, t_\alpha$ and an

external magnetic field (see (2)). An expression for the differential conductance of the two-channel Kondo system, based on the Kubo formula, appears in [25]. For pedagogical reasons and completeness of presentation we derive here these results using the Keldysh approach.

A.1 Differential Conductance $G(0,T)$

To find the conductance we start from a general formula for the finite bias current in terms of the local spectral density of the small dot[6] [26].

$$I = \frac{e}{h} \sum_{s=\uparrow,\downarrow} \int \mathrm{d}\varepsilon \left[f_l(\varepsilon) - f_r(\varepsilon) \right] \left[\frac{2\Gamma_l \Gamma_r}{\Gamma_l + \Gamma_r} A_s(\varepsilon) \right] . \tag{43}$$

We assume that $\Gamma_{l(r)} = \pi\nu \left| t_{l(r)} \right|^2$ does not depend on spin, and that the density of states ν is identical for the two leads, and we denote the Fermi function in the left (right) lead by $f_{l(r)}$. The spectral density of spin s, $A_s(\varepsilon)$ is related to the retarded Green's function of the dot $G^{\mathrm{ret}}_{ds}(\varepsilon)$ through:

$$A_s(\varepsilon) = -2 \ \mathrm{Im} \left[G^{\mathrm{ret}}_{ds}(\varepsilon) \right] . \tag{44}$$

In the general case $A_s(\varepsilon)$ may depend upon the chemical potential at the left and right leads, however there are two simple cases for which we can express the spectral density in terms of the $\mathcal{T}$-matrix at equilibrium. Note: The $\mathcal{T}$ matrix is defined below in (48).

When the bias V is small we find the differential conductance by expanding the Fermi function, and assuming that the electrochemical potential difference between the left and right lead is $\mu_l - \mu_r = eV$:

$$G(0,T) = \left. \frac{\mathrm{d}I}{\mathrm{d}V} \right|_{V=0} = -\frac{e^2}{h} \sum_{s=\uparrow,\downarrow} \int \mathrm{d}\varepsilon f'(\varepsilon) \left[\frac{2\Gamma_l \Gamma_r}{\Gamma_l + \Gamma_r} A_s(\varepsilon) \right] . \tag{45}$$

In this case $A_s(\varepsilon)$ is the *equilibrium* spectral density, since both leads are at the same electrochemical potential. To find the relation between A_s and the $\mathcal{T}$-matrix, we use the relation (at equilibrium)

$$\hat{\mathcal{G}} \equiv \hat{\mathcal{G}}_0 + \hat{\mathcal{G}}_0 \hat{t} \hat{\mathcal{G}}_0 \mathcal{G}_d \qquad \text{and} \qquad \hat{t} = \begin{pmatrix} |t_l|^2 & t_l t_r^* \\ t_r t_l^* & |t_r|^2 \end{pmatrix} , \tag{46}$$

where the components of the matrix $\hat{\mathcal{G}}$, $\mathcal{G}_{ij}$ are the propagators associated with annihilating an electron in lead i and creating an electron in lead j. At

[6] Notice that (43) is valid only when the coupling matrix to the left lead is proportional to the coupling matrix to the right lead. In our case we have only one state in the dot and therefore this condition holds.

energy scales below E_c the electrons from the large dot cannot be transferred to the leads so that its Green's function $\mathcal{G}_c$ is decoupled from the equation. The matrix $\hat{\mathcal{G}}_0 = \mathcal{G}_0\mathbf{I}$, with $\mathcal{G}_0$ denoting the free fermions Green's function and $\mathbf{I}$ is the identity matrix in the space of left and right leads.

A rotation in the left-right space [31] with $\hat{u} = \exp(-\mathrm{i}\theta\tau^y)\exp(-\mathrm{i}\varphi\tau^z)$, $\cos(\theta) = |t_l|/t_\psi$, $t_\psi^2 = |t_l|^2 + |t_r|^2, 2\varphi = \arg(t_l t_r^*)$ diagonalizes the matrices and gives:

$$\hat{u}^\dagger\hat{\mathcal{G}}\hat{u} = \begin{pmatrix} \mathcal{G}_0 + \mathcal{G}_0 t_\psi^2 \mathcal{G}_d \mathcal{G}_0 & 0 \\ 0 & \mathcal{G}_0 \end{pmatrix} \equiv \begin{pmatrix} \mathcal{G}_\psi & 0 \\ 0 & \mathcal{G}_\phi \end{pmatrix} . \tag{47}$$

Using (47) and the standard definition of the $\mathcal{T}$-matrix:

$$\mathcal{G} = \mathcal{G}_0 + \mathcal{G}_0\mathcal{T}\mathcal{G}_0 \tag{48}$$

we find

$$\mathcal{T}_\phi = 0, \;\; \mathcal{G}_d = \mathcal{T}_\psi/t_\psi^2 . \tag{49}$$

Introducing into (49) the spin index, combining it with the definition of the spectral density in (44), the result for the conductance at zero bias in (45) and the relation $\Gamma_{l(r)} = \pi\nu\left|t_{l(r)}\right|^2$ we find:

$$\begin{aligned} G(0,T) &= \left.\frac{\mathrm{d}I}{\mathrm{d}V}\right|_{V=0} \\ &= \pi\nu\frac{e^2}{h}\sin^2(2\theta)\sum_{s=\uparrow,\downarrow}\int \mathrm{d}\varepsilon f'(\varepsilon)\mathrm{Im}\;\mathcal{T}_\psi^s(\varepsilon) , \\ \sin(2\theta) &= \frac{2\,|t_l|\,|t_r|}{|t_l|^2 + |t_r|^2} . \end{aligned} \tag{50}$$

This result is almost identical to the result of [31] that was derived using the Kubo formalism. In the case at hand after the rotation (see (47)) we have *two* effective channels (ψ and c) connected to the dot. The conductance between the left and right leads is determined by the $\mathcal{T}$-matrix of channel ψ.

A.2 Differential Conductance $G(V,T)$

At finite bias between the left and the right leads of Fig. 1, it is impossible to perform the rotation of (47) because the left and the right leads are at different electrochemical potentials. In other words, the spectral function $A_s(\epsilon)$ of the small dot depends on the non-equilibrium potential in a non-trivial way. If however the tunneling matrix element to one of the leads is small (we assume without loss of generality that it is the left lead, i.e., $t_l \ll t_r, t_c$) then the small dot will be in equilibrium with the large dot and the right lead. Even at finite bias V, tunneling from the left lead will just probe the dot, but will not influence its properties, in analogy to scanning tunneling spectroscopy of electrons on surfaces.

More formally we expand (43) in $|t_l/t_r|$, giving $2\Gamma_l\Gamma_r/(\Gamma_l+\Gamma_r) \to 2\Gamma_l$. The Green's function of the small dot is given now by $\mathcal{G}_d = \mathcal{T}_r/|t_r|^2$. Using the relation (44) between the spectral density and the local Green function we find:

$$I = -\pi\nu\frac{e}{h}\frac{4\left|t_l\right|^2}{\left|t_r\right|^2}\sum_{s=\uparrow,\downarrow}\int \mathrm{d}\varepsilon\left[f_l\left(\varepsilon\right)-f_r\left(\varepsilon\right)\right]\mathrm{Im}\ \mathcal{T}_r(\varepsilon)\ . \tag{51}$$

Since $\mathcal{T}_r$ does not depend on the electrochemical potential in the left lead, eV, we have for $t_l \ll t_r$

$$\begin{aligned} G(V,T) &= \frac{\mathrm{d}I}{\mathrm{d}V} \\ &= \pi\nu\frac{e^2}{h}\frac{4\left|t_l\right|^2}{\left|t_r\right|^2}\sum_{s=\uparrow,\downarrow}\int \mathrm{d}\varepsilon f'\left(\varepsilon-eV\right)\mathrm{Im}\ \mathcal{T}_r(\varepsilon)\ . \end{aligned} \tag{52}$$

B A Few Definitions

B.1 The Function $I_1(x)$

The function $I_1(x)$ is defined as

$$\begin{aligned} I_1(x) &= \int_0^1 \mathrm{d}u\frac{4}{\pi}\frac{1}{\sqrt{u\left(1-u\right)^3}}\left[\cos\left(x\log u\right)F(u)-1\right]\ , \\ F(u) &= \frac{1-\sqrt{u}}{2}E\left(\frac{-4\sqrt{u}}{\left(\sqrt{u}-1\right)^2}\right) \end{aligned} \tag{53}$$

with $E(x)$ the complete elliptic function.

B.2 The Function $I_2(x)$

The function $I_2(x)$ is given by:

$$\begin{aligned} I_2(x) &= \frac{\pi}{2}\int \mathrm{d}y\frac{1}{\cosh^2\left(\pi(y-x)\right)}I_1(y) \\ &= \int_0^1 \mathrm{d}u\frac{4}{\pi}\frac{1}{\sqrt{u(1-u)^3}}\left[\frac{\left|\log u\right|\sqrt{u}}{1-u}F(u)\cos(x\log u)-1\right]\ . \end{aligned} \tag{54}$$

The numerical calculation of $I_2(x)$ is plotted in Fig. 3. The asymptotic limits are:

$$I_2(x) \cong \begin{cases} -4/3-3.99\ x^2 & \text{for } x \ll 0 \\ -\left(4\sqrt{\frac{2}{\pi}}\right)\sqrt{x} & \text{for } x \gg 1\ . \end{cases} \tag{55}$$

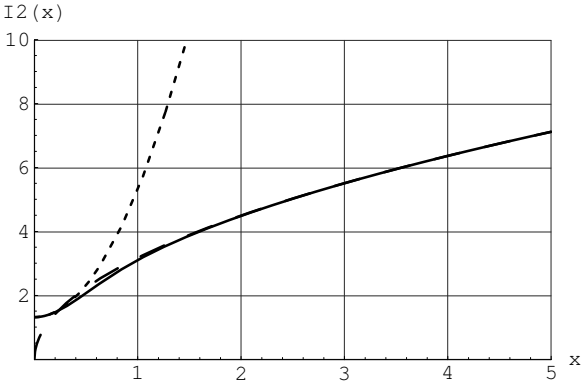

Fig. 3. The function $-I_2(x)$ of (54) is plotted in thick line. The dashed lines represent the asymptotic forms for at $x \ll 1$ and $x \gg 1$ [see (55)]

References

1. D. Goldhaber-Gordon, H. Shtrikman, D. Mahalu, D. Abusch-Magder, U. Meirav, M. A. Kastner: Nature **391**, 156 (1998)
2. J. Nygard, D. H. Cobden, P. E. Lindelof: Nature **408**, 342 (2000)
3. J. Park, A. N. Pasupathy, J. L. Goldsmith, C. Chang, Y. Yaish, J. R. Petta, M. Rinkoski, J. P. Sethna, H. D. Abruna, P. L. McEuen, D. C. Ralph: Nature **417**, 722 (2002)
4. Wenjie Liang, Marc Bockrath, Hongkun Park: Phys. Rev. Lett. **88**, 126801 (2002)
5. J. Kondo: Prog. Theor. Phys. **32**, 37 (1964)
6. J. M. Luttinger: Journal of Mathematical Physics **4**, 1154 (1963); S. Tomonaga: Prog. Theor. Phys. (Kyoto) **5**, 544 (1950)
7. R. B. Laughlin: Phys. Rev. Lett. **50**, 1395 (1983)
8. J. Bardeen, L. N. Cooper, J. R. Schrieffer: Phys. Rev. **108**, 1175 (1957)
9. D.L. Cox, A. Zawadowski: Advances in Physics **47**, 599 (1998); cond-mat/9704103 (1997)
10. Ph. Nozieres, A. Blandin: J. Phys. (Paris) **41**, 193 (1980)
11. A. Zawadowski: Phys. Rev. Lett. **45**, 211 (1980)
12. Ian Affleck, Andreas W. W. Ludwig: Phys. Rev. B **48**, 7297 (1993)
13. C. L. Seaman, M. B. Maple, B. W. Lee, S. Ghamaty, M. S. Torikachvili, J.-S. Kang, L. Z. Liu, J. W. Allen, D. L. Cox: Phys. Rev. Lett. **67**, 2882 (1991)
14. D. C. Ralph, R. A. Buhrman: Phys. Rev. Lett. **69**, 2118 (1992)
15. D. C. Ralph, A. W. W. Ludwig, Jan von Delft, R. A. Buhrman: Phys. Rev. Lett. **72**, 1064 (1994)
16. T. Cichorek, A. Sanchez, P. Gegenwart, F. Weickert, A. Wojakowski, Z. Henkie, G. Auffermann, S. Paschen, R. Kniep, F. Steglich: Phys. Rev. Lett. **94**, 236603 (2005)

17. In dilute rare earth/actinide alloys, thermodynamic evidence is largely compatible with 2ck, but the inability to reconcile transport data with 2ck theory has left the final state of understanding inconclusive. In the putative atomic two-level system 2ck measured by Ralph et al. in narrow metal constrictions [14,15] and more recently by Cichorek et al. in bulk metallic glasses [16], the transport data nicely match the expected scaling behavior, but a fundamental controversy persists about whether the original (near-degenerate but strongly-coupled two-level system) model Hamiltonian can even exist in a real physical system
18. K. A. Matveev: Sov. Phys. **72**, 892 (1991); Zh. Eksp. Teor. Fiz. **99**, 1598 (1991)
19. Eugene H. Kim: cond-mat/0106575 (2001)
20. Yuval Oreg, David Goldhaber-Gordon: cond-mat/0203302 (2002)
21. R. M. Potok, I. G. Rau, Hadas Shtrikman, Yuval Oreg, D. Goldhaber-Gordon: *Observation of the two-channel Kondo effect* (2006)
22. R. M. Potok, I. G. Rau, Hadas Shtrikman, Yuval Oreg, D. Goldhaber-Gordon: Supplementary information for cond-mat/0610721 (2006); Potok et al: *Observation of the two-channel Kondo effect* (2006)
23. S. Florens, A. Rosch: Phys. Rev. Lett. **92**, 216601 (2004)
24. M. Pustilnik, L. Borda, L. I. Glazman, J. von Delft: Phys. Rev. B **69**, 115316 (2004)
25. L. I. Glazman, M. Pustilnik: cond-mat/0302159 (2003)
26. Y. Meir, N. S. Wingreen: Phys. Rev. Lett. **68**, 2512 (1992)
27. A. C. Hewson: *The Kondo Problem to Heavy Fermions*, (Campridge University Press, Cambridge 1993)
28. F. D. M. Haldane: Phys. Rev. Lett. **40**, 416 (1978)
29. J. R. Schrieffer, P. A. Wolff: Phys. Rev. **149**, 491 (1966)
30. P. W. Anderson: J. Phys. C **3**, 2436 (1970)
31. M. Pustilnik, L. I. Glazman: Phys. Rev. B **64**, 045328 (2001)

Kondo Physics in Artificial Molecules

K. Kikoin[1] and Y. Avishai[1,2,3]

[1] Department of Physics, Ben-Gurion University, Beer Sheva 84105, Israel.
kikoin@bgu.ac.il, yshai@bgumail.bgu.ac.il
[2] The Ilse Katz Center for meso and nano scale Science and Technology, Ben-Gurion University, Beer Sheva 84105, Israel.
[3] Department of Applied Physics, University of Tokyo, Hongo Bunkyo-ku Tokyo 113-8656, Japan.

1 Introduction: Kondo Mapping and Beyond

There are numerous models in the literature of condensed matter theory, whose significance for achieving progress in our understanding of nature goes far beyond the original aim of explaining specific experimental observations. One may mention in this context the Bardeen-Cooper-Schrieffer's explanation of the nature of electron pairing in superconductors, the Ginzburg-Landau equation intended for describing critical fluctuations, the concept of self-localization of excitations in a perfect crystal formulated by Deigen, Pekar and Toyozawa and various other seminal ideas. The explanation offered by J. Kondo for the puzzling shallow minimum in the temperature dependent resistivity of metals doped by magnetic impurities [1] is one of the most salient examples of this kind of scenario. To explain it, consider first Kondo's original idea, which was formulated within the framework of a well-established Hamiltonian describing exchange interaction between an impurity spin $\boldsymbol{S_r}$ located on a given site $\boldsymbol{r}$ and the spin density $\boldsymbol{s_r}$ pertaining to a Fermi sea of conduction electrons at this site. The latter is defined by the Fourier transform of the itinerant spin $\boldsymbol{s_{kk'}} = c^\dagger_{\boldsymbol{k}\sigma}\hat{\tau}c_{\boldsymbol{k}'\sigma'}$ projected on the impurity site $\boldsymbol{r}$, namely $\boldsymbol{s_r} = \sum_{\boldsymbol{kk'}} \boldsymbol{s_{kk'}} \exp[\mathrm{i}(\boldsymbol{k}-\boldsymbol{k}')\cdot\boldsymbol{r}]$. Here $\hat{\tau}$ is the vector of Pauli matrices for a spin 1/2. The so-called *sd*-exchange Hamiltonian is,

$$H_{sd} = \sum_{\boldsymbol{k},\sigma} \varepsilon_k c^\dagger_{\boldsymbol{k}\sigma} c_{\boldsymbol{k}\sigma} + J\boldsymbol{S_r}\cdot\boldsymbol{s_r} \tag{1}$$

where ε_k is the energy dispersion of the itinerant electrons and J is the exchange coupling constant. At first glance, it looks deceptively simple. However, Kondo noticed that the first correction to the impurity scattering amplitude of the conduction electrons beyond the Born approximation suffers an infrared

logarithmic divergence in energy or temperature. This results in a singular behavior of amplitudes for an antiferromagnetic sign of the exchange coupling (that is, $J > 0$), and renders perturbation theory inapplicable below a certain energy scale known as the Kondo temperature. Nearly two decades of incessant efforts to take this singularity properly into account and to find the ground state of the system had crowned with finding the exact solution both numerically (in a framework of Numerical Renormalization Group, NRG) [2] and analytically (by using the Bethe ansatz) [3, 4].

Soon after Kondo's breakthrough, theoreticians started to extend this promising conceptual framework for other physical situations and for more complex objects than simple localized moments. It was recognized that the Kondo mechanism should work also in systems exhibiting electron tunneling, where two metallic slabs are separated by thin dielectric layer, which forms a tunnel barrier for electrons moving from one slab to the other. It was shown [5,6] that the magnetic impurity located somewhere near the tunneling layer plays the same role in tunnel conductance as magnetic impurity immersed in a metal (and subject to exchange interaction with Fermi sea electrons) does for impurity resistance. It was also shown [7] that the Friedel-Anderson model [8,9] for resonance scattering of conduction electrons by the electrons occupying the $3d$ levels of transition metal impurities can be mapped on the exchange Hamiltonian (1) provided the strong Coulomb interaction in the $3D$ shell suppresses charge fluctuations on the impurity site. In the next stage of development orbital degrees of freedom were incorporated in the Kondo physics. The idea of this generalization is based on the fact that the magnetic impurity being put in the center of coordinates imposes its point symmetry on the otherwise translationally invariant crystal, and the appropriate description of scattered waves should exploit the formalism of partial wave expansion (either in spherical waves [8] or in cubic harmonics [10]). Based on this idea, the generalized Schrieffer-Wolff model was proposed [11, 12], where the magnetic impurity is described as an effective N-component moment, but the exchange scattering is not restricted by the usual spin selection rule $\Delta m = 0, \pm 1$ for the projection m of this moment. Another version of this model allots the impurity and band states both by spin and orbital (referred to as "color" in the general case) index. In case of spin $s = 1/2$ and N colors the symmetry of the impurity is $SU(2N)$ [13, 14]. Appearance of additional degrees of freedom in the Kondo Hamiltonian might lead to the interesting scenario of overscreened Kondo effect, which arises when the number of conduction electron "colors" exceed that of impurity moment [15]. This effect is characterized by the non-Fermi-liquid low-temperature thermodynamics unlike the standard Kondo effect, which only modifies (although radically)) the Fermi liquid properties of undoped metal [16].

Another direction of expanding the Kondo physics is realized in mapping the Kondo or Anderson Hamiltonian on those of other (not necessarily magnetic) systems. A great variety of such generalizations is possible because some quantum systems may be described by a pseudospin, provided their

low-energy states are characterized as an effective two-level system (TLS) and external perturbations allow transitions between these levels. The first example of such generalization was suggested by the Hamiltonian describing tunneling between a Fermi sea electrons and an atom sitting in a double-well potential [17]. Another possibility of this sort arises when the crystal field splitting is involved in formation of the low-energy states of impurity atom (quadrupolar Kondo effect) [18, 19]. This type of "exotic" Kondo system was surveyed in a detailed review [20]. One should mention also the possibility of involving orbital degrees of freedom in the formation of Kondo-resonance for an adsorbed transition metal impurity, where the orbital degeneracy is lifted by the surface effects, e.g., the potential of atomic step edge [21].

A powerful incentive for further extension of the realm of Kondo physics has been offered in 1988, when the idea of underbarrier tunneling in presence of Kondo center was extended on the tunneling between metallic electrodes and nanoobjects like quantum dots or small metallic grains [22, 23]. Such nanoobject may serve as a Kondo center (a localized moment) provided (i) the electron spectrum is discrete due to spatial quantization, so that the level spacing $\delta\epsilon$ exceeds the tunneling rate, (ii) Coulomb blockade prevents charge fluctuations and (iii) the electron occupation number is odd, so that the effective spin of the nanoobject is 1/2. In this case the tunneling Hamiltonian may be mapped on the effective spin Hamiltonian (1), and the Kondo-like singularity arises as a zero-bias anomaly (ZBA) in tunnel transparency. This theoretical prediction was confirmed ten years later [24–26] in the experiments on planar quantum dots. Many experimental and theoretical studies then followed this experimental breakthrough ever since.

In the course of developing this new realm of condensed matter physics, further possibilities of extending the Kondo physics were subsequently discovered and exploited. Original idea of using the charge fluctuations as a source of Kondo tunneling was proposed [27] in the interim between the theoretical prediction of Kondo tunneling and its experimental verification. It was shown in this paper that in a situation, where two charge states of the quantum dot with occupation $N, N+1$ are nearly degenerate, this dot behaves as a two-level system, where the fluctuating charge configuration plays part of pseudospin, whereas the real spin projections may be treated as channel indices. Later on this idea was further developed and modified [28, 29].

Another facet of Kondo physics in nanoobjects was unveiled, when the possibility of Kondo effect in quantum dots with even electron occupation number was considered in several theoretical publications [30–33]. In this case the quantum dot with a singlet ground state may become magnetically active due to external forces, and Kondo effect arises either at finite energy [30] or at finite magnetic field [31–33]. Later on it was recognized that in many cases the direct mapping of the original Kondo model into such system is impossible, because the *effective* symmetry of the pertinent nanoobject is neither $SU(2)$ nor $SU(2N)$. The aim of this review is to describe various physical situations where the underlying nanoobject possesses complex (and in

some sense unusual) symmetries which are characterized by non-compact Lie groups or combinations of such groups with discrete groups of finite rotations.

2 Surplus Symmetries

Among the sources of surplus symmetries which enrich Kondo physics of nanoobjects one may find both discrete and continuous rotations stemming from complicated geometrical configurations of complex quantum dots, as well as those induced by external fields used in practical realizations of nanodevices. In this short review we will refrain from description of great variety of these devices, which may be found in current literature (see e.g., [34–36]). Fortunately, most of the relevant physics may be exposed in a relatively simple situation of electron tunneling through multivalley quantum dots in contact with metallic electrodes (leads).

A multivalley dot is an island with electrons confined by electrostatic potential in such a way that the spatially quantized electrons are distributed between several valleys. These valleys are coupled with each other by capacitive interaction and tunneling channels. Up to now there are several realizations of quantum dots with two and three valleys (double quantum dot, DQD, and triple quantum dot, TQD, respectively). Experimentally, the first such realizations of DQDs go back to mid 90-es [37–39]. It was pointed out that these objects can be treated as some forms of artificial molecules with each constituent dot playing the role of an artificial atom [40–43].

Compared with DQD, fabrication of TQD is a much more difficult experimental task and the first experimental realizations of these nanoobjects appeared only recently. One may mention in this connection the realization of TQD with an "open" central valley [44]. The term "open" here means that the tunneling between the side dots and the adjacent leads is limited by strong Coulomb blockade (see below), whereas the central dot freely donates and accepts electrons to and from its own reservoir, so its role in the device is only to mediate indirect exchange between the two side dots. Another successful attempt to fabricate a TQD was recorded in response to a theoretical proposal [45] to use this geometry for realization of ratchet effect in tunneling through nanoobjects. In this realization [46] the charge fluctuations were suppressed by the Coulomb blockade mechanism in all three valleys. The feasibility of filling the TQD with 1, 2 and 3 electrons by changing the gate voltages was demonstrated quite recently [47].

Theoretical studies of electronics in the TQD geometry were also inspired by possible applications in the field of quantum information [48, 49]. Investigation of the Kondo physics in TQD [50] was motivated by the experimental observation of molecular trimers Cr_3 on gold sublayers [51] by means of the tunnel electron spectroscopy, which allows to observe Kondo-type ZBA in conductance. Later on, other properties of these trimers such as the two-channel Kondo effect [52, 53], and the interplay between the Kondo effect and

Aharonov-Bohm effect in tunnel spectroscopy [54] were considered theoretically.

Our main focus of interest is a theoretical modeling of a device consisting of a multivalley quantum dot, metallic electrodes and corresponding gates. The latter regulate the electron occupation of any particular valley and of the dot as a whole. These properties and others are predetermined by the geometrical position of the valleys relative to the metallic leads (source and drain) in the device. Basically, there are three types of such devices which are possible for describing electric circuits with a DQD connected with a source s and drain d. They may be refereed as sequential, parallel and T-shape connections (Fig. 1a, Fig. 1b and Fig. 1c respectively). The two small dots which combine to form the DQD may either be identical or may differ in their size. Besides, different gate voltages may be applied to different valleys.

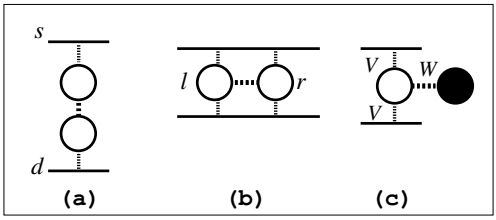

Fig. 1. Double quantum dot in **(a)** sequential, **(b)** parallel and **(c)** T-shape geometries. Filling black marks are the valley detached from the leads

In the most symmetrical configuration of identical dots the only additional symmetry which characterizes a DQD is the permutation symmetry P_2. In close analogy with Kondo effect in magnetically doped metals, one should consider the total symmetry of a device 'source + DQD + drain'. Then the symmetry group of the sequential configuration (Fig. 1a) contains the only discrete element, namely the s-d reflection axis, and this element adds nothing to the P_2 symmetry of the isolated DQD. The same statement is valid for the T-shape geometry (Fig. 1c). In the parallel geometry the system as a whole possesses two reflection axes, namely source-drain s-d and left-right l-r, where the indices l and r label two valleys of DQD.

Triple quantum dots (TQD) present theoreticians (and experimentalists as well) with a richer variety of geometrical configurations and possess more symmetries (Fig. 2). Similarly to DQD these trimers may be oriented both in sequential (vertical) and parallel (lateral) geometries (Fig. 2a and Fig. 2b, respectively). The natural generalization of the T-shape geometry presented in Fig. 1c is the cross geometry (Fig. 2c). Besides, TQD may be organized in a form of a triangle, which may form a closed or open element in an electronic circuit (Fig. 2d, Fig. 2e and Fig. 2f, respectively). In the two latter cases one deals with a three-terminal tunnel device. We will call the conformations shown in Fig. 2e and Fig. 2f as "ring" and "fork" configurations.

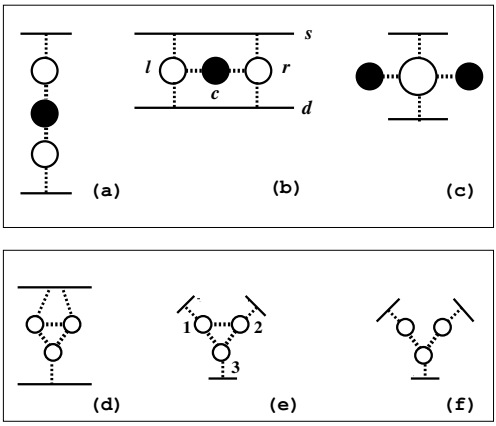

Fig. 2. Triple quantum dot in **(a)** sequential, **(b)** parallel, **(c)** cross-shape, **(d)** two-terminal triangular, **(e)** three-terminal triangular and **(f)** fork geometries

Now let us discuss the discrete symmetry elements characteristic for the above TQD configurations. If all three valleys are equivalent, the discrete symmetry of an isolated linear TQD is that of the permutation group P_3. The contact with the leads adds one more symmetry element, the *s-d* reflection, provided all three dots are coupled with the leads. We however consider the devices, where the central dot (filled black) is not coupled directly with the leads. Then the P_3 symmetry is lost. The only discrete symmetry element, namely *s-d* reflection is left in a vertical geometry (Fig. 2a), whereas both *s-d* and *l-r* reflections characterize a symmetry of lateral TQD (Fig. 2b). The same is valid for the cross geometry of Fig. 2c. The P_3 symmetry is inherent in the three-terminal configuration of Fig. 2e. In this case it is better to use the classification of discrete rotation group C_{3v}, which is isomorphic to the permutation group P_3. One may say [54, 55] that in the perfect triangular configuration the TQD imposes its C_{3v} symmetry on the device as a whole in close analogy with the Coqblin-Schrieffer-Cornut version [11, 12] of the conventional Kondo problem. In the geometries of Fig. 2d and Fig. 2f, the only element which is remained of the original symmetry of triangle is the *l-r* symmetry like in the cross geometry. *External magnetic field* applied perpendicularly to the plane of the triangle lowers the symmetry of a device by adding one more element, that is, chirality of the electron tunneling from the source to the drain.

Before considering the consequences of these symmetries for the Kondo physics, several introductory remarks about the derivation of the Kondo Hamiltonian are in order. In an early period of the theoretical studies, the problem of electron tunneling through short chains of quantum dots under strong Coulomb blockade restrictions was formulated in terms of the Mott-Hubbard picture [56, 57]. The theory was based on the idea that electrons

injected from the source *do not lose coherence* when propagating through the sequence of quantum dots until they leave the chain from the drain electrode. This approach is valid only for short enough "Hubbard chain", where the tunneling W between the adjacent valleys exceeds the tunneling V between the dot and the metallic leads. Using generalized Landauer method, one describes the tunnel transparency in terms of the Green functions of a nanoobject in contact with the leads [58]. Such a procedure starts with diagonalization of the Hamiltonian of nanoobject with subsequent calculation of renormalization of the spectrum of quantum dot due to tunnel contact with the leads. Early studies of these problems concentrated on the calculations of the Coulomb blockade peaks which arise with changing the occupation of the valleys (so-called Coulomb staircases). In terms of the Hubbard model, the Coulomb resonances are the Hubbard "minibands" [56], which arise as a result of collective Coulomb blockade [59] (Hubbard repulsion). It is known, however, that the spectral function of the Hubbard model contains also the central peak of predominantly spin origin. This peak is responsible for zero-bias anomalies in the tunnel conductance, which are at the center of our attention.

If the "chain" contains a single dot, one deals with the conventional mapping of the Anderson-like tunneling problem onto the Kondo-like cotunneling problem [22,23], so that the central peak is indeed the famous Abrikosov-Suhl resonance pinned to the Fermi level of the band electrons [3,4,60], which is responsible for the ZBA in tunnel conductance. DQD in sequential geometry is the first non-trivial generalization of the single dot case, and one of the primary tasks is to look whether some qualitative differences from the conventional Kondo effect arise because of combining the features of Anderson and Hubbard models in the effective tunneling Hamiltonian.

In any case, working in the above paradigm [56,57], one should derive the effective exchange Hamiltonian $H_{\rm ex}$ in accordance with the following procedure. The starting Hamiltonian is chosen in the same form as the original Anderson impurity Hamiltonian [9]:

$$H = H_{\rm band} + H_{\rm dot} + H_{\rm t} \tag{2}$$

where $H_{\rm dot}$ describes the properties of a chain detached from the lead *in terms of its eigenstates* $|\Lambda\rangle$:

$$H_{\rm dot} = \sum_{\Lambda} E_{\Lambda} |\Lambda\rangle\langle\Lambda| + Q(\hat{N} - \mathcal{N}) \; . \tag{3}$$

Here $\hat{N}$ is the operator of total electron number in the dot. The last term in (3) describes the Coulomb blockade mechanism: the total occupation of the dot $\mathcal{N}$ in neutral and charged states is fixed by the Coulomb blockade parameter Q entering the capacitive energy of a complex quantum dot as a whole. The eigenvalues E_{Λ} are found at fixed occupation $\mathcal{N}$ of DQD or TQD with all tunneling matrix elements W and interdot capacitive interactions Q' taken into account. The tunneling Hamiltonian $H_{\rm t}$ intermixes the states from

adjacent charge sectors $\mathcal{N},\mathcal{N}'$ due to injection or extraction of an electron from the complex dot:

$$H_{\rm t} = \sum_{\Lambda\Lambda'} \sum_{b=s,d} \sum_{k\sigma} V_{bk\sigma}^{\Lambda\Lambda'} c_{bk\sigma}^{\dagger} X^{\Lambda\Lambda'} + \text{h.c.}. \tag{4}$$

where $X^{\Lambda\Lambda'} = |\Lambda\rangle\langle\Lambda'|$ are the universal Hubbard operators [61]. The tunnel parameters $V_{bk\sigma}^{\Lambda\Lambda'}$ are usually approximated by a single parameter V. In the Hamiltonian (4) these configuration changing operators describe transitions between the states belonging to different charge sectors (one of this sectors corresponds to the neutral DQD and another belong to positively or negatively charged DQD. The index b enumerates the leads ($b = s, d$ in the two-terminal configurations)

The Hamiltonian $H_{\rm band}$ has the standard form

$$H_{\rm band} = \sum_{bk\sigma} \varepsilon_k c_{bk\sigma}^{\dagger} c_{bk\sigma} \ . \tag{5}$$

Comparing to the corresponding term in (1), this Hamiltonian contains one more index b. In the geometries with s-d reflection symmetry one may rotate the frame in such a way that the band operators are classified as even and odd operators relative to this reflection

$$c_{ek\sigma} = 2^{-1/2}(c_{sk\sigma} + c_{dk\sigma}) \ , \ c_{ok\sigma} = 2^{-1/2}(c_{sk\sigma} - c_{dk\sigma}) \ . \tag{6}$$

In case of single and double quantum dots this rotation usually excludes the odd combination from the tunneling Hamiltonian [22]. However in case of TQD it is not necessarily the case. We will return to this question in Sect. 4.

The cotunneling (exchange) Hamiltonian is usually obtained from (2) by means of the Schrieffer-Wolff (SW) canonical transformation [7], which excludes the states $|\Lambda\rangle$ belonging to the charge sectors $\mathcal{N} \pm 1$ from the effective Fock space. At fixed $\mathcal{N}$ we are left solely with spin degrees of freedom. In a conventional situation, the relevant symmetry is $SU(2)$ and the SW procedure results in an effective Hamiltonian (1). In case of even $\mathcal{N}$ and geometries including discrete rotations, the situation is more complicated, and the SW procedure intermixes the states $|\Lambda\rangle$ belonging to different irreducible representations of the Hamiltonian (3). The corresponding terms in the effective Hamiltonians may be expressed by means of the corresponding Hubbard operators $X^{\Lambda\Lambda'}$. In many cases, combinations of these operators form closed algebras which generate non-compact groups $SO(N)$ or $SU(N)$ with $N > 2$, describing the *dynamical symmetry* of complex quantum dots. Involvement of these dynamical symmetries turns the procedure of mapping the tunnel problem onto an effective exchange problem to be more complicated than in the simpler situations which were briefly described in Sect. 1. New features of the Kondo effect arising as a result of this procedure were described for the first time using the T-shaped DQD as an example [43]. Various manifestations of dynamical symmetries in physical problems are described in the

monographs [62,63]. Some mathematical aspects of the dynamical symmetries as applied to the Kondo problem may be found in the recent reviews [64,65].

3 Kondo Physics for Short Chains

Short chains represented in Figs. 1a-c, Figs. 2a-c are the most elementary objects, where many aspects of Kondo tunneling beyond the original paradigm [5,22] of Kondo mapping may be demonstrated. There is much in common between the linear DQD and TQD in vertical and lateral geometries, although there are some effects specific only for T-shaped and cross-shaped configurations (Fig. 1c and Fig. 2c), which will also be emphasized in this section.

3.1 Double Quantum Dots

We start a more detailed discussion of Kondo tunneling through artificial molecules with the case of DQD in the vertical geometry (Fig. 1a). Historically, this is the first generalization of a single dot problem (see above). At present, the tunneling through this simplest DQD is well-understood. Following an extensive theoretical discussion of Kondo physics in vertical quantum dot geometry [66–74], it turned out that the most general description of Kondo tunneling through vertical DQD may be given in terms of $SU(4)$ and $SO(4)$ symmetries of a low-energy multiplet in cases of odd and even electron occupation $\mathcal{N}$, respectively.

As was mentioned above, the microscopic description of Kondo tunneling is analyzed in a framework of the generalized Anderson Hamiltonian (2). All generic features of Kondo mapping are seen already in the most elementary cases of $\mathcal{N} = 1, 2$. DQD in these charge sectors can be treated as an artificial analog of the molecular ion H_2^+ and the neutral molecule H_2, respectively.

In case of $\mathcal{N} = 1$ the eigenstates of the Hamiltonian $H_{\rm dot}$ (3) are

$$E_{1,2} = \epsilon_d \mp W \ , \tag{7}$$

where ϵ_d is the discrete level position in the isolated valley of the DQD and W is the inter-valley tunneling. These two levels forming the "Hubbard miniband" in the $\mathcal{N} = 1$ charge sector correspond to even and odd combinations of the electron wave functions in the double-well confinement potential of the DQD. We are interested here not in the resonance tunneling through these levels, but in the ZBA connected with the Kondo effect. The characteristic energy T_K which scales the Kondo effect, should be compared with the level splitting W. However T_K itself depends on the level splitting, so that the comparison procedure should be performed self-consistently. In the limiting case $T_K(W) \ll W$, one may ignore the odd state E_2. Since the odd combination of the lead electron states in (6) is also excluded from the problem, we immediately see that in this case the mapping procedure (SW transformation)

reduces the tunneling problem to the case of single quantum dot with the Hamiltonian (1), where the exchange constant J is estimated as $J = V^2/E_C$, and E_C is expressed via addition and extraction energies, i.e., the energy costs to add or subtract an electron on/from the quantum dot:

$$E_C^{-1} = E_+^{-1} + E_-^{-1} \ , \qquad (8)$$
$$E^+ = \epsilon_d + Q - \epsilon_{\rm F} \ , \ E^- = \epsilon_{\rm F} - \epsilon_d \ .$$

The Kondo temperature is given by $T_K = D\exp(-1/2\rho J)$, where D is the characteristic energy scale for the band electrons in the leads and ρ is the density of states at the Fermi level $\epsilon_{\rm F}$.

In the opposite limit $T_K(W) \gg W$, one may neglect W when calculating T_K, so that the dot level acquires "orbital" degeneracy. Due to this quasi-degeneracy the effective exchange acquires an additional factor 2, so that the Kondo temperature in this case is $T_K = D\exp(-1/4\rho J)$. If fact, the difference between the two limiting cases is the manifestation of the $SU(4)$ symmetry, which characterizes the spin state of an electron in the double well potential. This symmetry will be described more strictly when we will turn to the case of lateral DQD. One should note that here we encountered the first manifestation of the complicated structure of spin multiplets in complex quantum dots, namely with non-universality of T_K: it crucially depends not only on the parameters of the Hamiltonian but also on the effective symmetry of the low-lying spin states involved in Kondo cotunneling. Interpolation between two limiting cases may be described in terms of gradual $SU(2) \to SU(4)$ crossover.

The study of the charge sector $\mathcal{N} = 2$ uncovers another important aspect of the Kondo mapping procedure. This sector corresponds to the half-filled Hubbard chain, where the single-electron tunneling is suppressed by Coulomb blockade (the interior of the Coulomb diamond in terms of the theory of single-electron tunneling [34, 35]). On the surface of it, the Kondo tunneling is also not achievable because the ground state of the DQD with $\mathcal{N} = 2$ is a spin singlet in close analogy with the case of hydrogen molecule H_2. However, the dynamical symmetry of DQD plays its part in this case as well.

Indeed, in case of strong Coulomb blockade $\beta \equiv W/Q \ll 1$ the spectrum of isolated DQD consists of two low-lying spin states $E_{S,T}$ and two charge transfer excitons $E_{e,o}$ (even singlet and odd triplet), with

$$E_S = 2\varepsilon - 2\beta W \ , \ E_T = 2\varepsilon \ , \qquad (9)$$
$$E_o = 2\varepsilon + Q \ , \ E_e = 2\varepsilon + Q + 2\beta W \ .$$

Only the low-energy singlet/triplet (S/T) pair is relevant for Kondo tunneling. Like in the case of $\mathcal{N} = 1$, the triplet state E_T is frozen provided the exchange gap $\Delta_{\rm ex} = 2\beta W$ essentially exceeds T_K. In the opposite case $\Delta_{\rm ex} \ll T_K$ the spin multiplet as a whole is involved in Kondo tunneling, and the Kondo-type ZBA may survive [69]. One may describe this phenomenon in terms of the theory of conventional two-site Kondo effect [75]. According to this theory, the

antiferromagnetic intersite exchange J_{12} competes with the single-dot Kondo temperature T_{K0}. At small J_{12} (small enough W in our case) each spin is screened independently and the Kondo-type ground state may be achieved. However, there exists a critical value J_c, so that at $J_{12} > J_c$ the two spins are locked into a singlet state and the Kondo effect does not apply.

The relevant dynamical symmetry is the $SO(4)$ symmetry of S/T manifold ("spin rotator" [43]). Again we postpone the derivation of the effective Hamiltonian for the Kondo tunneling and the discussion of its observable manifestations for the case of lateral DQD. Here one should note that in the charge sector $\mathcal{N} = 2$, the exchange gap Δ plays the same role as the charge-transfer gap W in case of $\mathcal{N} = 1$: the dependence $T_K(\Delta)$ is determined by a gradual symmetry crossover.

Now we turn to the parallel (lateral) geometry of Fig. 1b. A new element, which arises in this case is the possibility of *separate* channel for each dot. The new features brought by this additional quantum number were discussed in several publications [74, 76]. In case of two channels the conduction electrons retain additional "color" after rotation (6). We will enumerate the states corresponding to this color by the same indices $i, j = 1, 2$ as those used for the dots because they describe the odd and even states relative to the l-r reflection like those in the states (7) and (9).

We start once more with the charge sector $\mathcal{N} = 1$, where a single electron in DQD is distributed between two wells. It is well-known that the electron in a double well may be described by means of a pseudospin operator $\boldsymbol{T}$ with the following components:

$$T^z = \sum_\sigma \left(d^\dagger_{2\sigma} d_{2\sigma} - d^\dagger_{1\sigma} d_{1\sigma} \right) \ , \ T^+ = \sum_\sigma d^\dagger_{2\sigma} d_{1\sigma} \ , \ T^- = \sum_\sigma d^\dagger_{1\sigma} d_{2\sigma} \ . \quad (10)$$

This vector, together with four spin vectors $\boldsymbol{S}_{ij}$ with the following components,

$$S^z_{ij} = \frac{1}{2} \left(d^\dagger_{i\uparrow} d_{j\uparrow} - d^\dagger_{i\downarrow} d_{j\downarrow} \right) \ , \ S^+_{ij} = d^\dagger_{i\uparrow} d_{j\downarrow} \ , \ S^-_{ij} = d^\dagger_{i\downarrow} d_{j\uparrow} \quad (11)$$

form the set of 15 generators for the $SU(4)$ group.

To close the basis for the effective spin Hamiltonian, one has to introduce similar vector operators for the electrons in the leads, namely, the pseudospin operator $\boldsymbol{t}$ with components

$$t^z = \sum_{kk',\sigma} \left(c^\dagger_{2k\sigma} c_{2k'\sigma} - c^\dagger_{1k\sigma} c_{1k'\sigma} \right) \ , \ t^+ = \sum_{kk',\sigma} c^\dagger_{2k\sigma} c_{1k'\sigma} \ , \ t^- = \sum_{kk',\sigma} c^\dagger_{1k\sigma} c_{2k'\sigma} \quad (12)$$

and four spin operators $\boldsymbol{s}_{ij}$ with components

$$s^z_{ij} = \sum_{kk'} \left(c^\dagger_{ik\uparrow} c_{jk'\uparrow} - c^\dagger_{ik\downarrow} c_{jk'\downarrow} \right) \ , \ s^+_{ij} = \sum_{kk'} c^\dagger_{ik\uparrow} c_{jk'\downarrow} \ , \ s^-_{ij} = \sum_{kk'} c^\dagger_{ik\downarrow} c_{jk'\uparrow} \ . \quad (13)$$

The SW transformation carried out in terms of these operators results in the effective Hamiltonian,

$$H_{\rm eff} = H_{\rm band} + H_{\rm dot} + 2\sum_{ij} J_{ij} \boldsymbol{S}_{ij} \cdot \boldsymbol{s}_{ji} + 2K\boldsymbol{T} \cdot \boldsymbol{t} \; . \tag{14}$$

In the fully symmetric case which we describe here, all the effective exchange constants have the same value $J_{ij} = K \equiv J$. Then the Hamiltonian (14) may be reduced to a more compact and familiar form of exchange Hamiltonian in a fictitious magnetic field [74]

$$H_{\rm eff} = H_{\rm band} + J\boldsymbol{S} \cdot \boldsymbol{s} - \widetilde{\boldsymbol{B}} \cdot \boldsymbol{s} \; , \tag{15}$$

where

$$\boldsymbol{S} = \frac{1}{2} \sum_{\alpha\beta} f^{\dagger}_{\alpha} \hat{\Sigma}_{\alpha\beta} f_{\beta} \; , \;\; \boldsymbol{s} = \frac{1}{2} \sum_{kk'} \sum_{\alpha\beta} c^{\dagger}_{\alpha} \hat{\Sigma}_{\alpha\beta} c_{\beta} \; . \tag{16}$$

Spin fermion operators f_{α} and the matrix $\hat{\Sigma}$ involve 15 components, which are generators of the Lie algebra $SU(4)$, determined as

$$\{(\tau^{+}, \tau^{-}, \tau^{z}, I) \otimes \sigma^{+}, \sigma^{-}, \sigma^{z}, I\} - \{I \otimes I\}$$

where $\sigma^{\nu}(\tau^{\nu})$ are the Pauli matrices in the spin (pseudospin) space and I is the unit matrix. The fictitious magnetic field has only one non-zero component, namely the $\tau_z \otimes I$ component, and its magnitude is W, so that $\widetilde{\boldsymbol{B}} \cdot \boldsymbol{s} = WT^{z}$. Thus one concludes that in DQD with odd occupation the formally multichannel Kondo Hamiltonian (14) may be mapped onto the standard Hamiltonian (15) in the case of complete channel degeneracy, and the contribution of additional (permutation) degrees of freedom is described in this degenerate case by a fictitious magnetic field

In the charge sector $\mathcal{N} = 2$ the permutation symmetry degenerates into trivial unit transformation. As to the spin degrees of freedom, one may try to described them by means of two spin operators $\boldsymbol{s}_l$ and $\boldsymbol{s}_r$ using the above mentioned analogy with the two-site Kondo problem [75]. However, such approach [77] should be used with some caution. One should take into account that in a situation, where both triplet and singlet two-electron states are involved in Kondo effect, these spins are non-independent because the kinematical constraint is imposed on the S/T manifold by the Casimir operator $\mathcal{C} \neq \boldsymbol{s}_l^2 + \boldsymbol{s}_r^2$. In accordance with the prescriptions of the theory of dynamical symmetries [62–64], one should construct two operators

$$\boldsymbol{S} = \boldsymbol{s}_l + \boldsymbol{s}_r \; , \;\; \boldsymbol{R} = \boldsymbol{s}_l - \boldsymbol{s}_r \tag{17}$$

and impose on them the kinematic constraint

$$\mathcal{C} = \boldsymbol{S}^2 + \boldsymbol{R}^2 = 3 \; . \tag{18}$$

Then the three components (S^z, S^+, S^-) of the vector $\boldsymbol{S}$ describe the states within the spin triplet and transitions between them, whereas the three components (R^z, R^+, R^-) of the vector $\boldsymbol{R}$ describe transitions between the singlet S and the states with spin projections $\mu = 1, 0, -1$ of the triplet T. Six components of the vectors $\boldsymbol{S}$, $\boldsymbol{R}$ form a closed algebra

$$[S_\alpha, S_\beta] = \mathrm{i}e_{\alpha\beta\gamma}S_\gamma \ , \ [R_\alpha, R_\beta] = \mathrm{i}e_{\alpha\beta\gamma}S_\gamma \ , \ [R_\alpha, S_\beta] = \mathrm{i}e_{\alpha\beta\gamma}R_\gamma \ . \quad (19)$$

and form a set of generators of $SO(4)$ group. Here α, β, γ are Cartesian coordinate indices, and $e_{\alpha\beta\gamma}$ is the anti-symmetric Levi-Civita tensor. Two vector operators are orthogonal, $\boldsymbol{S} \cdot \boldsymbol{R} = 0$. Under these constraints, two vectors $\boldsymbol{s}_{l,r}$ are rather fictitious than real spin operators. More detailed discussion of interconnections between two representations as well as the derivation of these operators by means of the Hubbard operators $X^{\Lambda\Lambda'}$ may be found in [43]. If one tunneling channel couples this "spin rotator" with the reservoir of conduction electrons, then the dynamical group $SO(4)$ exhausts the spin degrees of freedom involved in Kondo tunneling. This scenario is realized in the T-shape geometry of Fig. 1c (see below).

More complicated is the mapping procedure for the two-channel Anderson Hamiltonian describing the DQD shown in Fig 1b. In this case the discrete symmetry is explicitly involved in the cotunneling process, so that the SW transformation give the exchange part of H_{eff} in the form

$$H_{\text{ex}} = 2\sum_{ij} J_{ij}\boldsymbol{S}_{ij} \cdot \boldsymbol{s}_{ji} + 2\sum_{ij} \tilde{J}_{ij}\boldsymbol{R}_{ij} \cdot \boldsymbol{s}_{ji} + 2K\boldsymbol{T} \cdot \boldsymbol{t} \ . \quad (20)$$

DQD with l-r reflection axis, where all the states are classified as even or odd states relative to this reflection, is described by this Hamiltonian may be qualified as a "double spin rotator" [78]. The corresponding dynamical symmetry group is $P_2 \times SO(4) \times SO(4)$.

Before turning to the physical aspects of Kondo tunneling through the objects of $SU(4)$ and $SO(4)$ symmetries, one should mention that the DQD in lateral geometry with two channels is described by the same basic two-level Anderson Hamiltonian as the single planar quantum dot with two levels, one of which is occupied and another is empty. The latter model was considered in many publications [31–33, 76, 77, 79–81] (see also the review papers [74, 82]). So the physical manifestations of "variable" symmetry are common for both systems. Fig. 3 illustrates the variation of T_K as a function of control parameter (the tunnel splitting W in case of $\mathcal{N} = 1$ and the exchange gap Δ_{ex} in case of $\mathcal{N} = 2$). In both cases the maximum value of T_K correspond to the degeneracy points.

This behavior may be easily understood within a perturbative (high-temperature) renormalization group (RG) approach. This approach is based on the study of flow diagrams describing the evolution of effective coupling parameters $J_a(\eta)$ due to reduction of the effective energy scale D of conduction electron kinetic energy [83] (label a enumerates the vertices in the exchange

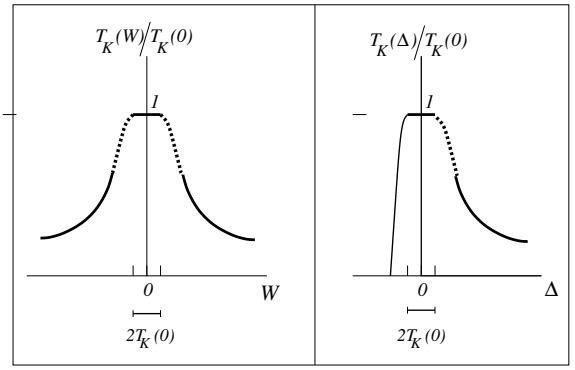

Fig. 3. Dependence of the Kondo temperature on the control parameter W and $\Delta = -\Delta_{\rm ex}$ for odd occupation (left panel) and even occupation (right panel)

Hamiltonian, $\eta = \ln D$ is the scaling variable). The general form of the system of scaling equations is

$$\frac{{\rm d}j_a}{{\rm d}\eta} = -\sum_b c_{ab} j_a j_b \ . \tag{21}$$

Here $j_a = \rho J_a$ are dimensionless coupling constants and c_{ab} are numerical coefficients. These equations should be solved under the boundary conditions $j(D_0) = \rho J_{a0}$, where D_0 and J_{a0} are the initial conduction bandwidth and the bare exchange integrals entering the Hamiltonian $H_{\rm ex}$. One may neglect the splitting of energy terms in the system (21) when $D \sim D_0 \gg (W, \Delta_{\rm ex})$ and all coupling constants J_a evolve together. When the scale D is reduced down to the splitting energy W or $\Delta_{\rm ex}$, the coupling parameters related to the upper level in the multiplets (7) or (9) stop to evolve, because the corresponding degrees of freedom are quenched below these energies. These parameters are K both in the Hamiltonians (14) and (20), and the exchange vertices $\tilde{J}_{ij}$ are related to the singlet-triplet transitions in (20). In a degenerate model (15), T_K is partially suppressed by the fictitious magnetic field ($\widetilde{\boldsymbol{B}}$). As a result the temperature T_K obtained as a solution of these equations becomes explicit function of the splitting energy. Its behavior is illustrated in the left and right panels of Fig. 3 for the charge sectors $\mathcal{N} = 1, 2$, respectively. In case of odd occupation $\mathcal{N} = 1$ the point $W = 0$ corresponds to two equivalent dots decoupled from each other. The system remains quasi-degenerate till $W < T_K(0) = D\exp(4/\rho J_0)$. With further increase of the level splitting, the contribution of the state E_2 in (7) to the Kondo tunneling diminishes. In the asymptotic regime $|W|/T_K(0) \gg 1$, the evolution of T_K is described by the asymptotic equation [33, 74]

$$T_K(W) = \frac{T_K(0)^{\gamma+1}}{|W|^\gamma} \ . \tag{22}$$

This asymptotic curves are shown bold in Fig. 3. The exponent depends on the model parameter and detailed geometry of quantum dot [33, 74, 77, 78].

Similar asymptotic curve describes also the evolution of T_K for even occupation $\mathcal{N} = 2$ with $|W|$ substituted for $\Delta_{\rm ex}$. But in this case such behavior is characteristic only for $\Delta_{\rm ex} < 0$, where the ground state is spin triplet. T_K is maximal in the "critical point" of S/T crossover. On the triplet side of S/T transition, it diminishes due to gradual quenching of triplet-singlet excitations described by the operator $\mathbf{R}$ in the Hamiltonian (20). On the singlet side of S/T crossover, the Kondo temperature falls down steeply as $\Delta_{\rm ex} > 0$. The singlet-triplet crossover driven by some external parameter (magnetic field, gate voltage, etc.) in DQD with even occupation is the most salient effect predicted and observed in these systems.

Additional information unveiling specific features of Kondo screening in complex quantum dots can be extracted from the temperature and magnetic field dependence of tunnel conductance $G_{\max}$ at given W or $\Delta_{\rm ex}$. Its behavior was discussed time and again [71, 76, 77, 81, 82]. The most interesting is the behavior of conductance in the charge sector $\mathcal{N} = 2$.

On the triplet side of S/T crossover $G_{\max}(T)$ increases with decreasing T. At high $T \gg T_K(|\Delta_{\rm ex}|)$, it grows logarithmically,

$$G_{\max}(T)/G_0 \sim \ln^{-2}(T/T_K) \tag{23}$$

where $G_0 = 2e^2/h$ is the limiting value of tunnel conductance in a Kondo regime (unitarity limit, where the sum of all phase shifts on the Fermi levels $\delta_{i\sigma}$ equals $\pi/2$). This limiting value quantify the tunnel conductance at $T = 0$. According to the Friedel sum rule [8, 60], generalized for the two-channel geometry [79, 80] one has

$$G(T=0) = G_0 \sin^2\left[\frac{\pi}{2}(\bar{n}_2 - \bar{n}_1)\right] \tag{24}$$

where $\bar{n}_{(2,1)}$ is the change in number of electrons under the Fermi level due to Kondo screening in two channels (even and odd). In terms of the phase shifts the argument of the sine function in the r.h.s. of (24) reads $\frac{1}{2}\sum_{i\sigma}\delta_{i\sigma}$. This means that the demand of spin rotation invariance and l-r symmetry means that the phase shift in each channel reaches $\pi/4$ at $T = 0$. According to the theory of Kondo scattering [16], the deviation from the unitarity limit is of standard Fermi-liquid character

$$G(T)/G_0 = \left[1 - \pi\,(T/T_K)^2\right] \tag{25}$$

at $T \ll T_K$. The monotonic interpolation [82] between two limiting temperature regimes (23) and (25) may be violated due to multistage Kondo effect. We will discuss this regime, when considering the case of triple quantum dots (see below).

The influence of external magnetic field B on the tunnel conductance at low $T \ll B$ is easily tractable [82]: in accordance with the general theory of

Kondo effect in presence of Zeeman splitting of the levels in the dot, in this regime T should be substituted for B in the asymptotic equations (23) and (25). At low $T \ll T_K$ additional information may be obtained by means of the NRG method [81]. Due to the loss of spin rotation symmetry, the phase shifts $\delta_{i\sigma}$ become explicit functions of magnetic field. In a single-channel case this dependence is scaled by T_K, namely $\delta_\sigma(B) = \delta(0) + \sigma(B/T_K)$. In a two-channel regime, level splitting W between the dot levels enters this dependence explicitly. A simple equation describing this dependence was derived analytically [80],

$$\delta_{i\sigma} = \delta(0) + \sigma(B/T_i) + (-1)^i(B/T_i') + O(B^3) \; . \tag{26}$$

Here the parameters T_i and T_i', which scale the field B depend on W. NRG calculations [81] show great variety of magnitoconductance curves $G(B, W)$.

On the singlet side of S/T crossover the ground state of DQD is non-magnetic, so that there is no room for Kondo-type ZBA at $T = 0$ (see the right panel of Fig. 3). However, at $T > \Delta_{\rm ex}$ the triplet state is still involved in perturbative scaling (21), so that $G(T)$ grows with decreasing T in accordance with (23). The T/S crossover occurs at $T \sim \Delta_{\rm ex}$, and conductance starts to fall with further decrease of T ending with exponentially small value of $G(T = 0)$ [76]. Similar effect should be observed in the behavior of G as a function of source-drain voltage v_{ds}. At $ev_{ds} > \Delta_{\rm ex}$ the DQD shows Kondo tunneling, but with lowering bias the triplet channel is quenched, and conductance shows up a zero field dip instead of zero field peak. Apparently this type of crossover was observed [84] in a two-orbital planar dot with even occupation.

T-Shape Geometry

In a T-shape geometry one of two dots is detached from the leads (Fig. 1c). In the first experimental device of this type [38], the role of the side (right) dot was to control single-electron tunneling through the left dot. Recently several new effects related to the Kondo-regime were discovered in this geometry.

First, it was shown [43] that the T-shape double dot with even occupation $\mathcal{N} = 2$ demonstrates the properties of spin rotator with $SO(4)$ symmetry. The simplest form of the Hamiltonian (20), namely

$$H_{\rm ex} = 2J\boldsymbol{S} \cdot \boldsymbol{s} + 2\tilde{J}\boldsymbol{R} \cdot \boldsymbol{s} \tag{27}$$

was derived just for this model. Besides, it was shown that in the asymmetric T-shape DQD, where the Coulomb blockade in the right dot is sufficiently larger than in the left dot, the S/T crossover may occur because of the many-body logarithmic renormalization [85, 86], which is determined by the renormalization group invariants E_Λ^*, namely

$$E_\Lambda^* = E_\Lambda(D) - \pi^{-1}\Gamma_\Lambda \ln(\pi D/\Gamma_\Lambda) \; . \tag{28}$$

The level crossing is possible because the inequality $\Gamma_T > \Gamma_S$ for the tunneling rates Γ_Λ is realized for asymmetric DQD, so that the renormalization of the triplet level E_T is stronger than that of the singlet state E_S. Similar effect may be achieved in a symmetric DQD by means of the gate voltage applied to the side dot (later on the possibility of inducing the S/T crossover by means of the gate voltage was found also for two-orbital planar dots [87]).

Another interesting possibility arises at the odd occupation $\mathcal{N} = 1$, where the strong Coulomb blockade exists only in the side dot [88–91]. It was found that in this case the Kondo resonance arises on the background of otherwise non-correlated transport between the electrodes via the right dot. The interference between the resonance scattering and free propagation is known as Fano effect discovered originally in the optical absorption spectra of free atoms [92]. The Fano effect in atoms is described by the same Hamiltonian as the Anderson impurity hybridization [9], so there is no wonder that similar effect was found in the resonance impurity scattering in metals [93]. Since the single electron transport through the quantum dot is also described by the Anderson Hamiltonian (2), the Fano effect ubiquitous in resonance scattering was expected in tunnel conductance of quantum dots, and the structures of characteristic Fano-type form were indeed observed in the tunneling spectra of planar quantum dot [94]. In the latter case the Fano effect arises due to the interference between the resonance level in the quantum dot and the band continuum in the leads. The corresponding contribution to the conductance has the form

$$G_{\rm Fano}(\epsilon) = G_0 \frac{(\varepsilon + q)^2}{\varepsilon^2 + 1} \,, \tag{29}$$

where $\varepsilon = 2(\epsilon - \epsilon_d))/\Gamma$, q is the so-called asymmetry parameter predetermined by the spectral characteristics of the lead electrons, $\Gamma = \pi\rho V^2$ is the tunneling rate. In simple terms, Fano effect is nothing but modification of the spectral density of conduction electrons due to its repulsion from the resonance level superimposed on the continuous spectrum.

In a T-shape geometry, the Fano effect arises due to superposition of the *Abrikosov-Suhl resonance* created by the Kondo effect in the right dot on the continuous tunneling spectra of the system 'source – left dot – drain'. As was noticed by Kang et al. [88], this Fano-Kondo effect looks as an anti-resonance in conductance: instead of the standard Friedel-Langer formula (24), one has

$$G(T = 0) = G_0 \cos^2 \left[\frac{\pi}{2} \bar{n}_r \right] \,, \tag{30}$$

where n_r stands for the average occupation number of the right dot. Such "unitarity limit" for conductance means that in the T-shape geometry, the Kondo effect in the right dot results in complete suppression of resonance tunneling through the left dot and a dip arises in the tunnel conductance instead of the usual peak. As was mentioned above, the phase shifts in the Kondo regime are spin dependent, so the modification of tunnel spectra due to

destructive Kondo-Fano interference is also spin-dependent, and the T-shape dot in this regime may work as a spin filter in external magnetic field [95].

One more advantage of the TQD geometry is the possibility to approach the two-channel Kondo effect [96, 97]. Such possibility arises in the geometry of Fig. 1c, when the Coulomb blockade is strong in the left dot and the right dot is big enough so that the level spacing is less than the tunnel rate, but the Coulomb blockade is still strong and fixes the electron occupation number. In such conditions the way opens to overcome the main difficulty in realizing a physical system that materializes the two-channel Kondo model. The necessary precondition for this regime is in creating two separate channels that equally screen the spin [15]. In conventional setups an electron from one channel that hops on the dot may hop to the other channel and thus cause mixing between the channels. This mixing lead eventually to two "eigen channels" with one channel coupled stronger than the other one. The channel with the stronger coupling fully screen the spin and the other channel is decoupled, and we thus have again the single-channel Kondo case. It was suggested [96] to overcome this mixing problem by using a large quantum dot as an additional channel. Then, the free leads form one channel [even one in accordance with our classification (6)] and the large dot forms the second channel. The channels do not mix as transfer of electrons between them charges the large dot. As a result, the SW transformation maps the original Anderson Hamiltonian for a T-shape quantum dot onto the two-channel exchange Hamiltonian

$$H_{2\mathrm{ch}} = \sum_{\gamma,k} \epsilon_{\gamma,k} c^{\dagger}_{\gamma,k} c_{\gamma,k} + \Sigma_{\gamma} J_{\gamma} \boldsymbol{S} \cdot \boldsymbol{s}_{\gamma} + BS^{z} \ . \tag{31}$$

Here the channel index γ stands for the even lead-dot channel and the states in the large dot r. The tunnel conductance in the 2-channel non-Fermi-liquid regime is realized at $J_2 = J_r$. It demonstrates specific temperature dependence. In accordance with predictions of conformal field theory [97, 98],

$$G_{\mathrm{nfl}}(T) = \frac{G_0}{2}\left(1 - \sqrt{\pi T/T_K}\right) \ . \tag{32}$$

Practically, the zero temperature limit cannot be achieved because the Kondo screening in the large dot is quenched due to discreteness of its electron spectrum at $T \sim \delta\epsilon$.

Unlike the single-channel case, the magnetic field is a relevant parameter in the two-channel case [99]. Introducing the channel anisotropy parameter $\Delta_{\mathrm{ch}} = J_r - J_2 \ll J_2$, which describes deviation from the 2-channel fixed point ($\Delta_{\mathrm{ch}} = 0$), the following equation for the magnitoconductance may be derived [97, 99]

$$G_{\mathrm{nfl}}(T, B) = G_0 \left[\frac{1}{2} + a\,\mathrm{sign}(\Delta_{\mathrm{ch}})\frac{B_{\Delta}}{B} - b\,\frac{B}{T_K}\ln\frac{T_K}{B}\right] \tag{33}$$

where $B_{\Delta} = \Delta_{\mathrm{ch}} T_K / J_2^2$, and a, b are positive numerical coefficients of the order of 1.

3.2 Triple Quantum Dots

In this section we consider new features of Kondo mapping which are characteristic for the chains consisting of three dots (Figs. 2a-c). Simple increase of the number of sites in the chain promises no new qualitative results, so we will discuss here the configurations where the central dot differs from two side dots in its size (and hereby by the magnitude of the Coulomb blockade parameter Q_c), but the latter dots are identical, so that the TQD retains its reflection symmetries. Both conceivable situations, namely $Q_c \gg Q_s$ and $Q_c \ll Q_s$ will be considered (the indices c, s are used in this section to indicate the physical quantities related to the central and side dots, respectively). In the first experimental realization of TQD [44] the former option was chosen, whereas the first theoretical study [78] was devoted to the second possibility.

In a TQD with "open" central dot and $Q_c \ll Q_s$, its role in formation of Kondo tunneling regime reduces to providing the channel for indirect RKKY-type mechanism of exchange between two localized spins formed in side dots. Thus, from the theoretical point of view this problem is in fact may be mapped onto that for a DQD (see Sect. 3.1) with specific exchange mechanism. The Kondo effect may be described in terms of two-site Kondo Hamiltonian, where the trend to interdot spin coupling competes with the trend to individual Kondo coupling between two side dots and the electrons in the leads [100,101]. One may mention in this connection that the phase diagram of effective two-site Kondo effect has been calculated in a framework of the model, where the pair of spins is coupled to the linear electron chain in a side geometry [102].

Qualitatively new features of Kondo mapping problem were found out in the opposite limiting case $Q_c \gg Q_s$. As follows from the general scheme of Kondo-mapping based on the dynamical symmetry of artificial molecule (Sect. 3.1), the form of effective exchange Hamiltonian (20) depends on the structure of spin multiplet of isolated multivalley quantum dot. Up to now only two possibilities were exploited: spin and orbital doublet for odd occupation $\mathcal{N} = 1$ and singlet-triplet pair for even occupation $\mathcal{N} = 2$. Study of TQD with odd and even occupations $\mathcal{N} = 3$ and $\mathcal{N} = 4$ give us new opportunities [78, 103].

To demonstrate these opportunities let us consider TQD in lateral geometry. Genesis of spin multiplets may be understood from a general set-up illustrated by Fig. 4. In case of $\mathcal{N} = 3$ three electrons are distributed over three dots in such a way that the state with doubly occupied central dot is suppressed by strong Coulomb blockade Q_c. In accordance with the Young tableaux for a system with l-r permutation symmetry, the spin multiplet consists of two doublets with $S = 1/2$ having even and odd symmetry relative to this permutation and one quartet $S = 3/2$ with full orbital symmetry. In case of $\mathcal{N} = 4$ the spin manifold consists of two spin singlets $S_{e,o}$ and two triplets $T_{e,o}$, both even and odd relative to l-r permutation. Varying the gate voltages v_{gl}, v_{gr} and playing with tunnel parameters $V_{l,r}$ and $W_{l,r}$, one may break l-r symmetry (Fig. 4) and change the singlet-triplet splitting, so that the spin

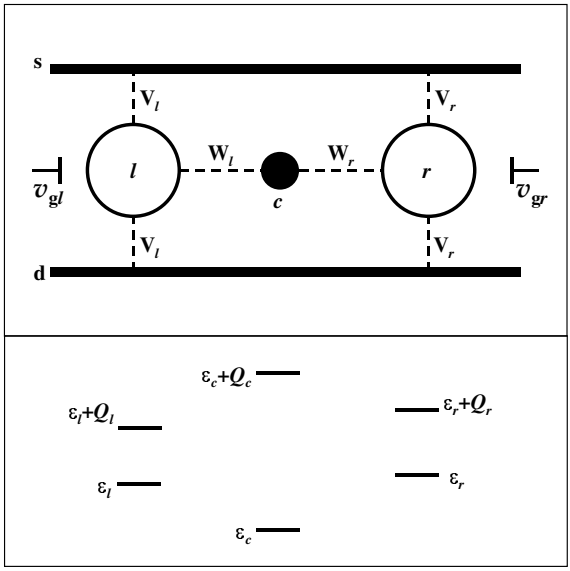

Fig. 4. TQD in parallel geometry and energy levels of each dot $\varepsilon_a - ev_{ga}$ at $W_a = 0$

states are classified as $S_{l,r}$ and $T_{l,r}$. The relative positions of energy levels in spin multiplets evolve as a function of model parameters and various types of level crossings occur (see [78] for detailed calculations). Similar situation arises for vertical configuration of Fig. 2a.

In accordance with general theory of dynamical symmetries [64], quasi-degeneracy of low-lying states in spin multiplets within the energy scale $\sim T_K$ generates special symmetries of TQD. For example, if the multiplet of low-lying states consists of two singlets and one triplet, the relevant dynamical symmetry is $SO(5)$. If this multiplet is formed by two triplets and one singlet, the corresponding symmetry is $SO(7)$, etc. The methods of constructing the generators for these groups are described in details in the reviews [64, 65]. As a result, an unique opportunity arises to change the value of index n characterizing the symmetry $SO(n)$ of TQD by varying the gate voltages and other experimentally controllable parameters of a device. The phase diagram of vertical TQD with $\mathcal{N} = 4$ calculated in [78] is presented in Fig. 5. This diagram shows great variety of phases with different symmetries from the most symmetric one $P_2 \otimes SO(4) \otimes SO(4)$ to conventional $SO(3)$ phase where the ground state of TQD is spin triplet, or non-Kondo singlet ground state (shaded areas). Each phase is characterized by its own T_K, and this means that the ZBA in conductance should follow the change of the Kondo temperature, so that each crossover from one symmetry to another is accompanied by the abrupt change of conductance at given temperature.

At $\mathcal{N} = 3$ we meet a somewhat unexpected situation where Kondo tunneling in a quantum dot with *odd* occupation demonstrates the exchange Hamil-

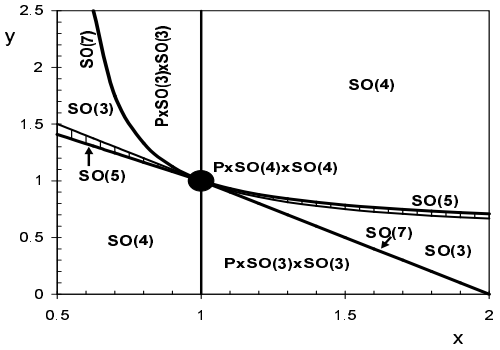

Fig. 5. Dynamic symmetries in TQD. Phase diagram in the coordinates $x = \Gamma_l/\Gamma_r$ and $y = (\varepsilon_l - \varepsilon_c)/(\varepsilon_r - \varepsilon_c)$

tonian of a quantum dot with *even* occupation. The reason for this scenario is the specific structure of the wave function of TQD with $N = 3$. The corresponding wave functions are vector sums of states composed of a "passive" electron sitting in the central dot and singlet/triplet (S/T) two-electron states in the l, r dots. Then using certain Young tableaux [78], one concludes that the spin dynamics of such TQD is represented by the spin 1 operator $\boldsymbol{S}$ corresponding to the $l - r$ triplet, the corresponding R-operator $\boldsymbol{R}$ and the spin 1/2 operator $\boldsymbol{s}_c$ of a passive electron in the central well. The latter does not enter the effective Hamiltonian $H_{\rm ex}$ but influences the kinematic constraint via Casimir operator $\mathcal{C} = \boldsymbol{S}^2 + \boldsymbol{R}^2 + \boldsymbol{s}_c^2 = 15/4$. The dynamical symmetry is therefore $SO(4) \otimes SU(2)$, and only the $SO(4)$ subgroup is involved in Kondo tunneling. Similar situation, although for different reasons is realized in fork and cross geometries (see below).

Remarkable symmetry reduction occurs in external magnetic field [78]. First example of such reduction was found in a situation where the exchange splitting of S/T multiplet (symmetry group $SO(4)$) is compensated by the Zeeman splitting [31], so that the up spin projection $|T1\rangle$ of triplet forms a pseudospin with singlet $|S\rangle$ and the symmetry reduction $SO(4) \to SU(2)$ takes place. In case of TQD with $SO(5)$ symmetry, due to the same compensation the system may be left in a subspace $\{T1_l, S_l, S_r\}$. The symmetry reduction in this case is $SO(5) \to SU(3)$, and the Anderson Hamiltonian is mapped on a very specific *anisotropic* Kondo Hamiltonian involving only operators $\boldsymbol{R}_i$,

$$H_{\rm ex} = \sum_{ij} \sum_{\mu\nu} J_{ij}^{\mu\nu} R_i^\mu s_j^\nu \tag{34}$$

where μ, ν are Cartesian components of scalar product. Here the Kondo effect is described exclusively in terms of dynamical symmetry.

Another non-standard manifestation of Kondo mapping for linear TQD is the possibility of two-channel Kondo effect in vertical geometry of Fig. 2a with $Q_c \gg Q_s$ at $\mathcal{N} = 3$ with preserved s-d mirror symmetry [103]. The

strong Coulomb blockade in central dot prevents direct s-d tunneling. The cotunneling is possible only because the wave functions of electrons centered on the side dots have small tails on the central dot. It is crucially important that the standard rotation (6) does not eliminate the odd channel from the tunneling Hamiltonian in TQD. In the situation, where the ground state of TQD is the spin doublet with even parity $|D_e\rangle$, the SW transformation for the original Anderson Hamiltonian results in anisotropic two-channel exchange Hamiltonian,

$$H_{\rm ex} = J_s \boldsymbol{S}\cdot\boldsymbol{s}_s + J_d \boldsymbol{S}\cdot\boldsymbol{s}_d + J_{sd}\boldsymbol{S}\cdot(\boldsymbol{s}_{sd}+\boldsymbol{s}_{ds})\ . \tag{35}$$

Due to the presence of non-diagonal vertex J_{sd} the incurable orbital anisotropy arises: the tunneling through two channels is controlled by the parameters $J_\pm = (J_s + J_d)/2 \pm \sqrt{(J_s - J_d)^2/4 + J_{sd}^2}$. In accordance with the theory of two-channel Kondo effect [15], this anisotropy makes the 2-channel fixed point unattainable, but due to strong Coulomb blockade in central dot $J_{sd} =\sim V^2W^2/Q_c\varepsilon_d^2$ is extremely small, so that one may approach the fixed point close enough, and the predecessor of 2-channel regime may be observed experimentally as a dip in conductance G as a function of the difference of gate voltages $v_{gs} - v_{gd}$ applied to the side dots (this difference controls the degree of channel anisotropy).

TQD in Cross and Fork Geometry

To complete the studies of linear artificial molecules we consider in this section the configurations shown in Fig. 2c and Fig. 2f. We have seen above that one may meet the situation, where the linear TQD with odd occupation and half-integer spin demonstrates the Kondo physics characteristic for even occupation with integer spin due to the fact that one of the electrons in the dot does not participate in tunneling. Here we will discuss two more mechanisms of such "disguise" [104].

One of these mechanisms is realized in *cross geometry* at occupation $\mathcal{N} = 3$ under condition $Q_s \gg Q_d$ for Coulomb blockade parameters. In this case two side electrons are passive: the tunneling between source and drain occurs through the central dot. However, these passive electrons influence Kondo mapping because they are responsible for the *parity* of the 3-electron wave function relative to the l-r reflection. Diagonalization of the low-energy spin states shows that it consists of three spin doublets and one spin quartet, and the lowest state in this manifold is the doublet D_u, which is odd relative to the mirror reflection (see [104] for details). Although the wave functions of two passive electrons do not enter explicitly in the indirect exchange integral arising due to SW transformation, this integral changes its sign due to odd parity of the state D_u. Thus, in contrast to the standard paradigm of Kondo mapping, the effective exchange Hamiltonian corresponds to ferromagnetic coupling, which is irrelevant to Kondo effect, and the Kondo-type ZBA does

not arise in this case in spite of the fact that the net spin of quantum dot is 1/2.

However, this is not the end of the story. The excited states in the spin multiplet which are Kondo active, influence the tunnel transparency and conductance at finite temperature and finite energies of incident electron due to non-trivial dynamical symmetry of TQD described above. The states involved in the Kondo effect are the even spin doublet D_g and the quartet Q, so that the overall dynamical symmetry of TQD in cross geometry is $SU(2) \otimes SU(2) \otimes SU(2)$. We meet here the situation, which reminds the "two-stage" Kondo effect in DQD with $\mathcal{N} = 2$, on the singlet side of S/T crossover (see Fig. 3b and subsequent discussion). Here, however, there are three stages of Kondo screening, where the states Q and D_g are quenched one after another with decreasing energy or temperature and the Kondo screening eventually stops at zero T. Besides, the hierarchy of tunneling rates $\Gamma_Q > \Gamma_{Dg} > \Gamma_{Du}$ exists in this charge sector, so the level crossing controlled by the parameters of TQD is possible in accordance with (28). This level crossing is shown in Fig. 6. Here the scaling variable is chosen in the form $\eta = \ln(\pi D/\Gamma_Q)$. The value of $\bar{D}$ is determined from the crossover condition $\bar{D}_\Lambda \approx E_\Lambda(\bar{D}_\Lambda)$, where the renormalization (28) changes for the SW regime with fixed charge, $\bar{D}_0$ is the initial value of scaling variable.

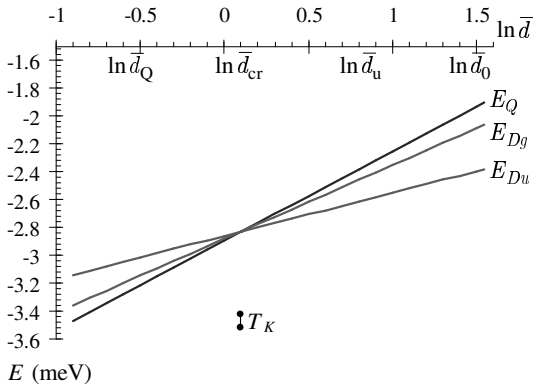

Fig. 6. Flow diagram for the levels E_Λ determined by the scaling invariant (28). $\Lambda = D_u,\ D_g,\ Q$, and $\bar{d} = \pi\bar{D}/\Gamma_Q$. Energy is measured in meV units (see text for further explanations)

Three points ($\ln \bar{D}_u,\ \ln \bar{D}_Q,\ \ln \bar{D}_{cr}$,) on the abscissa axis correspond to three values of the control parameters where the crossover to the SW regime occurs for the ground states E_{D_u}, E_Q and the completely degenerate ground state, respectively. By means of appropriate variation of the control parameters, the system may be transformed from a non-Kondo regime with the ground state E_{Du} to the underscreened Kondo regime with the ground state

E_Q and spin $S = 3/2$. In accordance with the general theory of Kondo mapping (Sect. 3.1), T_K is maximum in the point of maximum degeneracy. Evolution of T_K is shown in Fig. 7, which should be compared with Fig. 3b. Unlike

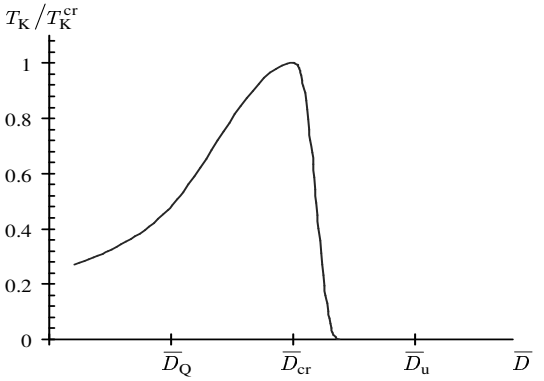

Fig. 7. Evolution of T_K as a function of a control parameter $\bar{D}$

the case of singlet/triplet crossover in DQD with $\mathcal{N} = 2$, here one deals with the crossover from the non-Kondo spin doublet D_u to the Kondo spin quartet Q via the highly degenerate region of $SU(2) \otimes SU(2) \otimes SU(2)$ symmetry.

Let us turn to the *fork geometry* shown in Fig. 2f. In this geometry nontrivial Kondo physics arises already in the simplest case of odd occupation $\mathcal{N} = 1$ in a situation with the *l*-*r* mirror symmetry. The fork may be considered as a "quantum pendulum" [49, 105]. Three one-electron eigenvalues are

$$E_{D_{b,a}} = \epsilon_c \mp 2W^2/\Delta \ , \ E_{D_n} = \epsilon_s \ , \tag{36}$$

($\Delta = \max\{|\epsilon_s - \epsilon_c|, |Q_c - Q_s|\}$). The eigenstates are classified as a non-bonding spin doublet D_n (odd combination of the wave functions centered in the sites 1,2) and bonding/antibonding pair $D_{b,a}$ of corresponding even combination with the state centered in the site 3 (see [104] for details). The latter pair is the analog of resonant valence bonds (RVB) known in "natural" molecules. To describe this pendulum one should introduce the pseudospin vector $\boldsymbol{T}$ defined in (10) and work with the Hamiltonian (14). Like in the cross geometry, the level crossing effect as a function of control parameter takes place (Fig. 8).

Here T_K is non-zero on both sides of the crossover and its evolution is described by the bell-like curve similar to that in Fig. 3a (although slightly asymmetric). However the tunnel conductance is drastically influenced by the pendulum structure of the electron wave function. In the three-terminal fork geometry, one should consider separately the situations, where the bias voltage is applied between the leads 1 and 2 and between the leads 1 and 3. We define the corresponding components of tunnel conductance as G_{22} and G_{33}, respectively. The Kondo anomaly in G_{22} is predetermined by the RVB

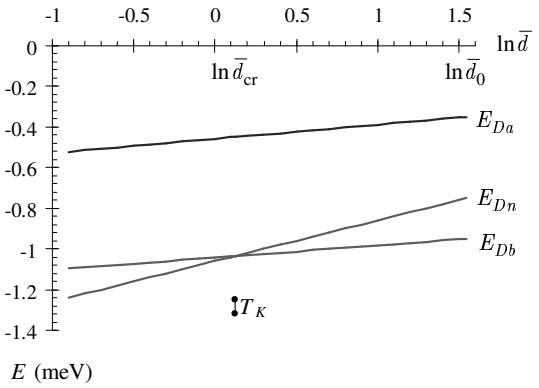

Fig. 8. Flow diagram for the levels E_Λ of the TQD in fork geometry

pair, and the ZBA roughly follows the evolution of T_K through the crossover. More peculiar behavior is expected in 1-3 channel because the non-bonding state $D_{n\sigma} = 2^{-1/2}(d^\dagger_{1\sigma} - d^\dagger_{2\sigma})$ is detached from the lead 3. As a result the Kondo contribution to G_{33} manifests itself as a finite bias anomaly (FBA) in a situation where the ground state of TQD is E_{D_n}. Tunnel conductance as a function of bias voltage in both channels is illustrated in Fig. 9. The dip in the curve b on the right panel reminds similar dip in the tunnel conductance of DQD with $\mathcal{N} = 2$ on the singlet side of S/T crossover [76, 84].

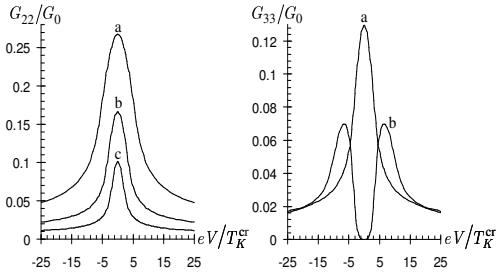

Fig. 9. Left panel: Tunnel conductance in the channel '2–1' for **(a)** $\bar{D} = \bar{D}_{cr}$, **(b)** $\bar{D} > \bar{D}_{cr}$ and **(c)** $\bar{D} < \bar{D}_{cr}$. Right panel: Tunnel conductance in the channel '3–1' for **(a)** $\bar{D} > \bar{D}_{cr}$ and **(b)** $\bar{D} < \bar{D}_{cr}$

Thus, we see that some manifestations of the Kondo effect in TQD with odd occupation may mimic those for DQD with even occupation due to specific influence of the mirror reflection on the structure of the electron wave functions in trimers.

4 Kondo Physics for Small Rings

In previous sections we discussed the Kondo effect in short "Hubbard chains" in contact with metallic reservoirs. Meanwhile, the Hubbard-like objects were studied also in closed ring geometries as well (see, e.g., [106,107]). Experimentally Kondo effect on closed rings was observed on gutter-like dots in planar geometry [108,109], but the impact to theoretical investigation of Kondo effect in ring-like nanoobjects [50] was given by experimental studies of Co trimers adsorbed on metallic sublayer [51]. It was shown that the basic symmetry of equilateral triangular triple quantum dot (TTQD, Fig. 2e) with odd occupation is $SU(4)$ due to the interplay between the spin and orbital degrees of freedom, similar to that in two-orbital DQD (see Sect. 3.1). Special attention was paid to the case $\mathcal{N} = 3$ which models triangular Co trimer [53,110]. In this charge sector the effective spin Hamiltonian contains not only exchange interaction between the spins in the dots and adjacent leads, but also the two-site Heisenberg exchange between spins in the neighboring dots. Magnetic frustrations in triangular geometry affect the spin state and therefore influences the Kondo-type ZBA in tunneling spectra. These spectra were calculated by the NRG and quantum Monte-Carlo methods. Besides, it was found [53] that in case of complete channel isotropy the non-Fermi-liquid regime arises from the interplay of magnetic frustrations and Kondo physics.

Another phenomenon, which interplays with the Kondo physics is the Aharonov-Bohm oscillation of tunnel transparency in magnetic field directed perpendicularly to the plane of triangle. This effect may be seen already for the Hubbard ring with 1/3 occupation (TTQD with $\mathcal{N} = 1$), where there is no room for exchange interaction between spins localized in neighboring sites and concomitant magnetic frustrations. The starting point for solving the problem of interplay between Kondo and Aharonov-Bohm phenomena [54, 111] is the Anderson Hamiltonian (2) rather than the exchange Hamiltonian (14).

In accordance with the general scheme discussed in Sect. 2, one should start with the diagonalization of the Hamiltonian of 3-site Hubbard ring. The point symmetry of this equilateral triangle is C_{3v}. This group describes the "orbital" degrees of freedom, whereas the continuous spin symmetry is usual $SU(2)$ symmetry of spin 1/2. According to the irreducible representations of C_{3v} group, the spectrum of TTQD consists of three levels $\Lambda = DA, DE_{\pm}$. Here as usual D stands for spin doublet, A is the fully symmetric orbital singlet and $E_{\pm}$ are two components of orbital doublet. The energies of these states in out-of-plane magnetic field B are

$$E_{D\Gamma}(p) = \epsilon - 2W \cos\left(p - \frac{\Phi}{3}\right) , \tag{37}$$

such that for negative W and for $B = 0$, $p = 0,\ 2\pi/3,\ 4\pi/3$ correspond respectively to $\Lambda = A, E_{\pm}$ with the ground state DA, so that the orbital degrees of freedom are quenched at low temperature and the SW mapping procedure ends with conventional Kondo Hamiltonian (1). However, using

the magnetic field as a control parameter, one may initiate level crossing by varying the magnetic flux Φ through TTQD. This level crossing is shown on the upper panel of Fig. 10. In each crossing points and its nearest vicinity

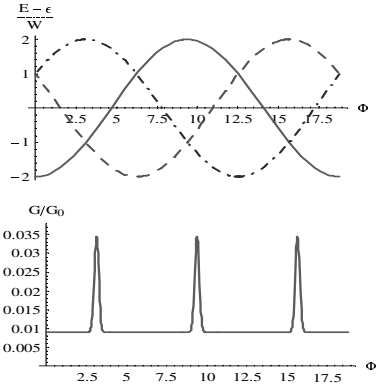

Fig. 10. Upper panel: Evolution of the energy levels E_A (solid line) and $E_\pm$ (dashed and dash-dotted line, respectively). Lower panel: corresponding evolution of conductance (in units of $G_0 = \pi e^2/\hbar$)

the dynamical symmetry of TTQD is SU(4), so that the symmetry crossovers $SU(2) \to SU(4) \to SU(2)$ occur at $\Phi = (n + \frac{1}{2})\Phi_0$, where Φ_0 is the quantum of magnetic flux. Each crossover is accompanied by the change of T_K from $\exp(-1/2J)$ to $\exp(-1/4J)$ and back (see Sect. 3.1). The ZBA peak in the two-terminal tunnel conductance changes accordingly (Fig. 10, lower panel).

Although formally there are three tunneling channels, the non-Fermi-liquid Kondo regime cannot arise, because the non-diagonal components J_{ij} appear in the exchange Hamiltonian (14). Further diagonalization should be done by means of rotating frameworks for Bloch electrons. This diagonalization introduces irremovable channel anisotropy, so that the 2-channel non-Fermi-liquid fixed point cannot be achieved unlike the case considered in [53], where the channel isotropy was postulated from the very beginning.

In order to realize the Aharonov-Bohm interference, one should use the two-terminal geometry shown in Fig. 2d. In this case there are two paths (1-3) and (2-3) for single electron tunneling between source and drain. Interference of two waves in the drain results in famous Aharonov-Bohm oscillations. The field B affects the lead-dot hopping phases. In the chosen gauge, the hopping integrals are modified as, $W \to W\exp(\mathrm{i}\Phi_1/3)$, $V_{1,2} \to V_s \exp[\pm\mathrm{i}(\Phi_1/6 + \Phi_2/2)]$, where $\Phi_{1,2}$ are magnetic fluxes through the upper and lower loop of the device. As a result the exchange Hamiltonian reads

$$H = J_s \boldsymbol{S}\cdot\boldsymbol{s}_s + J_d \boldsymbol{S}\cdot\boldsymbol{s}_d + J_{sd}\boldsymbol{S}\cdot(\boldsymbol{s}_{sd} + \boldsymbol{s}_{ds}) + K\boldsymbol{T}\cdot\boldsymbol{t} \qquad (38)$$

(the latter term becomes actual when the magnetic field induces level crossing in accordance with Fig. 10). Magnetic flux enters the coupling constants J_s, J_d,

J_{sd}, K via SW transformation. As a result the constant $J_{sd}(\Phi_1, \Phi_2)$ turns into zero at some values of magnetic flux, so that the Aharonov-Bohm interference completely blocks Kondo transparency. Thus TTQD serves simultaneously as a Kondo "pass valve" and as an Aharonov-Bohm interferometer. It should be stressed that both the continuous spin degrees of freedom and discrete "rotations" of triangle are involved in these two phenomena in TTQD.

5 Concluding Remarks

Among many aspects of Kondo tunneling through complex quantum dots we have chosen for this review only the symmetry related properties predetermined by the structure of the low-lying states in the spin multiplet characterizing the fixed charge sector of complex quantum dot. New features, which are introduced by the dynamical symmetries in the Kondo physics are the multistage process of Kondo screening, symmetry crossovers driven by experimentally tunable control parameters, interplay between continuous spin rotation symmetry and discrete point symmetry of nanodevices. The main tool of experimental monitoring of variable symmetries is the study of temperature and magnetic filed dependence of zero- and finite-bias anomalies in tunnel conductance.

Among other facets of Kondo effect in nanostructures one should mention non-equilibrium Kondo effect at finite bias and under light illumination, where both quantum dots and leads are far enough from thermodynamic balance. Under this conditions such phenomena as spin relaxation, dephasing and decoherence influence the tunnel transport in Kondo regime. Real atoms and molecules also may be included in electric circuit by means of advanced experimental techniques (scanning tunnel spectroscopy, break-junction method etc). In this case phonon- and photon-assisted processes should be taken into account, which result in interplay of Kondo resonance tunneling with various "polaronic" and "excitonic" effects. Besides, Kondo processes may be included in nano-electro-mechanical shuttling, so that Kondo physics of movable objects becomes one of the challenging items on the agenda.

References

1. J. Kondo: Prog. Theor. Phys. **32**, 37 (1964)
2. K. G. Wilson: Rev. Mod. Phys. **47**, 773 (1975)
3. A. M. Tsvelik, P. B. Wiegmann: Adv. Phys. **3**2, 453 (1983)
4. N. Andrei, K. Furuya, J. H. Lowenstein: Rev. Mod. Phys. **55**, 331 (1983)
5. J. Appelbaum: Phys. Rev. Lett. **17**, 91 (1966)
6. P. W. Anderson: Phys. Rev. Lett. **17**, 95 (1966)
7. J. R. Schrieffer, P. A. Wolff: Phys. Rev. **149**, 491 (1966)
8. J. Friedel: Can. J. Phys. **34**, 1190 (1956)
9. P. W. Anderson: Phys. Rev. **124**, 41 (1961)

10. K. S. Dy: Phys. Status Solidi (b) **81**, K111 (1977)
11. B. Coqblin, J. R. Schrieffer: Phys. Rev. **185**, 847 (1969)
12. B. Cornut, B. Coqblin: Phys. Rev. B **5**, 4541 (1972)
13. I. Affleck, J. Marston: Phys. Rev. Lett. **37**, 3774 (1988)
14. N. Read, S. Sachdev: Nucl. Phys. B **316**, 609 (1989)
15. P. Nozieres, A. Blandin: J. Phys. (Paris) **41**, 193 (1980)
16. P. Nozieres: J. Low Temp. Phys. **17**, 31 (1974)
17. A. Zawadowski, K. Vladar: Solid State Commun. **35**, 217 (1980)
18. D. L. Cox: Phys. Rev. Lett. **59**, 1240 (1987)
19. S. E. Barnes: Phys. Rev. B **37**, 3671 (1988)
20. D. L. Cox, A. Zawadowski: Adv. Phys. **47**, 943 (1998)
21. A. K. Zhuravlev, V. Yu. Irkhin, M. I. Katsnelson et al: Phys. Rev. Lett. **93**, 236403 (2004)
22. L. I. Glazman, M. E. Raikh: JETP Lett. **47**, 452 (1988)
23. T. K. Ng, P. A. Lee: Phys. Rev. Lett. **61**, 1768 (1988)
24. D. Goldhaber-Gordon, H. Shtrikman, D. Mahalu et al: Nature **391**, 156 (1998)
25. S. M. Cronenwett, T. H. Oosterkamp, L. P. Kouwenhoven: Science **281**, 540 (1998)
26. F. Simmel, R. H. Blick, J. P. Kotthaus et al: Phys. Rev. Lett. **83**, 804 (1999)
27. K. A. Matveev: Zh. Eksp. Teor. Phys. **9**8, 1598 (1991) [Sov. Phys. JETP **72**, 892 (1991)]; Phys. Rev. B **51**, 1743 (1995)
28. E. Lebanon, A. Shiller: Phys. Rev. B **64**, 245338 (2001)
29. C. J. Bolech, N. Shah: Phys. Rev. Lett. **95**, 036801 (2005)
30. K. Kikoin, Y. Avishai: Phys. Rev. B **62**, 4647 (2000)
31. M. Pustilnik, Y. Avishai, K. Kikoin: Phys. Rev. Lett. **84**, 1756 (2000)
32. D. Giuliano, A. Tagliacozzo: Phys. Rev. Lett. **84**, 4677 (2000)
33. M. Eto, Yu. Nazarov: Phys. Rev. Lett. **85**, 1306 (2000)
34. *Single Charge Tunneling*, ed by H. Grabert, M. H. Devoret (Plenum, New York 1992)
35. *Mesoscopic Electron Transport*, ed by L. L. Son, L. P. Kouwenhoven, G. Schön (Kluwer, Dordrecht 1997)
36. W. G. van der Wiel, S. De Franceschi, J. M. Enselman et al: Rev. Mod. Phys. **75**, 1 (2003)
37. F. Hofmann, T. Heinzel, D. A. Wharam et al: Phys. Rev. B **51**, 13872 (1995)
38. L. W. Molenkamp, K. Flensberg, M. Kemerlink: Phys. Rev. Lett. **75**, 4282 (1995)
39. C. Livermore, C. H. Crouch, R. M. Westervelt et al: Science **274**, 1382 (1996)
40. J. J. Palacios, P. Hawrilak: Phys. Rev. B **51**, 1769 (1995)
41. D. Loss, D. P. DiVincenzo: Phys. Rev. A **57**, 120 (1998)
42. B. Partoens, F. M. Peeters, Phys. Rev. Lett. **84**, 4433 (2000)
43. K. Kikoin, Y. Avishai: Phys. Rev. Lett. **86**, 2090 (2001); Phys. Rev. B **65**, 115329 (2002)
44. N. J. Craig, J. M. Taylor, E. A. Lester et al: Science **304**, 565 (2004)
45. M. Stopa: Phys. Rev, Lett. **88**, 146802 (2002)
46. A. Vidan, R. M. Westervelt, M. Stopa et al: Appl. Phys. Lett. **85**, 3602 (2004)
47. L. Gaudreau, S. Studenikin, A. Sachrajda et al: cond-mat/06015967 (2006)
48. T. Tanamoto: Phys. Rev. A **61**, 022305 (2000)
49. D. S. Saraga, D. Loss: Phys. Rev. Lett. **90**, 166803 (2003)
50. G. Zaránd, A. Brataas, D. Goldhaber-Gordon: Solid State Commun. **126**, 463 (2003)

51. T. Jamneala et al: Phys. Rev. Lett. **87**, 256804 (2001)
52. B. Lazarovits, P. Simon, G. Zaránd et al: Phys. Rev. Lett. **95**, 077202 (2005)
53. K. Ingersent, A. W. Ludwig, J. Affleck: Phys. Rev. Lett. **95**, 257204 (2005)
54. T. Kuzmenko, K. Kikoin, Y. Avishai: Phys. Rev. Lett. **96**, 046601 (2006)
55. T. Kuzmenko, K. Kikoin, Y. Avishai: Physica E **29**, 334 (2005)
56. C. A. Stafford, S. Das Sarma: Phys. Rev. Lett. **72**, 3590 (1994)
57. C. Klimek, G. Chen, S. Datta: Phys. Rev. B **50**, 2316 (1994)
58. Y. Meir, N. S. Wingreen: Phys. Rev. Lett. **68**, 2512 (1992)
59. J. M. Golden, B. I. Halperin: Phys. Rev. B **53**, 3893 (1996)
60. D. C. Langreth, Phys. Rev. **150**, 516 (1966)
61. J. Hubbard, Proc. Roy. Soc. A **285**, 542 (1965)
62. M. J. Englefield: *Group Theory and the Coulomb Problem* (Wiley, New York 1972)
63. I. A. Malkin, V. I. Man'ko: *Dynamical Symmetries and Coherent States of Quantum Systems* (Fizmatgiz, Moscow 1979) [in Russian]
64. K. Kikoin, Y. Avishai, M. N. Kiselev: In: *Nanophysics, Nanoclusters, Nanodevices* (Nova Science, New York 2006)
65. M. N. Kiselev: Int. J. of Mod. Phys. B **20**, 381 (2006)
66. T. Ivanov: Europhys. Lett. **40**, 183 (1997)
67. T. Pohjola, J. König, M. M. Salomaa et al: Europhys. Lett. **40**, 189 (1997)
68. T. Aono, M. Eto, K. Kawamura: J. Phys. Soc. Jpn. **67**, 1860 (1998)
69. A. Georges, Y. Meir: Phys. Rev. Lett. **82**, 3508 (2002)
70. V. N. Golovach, D. Loss: Europhys. Lett. **62**, 83 (2003)
71. L. Borda, G. Zarand, W. Hofstetter et al: Phys. Rev. Lett. **90**, 026602 (2003)
72. Y. Tanaka, N. Kawakami: J. Phys. Soc. Jpn. **73**, 2795 (2004); Phys. Rev. B **72**, 085304 (2005)
73. R. Sakano, N. Kawakami: Phys. Rev. B **73**, 155332 (2005)
74. M. Eto: J. Phys. Soc. Jpn. **74**, 95 (2005)
75. B. A. Jones, C. Varma: Phys. Rev. B **40**, 324 (1989)
76. W. Hofstetter, H. Schoeller: Phys. Rev. Lett. **88**, 016803 (2002)
77. M. Pustilnik, L. I. Glazman: Phys. Rev. Lett. **85**, 2993 (2000); Phys. Rev. B **64**, 045328 (2001)
78. T. Kuzmenko, K. Kikoin, Y. Avishai: Phys. Rev. Lett. **89**, 156602 (2002); Phys. Rev. B **69**, 195109 (2004)
79. W. Izumida, O. Sakai, Y. Shimizu: J. Phys. Soc. Jpn. **67**, 2444 (1998)
80. M. Pustilnik, L. I. Glazman, W. Hofstetter: Phys. Rev. B **68**, 161303 (2003)
81. W. Hofstetter, G. Zarand: Phys. Rev. B **69**, 235301 (2004)
82. M. Pustilnik, L. I. Glazman: J. Phys.: Cond. Mat. **16**, R513 (2004)
83. P. W. Anderson: J. Phys. C **3**, 2436 (1980)
84. W. G. van der Wiel, S. de Francheschi, T. Fujisawa et al: Science **289**, 2105 (2000)
85. A. F. Barabanov, K. A. Kikoin, L. A. Maksimov: Teor. Mat. Fiz. **20**, 364 (1974)
86. F. D. M. Haldane: Phys. Rev. Lett. **40**, 416 (1978)
87. A. Kogan, G. Granger, M. A. Kastner et al: Phys. Rev. B **67**, 113309 (2003)
88. K. Kang, S. Y. Cho, J-J. Kim et al: Phys. Rev. B **63**, 113304 (2001)
89. T-S. Kim, S. Hershfield: Phys. Rev. B **63**, 245326 (2001)
90. Y. Takazawa, Y. Imai, N. Kawakami: J. Phys. Soc. Jpn. **71**, 2234 (2002)
91. M. E. Torio, K. Hallberg, A. H. Ceccatto et al: Phys. Rev. B **65**, 085302 (2002)
92. U. Fano: Phys. Rev. **124**, 1866 (1961)

93. A. Shibatani, Y. Toyozawa: J. Phys. Soc. Jpn. **25**, 335 (1968)
94. J. Göres, D. Goldhaber-Gordon, S. Heemeyer et al: Phys. Rev. B **62**, 2188 (2000)
95. M. E. Torio, K. Hallberg, S. Flach et al: Eur. Phys. J. B **37**, 399 (2004)
96. Y. Oreg, D. Goldhaber-Gordon: Phys. Rev. Lett. **90**, 136602 (2003)
97. M. Pustilnik, L. Borda, L. I. Glazman et al: Phys. Rev. B **69**, 115316 (2004)
98. I. Affleck, A. W. W. Ludwig: Phys. Rev. B **48**, 7297 (1993)
99. I. Affleck, A. W. W. Ludwig, H. B. Pang et al: Phys. Rev. B **45**, 7918 (1992)
100. P. Simon, R. Lopez, Y. Oreg: Phys. Rev. Lett. **94**, 086602 (2005)
101. M. G. Vavilov, L. I. Glazman: Phys. Rev. Lett. **94**, 086805 (2005)
102. M. Vojta, R. Bulla, W. Hofstetter: Phys. Rev. B **65**, 140405 (2002)
103. T. Kuzmenko, K. Kikoin, Y. Avishai: Europhys. Lett. **64**, 218 (2003)
104. T. Kuzmenko, K. Kikoin, Y. Avishai: Phys. Rev. B **73**, 235310 (2006)
105. K. Le Hur, P. Recher, E. Dupont et al: Phys. Rev. Lett. **96**, 106803 (2006)
106. N. Yu, M. Fowler: Phys. Rev. B **45**, 11795 (1992)
107. J. D. Lee: Phys. Rev. B **67**, 153108 (2003)
108. A. Fuhrer, T. Ihn, K. Ensslin et al: Phys. Rev. Lett. **93**, 176803 (2004)
109. B. Leturcq, L. Schmid, K. Ensslin et al: Phys. Rev. Lett. **95**, 126603 (2005)
110. V. V. Savkin, A. N. Rubtsov, M. I. Katsnelson et al: Phys. Rev. Lett. **94**, 026402 (2005)
111. T. Kuzmenko, K. Kikoin, Y. Avishai: Physica E **29**, 334 (2005)

Low Temperature Decoherence and Relaxation in Charge Josephson Junction Qubits

Alex Grishin[1], Igor V. Yurkevich[2], and Igor V. Lerner[3]

[1] School of Physics and Astronomy, University of Birmingham, UK.
ag@th.ph.bham.ac.uk
[2] School of Physics and Astronomy, University of Birmingham, UK.
ivy@th.ph.bham.ac.uk
[3] School of Physics and Astronomy, University of Birmingham, UK.
ivl@th.ph.bham.ac.uk

1 Introduction

Research interest in controllable two-level systems, which have been enthusiastically called quantum bits or qubits, has grown enormously during the last decade. Behind a huge burst of activity in this field stands an idea of what is possible in principle but extremely difficult to achieve instrumentally - the fascinating idea of quantum computing. The very principle of quantum superposition allows many operations to be performed on a quantum computer in parallel, while an ordinary 'classical' computer, however fast, can only handle one operation at a time. The enthusiasm is not held back by the fact that exploiting quantum parallelism is by no means straightforward, and there exist only a few algorithms (e.g., [1, 2]) for which the quantum computer (if ever built) would offer an essential improvement in comparison with its 'classical' counterpart. Even if other uses of quantum computing prove limited (which might or might not be the case), its existence would most certainly lead to a breakthrough in simulations of real physical many-particle systems.

Whether or not the ultimate goal of building a working quantum computer is ever achieved, both experimental and theoretical studies of properties of single or entangled qubits are flourishing. One of the most compelling reasons for this is an exciting overlap of the possibility of a future technological breakthrough and the reality of research in fundamentals of quantum mechanics. There exist various experimental realizations of the qubit, amongst which solid state qubits are of particular interest as they provide one of the most promising routes to implementing a scalable set of qubits, which is one of the minimal requirements for implementing quantum computations. However, any solid state qubit, albeit representing effectively a two-level system, comprises

a huge number of internal degrees of freedom whose unavoidable coupling to the environment leads to loss of coherence. Quantum computations require a set of fully or at least partially entangled states on which some unitary operations are performed. Decoherence would make evolution of states non-unitary and would lead to the loss of entanglement between the states. Thus, the loss of coherence before a sufficient amount of quantum operations was performed would be the major impediment in using solid-state qubits in quantum computations. It is believed that tens of thousands of unitary operations are required for quantum computation to become a reality [3] so that sufficiently long decoherence times (much longer than those currently archived in the best solid state qubits) should be achieved experimentally. It necessitates a better theoretical understanding of realistic mechanisms of decoherence.

In this article, after illustrating in a simple way the main features of decoherence in a generic qubit coupled to the environment (Sect. 2), we will consider onset of decoherence in one particular realization of qubit, namely charge Josephson junction (JJ) qubit. First, we will describe briefly what is the JJ qubit (Sect. 3) and then focus on a mechanism widely believed to be responsible for one of the main channels of decoherence for the charge JJ qubit, namely its inevitable coupling to fluctuating background charges (Sect. 4). This model has been thoroughly investigated in all regimes [4–7] and has a tutorial advantage of being exactly solvable within a fully quantum approach [7] and showing a rich variety of different regimes with a non-trivial dependence on temperature and on (unfortunately ill known experimentally) the strength of coupling between the qubit and the fluctuating charges.

In Sect. 5 we will offer a solution for decoherence rate in this model which is formally exact at an arbitrary temperature T. In the 'high-T regime' (which could still correspond to rather low temperatures), the decoherence rate saturates and becomes T-independent, while at low temperature it turns out to be linear in T and behave non-monotonically as a function of the coupling strength between qubit and the environment. In conclusion we will also consider the relaxation rate, albeit only perturbatively with respect to the coupling strength, and demonstrate that the model can qualitatively explain the experimentally observed [8] quasi-linear behavior of the spectral density of noise with humps at certain frequencies.

2 Coupling to the Environment and Decoherence

Before considering a realistic solid-state qubit, we start with illustrating what is the loss of coherence in a generic qubit. Such a qubit is a two-level system so that its Hamiltonian can be mapped to that of spin $\frac{1}{2}$ and written

$$\hat{H}_0 = \tfrac{1}{2} B_z \hat{\sigma}_z - \tfrac{1}{2} B_x \hat{\sigma}_x \ . \tag{1}$$

where $\boldsymbol{B}$ is an effective 'magnetic field' (measured here in energy units). States of the qubit can be described in terms of its density matrix,

$$\hat{\rho}(t) = \sum_{i,j=\uparrow,\downarrow} |i\rangle \, \rho_{ij}(t) \, \langle j| \ . \tag{2}$$

When the qubit (or any system) is in a pure state, one can always find a basis where $\hat{\rho} = \sum_i |i\rangle \langle i|$. In any other (rotated) basis, the density matrix of a pure state obeys $\hat{\rho}^2 = \hat{\rho}$. In a rotated basis, the density matrix performs a unitary evolution described by the Heisenberg equation of motion,

$$\frac{\partial \hat{\rho}(t)}{\partial t} = -\mathrm{i} \left[\hat{H}_0, \hat{\rho}(t) \right] \ , \tag{3}$$

whose solution is $\hat{\rho}(t) = \hat{U} \hat{\rho}(0) \hat{U}^\dagger$ where in this trivial case the evolution operator is $\hat{U} = \mathrm{e}^{-\mathrm{i} H_0 t}$. In a semiclassical language, such an evolution is

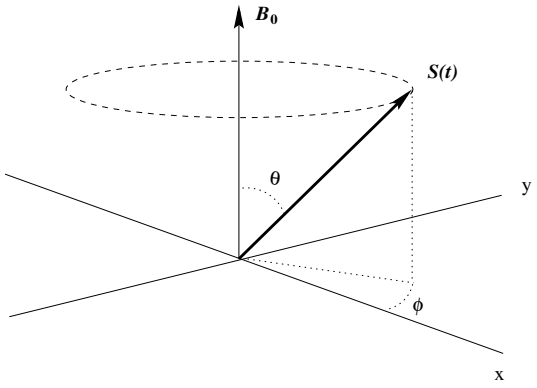

Fig. 1. Bloch representation

convenient to visualize using the Bloch sphere representation, shown in Fig. 1. There the spin evolution is parameterized by two angles on a unit sphere, θ and ϕ. For the simple system described by (1)–(3) above, such an evolution is simply rotating around z axis, provided that the 'field' $\boldsymbol{B} = (B_z, 0)$, and the initial state of the spin was diagonal in a different basis. The Bloch angles are related to the density matrix by

$$\rho(t) = \frac{1}{2} \left(1 + \boldsymbol{S}(t) \cdot \boldsymbol{\sigma} \right) \ . \tag{4}$$

An inevitable coupling to the environment (bath) can be schematically described by the Hamiltonian

$$\hat{H} = \hat{H}_0 + \hat{V}_{\mathrm{coupl}} + \hat{H}_{\mathrm{bath}} \ . \tag{5}$$

Now it is the full density matrix of the qubit + bath which obeys the Heisenberg equation of motion. As we are interested in states measured on the qubit,

we need to introduce a so-called 'reduced' density matrix, $\hat{\rho}^q = \mathrm{Tr}_{\mathrm{bath}}\,\hat{\rho}$ where the trace is taken over all the bath states with the proper Gibbs weight. We will describe a consistent way of performing such a trace in Sect. 5. Here we want just to illustrate how the coupling to the environment makes the time evolution non-unitary and leads to the loss of coherence.

Let us consider a model in which there is a minimal, so-called longitudinal coupling of the qubit to the bath,

$$V_{\mathrm{coupl}} = \hat{X}\hat{\sigma}_z \,. \tag{6}$$

Let us further assume that the qubit was prepared in a pure state but in a different basis, so that its density matrix has both diagonal and off-diagonal elements,

$$\hat{\rho} \equiv \begin{pmatrix} \rho_{11} & \rho_{12} \\ \rho_{21} & \rho_{22} \end{pmatrix} = \begin{pmatrix} n & f \\ f^* & 1-n \end{pmatrix} \,,$$

with $|f|^2 = n - n^2$ at $t = 0$ assuming that the qubit was prepared in a pure state, $\hat{\rho}^2 = \hat{\rho}$. Let finally switch off the x component of the 'magnetic field' $\boldsymbol{B}$. In the absence of coupling, the spin that represents our qubit on the Bloch sphere of Fig. 1 would simply rotate at a constant frequency $\omega_0 \propto B_z$ around z axis. The longitudinal coupling results in the appearance of an additional, fluctuating time dependent component of the field, $B_z(t)$. This fluctuating field results from the thermal noise in the bath and inevitably destroys coherence as we will now show.

The spin rotation is described in a semiclassical language by the Landau-Lifshitz equation,

$$\frac{\mathrm{d}\boldsymbol{S}}{\mathrm{d}t} = \boldsymbol{S} \times \boldsymbol{B} \,.$$

In the absence of the transverse component of the magnetic field, the diagonal elements of the density matrix, related to $\boldsymbol{S}$ by Eq. (4), remain constant. In the presence of the environment-induced time-dependent component of the longitudinal magnetic field, $\boldsymbol{S}$ rotates with a changing frequency resulting in the off-diagonal elements of $\hat{\rho}$ acquiring the following time-dependence:

$$\rho_{12}(t) = \tfrac{1}{2}\left(S_x + \mathrm{i}S_y\right) = \rho_{12}(0)\mathrm{e}^{\mathrm{i}\omega_0 t + \mathrm{i}\int_0^t B(\tau)\mathrm{d}\tau} \,. \tag{7}$$

The contributions due to the fluctuating field, $B_z(t)$, should be averaged over time. It is natural to assume that fluctuations of B_z are Gaussian, as they are due to a very large number of degrees of freedom in the bath. For Gaussian fluctuations, one can use the standard averaging formula,

$$\left\langle \mathrm{e}^{\mathrm{i}\varphi} \right\rangle = \mathrm{e}^{-\frac{1}{2}\left\langle \varphi^2 \right\rangle} \,,$$

so that

$$\begin{aligned}\langle\rho_{12}(t)\rangle &= \rho_{12}(0)\mathrm{e}^{\mathrm{i}\omega_0 t}\left\langle \mathrm{e}^{\mathrm{i}\int_0^t \mathrm{d}\tau B(\tau)}\right\rangle \\ &= \rho_{12}(0)\mathrm{e}^{\mathrm{i}\omega_0 t}\exp\left[-\frac{1}{2}\int_0^t \mathrm{d}\tau\int_0^t \mathrm{d}\tau'\Big\langle B(\tau)B(\tau')\Big\rangle\right] \\ &= \rho_{12}(0)\mathrm{e}^{\mathrm{i}\omega_0 t}\exp\left[-t\int_{-t}^t \mathrm{d}\tau\Big\langle B(0)B(\tau)\Big\rangle\right] \equiv \rho_{12}(0)\mathrm{e}^{\mathrm{i}\omega_0 t}\mathrm{e}^{-\Gamma_2 t}\ . \quad (8)\end{aligned}$$

Here we have semi-formally defined the decoherence rate Γ_2 as the rate of relaxation of the off-diagonal part of the density matrix, using that the correlation function of the fluctuating field B_z depends only on the time difference. In this definition, Γ_2 can still be time-dependent. However, for large enough t, longer than the longest relaxation time for thermal noise in the bath which makes B_z fluctuating, Γ_2 should saturate at some limiting value. Thus, making a bit more formal definition, we arrive at

$$\Gamma_2 \equiv \frac{1}{T_2} = \lim_{t\to\infty}\int_{-t}^t \mathrm{d}\tau\Big\langle B(0)B(\tau)\Big\rangle = S_B\left(\omega\approx 0\right)\ , \quad (9)$$

where $S_B(\omega)$ is the noise power spectrum of the fluctuating field B,

$$S_B(\omega) \equiv \int_{-\infty}^{\infty}\Big\langle B(0)B(t)\Big\rangle\,\mathrm{e}^{-\mathrm{i}\omega t}\mathrm{d}t\ . \quad (10)$$

The above calculation should be considered just as an illustration as we have just introduced Gaussian fluctuations of the 'magnetic field' B_z. However, it is very easy both to make this sort of calculations fully quantum-mechanical, and to generalize the model beyond the longitudinal coupling .

To show that the effective coupling to the bath leads to the thermal noise that directly results in the appearance of decoherence given by (9), one needs to model the bath. The most standard theoretical approach to such modeling and thus to decoherence by the environment is based on spin-boson models [9, 10], where the environment is modelled as a set of harmonic oscillators,

$$\hat{H}_{\mathrm{bath}} = \sum_{\boldsymbol{k}}\omega_{\boldsymbol{k}} b^\dagger_{\boldsymbol{k}} b_{\boldsymbol{k}}\ . \quad (11)$$

The linear coupling in (6) should be understood as coupling to all the oscillator degrees of freedom,

$$V_{\mathrm{coupl}} = \hat{X}\hat{\sigma}_z \equiv \sum_{\boldsymbol{k}}\left(\lambda_{\boldsymbol{k}} b^\dagger_{\boldsymbol{k}} + \mathrm{h.c.}\right)\hat{\sigma}_z\ . \quad (12)$$

The fluctuation-dissipation theorem allows one to express the noise power for the coupling operator $\hat{X}(t)$ (in the interaction picture) as

$$S_X(\omega) = \left\langle X^2_\omega\right\rangle + \left\langle X^2_{-\omega}\right\rangle = 2J(\omega)\coth\frac{\hbar\omega}{2T}\ , \quad (13)$$

where $J(\omega)$ depends on the density of oscillator states, i.e., on the spectrum $\omega_{\boldsymbol{k}}$ in (11), and on the coupling $\lambda(\omega)$ in (12). The decoherence rate is still given by (9), where $S_X(\omega)$ should be substituted for $S_B(\omega)$. For the most typical, Ohmic model, $J(\omega) \propto \omega$ so that $\Gamma_2 \propto T$. As we will show later, such a linear dependence on temperature is characteristic for a realistic model to be considered later in this article, but only for sufficiently low T.

It is straightforward to generalize our considerations beyond the longitudinal model. To this end one should take into account the σ_x-proportional contribution in the qubit action – this would be general enough even if one leaves the coupling as it stands in (6) or (12).

3 Charge Josephson Junction Qubit

Not any two-level system could serve as a qubit. There is a set of requirements that such a system must satisfy in the first place. In 1997 David DiVincenzo [11] formulated five criteria which have to be satisfied by a physical system considered as a candidate for quantum computation, conditions which are widely known as the 'DiVincenzo checklist':
i) One needs well-defined two-state quantum systems (qubits);
ii) One should be able to prepare the initial state of the qubits with sufficient accuracy;
iii) A long phase time coherence is needed, sufficient to allow for a large number ($\sim 10^4$) of coherent manipulations;
iv) Control over the qubit's Hamiltonian is required to perform the necessary unitary transformations;
v) A quantum measurement is needed to read out the quantum information.
At the present time any known physical system falls short of the requirements stipulated in the above list. Currently lots of different possibilities of building a quantum computer are investigated including nuclear spins, quantum dots, Josephson junctions, trapped ions, optical lattices, electrons on liquid helium and some others with each of them being quite far from satisfying this or that condition of the above set.

A route which seems to be one of the most promising is to build a system of Josephson junction (JJ) qubits: the qubits which are based on Josephson junctions and utilize the charge and flux degrees of freedom. The main advantages of the superconducting devices involving Josephson junctions are a) a relative easiness to manufacture - the lithographic methods used to fabricate them are well-established; b) controllability - gate voltages controlling charge Josephson qubits can be adjusted with a very high degree of accuracy allowing, in particular, the high-fidelity preparation of the initial state of the qubit; and finally c) measurability - the techniques for almost non-invasive measurements which can be used for this type of qubits are quite advanced. It is point iii) in the DiVincenzo checklist - decoherence - which causes the greatest worry for the JJ qubit, as well as for any other realization of a solid

state qubit. We stress again that 'spin up' and 'spin down' states in solid state qubits are formed by a substantial number of electrons (which is not the case for nuclear spins which can be effectively decoupled from the external world). All degrees of freedom interact with the environment thus causing the loss of quantum coherence. Whether this problem of decoherence can or cannot be satisfactorily solved is most likely to determine the future of JJ qubits.

The JJ qubit is made of superconducting islands separated by a Josephson junction; the scheme is depicted in Fig. 2. In the scheme, 'SC' marks the

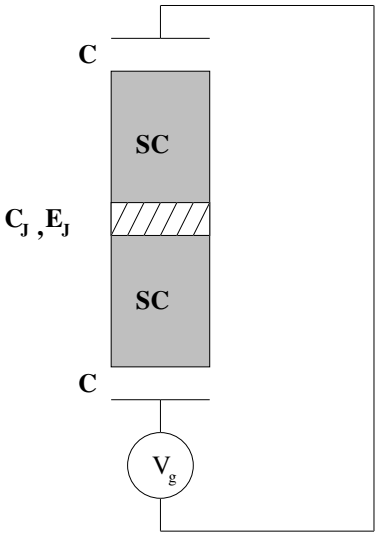

Fig. 2. Scheme for charge JJ qubit

superconducting islands and the hatched area is the Josephson junction with capacity C_J and Josephson coupling E_J. The superconductors are separated from the leads by capacitors C, which do not allow any tunneling, and these capacitors are biased by some controllable gate voltage V_g. Cooper pairs can tunnel through the Josephson junction thus changing the total charge of each island.

The charge JJ qubit is a two-level system in which states are different by the charge of a single Cooper pair. The equilibrium number of Cooper pairs in each island is controlled by the gate voltage. When it is tuned to be close to half integer $\sim n_0 + 1/2$, the states with n_0 and $n_0 + 1$ Cooper pairs (circled in Fig. 3) can be arbitrarily close to each other in energy while the distance to other energy levels (crossed in Fig. 3) is of the order of the charging energy that – in the temperature units - could be tens or even a hundred degrees of Kelvin which is much higher than typical experimental temperatures. We assume that the Josephson coupling energy is much smaller than the charging energy so that the states are really discriminated by charge. (In the opposite

limit, a so called flux JJ qubit can be – very successfully – built). Thus we effectively separate all the charge states with $n < n_0$ or $n > n_0 + 1$ and thus have a two-level system, whose states are made of a macroscopic number of electrons. In future we will be referring to these as states 'spin up' and 'spin down'.

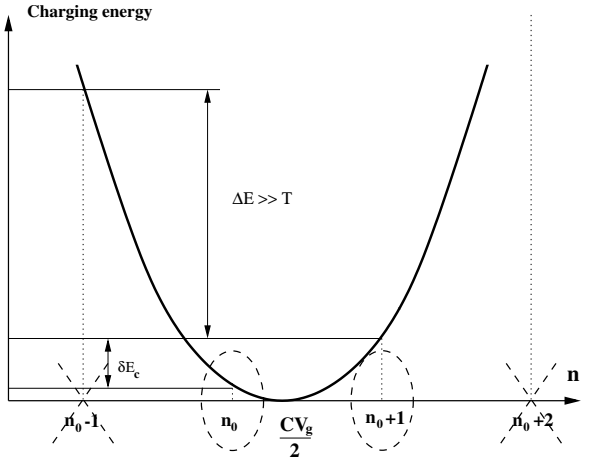

Fig. 3. Two charge states

More formally, the total energy of the system which consists of charging and Josephson contributions can be written as follows:

$$E = \frac{1}{2C_J + C}\left(2|e|n - \frac{CV_g}{2}\right)^2 - E_J \cos\Theta\,, \tag{14}$$

where Θ is the superconducting order-parameter phase shift between the islands. In properly chosen units, the pair of n and Θ are canonically conjugated coordinate and momentum since they generate correct Hamiltonian equations of motion. Differentiating with respect to momentum $\mathrm{d}n/\mathrm{d}t = \partial E/\partial\Theta = E_J \sin\Theta$ produces the correct equation for the Josephson current (in the units chosen E_J coincides with the critical current), and differentiation with respect to coordinate $\mathrm{d}\Theta/\mathrm{d}t = -\partial E/\partial n = (CV_g - 2n)/(2C_J + C) = V_J$ produces the correct value for the voltage drop V_J across the junction. The quantization of (14) then gives:

$$\hat{H} = \frac{1}{2C_J + C}\left(\hat{n} - \frac{CV_g}{2}\right)^2 - E_J \cos\hat{\Theta}, \qquad \left[\hat{n}, \hat{\Theta}\right] = \mathrm{i}\,. \tag{15}$$

After taking the projection of the Hamiltonian (15) onto two states of interest, in this 'spin up' and 'spin down' basis the charging part can be

written as $(\delta E_c/2)\hat{\sigma}_z$, where δE_c is the charging energy difference between the two states:

$$\delta E_c(V_g) = \frac{2}{2C_J + C}\left(n_0 + \frac{1}{2} - \frac{CV_g}{2}\right) . \tag{16}$$

The commutator between $\hat{n}$ and $\hat{\Theta}$ dictates the commutation relations $[\hat{n}, \mathrm{e}^{\pm \mathrm{i}\hat{\Theta}}] = \mp \mathrm{e}^{\pm \mathrm{i}\hat{\Theta}}$ from which it immediately follows that in the 'spin up' and 'spin down' basis $\cos\hat{\Theta}$ is given by $\hat{\sigma}_x/2$. Summarizing the above we substitute the Hamiltonian (15) with its projection:

$$\hat{H} = \frac{\delta E_c(V_g)}{2}\,\hat{\sigma}_z - \frac{E_J}{2}\,\hat{\sigma}_x . \tag{17}$$

As the result we have a controllable two-level system, that is a qubit. Control is exercised by changing the gate voltage V_g on which the $\hat{\sigma}_z$ coefficient δE_c depends (16). If δE_c is maintained large $\delta E_c \gg E_J$, which is characteristic for charge JJ qubit, then the Josephson part of the Hamiltonian is irrelevant and evolution amounts to acquiring phase shift between 'spin up' and 'spin down' states in the initial mixture. On the contrary, tuning charging part to the degeneracy point $\delta E_c = 0$ stimulates the spin flip process. It can be shown that the ability to switch for arbitrary time to the above two regimes is enough in order to provide all necessary one qubit quantum operations or quantum gates.

In building a quantum computer, one and two-qubit gates will be required. The details of how to build them are beyond the scope of this article and can be found in the comprehensive review [12] of Makhlin et al. But what is crucial for any quantum computation is to maintain entanglement between the qubits. This will be undermined by inevitable decoherence. Therefore, understanding its mechanisms and studying how it sets in is crucial for future progress in this area.

4 Decoherence - Fluctuating Background Charges Model

4.1 The Model

The main problem with charge JJ qubits (as with any other solid state qubits) is that they lose quantum coherence too quickly due to unwanted, but unavoidable, coupling to the environment. Recent experiments [13–17] show that hundreds of elementary quantum operations can be achieved before coherence is destroyed. This is at least two orders of magnitude short [3] of a rough estimate of what is required for non-trivial quantum calculations. Thus the problem of decoherence is currently the major obstacle for the progress of qubits of this type, and requires a close theoretical examination.

Spin-boson models give a general description but will not offer us any insight into how physical parameters of the problem affect decoherence. To

obtain this we need a physical model describing the real processes which decohere the qubit. It is widely believed [13, 14] that in charge JJ qubits the main contribution to decoherence comes from coupling of the qubit to some charge fluctuators present in the environment. The fluctuating background charges (FBC) model, suggested in this context in [4] and then also used in [5–7, 18], is a possible microscopic model for such a dominant channel of decoherence.

The schematic picture of how the qubit interacts with the FBC is shown in Fig. 4. Impurities which contribute to decoherence sit on the substrate where

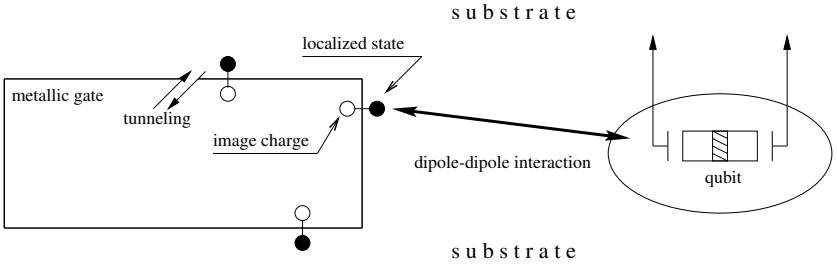

Fig. 4. Interaction between the qubit and fluctuating background charges

the nanocircuit was grown on, close enough to a metallic lead to allow tunneling to and from the lead. Solid circles in Fig. 4 are charged impurities and transparent circles are their image charges. These dipoles interact electrically with the qubit, which in turn behaves like a dipole, since one of the superconducting islands has some number of excessive Cooper pairs and the other one lacks the same number. The interaction depends on the number of such extra pairs and thus is different for the two charge states of the qubit, enforcing dependence of the coupling term on $\hat{\sigma}_z$. Although a static dipole-dipole interaction could only shift the qubit states without any loss of coherence, the tunneling between the impurities and the metallic lead makes charges on the impurities to fluctuate, effectively creating a random time-dependent field on the qubit. This random field causes decoherence. This is similar to the generic case considered in Sect. 2: the time-dependent electric field coupled to σ_z plays the role of the random 'magnetic field' B_z in (8). In the following section, we will show that within the Fermi golden rule approximation (valid for a weak coupling between the qubit and the FBC), both decoherence and relaxation can be calculated in a way similar to that schematically developed in Sect. 2, with an additional advantage of calculating the noise power-spectrum within a fully microscopical model. In this section, we restrict considerations to the longitudinal model of pure decoherence. Its advantage is that one can go well beyond the Fermi golden rule, and calculate the decoherence rate (but not the relaxation!) non-perturbatively and within a fully quantum description valid

for any temperature, including $T \to 0$ (in which limit there is no decoherence in the absence of relaxation).

The Hamiltonian corresponding to the described model can be written as

$$\hat{H} = \frac{\delta E_c}{2}\,\hat{\sigma}_z - \frac{E_J}{2}\,\hat{\sigma}_x + \hat{\sigma}_z\hat{V} + \hat{H}_{\rm B}\ ; \qquad \hat{V} = \frac{1}{2}\sum_i v_i\,\hat{d}_i^\dagger\hat{d}_i\ \ ;$$
$$\hat{H}_{\rm B} = \sum_i \varepsilon_i^0\hat{d}_i^\dagger\hat{d}_i\ + \sum_{i,\boldsymbol{k}}\left[t_{\boldsymbol{k}i}\hat{c}_{\boldsymbol{k}}^\dagger\hat{d}_i\ + \text{h.c}\right] + \sum_{\boldsymbol{k}}\varepsilon_{\boldsymbol{k}}\hat{c}_{\boldsymbol{k}}^\dagger\hat{c}_{\boldsymbol{k}}\ . \tag{18}$$

Here $\hat{d}_i$, $\hat{d}_i^\dagger$ are the operators of annihilation and creation of an electron on the i-th impurity; $\hat{c}_{\boldsymbol{k}}$, $\hat{c}_{\boldsymbol{k}}^\dagger$ are the operators of the conduction electrons in the metal; $t_{\boldsymbol{k}i}$ are the hybridization amplitudes; ε_i^0 is the energy of the localized state on the i-th impurity; $\varepsilon_{\boldsymbol{k}}$ are the energies of the conducting electrons; v_i is the coupling strength between the qubit and the $i^{\rm th}$ impurity.

The charge JJ qubit corresponds to $\delta E_c \gg E_J$, which is always the case for the small enough capacities (i.e., small superconducting islands and the JJ junction in Fig. 2). A non-trivial circuit is supposed to consist of many qubits, and most of the time each particular qubit is in an idle regime for which the condition above is fulfilled. During relatively short times necessary for operations involving spin-flips, the longitudinal model, $E_J = 0$, is not appropriate. However, if there is any hope for a working qubit, the onset of decoherence should only happen in the idle regime.

Apart from temperature, there are three other parameters of dimensions of energy for each fluctuator: the coupling strength v_i, tunneling rate γ_i ($\gamma_i = 2\pi\sum_{\boldsymbol{k}}|t_{\boldsymbol{k}i}|^2\delta(\omega-\varepsilon_{\boldsymbol{k}})$), and fluctuator energy ε_i (note that the bare energy ε_i^0 is renormalized by hybridization). All three parameters are broadly distributed, and temperature can be considered 'low' for some of them and 'high' for others. In Sect. 5 we will show that these two regimes are defined as follows:

$$\begin{cases} T \ll \min\left|\varepsilon_i \pm \frac{1}{2}\sqrt{v_i^2-\gamma_i^2}\right| \quad \text{or} \quad T \ll \gamma_i & \text{low}-\text{T regime} \\ \\ T \gg \max\left|\varepsilon_i \pm \frac{1}{2}\sqrt{v_i^2-\gamma_i^2}\right| \text{ and } \quad T \gg \gamma_i & \text{high}-\text{T regime} \end{cases} . \tag{19}$$

In the high-temperature regime, the decoherence rate can be calculated classically [4,5] which gives for one fluctuator (omitting the index i)

$$\Gamma_2 = \frac{\gamma}{2}\left[1 - \Re\text{e}\sqrt{1-\frac{v^2}{\gamma^2}}\right] = \begin{cases} \gamma/2\,, & v > \gamma \\ \\ v^2/4\gamma\,, & v \ll \gamma \end{cases} . \tag{20}$$

The high temperature treatment of the model would be justified if the high energy impurities, $|\varepsilon_i| \gtrsim T$, were frozen. However, the same hybridization γ which makes impurity charges to fluctuate, broadens their energy positions. Such a broadening creates a Lorentzian tail which is, of course, power-law suppressed but nevertheless allows for a contribution from the energetically

remote impurities. Moreover, even impurities with $|\varepsilon_i| \ll T$ may not satisfy the high-T inequality (19) and contribute in a non-classical way. We will present below some arguments that the impurities most relevant for decoherence are likely to to be in the low-temperature regime.

4.2 Why Low Temperature?

Let us first illustrate why the hybridization leads to the Lorentzian tail. To simplify notation, we will be considering a single impurity. The quadratic Hamiltonian of the bath $\hat{H}_{\rm B}$ can be diagonalized by means of some linear transformation [19], after which it becomes $\hat{H}_{\rm B} = \sum_n \epsilon_n \hat{\alpha}_n^\dagger \hat{\alpha}_n$. The interaction part $\hat{V}$ gets transformed into a sum over all exact states: $\hat{V} = (v/2)\sum_{n,m} u_n^* u_m \hat{\alpha}_n^\dagger \hat{\alpha}_m$, where coefficients u_n are related to the density of states (DoS) $\nu_\varepsilon(\omega)$ on the impurity broadened by hybridization with the conduction band: $|u_n|^2 = \nu_\varepsilon(\epsilon_n)\delta$; here δ is the level spacing in the conduction band, and the DoS is given by the Lorentzian

$$\nu_\varepsilon(\omega) = \frac{1}{\pi}\frac{\gamma/2}{(\omega-\varepsilon)^2+\gamma^2/4}\,. \tag{21}$$

In this notation, low temperature manifests itself in smallness of the interaction term $\hat{V}$, to which only n and m in the vicinity of the Fermi level will effectively contribute (see Fig. 5). Since the smallness of u coefficients is controlled by the Lorentzian (21) (rather than distribution function), the

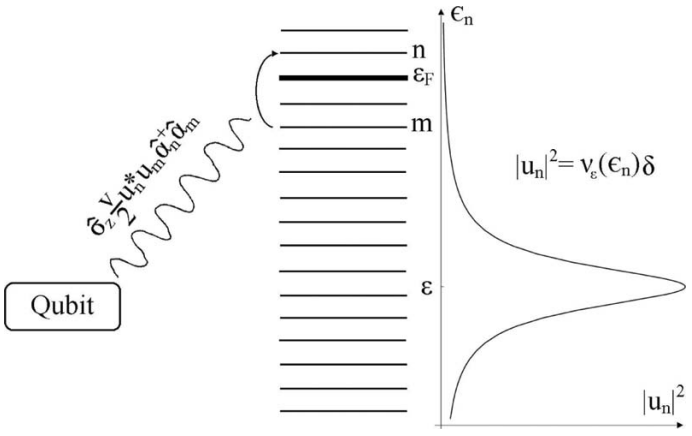

Fig. 5. Qubit is coupled to the transitions between all exact states of the bath Hamiltonian $\hat{H}_{\rm B}$, with weight being proportional to $u_n^* u_m$. Coefficients u_n^*, u_m are governed by the Lorentzian $|u_n|^2 = \nu_\varepsilon(\epsilon_n)\delta$

effect of such a high energy impurity can only be power law suppressed.

The second observation concerns very stringent geometrical restrictions on where an 'effective' impurity can be and a broad distribution of energies of impurities. In order to allow tunneling to and from the impurity, it should not be too far away from the metallic gate, which implies that the total volume available for 'effective' impurities cannot exceed $1\,\mu\mathrm{m} \times (10\lambda_\mathrm{F})^2$, where $1\,\mu m$ is a characteristic size of a metallic gate and λ_F is the Fermi wavelength in the metal. If we are hoping to find at least one low energy $|\varepsilon_i| < T$ impurity in this volume, the total number of impurities there should be of order of D/T, where D is the width of the distribution function of energies. For temperature which varies in experiments between 30 mK and 50 mK, and for D having a typical chemical value of 1 eV the above ratio is of order of 10^5. Having that many impurities in a volume which in total accommodates $\sim 10^6$ atoms is completely unrealistic.

So, the contribution of high energy impurities is suppressed but not strongly enough in order to exclude them from consideration, and low energy impurities effectively do not exist. If for a particular sample there happened to be one, this sample will exhibit orders of magnitude stronger decoherence, and should better be discarded.

The condition $|\varepsilon_i| \gg T$ enforces low temperature regime which requires a full quantum mechanical treatment. Such a treatment, valid at any temperature, will be offered in the next Section.

5 Exact Solution for FBC Model at Arbitrary Temperature

5.1 Calculations

To calculate the decoherence rate we will need to know how the off-diagonal elements of the reduced density matrix of the qubit decay with time. The dynamics of the full density matrix of the whole system qubit + bath is given by the Heisenberg equation of motion:

$$\frac{\partial \hat{\rho}(t)}{\partial t} = -\mathrm{i}\left[\hat{H}, \hat{\rho}(t)\right] , \tag{22}$$

where $\hat{H}$ is the Hamiltonian given by the (18) with Josephson energy being taken equal to zero. Since in the longitudinal ($E_J = 0$) model the interaction term commutes with the qubit Hamiltonian, one can trace out the bath degrees of freedom thus yielding the following result for the reduced density matrix of the qubit:

$$\hat{\rho}^{\,(q)}(t) = \begin{pmatrix} \rho_{11}^{(q)}(0) & \rho_{12}^{(q)}(0)\,\mathrm{e}^{-\mathrm{i}\delta E_c t} D(t) \\ \rho_{21}^{(q)}(0)\,\mathrm{e}^{\,\mathrm{i}\delta E_c t} D^*(t) & \rho_{22}^{(q)}(0) \end{pmatrix} , \tag{23}$$

where the standard separable initial condition for the full density matrix, $\hat{\rho}(0) = \hat{\rho}^{(q)}(0) \otimes \hat{\rho}_B$; $\hat{\rho}_B = \mathrm{e}^{-\beta \hat{H}_\mathrm{B}} / \operatorname{Tr} \mathrm{e}^{-\beta \hat{H}_\mathrm{B}}$, was assumed. The diagonal elements of the reduced density matrix do not evolve since there are no spin flip processes present, and the dynamics of the off-diagonal elements is governed by the decoherence function $D(t)$:

$$D(t) = \left\langle \mathrm{e}^{\mathrm{i}(\hat{H}_\mathrm{B}+\hat{V})t} \mathrm{e}^{-\mathrm{i}(\hat{H}_\mathrm{B}-\hat{V})t} \right\rangle , \tag{24}$$

where averaging should be performed with the bath Hamiltonian $\hat{H}_\mathrm{B}$. At long times t the decoherence function must decay exponentially, $D(t) \sim \mathrm{e}^{-\Gamma_2 t}$, so that the decoherence rate Γ_2 is defined as

$$\Gamma_2 = -\Re\mathrm{e} \lim_{t\to\infty} t^{-1} \ln D(t) , \tag{25}$$

in the agreement with equation (9).

The average in (24) can be represented as the following functional integral with the Grassmann fields ξ and η defined on the Keldysh contour:

$$D(t) = Z^{-1} \int \mathcal{D}\xi^* \mathcal{D}\xi \mathcal{D}\eta^* \mathcal{D}\eta \, \mathrm{e}^{\mathrm{i}S_0[\xi]} \mathrm{e}^{\mathrm{i}S_0[\eta]} \mathrm{e}^{\frac{\mathrm{i}}{2}\int\limits_{c_K} \sum\limits_i v_i(t')\xi_i^*(t')\xi_i(t')\mathrm{d}t'} \times$$
$$\times \, \mathrm{e}^{-\mathrm{i}\int\limits_{c_K} \sum\limits_{\boldsymbol{k},i} \left[t_{\boldsymbol{k}i}\eta_{\boldsymbol{k}}^*(t')\xi_i(t') + t_{\boldsymbol{k}i}^*\xi_i^*(t')\eta_{\boldsymbol{k}}(t')\right]\mathrm{d}t'} . \tag{26}$$

Here Z is the same functional integral but with $v_i(t') \equiv 0$, the fields ξ_i, ξ_i^* correspond to the localized state on the i-th impurity, $\eta_{\boldsymbol{k}}$, $\eta_{\boldsymbol{k}}^*$ to the conduction electrons, the impurity $S_0[\xi]$ and free electrons $S_0[\eta]$ actions are given by:

$$S_0[\xi] = \int\limits_{c_K} \sum_i \xi_i^*(t')(\mathrm{i}\partial_{t'} - \varepsilon_i^0)\xi_i(t')\mathrm{d}t'$$
$$S_0[\eta] = \int\limits_{c_K} \sum_{\boldsymbol{k}} \eta_{\boldsymbol{k}}^*(t')(\mathrm{i}\partial_{t'} - \varepsilon_{\boldsymbol{k}})\eta_{\boldsymbol{k}}(t')\mathrm{d}t' , \tag{27}$$

and the Keldysh time-dependent coupling $v_i(t')$ is zero everywhere on the contour apart from the interval $(0, t)$ where on the upper branch it takes the value of v_i, and on the lower branch it is equal to $-v_i$ (see Fig. 6). Since the action $S_0[\eta]$ is quadratic and the hybridization term is linear in η fields, it is straightforward to integrate out the conduction electrons:

$$D(t) = Z_0^{-1} \int \mathcal{D}\xi^* \mathcal{D}\xi \, \mathrm{e}^{\mathrm{i}S_0[\xi]} \mathrm{e}^{-\mathrm{i}\iint\limits_{c_K} \sum\limits_{ij} \xi_i^*(t_1)\Sigma_{ij}(t_1,t_2)\xi_j(t_2)\mathrm{d}t_1\mathrm{d}t_2}$$
$$\mathrm{e}^{\frac{\mathrm{i}}{2}\int\limits_{c_K} \sum\limits_i v_i(t_1)\xi_i^*(t_1)\xi_i(t_1)\mathrm{d}t_1} . \tag{28}$$

Again, here Z is the same functional integral but with $v_i(t') \equiv 0$. All the information about the conduction electrons is now encoded in the self-energy matrix which is defined on the contour:

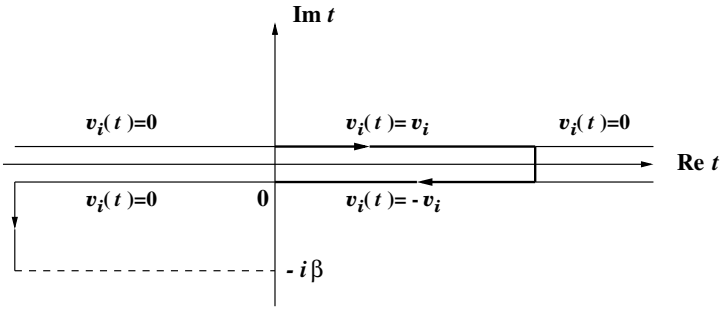

Fig. 6. Dependence of $v_i(t)$ on time along the Keldysh contour

$$\Sigma_{ij}(t_1,t_2)=\sum_{\boldsymbol{k}} t^*_{\boldsymbol{k}i}t_{\boldsymbol{k}j}g_{\boldsymbol{k}}(t_1,t_2)\,,\qquad t_1,t_2\in c_K \tag{29}$$

with $g_{\boldsymbol{k}}(t_1,t_2)$ being the Green function of free electrons. The full action in (28) is quadratic and therefore the corresponding functional integral can be written in a standard symbolic trace-log notation:

$$D(t)=\mathrm{e}^{\,\mathrm{Tr}\ln\left[\hat{1}+\hat{G}\hat{U}\right]}\,, \tag{30}$$

where both $\hat{G}$ and $\hat{U}$ are matrices in the space of impurity indices, and both depend on two times along the Keldysh contour. The symbol Tr here stands for trace with respect to the impurity indices and for integration along the Keldysh contour over all the times involved. The Green function $\hat{G}$ obeys the following integro-differential equation:

$$\int\limits_{c_K}\mathrm{d}t'\sum_m\left[\delta_{im}\delta(t_1,t')\left(\mathrm{i}\frac{\partial}{\partial t'}-\varepsilon^0_m\right)-\Sigma_{im}(t_1,t')\right]G_{mj}(t',t_2)=\delta_{ij}\delta(t_1,t_2)\,, \tag{31}$$

with δ-function $\delta(t_1,t_2)$ being defined on the contour. The impurity matrix $\hat{U}$ is made out of the time-dependent coupling $v_i(t')$:

$$U_{ij}(t_1,t_2)=\delta_{ij}\delta(t_1,t_2)\frac{v_i(t_2)}{2}\,. \tag{32}$$

Being written explicitly, the symbolic expression in (30) becomes

$$D(t)=\mathrm{e}^{-\sum\limits_{n=1}^{\infty}\frac{(-1)^n}{n}\,\mathrm{tr}\int\limits_{c_K}\cdots\int\limits_{\text{2n integrals}}\mathrm{d}t_1\mathrm{d}t'_1\ldots\mathrm{d}t_n\mathrm{d}t'_n\hat{G}(t_1,t'_1)\hat{U}(t'_1,t_2)\ldots\hat{G}(t_n,t'_n)\hat{U}(t'_n,t_1)} \tag{33}$$

where tr stands for a trace over impurity indices only.

Since the matrix $\hat{U}(t_1,t_2)$ as function of its both times is non-vanishing only between 0 and t, all the contour integrals in (33) can be converted into

ordinary integrals from 0 to t by introducing the Keldysh structure for both matrices $\hat{G}$ and $\hat{U}$. Both arguments of $\hat{G}(t_1,t_2)$ can either be on the upper or on the lower branch of the Keldysh contour leaving four choices which generate the following Keldysh structure:

$$G_{ij}(t_1,t_2)_{t_{1,2}\in c_K} \mapsto \check{\mathcal{G}}_{ij}(t_1-t_2)_{t_{1,2}\in(-\infty,+\infty)} = \begin{pmatrix} \mathcal{G}_{ij}(t_1-t_2) & \mathcal{G}^{<}_{ij}(t_1-t_2) \\ \mathcal{G}^{>}_{ij}(t_1-t_2) & \widetilde{\mathcal{G}}_{ij}(t_1-t_2) \end{pmatrix} \tag{34}$$

where $\mathcal{G}_{ij}(t_1-t_2)$ and $\widetilde{\mathcal{G}}_{ij}(t_1-t_2)$ are the time-ordered and anti-time-ordered Green functions. Using the Keldysh structure for the $\hat{U}$-matrix,

$$U_{ij}(t_1,t_2)_{t_{1,2}\in c_K} \mapsto \tfrac{1}{2}\delta_{ij}\,\hat{\sigma}_z\,\delta(t_1-t_2)\,v_i(t_2)_{t_{1,2}\in(-\infty,+\infty)}\ ,$$

and the general rule,

$$\int\limits_{c_K} A(t_1,t')B(t',t_2)\mathrm{d}t'\Big|_{t_{1,2}\in c_K} \mapsto \int\limits_{-\infty}^{+\infty} \check{A}(t_1,t')\hat{\sigma}_z\check{B}(t',t_2)\mathrm{d}t'\Big|_{t_{1,2}\in(-\infty,+\infty)}\ ,$$

we rewrite the expression (33) for the decoherence function as follows:

$$D(t) = \mathrm{e}^{-\sum\limits_{n=1}^{\infty}\frac{(-1)^n}{n2^n}\,\mathrm{Tr}\int\limits_0^t\ldots\int\limits_0^t \mathrm{d}t_1\mathrm{d}t_2\ldots\mathrm{d}t_n\check{\mathcal{G}}(t_1-t_2)(\hat{v}\otimes\hat{1})\check{\mathcal{G}}(t_2-t_3)(\hat{v}\otimes\hat{1})\ldots\check{\mathcal{G}}(t_n-t_1)(\hat{v}\otimes\hat{1})}\ . \tag{35}$$

Now the symbol Tr stands for the trace over both impurity and Keldysh matrix indices, while $(\hat{v}\otimes\hat{1})$ denotes the matrix with elements v_i on the main diagonal in the impurity space and which is the unity matrix in the Keldysh space.

It is convenient to do a standard rotation [20] in Keldysh space:

$$\hat{\mathcal{G}}_{ij} = \hat{L}\hat{\sigma}_z\check{\mathcal{G}}_{ij}\hat{L}^{\dagger}; \qquad \hat{L} = \frac{1}{\sqrt{2}}(\,\hat{1} - \mathrm{i}\hat{\sigma}_y)\ , \tag{36}$$

which converts the four-element matrix in (34) into the three-element one of retarded, advanced and Keldysh Green functions:

$$\hat{\mathcal{G}}_{ij} = \begin{pmatrix} G^R_{ij} & G^K_{ij} \\ 0 & G^A_{ij} \end{pmatrix}\ . \tag{37}$$

In the rotated basis the expression (35) for the decoherence function becomes:

$$D(t) = \mathrm{e}^{-\sum\limits_{n=1}^{\infty}\frac{(-1)^n}{n2^n}\,\mathrm{Tr}\int\limits_0^t\ldots\int\limits_0^t \mathrm{d}t_1\mathrm{d}t_2\ldots\mathrm{d}t_n(\hat{v}\otimes\hat{\sigma}_x)\hat{\mathcal{G}}(t_1-t_2)(\hat{v}\otimes\hat{\sigma}_x)\hat{\mathcal{G}}(t_2-t_3)\ldots(\hat{v}\otimes\hat{\sigma}_x)\hat{\mathcal{G}}(t_n-t_1)} \tag{38}$$

Since due to the time translation invariance each $\hat{G}$ depends only on the difference of its time arguments, the n^{th} order integrand depends on $n-1$ differences in times, while the integration over the last time variable produces the overall factor of t. The region of integration becomes, at arbitrary time, quite complicated. However, when time t is much bigger than the characteristic time on which $\hat{\mathcal{G}}(\tau)$ function decays (which is true if $t \gg \gamma_i^{-1}, T^{-1}$), all the integrals over the time differences can be extended to the entire axis. Then the integral has a convolution structure in time and, performing a Fourier transform and a straightforward summation that restores the logarithm, it finally reduces to

$$D(t) = \mathrm{e}^{\int\limits_{-\infty}^{+\infty} \frac{\mathrm{d}\omega}{2\pi} t\, \mathrm{Tr}\ln\left[\hat{1}+\frac{1}{2}(\hat{v}\otimes\hat{\sigma}_x)\hat{\mathcal{G}}(\omega)\right]} . \tag{39}$$

Using the definition of decoherence rate (25) and calculating trace in the Keldysh space explicitly we have:

$$\Gamma_2 = -\Re\mathrm{e} \int\limits_{-\infty}^{+\infty} \frac{\mathrm{d}\omega}{2\pi}\ \mathrm{tr}\ln\left[\hat{1} + \frac{\hat{v}}{2}\,\hat{G}^K(\omega) - \frac{\hat{v}}{2}\,\hat{G}^R(\omega)\frac{\hat{v}}{2}\,\hat{G}^A(\omega)\right] , \tag{40}$$

where tr refers only to the impurity matrix indices.

It should be stressed that the analytical structure of the expression in the r.h.s. of (40) ensures that there is no decoherence at zero temperature. Indeed, using the standard relation [20] between the components of Keldysh Green functions at equilibrium, $\hat{G}^K(\omega) = \tanh(\beta\omega/2)[\hat{G}^R(\omega) - \hat{G}^A(\omega)]$, which at $T = 0$ reduces to $\hat{G}^K(\omega) = \mathrm{sgn}(\omega)[\hat{G}^R(\omega) - \hat{G}^A(\omega)]$, we have:

$$\Gamma_2(T=0) = -\Re\mathrm{e} \int\limits_{-\infty}^{+\infty} \frac{\mathrm{d}\omega}{2\pi}\,\mathrm{tr}\ln\left[\left(\hat{1} + \frac{\hat{v}}{2}\,\hat{G}^R(\omega)\right)\left(\hat{1} - \frac{\hat{v}}{2}\,\hat{G}^A(\omega)\right)\right] . \tag{41}$$

Separating the retarded and advanced parts in the expression above, and then expanding the logarithm in both of them, we see that starting from the second order all the integrals are convergent at $\omega \to \infty$ and therefore equal to zero since the retarded (advanced) Green function is analytic in the upper (lower) half plane. Combining the first order contributions from both parts we see that it is purely imaginary and therefore also vanishes.

Assuming that the impurity states are well localized, the tunneling amplitudes $t_{\boldsymbol{k}i}$ can be written as $t_{\boldsymbol{k}i} = t_i \mathrm{Vol}^{-1/2}\mathrm{e}^{\mathrm{i}\boldsymbol{k}\boldsymbol{r}_i}$, where $\boldsymbol{r}_i$ is the position of the i-th impurity. This in turn means that the self-energy matrix (29) has the following representation:

$$\hat{\Sigma}_{ij}(t_1 - t_2) = t_i^* t_j\, \hat{g}(t_1 - t_2; \boldsymbol{r}_j - \boldsymbol{r}_i) . \tag{42}$$

Since the characteristic distance l between the impurities is much bigger than the Fermi wavelength, $l \gg \lambda_{\mathrm{F}}$, all the off-diagonal elements of the self-energy matrix will be averaged out while diagonal can be represented as

$$\Sigma_{ii}^{A/R}(\omega) = \left[\sum_{\boldsymbol{k}} \frac{|t_{\boldsymbol{k}i}|^2}{\omega - \varepsilon_{\boldsymbol{k}}} \pm \mathrm{i}\pi \sum_{\boldsymbol{k}} |t_{\boldsymbol{k}i}|^2 \delta(\omega - \varepsilon_{\boldsymbol{k}})\right] = \left[\frac{\alpha_i(\omega)}{2} \pm \mathrm{i}\frac{\gamma_i(\omega)}{2}\right] . \tag{43}$$

This allows one to solve (the Keldysh rotated analogue of) the equation (31):

$$G_{ij}^{A/R}(\omega) = \frac{\delta_{ij}}{\omega - \varepsilon_i^0 - \Sigma_{ii}^{A/R}(\omega)} = \frac{\delta_{ij}}{\omega - \varepsilon_i \mp \mathrm{i}\gamma_i/2} , \tag{44}$$

where both the real and imaginary parts of the self-energy matrix were assumed ω-independent, and the real part was absorbed by renormalization of the energy levels $\varepsilon_i = \varepsilon_i^0 + \alpha_i/2$. In such a diagonal approximation for the self-energy, the contributions from all the impurities are independent and the total decoherence rate is given by their sum. Substituting (44) for the Green functions into (40), subtracting (to improve the integral convergency) the identically zero expression (41) for Γ_0 and taking the real part of the resulting expression, we obtain the following contribution of a single impurity at energy $\varepsilon_j \equiv \varepsilon$ to the decoherence rate:

$$\Gamma_2(T) = -\int_{-\infty}^{+\infty} \frac{\mathrm{d}\omega}{4\pi} \ln\left[1 - \frac{v^2\gamma^2 \cosh^{-2}(\omega/2T)}{v^2\gamma^2 + 4\left[(\omega-\varepsilon)^2 + \gamma^2/4 - v^2/4\right]^2}\right] . \tag{45}$$

5.2 Discussion of the Results

It is convenient to introduce an auxiliary 'spectral function' $\Lambda(\omega)$,

$$\Lambda(\omega) = \frac{1}{1 + 4(\omega-\varepsilon_+)^2(\omega-\varepsilon_-)^2/v^2\gamma^2} , \qquad \varepsilon_\pm = \varepsilon \pm \frac{1}{2}\sqrt{v^2 - \gamma^2} , \tag{46}$$

thus rewriting (45) as follows:

$$\Gamma_2(T) = -\int_{-\infty}^{+\infty} \frac{\mathrm{d}\omega}{4\pi} \ln\left[1 - \Lambda(\omega) \cosh^{-2}\frac{\omega}{2T}\right] . \tag{47}$$

Behavior of $\Lambda(\omega)$ is qualitatively different depending on whether $g \equiv v/\gamma < 1$ or $g > 1$. In the former case $\Lambda(\omega)$ is peaked at $\omega = \varepsilon$, while in the latter the peak splits into two Lorentzians which are centered at $\omega = \varepsilon_\pm$ and travel in different directions away from the $\omega = \varepsilon$ point as v increases. This signals a qualitative change between the weak coupling ($g \ll 1$) and the strong coupling ($g \gg 1$) regimes.

The decoherence rate is determined by the overlap of the temperature function, $\cosh^{-2}(\omega/2T)$, and $\Lambda(\omega)$ in (47). The former is centered at $\omega = 0$ and exponentially decays for $|\omega| > T$, while the latter has the power-law decay away from the above described peaks. If the temperature function is wide and

the Lorentzian peaks are sitting inside T, the overlap is solely controlled by $\Lambda(\omega)$ as if the temperature function was equal to 1 everywhere. In the opposite case when the temperature function is narrow either with respect to the distance to the Λ-function peaks or to their width, the overlap is controlled by $\cosh^{-2}(\omega/2T)$ and $\Lambda(\omega)$ can be replaced by $\Lambda(\omega = 0)$. We will be calling these two cases high-temperature ($T \gg \max|\varepsilon_{\pm}|, \gamma$) and low-temperature ($T \ll \min|\varepsilon_{\pm}|$ or $T \ll \gamma$) regimes correspondingly.

In the high-temperature regime, when either $T \gg \max|\varepsilon_{\pm}|$ or $T \gg \gamma$, the decoherence rate Γ_2 saturates [4, 5] at ε-independent value:

$$\Gamma_2^{(\mathrm{high})} = -\int\limits_{-\infty}^{+\infty} \frac{\mathrm{d}\omega}{4\pi} \ln\left[1 - \Lambda(\omega)\right] = \frac{\gamma}{2}\left[1 - \Re\mathrm{e}\sqrt{1 - g^2}\right] . \tag{48}$$

In the low-T regime, when either $T \ll \min|\varepsilon_{\pm}|$ or $T \ll \gamma$, the decoherence rate becomes linear in temperature:

$$\begin{aligned}\Gamma_2^{(\mathrm{low})}(T) &= -\int\limits_{-\infty}^{+\infty} \frac{\mathrm{d}\omega}{4\pi} \ln\left[1 - \Lambda(\omega = 0)\cosh^{-2}\frac{\omega}{2T}\right] \\ &= \frac{T}{\pi}\arctan^2\left(\frac{2g}{4\varepsilon^2/\gamma^2 - g^2 + 1}\right) . \end{aligned} \tag{49}$$

When both the coupling strength v and energy ε are smaller than the tunneling rate γ, the regime change occurs in the vicinity of $T \sim \gamma$ point (the upper curve on the insert in Fig. 7). If γ is smaller than either of the other two parameters: $\gamma \ll A = \max\{\varepsilon, v/2\}$, then the high- and low-temperature asymptotes parametrically mismatch. In this case there is a crossover between the two regimes which is exponentially fast and occurs in the logarithmically narrow interval of temperatures:

$$\Gamma_2^{(\mathrm{x-over})}(T) \sim \Gamma_2^{(\mathrm{high})}\mathrm{e}^{-A/T} , \qquad \frac{A}{\ln A/\gamma} \lesssim T \lesssim A . \tag{50}$$

The temperature dependence of the decoherence rate is shown in Fig. 7 for the case of different couplings and fluctuator energies. We should stress that although a linear in T behavior similar to that in the (49) would also follow from the spin-boson models with the ohmic spectral function, as briefly described in Sect. 2, only a full quantum mechanical treatment can result in a non-trivial T dependence depicted in Fig. 7.

Let us also stress that, although at high temperatures decoherence is always stronger for a strongly coupled fluctuator, this does not have to be the case as temperature is lowered, since in the crossover region the $g \gg 1$ fluctuators undergo stronger suppression than their $g \ll 1$ counterparts. Such a non-monotonic dependence of the decoherence rate Γ_2 on the coupling strength v outside the classical region of high T deserves special attention.

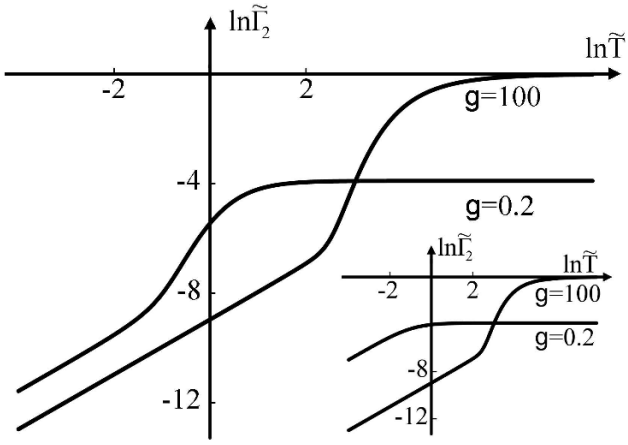

Fig. 7. Dependence of the decoherence rate on temperature, (47), for strong and weak coupling. The main picture shows a fluctuator with $\nu_\varepsilon(\omega)$ centered at $\tilde{\varepsilon} = 3$, the insert shows $\tilde{\varepsilon} = 0$. $\tilde{T}, \tilde{\Gamma}_2, \tilde{\varepsilon}$ are measured in units of $\gamma/2$

When we start with a small value of v, $v < \gamma$, the peak of $\Lambda(\omega)$ grows as v increases which leads to higher decoherence rate Γ_2. This proceeds on until the $v = \gamma$ point is reached when the peak stops growing but splits into two instead. One of those peaks then starts moving towards the origin which increases the overlap between the temperature function and $\Lambda(\omega)$, so Γ_2 keeps on growing. The maximum possible overlap is achieved when the incoming Lorentzian is centered at $\omega = 0$ which happens at the $v = \sqrt{4\varepsilon^2 + \gamma^2}$ point. If the coupling strength is increased beyond this point, the peak of $\Lambda(\omega)$ becomes de-tuned from the peak of the temperature function and the decoherence rate goes down. Since $\Lambda(\omega)$ plays the role of a spectral function, the positions of its maxima $\varepsilon_\pm$ give us the energies of charge states of the system. Pushing them away from the Fermi level leads to freezing them out so that the impurity is no longer a fluctuator and thus does not contribute to the decoherence rate. Note that such a strong coupling is still not a good news for the qubit operation as the qubit and the impurity form together a four-level system which, although remains coherent, does not operate as intended.

The above described behavior of the decoherence rate as a function of coupling strength is depicted in Fig. 8. Note that the maximum of Γ_2 has a cusp which is not smeared by temperature (but would be smeared by including the σ_x part into the Hamiltonian): Only at a rather high temperature ($\tilde{T} \sim 100$) the decoherence rate practically saturates at its classical limit $\Gamma_2^{(\mathrm{high})} = \gamma/2$.

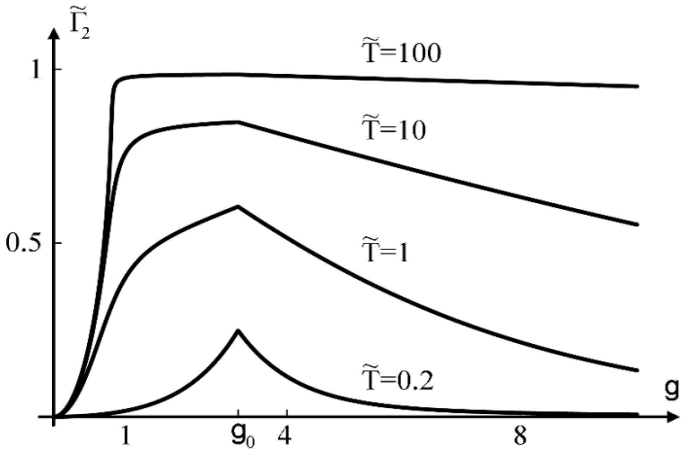

Fig. 8. Non-monotonic dependence of the decoherence rate on the coupling strength at different temperatures for $\tilde{\varepsilon} = 3$; here $g_0 = \sqrt{\tilde{\varepsilon}^2 + 1}$

6 Relaxation in the FBC Model

Apart from decoherence, coupling of a qubit to the environment also leads to relaxation processes, i.e., to a decay of diagonal elements of the density matrix (23), as outlined in Sect. 2. In the model under consideration, equation (18), the presence of the $\hat{\sigma}_x$-part inevitably induces relaxation. Such a relaxation would also destroy coherence but the relaxation rate Γ_1 is typically smaller than the decoherence rate Γ_2 and therefore the time available for quantum operations is going to be determined by the latter. Experiments on the relaxation rate are of special interest as they can provide valuable information about the degrees of freedom in the environment to which qubit is coupled.

It is convenient to diagonalize the qubit part of the Hamiltonian (18), transforming simultaneously the interaction part $\hat{\sigma}_z$ into a sum of $\hat{\tau}_z$ and $\hat{\tau}_x$ terms - the Pauli matrices in the new basis:

$$\hat{H} = \frac{\Delta E}{2}\,\hat{\tau}_z + (\cos\Theta\,\hat{\tau}_z + \sin\Theta\,\hat{\tau}_x)\,\hat{V} + \hat{H}_{\rm B}\,; \qquad \text{where}$$
$$\Delta E = \sqrt{(\delta E_c)^2 + E_J^2}\,, \quad \cos\Theta = \frac{\delta E_c}{\Delta E}\,, \quad \sin\Theta = \frac{E_J}{\Delta E}\,. \tag{51}$$

As it is well-known [10], in Born-Markov rotating-wave approximation the decoherence rate Γ_1 is given by:

$$\Gamma_1 = \sin^2\Theta\, S(\Delta E)\,, \qquad S(\Omega) = \left\langle \left\{ \hat{V}(t) - \langle \hat{V} \rangle, \hat{V}(t') - \langle \hat{V} \rangle \right\} \right\rangle_{\Omega}\,, \tag{52}$$

where $S(\Omega)$ is the spectral density of noise. (The appropriate derivation is outlined in Sect. 2). The symbol $\langle \ldots \rangle_{\Omega}$ denotes the Fourier transform of the bath-averaged expression with respect to $t - t'$ at the frequency Ω.

In the experiment of Astafiev et al. [8] the authors determined the spectral density of noise in a wide window of frequencies, by changing ΔE and measuring the relaxation rate. The experimental data for $S(\Omega)$ is presented in Fig. 9. Although the data points are somewhat scattered, the authors nevertheless

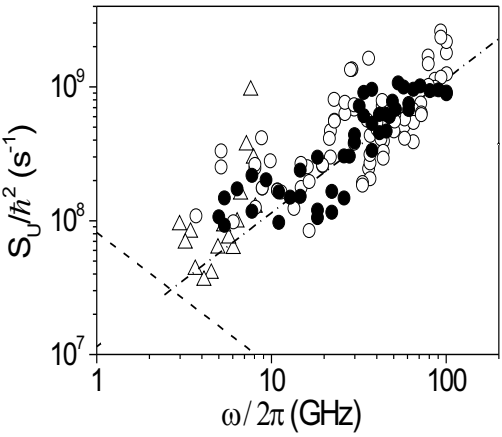

Fig. 9. The experimental data by Astafiev et al. [8]: the noise spectral density S deduced from measurements of Γ_1; open and closed symbols correspond to different samples while circles and triangles – to different measurement regimes; $T \approx 50\,\mathrm{mK}$

claim that they observed a linear behavior of spectral density of noise with peaks at 7 and 23 GHz.

Below we will demonstrate that qualitatively the same behavior is expected from $S(\Omega)$ in the FBC model at low temperatures. A contribution to the spectral density of noise from one impurity is given by

$$S(\Omega) = \frac{v^2}{4} \left\langle \left\{ \hat{d}^\dagger(t)\hat{d}(t) - \langle \hat{d}^\dagger \hat{d} \rangle \, , \, \hat{d}^\dagger(t')\hat{d}(t') - \langle \hat{d}^\dagger \hat{d} \rangle \right\} \right\rangle_\Omega =$$

$$= \frac{\pi v^2}{2} \coth \frac{\Omega}{2T} \int\limits_{-\infty}^{+\infty} \mathrm{d}\omega \, \nu_\varepsilon(\omega)\nu_\varepsilon(\omega+\Omega) \left[n_\mathrm{F}(\omega) - n_\mathrm{F}(\omega+\Omega) \right] \, , \qquad (53)$$

where the density of states $\nu_\varepsilon(\omega)$ is given by (21), and $n_\mathrm{F}(\omega)$ is the Fermi distribution function.

The frequencies accessed in the experiment [8] are in the window $T \ll \Omega < 100T$. Getting back to the rough estimates for the probability of having an effective impurity with energy smaller than temperature (the end of Sect. 4), we see that even a scenario with $|\varepsilon_i| \gg 100T$ for all effective impurities is quite likely. If this is the case then the condition $T \ll \Omega \ll |\varepsilon|$ holds and it is clear that the spectral density of noise $S(\Omega)$ exhibits linear behavior at such

frequencies. As the thermal factor is practically equal to 1 within the range $-\Omega < \omega < 0$ and zero otherwise, the smearing of this step is of the order of temperature. Far to the right (assuming ε to be positive) we have two peaks in DoS at $\varepsilon - \Omega$ and ε. The main contribution to the integral (53) is coming from the tails of ν_ε at the interval $-\Omega < \omega < 0$, rather than from the thermal function tail in the vicinity of DoS peaks which is exponentially suppressed. The latter contribution would only win if γ itself was exponentially small, but the corresponding impurity is obviously non-effective. If γ is not too small, we find

$$S(\Omega) = \frac{\pi v^2}{2}\, \nu_\varepsilon^2(0)\Omega\,, \qquad T \ll \Omega \ll \varepsilon\,. \tag{54}$$

The summation of contributions (54) coming from all the effective impurities produces a linear in Ω function, which can be a possible explanation for a linear (but very noisy) trend in $S(\Omega)$ observed in [8].

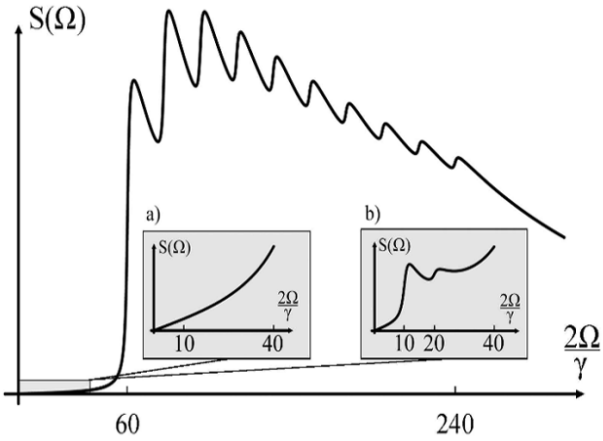

Fig. 10. The noise spectral density $S(\Omega)$ in arbitrary units at $T = 0$ for the case of ten impurities with energies uniformly distributed between $\tilde{\varepsilon} = 60$ and $\tilde{\varepsilon} = 240$. The inserts zoom the bottom left corner (the grey area on the main figure for $\tilde{\varepsilon} \leq 40$) when there is either a) no impurity in this interval or b) two impurities ($\tilde{\varepsilon}_1 = 10$, $\tilde{\varepsilon} = 20$) with coupling constants $v_{1,2}$ 50 times weaker than those for $60 \leq \tilde{\varepsilon} \leq 240$

Humps in Fig. 9 can be coming from less effective but more numerous impurities which makes the situation $T \ll |\varepsilon| < 100T$ possible. At the point $\Omega \sim \varepsilon$ the left one of the DoS double-Lorentzian overlaps with the thermal function thus producing a peak in $S(\Omega)$. As Ω further increases, the overlap between the left Lorentzian and the thermal function stays essentially the same but the region where it occurs drifts away to the left from the second Lorentzian which is stationary positioned at $\omega = \varepsilon$. Qualitatively the picture

of a linear behavior of $S(\Omega)$ followed by a peak at $\Omega \sim \varepsilon$ is the same for any relation between γ and T, but the form of the peak is different.

With a purely illustrative purpose we show in Fig. 10 a picture corresponding to the following scenario: all the tunneling rates are the same, $\gamma_i = \gamma$, obeying $\gamma \gg T$ (which effectively allows one to put T to zero), and several impurities are uniformly distributed in some interval of energies. All these impurities contribute to the linear behavior on the left of this interval [the grey area zoomed in the insert a)], while humps might be due to fluctuators with smaller energies weakly coupled to the qubit which could not break the general linear trend [the grey area zoomed in the insert b)]. The behavior of $S(\Omega)$ in the grey area qualitatively resembles the experimental data.

7 Conclusions

In this review, we have described some essential features of loss of coherence by a qubit coupled to the environment. We have first presented well-known semiclassical arguments that relate both decoherence and relaxation to the environmental noise. Then we have shown that models with pure decoherence (when there is no relaxation in qubit states as the part of coupling to the environment that excludes flipping of the states) can be exactly solvable. As an example, we have treated in detail the model of fluctuating background charges [4–7] which is believed to describe the most important channel for decoherence for the charge Josephson junction qubit. Following our earlier treatment [7], we have shown that the decoherence rate saturates at 'high' temperatures [4,5] while becoming linear in T at low temperatures and showing in all regimes a non-monotonic behavior as a function of the coupling of the qubit to the fluctuating background charges. We have also considered, albeit only perturbatively, the qubit relaxation by the background charges and demonstrated that the quasi-linear behavior of the spectral density of noise deduced from the measurements of the relaxation rate can be qualitatively explained within this model in the low temperature regime.

Acknowledgments

We thank B. L. Altshuler, Y. M. Galperin, R. Fazio and A. Shnirman for useful comments. One of us (I.V.L.) is thankful to Professor S. N. Karmakar for warm hospitality extended to him during this workshop at Saha Institute of Nuclear Physics in Kolkata. This work was supported by the EPSRC grant GR/R95432.

References

1. P. W. Shor: SIAM Journal on Computing **26**, 1484 (1997)

2. L. K. Grover: Phys. Rev. Lett. **79**, 325 (1997)
3. D. P. DiVincenzo, D. Loss: J. Magn. Magn. Mater. **200**, 202 (1999); Phys. Rev. B **71**, 035318 (2005)
4. E. Paladino, L. Faoro, G. Falci, R. Fazio: Phys. Rev. Lett. **88**, 228304 (2002)
5. Y. M. Galperin, B. L. Altshuler, D. V. Shantsev: cond-mat/0312490 (2003)
6. Y. M. Galperin, B. L. Altshuler, J. Bergli, D. V. Shantsev: Phys. Rev. Lett. **96**, 097009 (2006)
7. A. Grishin, I. V. Yurkevich, I. V. Lerner: Phys. Rev. B **72**, 060509 (2005)
8. O. Astafiev, Y. A. Pashkin, Y. Nakamura, T. Yamamoto, J. S. Tsai: Phys. Rev. Lett. **93**, 267007 (2004)
9. A. J. Leggett, S. Chakravarty, A. T. Dorsey, M. P. A. Fisher, A. Garg, W. Zwerger: Rev. Mod. Phys. **59**, 1 (1987)
10. H. P. Breuer, F. Petruccione: *The Theory of Open Quantum Systems* (Oxford University Press, Oxford 2002)
11. D. P. DiVincenzo: In: *Mesoscopic Electron Transport, NATO ASI Series E: Applied Sciences*, vol 345, ed by L. Kouwenhoven et al. (Kluwer, Dordrecht 1997) p. 657
12. Y. Makhlin, G. Schon, A. Shnirman: Rev. Mod. Phys. **73**, 357 (2001)
13. Y. Nakamura, Y. A. Pashkin, J. S. Tsai: Nature **398**, 786 (1999)
14. Y. A. Pashkin, T. Yamamoto, O. Astafiev, Y. Nakamura, D. V. Averin, J. S. Tsai: Nature **421**, 823 (2003)
15. C. H. van der Wal, A. C. J. ter Haar, F. K. Wilhelm, R. N. Schouten, C. J. P. M. Harmans, T. P. Orlando, S. Lloyd, J. E. Mooij: Science **290**, 773 (2000)
16. D. Vion, A. Aassime, A. Cottet, P. Joyez, H. Pothier, C. Urbina, D. Esteve, M. H. Devoret: Science **296**, 886 (2002)
17. Y. Yu, S. Y. Han, X. Chu, S. I. Chu, Z. Wang: Science **296**, 889 (2002)
18. L. Faoro, J. Bergli, B. L. Altshuler, Y. M. Galperin: Phys. Rev. Lett. **95**, 046805 (2005)
19. G. D. Mahan: *Many-Particle Physics* (Plenum, New York 1990)
20. J. Rammer, H. Smith: Rev. Mod. Phys. **58**, 323 (1986)

Low-Energy Physical Properties of Edge States in Nanographite Systems

Katsunori Wakabayashi

Department of Quantum Matter, AdSM, Hiroshima University, Higashi-Hiroshima 739-8530, Japan. kwaka@hiroshima-u.ac.jp

1 Introduction

Recent advances in nanotechnology make it possible to fabricate ultra small artificial physical systems like quantum dot, quantum interferometer, quantum wire, etc. in which quantum effects are experimentally observable. Both from the perspective of fundamental physics or potential applications, these artificial systems have generated a lot of excitement as they enabled the realization of a remarkable variety of physical phenomena such as the quantum Hall effect, ballistic transport, Aharonov-Bohm effect, universal conductance fluctuation, Kondo effect [1] etc. arising out of the quantum effects. Among such artificial systems, the nanoscopic carbon systems like carbon nanotubes [2–4] and nanographite [5–7] have received enormous attention not only for their intriguing form, but also for their unusual physical properties. In these systems, the geometry of sp^2 carbon networks crucially affects the electronic states near Fermi surface [8–10]. Studies with scanning tunneling microscopy and spectroscopy have confirmed the connection between the electronic states of single wall carbon nanotubes (SWCN) and their geometry [11, 12]. The nanometer-sized systems with open boundaries called "nanographites," display many unusual features, where the shapes of their edges have very important role in determining the electronic structure of these systems [5]. The edges of the nanographite systems can have two typical shapes, the armchair and the zigzag edges. Zigzag edges leads to localized edge states near Fermi level, while such localized states are completely absent in nanographite systems with armchair edges. These edge states correspond to the non-bonding molecular orbitals which generate large peaks in the density of states and have pronounced effect on the electronic, magnetic and transport properties of nanographite systems [5,13–18]. Recent experiments on nanographites have reported direct observation of the edge states [19], the paramagnetic response at low-temperature [20, 21], and the ferromagnetic response at the graphite

edges [22, 23]. The purpose of this article is to elucidate the role of the edge states on the low-energy physical properties of nanographite systems.

2 Electronic States of Nanographite Ribbons

There are two typical shapes for a graphite edge, the *armchair* and the *zigzag* edges as illustrated in Fig. 1(a). The two edges have 30 degrees difference in their cutting directions. Here we briefly discuss how the graphite edges drastically change the π-electronic structures. Especially, a zigzag edge provides the localized edge states, while an armchair edge does not give such localized states. We hereafter term these states as the "edge states" [5].

The graphite ribbons with zigzag and armchair edges are usually described by the simple geometrical structures as depicted in Fig. 1(b) and Fig. 1(c), respectively. The width N of graphite ribbon is taken as the number of dimer lines (two carbon sites) or the number of zigzag lines corresponding to the armchair and the zigzag ribbons, respectively. It is assumed that all dangling bonds at the edges are terminated by hydrogen atoms, and thus they have no contribution to the electronic states near the Fermi level. Throughout this article, it is also assumed that graphite ribbons have the translational invariance, and the unit cells can be defined as illustrated in Fig. 1(b) and Fig. 1(c). We employ the single-orbital tight-binding model for the π-electron network. This model has been successfully used for the calculation of electronic structure of fullerene molecules, carbon nanotubes and other carbon-related materials [9, 24]. The tight-binding Hamiltonian is given by,

$$H = \sum_{\langle i,j \rangle} t_{ij} c_i^\dagger c_j \ , \tag{1}$$

where $c_i^\dagger$ creates an electron at site i. Here $\langle i, j \rangle$ denotes the nearest neighbor sites. The transfer integrals between all the nearest neighbor sites are taken as t for simplicity. This model Hamiltonian adequately describes the intrinsic difference between the electronic structures of the topologically inequivalent armchair and zigzag graphite ribbons. The value of t is considered to be about 3.0 eV in a graphite ribbon [25].

Prior to our discussion of the π-electronic states of graphite ribbons, we first briefly review the π-band structure of a graphite sheet [24]. To diagonalize the Hamiltonian of a graphite sheet, we use a basis of two-component spinor $\boldsymbol{c}_{\boldsymbol{k}}^\dagger = (c_{A\boldsymbol{k}}^\dagger, c_{B\boldsymbol{k}}^\dagger)$, which is the Fourier transform of $(c_{i\in A}^\dagger, c_{i\in B}^\dagger)$. Let $\boldsymbol{a_1}$, $\boldsymbol{a_2}$, $\boldsymbol{a_3}$ be the displacement vectors from a B site to its three nearest-neighbor A sites, so defined that $\hat{\boldsymbol{z}} \cdot \boldsymbol{a_1} \times \boldsymbol{a_2}$ is positive (see Fig.2(a)). Here $\hat{\boldsymbol{z}}$ is a unit vector normal to the graphite sheet. In this representation Hamiltonian can be expressed as $H = \sum_{\boldsymbol{k}} \boldsymbol{c}_{\boldsymbol{k}}^\dagger H_{\boldsymbol{k}} \boldsymbol{c}_{\boldsymbol{k}}$, where $H_{\boldsymbol{k}}$ is given by,

$$H_{\boldsymbol{k}} = -t \sum_{i=1}^{3} (\cos(\boldsymbol{k} \cdot \boldsymbol{a_i}) \hat{\sigma}_x + \sin(\boldsymbol{k} \cdot \boldsymbol{a_i}) \hat{\sigma}_y) \ . \tag{2}$$

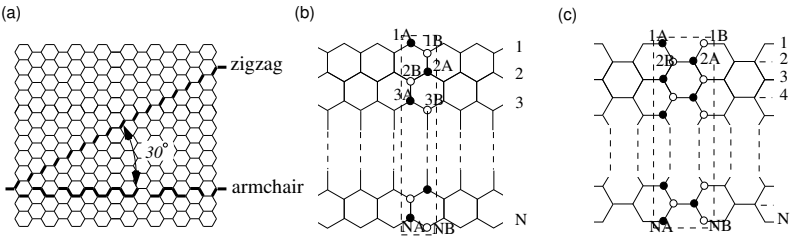

Fig. 1. (a) The typical graphite edge, *armchair* and *zigzag*. The structure of graphite ribbons with **(b)** zigzag edges and **(c)** armchair edges. The rectangle with the dashed lines is the unit cell

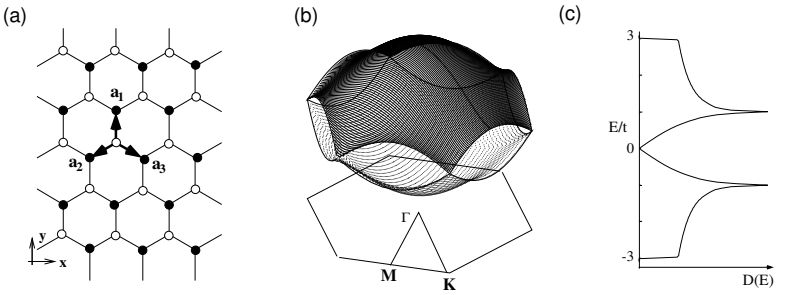

Fig. 2. (a) Graphite sheet in real space, where the black (white) circles mean the A(B)-sublattice sites. **(b)** The π band structure and **(c)** the density of states of graphite sheet. The valence and conduction bands make contact at the degeneracy point K

Here $\hat{\boldsymbol{\sigma}} = (\hat{\sigma}_x, \hat{\sigma}_y, \hat{\sigma}_z)$ are the Pauli matrices. The energy eigenvalues for this Hamiltonian are $E^{\pm}_{\boldsymbol{k}} = \pm t|\sum_{i=1}^{3} \exp(\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{a_i})|$ and only the $E^{-}_{\boldsymbol{k}}$–band will be completely filled as each carbon site has one π-electron on an average. Hereafter, we call $E^{-}_{\boldsymbol{k}}$ $[E^{+}_{\boldsymbol{k}}]$ as the valence [conduction] band.

In Fig. 2(b) and Fig. 2(c), the energy dispersion of π-bands in the 1st Brillouin Zone (BZ) and the density of states (DOS) are depicted, respectively. Near the $\boldsymbol{\Gamma}$ point, both valence and conduction bands have quadratic form in k_x and k_y, i.e., $E_{\boldsymbol{k}} = \pm(3 - 3|\boldsymbol{k}|^2/4)$. At the $\boldsymbol{M}$ points, the middle of the sides of hexagonal BZ, we have the saddle points of energy dispersion and here density of states diverges logarithmically. Near the $\boldsymbol{K}$ points, the corners of the hexagonal 1st BZ, the energy dispersion is linear in the magnitude of wave vector, and explicitly we have $E_{\boldsymbol{k}} = \pm 3ta|\boldsymbol{k}|/2$, where the density of states also linearly depends on energy. Here $a = |\boldsymbol{a_i}|$ $(i = 1, 2, 3)$. The Fermi energy is located at the $\boldsymbol{K}$ points and there is no energy gap at these points, since $E_{\boldsymbol{k}}$ vanishes at these points by the hexagonal symmetry. Thus, the graphite sheet is a zero-gap semi-conductor on this 2D model.

The calculated band structures of the armchair ribbons are shown in Figs. 3(a)-(c) for three different ribbon widths together with the density of states. The wave number k is normalized by the length of the primitive translation vector of each graphite ribbon, and the energy E is scaled by the transfer integral t. The top of the valence band and the bottom of the conduction band are located at $k = 0$. It should be noted that the ribbon width decides whether the system is metallic or semiconducting . As shown in Fig. 3(b), the system is metallic when $N = 3M - 1$, where M is an integer. For the semiconducting ribbons, the direct gap decreases with increasing ribbon width and tends to zero in the limit of very large N.

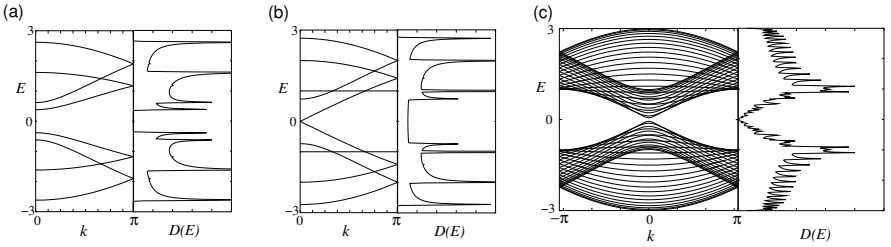

Fig. 3. Energy band structure $E(k)$ and density of states $D(E)$ of armchair ribbons of various widths [**(a)** $N = 4$, **(b)** 5 and **(c)** 30]

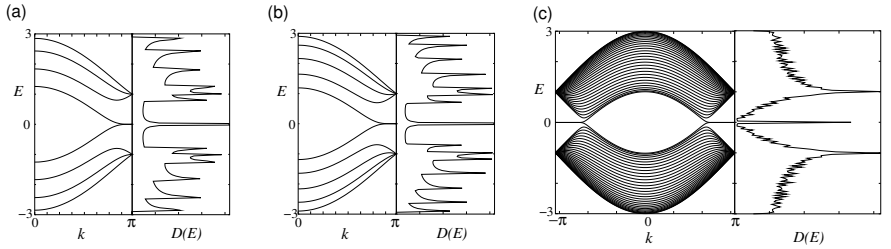

Fig. 4. Energy band structure $E(k)$ and density of states $D(E)$ of zigzag ribbons of various widths [**(a)** $N = 4$, **(b)** 5 and **(c)** 30]

For zigzag ribbons, however, a remarkable new feature arises in the band structure, as shown in Figs. 4(a)-(c). We see that the highest valence band and lowest conduction band are always degenerate at $k = \pi$. It is found that the degeneracy of the central bands at $k = \pi$ does not originate from the intrinsic band structure of the 2D graphite sheet. These two special central bands get flatten as we increase ribbon width. A pair of almost flat bands appears within

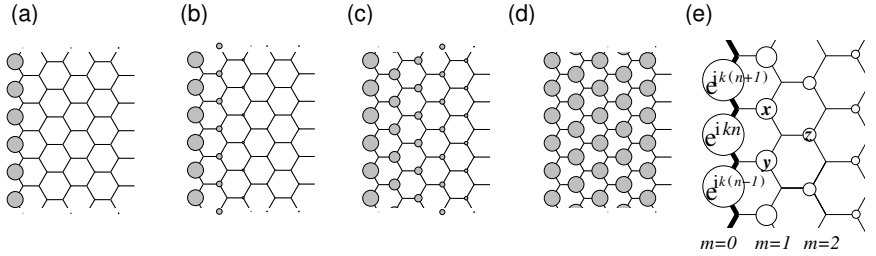

Fig. 5. Charge density plot form analytic solution of the edge states in a semi-infinite graphite, when **(a)** $k=\pi$, **(b)** $8\pi/9$, **(c)** $7\pi/9$ and **(d)** $2\pi/3$. **(e)** An analytic form of the edge state for a semi-infinite graphite sheet with a zigzag edge, emphasized by bold lines. Each carbon site is specified by a location index n on the zigzag chain and by a chain order index m from the edge. The magnitude of the charge density at each site, such as x, y and z, is obtained analytically (see text). The radius of each circle is proportional to the charge density on each site, and the drawing is made for $k = 7\pi/9$

the region of $2\pi/3 \leq |k| \leq \pi$, where the bands sit in the vicinity of the Fermi level. No such flat band is expected in the band structure of 2D graphite sheet.

The electronic states in the partial flat bands of the zigzag ribbons are found to be localized near the zigzag edges which become apparent from the charge density distribution [5, 13, 14, 26, 27]. The puzzle for the emergence of the edge states can be solved by considering a semi-infinite graphite sheet with a zigzag edge. First to have the analytic form for the edge states, in Fig. 5(a)-(d) we depict the distribution of charge density in the flat band states for some wave numbers, where amplitude is proportional to the radius. The wave function has non-bonding character, i.e., finite amplitudes only on one of the two sublattices that includes the edge sites. It is completely localized at the edge sites when $k=\pi$, and starts to penetrate gradually into the inner sites as k deviates from π and the state eventually becomes extended at $k=2\pi/3$.

Considering the translational symmetry, we can start constructing the analytic solution for the edge state by letting the Bloch components of the linear combination of atomic orbitals (LCAO) wave function be ..., $e^{ik(n-1)}$, e^{ikn}, $e^{ik(n+1)}$, ... on successive edge sites, where n denotes a site location on the edge. Then the mathematical condition necessary for the wave function to be exact for $E = 0$ is that the total sum of the components of the complex wave function over the nearest-neighbor sites should vanish. In Fig. 5(e), the above condition is $e^{ik(n+1)} + e^{ikn} + x = 0$, $e^{ikn} + e^{ik(n-1)} + y = 0$ and $x + y + z = 0$. Therefore, the wave function components x, y and z are found to be $D_k e^{ik(n+1/2)}$, $D_k e^{ik(n-1/2)}$, $D_k^2 e^{ikn}$, respectively. Here $D_k = -2\cos(k/2)$. Thus we see that the charge density is proportional to $D_k^{2(m-1)}$ at each non-nodal site of the m-th zigzag chain from the edge. Then the convergence condition $|D_k| \leq 1$ must be satisfied otherwise the wave function would diverge in a

semi-infinite graphite sheet. This convergence condition defines the region $2\pi/3 \leq |k| \leq \pi$ where the flat band appears.

From the analytic expression of the edge state, we can derive an analytic expression for the density of states (DOS) near Fermi energy. We see that the edge state penetrate into the inner sites when wave number changes from π to $2\pi/3$. If we consider the graphite ribbons of width N, two edge states which come from both the edges will overlap with each other and develop the bonding and anti-bonding interaction. Since the magnitude of overlap becomes larger as the wave number approaches $2\pi/3$, the band gap between the bonding and anti-bonding state formed by the two edge states gets larger near $k = 2\pi/3$. Then the partly flat bands acquire a slight dispersion and it depends on the ribbon width N. The energy dispersion can be calculated from the overlap of two edge states. The amplitude of the edge state which penetrates from the first zigzag line is given by $\psi_n = D_k^{n-1} \equiv \psi_A$, which is located only on the A-sublattice. On the other hand, the amplitude of the edge state which penetrate from N^{th} zigzag line, is given by $\psi_{N-n} = D_k^{n-1} \equiv \psi_B$, which is located only on the B-sublattice. Using the tight-binding Hamiltonian, the overlap matrix element between the two edge states can be easily calculated, and we have

$$\langle \psi_A | H | \psi_B \rangle = -2tND_k^{N-1} \left(1 + \frac{D_k}{2} \right) \equiv \eta_k \ . \tag{3}$$

Then the energy eigenvalues for the edge states in the lowest order perturbation in overlap is given by the following eigenvalue problem,

$$\begin{pmatrix} 0 & \eta_k \\ \eta_k & 0 \end{pmatrix} \begin{pmatrix} \psi_A \\ \psi_B \end{pmatrix} = \epsilon \begin{pmatrix} \psi_A \\ \psi_B \end{pmatrix} \ . \tag{4}$$

The energy eigenvalues are $\epsilon = \pm\eta_k$, and they have the form $\epsilon \sim k^N$ around $k = \pi$. Therefore, the DOS corresponding to the edge states has the form,

$$\rho(\epsilon) = \frac{\partial k}{\partial \epsilon} \sim \frac{1}{N} \epsilon^{\alpha} \ , \tag{5}$$

where $\alpha = 1/N - 1$. Note that this DOS has a power-law dependence, which is different from the ordinary van Hove singularity of $\rho \sim 1/\sqrt{E}$ observed in one-dimensional system. It is found that the renormalized DOS is inversely proportional to the ribbon width, which has also been confirmed by numerical calculation [13].

In Fig. 6(a), we plot the energy gap $\Delta_{\rm a}$ of armchair ribbons at $k = 0$ and the energy gap $\Delta_{\rm z}$ of zigzag ribbons at $k = 2\pi/3$ with respect to W. It shows that the energy gaps are inversely proportional to the width of graphite ribbon. This result supports that the physical quantities related to energy gap can be scaled by the ribbon width. Fig. 6(b) shows the plot of N times the gap Δ as a function of N. We observe that $N\Delta$ becomes constant when N exceeds $N = 30$ ($N = 60$) for the zigzag (armchair) ribbon. The analytic

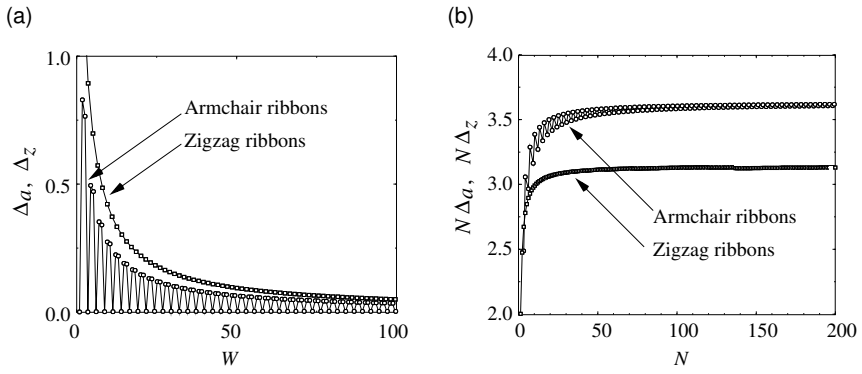

Fig. 6. (a) The width dependence of the energy gap of armchair ribbons at $k = 0$ ($\Delta_{\rm a}$) and those of the zigzag ribbons at $k = 2\pi/3$ ($\Delta_{\rm z}$). **(b)** The plot of $N\Delta_a$ ($N\Delta_{\rm z}$) versus N

expression for the ribbon width dependence of the energy gaps can be found in [14, 15, 18].

It should be noted that the same number N for both zigzag and armchair ribbons does not give the same ribbon width, when the ribbons are measured by the same unit of length. Therefore, we have introduced the following definition to compare physical quantities of zigzag and armchair ribbons with the same width W,

$$W = \begin{cases} \frac{3}{2}Na - a \equiv W_z & \text{zigzag ribbons} \\ (N-1)\frac{\sqrt{3}}{2}a \equiv W_a & \text{armchair ribbons} \end{cases} \tag{6}$$

where a is the lattice constant.

2.1 $\boldsymbol{k \cdot p}$ Equation

In the vicinity of $\boldsymbol{K}$ point of the 1st BZ of graphene sheet, the energy band has a conical structure, and we have a $\boldsymbol{k}$-linear dispersion relation or a massless spectrum. Since a graphene sheet is a half-filled system, the Fermi level is at the center of this spectrum. Thus the massless Dirac equation or the so-called $\boldsymbol{k \cdot p}$ equation becomes a powerful tool to describe the low-energy electronic properties of graphite or graphite-related materials. We use the $\boldsymbol{k \cdot p}$ equation [28–33] for studying the electronic states of a graphene system in weak magnetic fields and derive the half-integer quantized Hall conductance that has been observed experimentally [34]. In this subsection, we use $\sqrt{3}a$ as the unit of length instead of the lattice constant.

In the absence of magnetic field, $\boldsymbol{k \cdot p}$ equation at $\boldsymbol{K}$-point is given by,

$$\hat{H}\psi = \epsilon\psi, \qquad \psi = \begin{pmatrix} F_{\rm A}(\boldsymbol{r}) \\ F_{\rm B}(\boldsymbol{r}) \end{pmatrix} , \tag{7}$$

$$\hat{H} = \gamma \begin{pmatrix} 0 & \hat{k}_x \mp \mathrm{i}\hat{k}_y \\ \hat{k}_x \pm \mathrm{i}\hat{k}_y & 0 \end{pmatrix} . \tag{8}$$

Here upper (lower) sign of the Hamiltonian corresponds to $\boldsymbol{K}$ ($\boldsymbol{K}'$) point, and γ is the transfer integral. $F_{\mathrm{A}}^{\boldsymbol{K}}(\boldsymbol{r})$ ($F_{\mathrm{B}}^{\boldsymbol{K}}(\boldsymbol{r})$) represents the envelope function on the A(B)-sublattice of graphene. From the $\boldsymbol{k}\cdot\boldsymbol{p}$ equation, we can easily obtain the energy spectrum as

$$\epsilon = \pm\gamma|\boldsymbol{k}| . \tag{9}$$

Also, when we introduce the mass term as

$$\hat{H} \to \hat{H} + m\hat{\sigma}_z , \tag{10}$$

where m is mass, the energy eigenvalues are

$$\epsilon = \pm\gamma\sqrt{|\boldsymbol{k}|^2 + m^2} . \tag{11}$$

In Fig. 7(a), the energy spectrum of graphene is shown in the absence of magnetic field.

Next let us turn our attention to the electronic states of graphene sheet in a weak magnetic field, following McClure [31]. Here we take the Landau gauge, i.e., $\boldsymbol{A} = (0, Bx, 0)$. Then, the $\boldsymbol{k}\cdot\boldsymbol{p}$ Hamiltonian reads as

$$\hat{H}_{\pm} = \begin{pmatrix} 0 & \sqrt{2}t\hat{L}_{\mp}/l \\ \sqrt{2}t\hat{L}_{\mp}/l & 0 \end{pmatrix} , \tag{12}$$

$$\hat{L}_{\pm} = \frac{l}{\sqrt{2}}\left(\hat{k}_x \pm \mathrm{i}\hat{k}_y\right) , \tag{13}$$

where upper and lower signs correspond to $\boldsymbol{K}$ and $\boldsymbol{K}'$ points, respectively and $l = \sqrt{c\hbar/eB}$ is magnetic length.

This Hamiltonian supports a Landau level with a vanishing energy eigenvalue, which is responsible for the large diamagnetism of graphene and it will be discussed in the next section. Thus we have

$$\epsilon_{\mathrm{2D}} = 0, \quad \psi_{\boldsymbol{K}} = \frac{1}{\sqrt{L_y}} \exp\left(\mathrm{i}k_y y\right) \begin{pmatrix} 0 \\ h_0 \end{pmatrix} \quad \text{for } \boldsymbol{K}\text{−point}, \tag{14}$$

$$\epsilon_{\mathrm{2D}} = 0, \quad \psi_{\boldsymbol{K}'} = \frac{1}{\sqrt{L_y}} \exp\left(\mathrm{i}k_y y\right) \begin{pmatrix} h_0 \\ 0 \end{pmatrix} \quad \text{for } \boldsymbol{K}'\text{−point}, \tag{15}$$

where L_y and k_y are the system size and the wave vector, respectively, in the y direction. Other Landau levels and the eigenvalues for the $\boldsymbol{K}$-point are given by,

$$\epsilon_{\mathrm{2D}} = \pm\sqrt{2}t\frac{\sqrt{n+1}}{l}, \quad \psi_{\boldsymbol{K}} = \frac{1}{\sqrt{2L_y}} \exp\left(\mathrm{i}k_y y\right) \begin{pmatrix} \pm h_n \\ h_{n+1} \end{pmatrix} , \quad n = 0, 1, 2, \cdots , \tag{16}$$

while those for the $\boldsymbol{K}'$-point are,

$$\epsilon_{\rm 2D} = \pm\sqrt{2}t\frac{\sqrt{n+1}}{l}, \quad \psi_{\boldsymbol{K}'} = \frac{1}{\sqrt{L_y}}\exp\left(\mathrm{i}k_y y\right)\begin{pmatrix}\pm h_{n+1}\\ h_n\end{pmatrix}, \quad n = 0, 1, 2, \cdots, \tag{17}$$

where upper and lower signs correspond to conduction and valence bands, respectively, and h_n represents the harmonic oscillator eigenfunctions

$$h_n = \frac{\mathrm{i}^n}{\sqrt{2^n n!\sqrt{\pi}l}}\exp\left[-\frac{1}{2}\left(\frac{x+l^2k_y}{l}\right)^2\right]H_n\left(\frac{x+l^2k_y}{l}\right) \tag{18}$$

with $H_n(x)$ being the Hermite polynomial. These eigenfunctions have the properties:

$$\hat{L}_+\hat{L}_- h_n = nh_n, \quad \hat{L}_+ h_n = \sqrt{n+1}h_{n+1}, \quad \hat{L}_- h_n = \sqrt{n}h_{n-1}\,. \tag{19}$$

The Landau levels at $\boldsymbol{K}$ and $\boldsymbol{K}'$-points are degenerate and so far difference appears only in the wave functions.

In order to inspect closely the difference between $\boldsymbol{K}$ and $\boldsymbol{K}'$-points, let us consider the role of the mass term. The Landau level with finite mass can also be calculated in a similar manner. The energy eigenvalues for the $\boldsymbol{K}$-point are

$$\epsilon = -\mathrm{sgn}\,(B)m, \qquad n = 0 \tag{20}$$

$$\epsilon = \pm\sqrt{(m/\gamma)^2 + 2n/l^2}, \qquad n = 1, 2, \cdots \tag{21}$$

while those for the $\boldsymbol{K}'$-point are

$$\epsilon = \mathrm{sgn}\,(B)m, \qquad n = 0 \tag{22}$$

$$\epsilon = \pm\sqrt{(m/\gamma)^2 + 2n/l^2}, \qquad n = 1, 2, \cdots\,. \tag{23}$$

Thus the lowest Landau level (LLL) at $\boldsymbol{K}$-point shifts downward with the inclusion of mass, while the LLL at $\boldsymbol{K}'$-point shifts upward. In Fig. 7(b), the energy spectrum of graphene is shown in the presence of weak magnetic field.

It is now possible to explain the quantized Hall conductance of graphene. Let us assume that the Fermi energy is on $E = 0$. The LLL at $\boldsymbol{K}$-point is occupied by electrons, but that for the $\boldsymbol{K}'$-point remain unoccupied. Therefore, only the LLL with $n = 0$ is half-filled. Since each landau level can have $1/2\pi l^2$ electrons, the total number of electron at $\boldsymbol{K}$-point is

$$n_{\rm e}^{\boldsymbol{K}} = \sum_{i=-1} n_i + n_0 = \frac{N+1}{2\pi l^2}\,, \tag{24}$$

where n_i is the number of electrons in i-th Landau level. Similarly, the number of electrons at $\boldsymbol{K}'$-point is

$$n_{\rm e}^{\boldsymbol{K}'} = \sum_{i=-1} n_i = \frac{N}{2\pi l^2}\,. \tag{25}$$

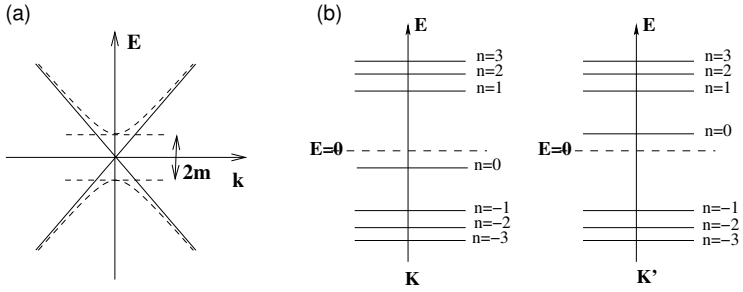

Fig. 7. (a) The energy spectrum of a graphene for $m = 0$ (solid line) and $m \neq 0$ (dashed line). **(b)** The Landau levels of graphene for $m \neq 0$

Therefore, the total number of electrons of the system is

$$n_{\rm e} = n_{\rm e}^{\boldsymbol{K}} + n_{\rm e}^{\boldsymbol{K}'} = \frac{2N+1}{2\pi l^2} \,, \tag{26}$$

and the quantized Hall conductivity is

$$\sigma_{xy} = (2N+1)\frac{e^2}{h} \,. \tag{27}$$

2.2 Electronic States in a Magnetic Field

In this subsection we study the electronic states of nanographites in a magnetic filed. The magnetic field $\boldsymbol{B}$ perpendicular to the graphite plane is incorporated within the transfer integral t_{ij} by means of the Peierls phase through the transformation $t_{ij} \to t_{ij} \exp\left[\mathrm{i}2\pi(e/ch)\int_i^j \mathrm{d}\boldsymbol{l}\cdot\boldsymbol{A}\right]$, where $\boldsymbol{A}$ is the vector potential and the integral is along the hopping path. Let ϕ be the magnetic flux passing through a single hexagonal ring of graphite sheet in the unit of the quantum flux ($\phi_0 = ch/e$). Since the single hexagonal ring has an area $S_{\rm hex} = 3\sqrt{3}a^2/2$, where $a = 1.42$ Å is the lattice constant of graphite, we have $\phi = BS_{\rm hex}/\phi_0$. Here $B = |\boldsymbol{B}|$ and for $\phi = 1$ it corresponds to 7.9×10^4 T. In this article, we treat only the weak magnetic field limit. The cyclotron radius (magnetic length), l, is given by $l = \sqrt{c\hbar/eB}$. We use Landau gauge with $\boldsymbol{A} = (0,\, Bx,\, 0)$, where we define the translational invariant direction of each ribbon as the y-axis, and the x-axis lies perpendicular to y-axis.

In ribbon-shaped systems, the ratio between the width of a ribbon and the cyclotron diameter effectively characterizes the electronic states of ribbon in a magnetic field. Therefore, we define the ratio between the ribbon width and the cyclotron diameter as the effective magnetic field $\tilde{B}$ for convenience, which is given by $\tilde{B} = (W/d)^2$, where $d = 2l$. In the case of $\tilde{B} < 1$, since the cyclotron diameter is larger than the ribbon width, the cyclotron motion of

electrons is impeded and the Landau levels are not formed. Secondly in the case of $\tilde{B} > 1$, since the cyclotron diameter is smaller than the ribbon width, the cyclotron motion of electrons is not disturbed except in the vicinity of the ribbon edges and the Landau levels appear. Finally in the case of $\tilde{B} \gg 1$, the cyclotron diameter is much smaller than the ribbon width and becomes of the order of the lattice constant so that the topology of the lattice becomes important.

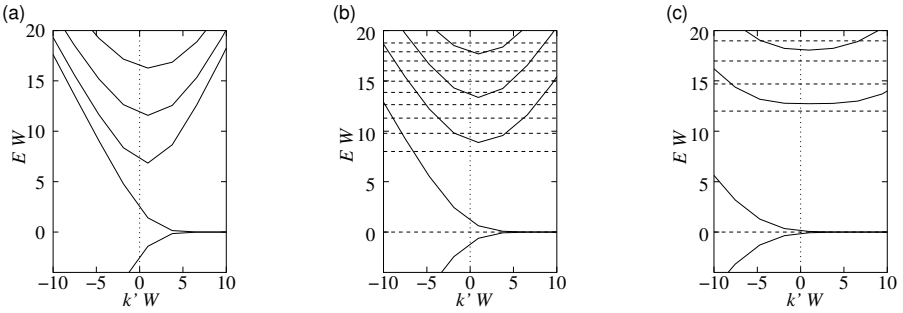

Fig. 8. The scaled energy band structures of zigzag ribbons near $k = 2\pi/3$ for **(a)** $\tilde{B} = 0$, **(b)** $\tilde{B} = 4$ and **(c)** $\tilde{B} = 16$. Bold lines are dispersions of zigzag ribbons. Here k' means the wave number measured from $k = 2\pi/3$. Dotted lines are the corresponding Landau levels of graphite sheet for comparison

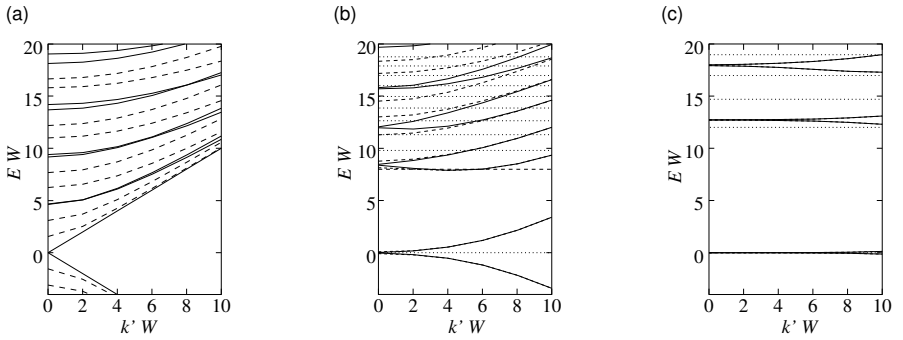

Fig. 9. The scaled energy band structures of armchair ribbons near $k = 0$ for **(a)** $\tilde{B} = 0$, **(b)** $\tilde{B} = 4$ and **(c)** $\tilde{B} = 16$. Bold lines are for metallic armchair ribbons ($N = 3m - 1$) and dashed lines are for semiconducting armchair ribbons ($N \neq 3m - 1$). Dotted lines are the corresponding Landau levels of graphite sheet for comparison

Now we show that the band dispersions near $E = 0$ can be scaled by the ribbon width W for sufficiently wide ribbons. In Fig. 8, the scaled band structures are shown for (a) $\tilde{B} = 0$, (b) $\tilde{B} = 4$ and (c) $\tilde{B} = 16$. Here the energy and the wave number k' are scaled by W, where k' means the wave number measured from $k = 2\pi/3$. In the figures, the bold lines are dispersions of zigzag ribbons, and the dotted lines are the corresponding Landau levels of graphite sheet. Similarly, the band dispersions of the armchair ribbons can be scaled and in Fig. 9 the scaled band structures are shown for (a) $\tilde{B} = 0$, (b) $\tilde{B} = 4$ and (c) $\tilde{B} = 16$. Bold lines are dispersions for the metallic armchair ribbons ($N = 3m - 1$) and dashed lines are those for the semiconducting armchair ribbons ($N \neq 3m - 1$). The dotted lines are the corresponding Landau levels of graphite sheet.

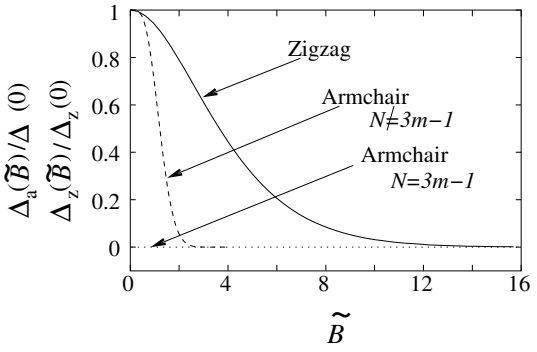

Fig. 10. Magnetic field dependence of the energy gap of armchair (zigzag) ribbons at $k = 0$ ($k = 2\pi/3$), $\Delta_{\rm a}(\tilde{B})$ ($\Delta_{\rm z}(\tilde{B})$). The energy gaps are normalized by the energy gap, $\Delta_{\rm a}(0)$ ($\Delta_{\rm z}(0)$), corresponding to $\tilde{B} = 0$

In Fig. 10, the magnetic field dependence of the energy gap for armchair and zigzag ribbons is shown. The energy gap of semiconducting armchair ribbons becomes negligible when effective magnetic field $\tilde{B}$ is much larger than 2. This condition is the same as that of an usual quantum wire. However, we need stronger magnetic field to collapse the energy gap (at $k = 2\pi/3$) of zigzag ribbons. This is due to the non-bonding character of the edge states. In the zero-field limit, $(1 - \Delta_{\rm a(z)}(\tilde{B})/\Delta_{\rm a(z)}(0))$ depends on $\tilde{B}^4$ for armchair ribbons, and $\tilde{B}^2$ for zigzag ribbons.

2.3 Bearded Edge

In this subsection, we discuss other important edge shapes having translational symmetry of zigzag axis, e.g., the *bearded edge*. Although this edge look rather artificial than the zigzag edge, they are interesting because they also show the

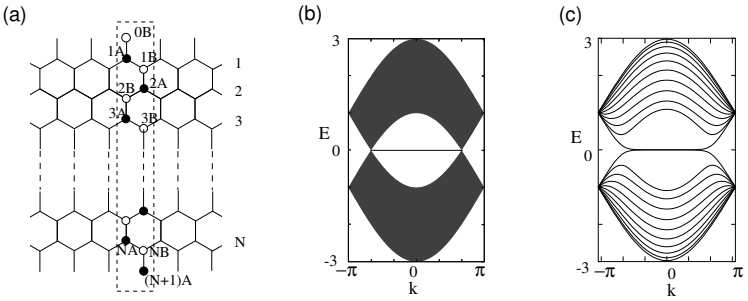

Fig. 11. (a) The structure of a graphite ribbon with two bearded edges. **(b)** The band structure of a semi-infinite graphite sheet with a bearded edge. **(c)** The band structure of bearded ribbon for $N = 10$

non-bonding edge localization.

A bearded edge is a zigzag edge with additional π-electron hopping bonds as shown in Fig. 11(a). This type of edge was first studied by Klein [35]. In Fig. 11(b), the band structure of a semi-infinite graphite sheet with a bearded edge is shown. Interestingly, a partial flat band appears in the region of $|k| \leq 2\pi/3$, which is just the opposite condition corresponding to the semi-infinite graphite sheet with a zigzag edge. The analytic form of this edge state can be derived in a similar way as described in the previous subsection for the derivation of the edge state of zigzag edges. If we consider the case of infinite N (Fig. 11(a)), the analytic solution can be written as,

$$\Psi_{n\mathrm{B}} = (1/D_k)^n \quad \text{and} \quad \Psi_{n\mathrm{A}} = 0, \tag{28}$$

where $D_k = -2\cos(k/2)$. The convergence condition of the wave function, $|1/D_k| \leq 1$, defines the region $|k| \leq 2\pi/3$ corresponding to the edge states. Since the penetration depth of the edge states get longer toward $k = \pm 2\pi/3$, we have small energy gaps near $k = \pm 2\pi/3$ in the case of finite width ribbons due to bonding and anti-bonding interaction between the two edge states.

Next we consider the ribbons having zigzag and bearded edge as shown in Fig. 12(a). Since in this ribbon $|N_\mathrm{A} - N_\mathrm{B}| = 1$, where $N_\mathrm{A}(N_\mathrm{B})$ means the number of sites belonging to the A(B)-sublattice, there is a flat band at $E = 0$ all over the 1st BZ as shown in Fig. 12(b). The analytic solution of this flat band can be easily understood by the combination of two edge states for the zigzag and bearded edges. In the region of $|k| < 2\pi/3$, the electrons are localized at the bearded edge, and in the region of $|k| > 2\pi/3$, the electrons are localized at the zigzag edge. At $k = \pm 2\pi/3$, the electrons are delocalized. It should be noted that this ribbon is insulating because the flat band does not carry current, and there are energy gaps between the flat band and the next subbands.

The important conclusion in this subsection is that an edge which is not parallel to an armchair axis provides edge states. Actually, graphite ribbons

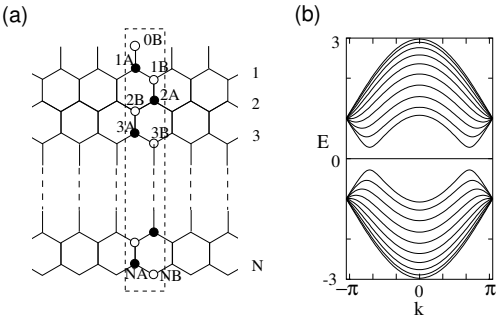

Fig. 12. (a) The structure of a graphite ribbon with a zigzag and a bearded edge and **(b)** the band structure for $N = 10$

with mixed armchair edges and zigzag edges also show localized states [13]. Thus it is considered that the edges whose cutting direction is not parallel to the armchair axis show edge localized states. The reason why an armchair edge does not provide an edge state can be understood from a comparison with the surface bound state of anisotropic superconductor, see [15, 18].

3 Magnetic Properties of Nanographites

In the following two subsections, the magnetic properties of nanographites are discussed in the absence and also in the presence of electron-electron interaction. It is well-known that bulk graphite shows a large anisotropic diamagnetic susceptibility, while aromatic molecules have only weak diamagnetism. This means that the orbital diamagnetic susceptibility is sensitive to the size of graphite fragments. On the other hand, the sharp peak in the density of states due to the edge states gives rather strong Pauli paramagnetic response, so that the competition between these two components occurs in nanographite systems. Since the edge states also give the possibility of spin polarization at low-temperature due to electron-electron interaction, we will discuss about the possibility of appearance of localized spins at the edges.

3.1 Orbital Magnetization and Susceptibility

The observed magnetic susceptibility χ is the sum of four components: (1) localized spin susceptibility $\chi_{\rm spin}$, (2) diamagnetic susceptibility due to the core electrons $\chi_{\rm core}$, (3) Pauli paramagnetic susceptibility $\chi_{\rm P}$ and (4) orbital diamagnetic susceptibility $\chi_{\rm orb}$ due to the cyclotron motion of the itinerant electrons. Since we neglect electron-electron interaction for the moment, $\chi_{\rm spin}$ can be neglected. Furthermore, $\chi_{\rm core}$ is unimportant for us, because it is small and basically temperature independent. On the other hand, the Pauli paramagnetic susceptibility is related to the DOS at the Fermi level, and becomes

an important component in zigzag nanographite ribbons where an enhanced density of states appears at the Fermi level. Note that $\chi_{\rm P}$ is negligible in armchair ribbons, aromatic molecules and graphite sheets, because their DOS are suppressed at the Fermi level. We will see below that since the DOS introduced by the edge states is sharply peaked at the Fermi energy, $\chi_{\rm P}$ exhibits a very pronounced temperature dependence which is nearly Curie-like. The diamagnetic contribution to susceptibility is very familiar from the magnetic properties of graphite sheets. It is due to the orbital cyclotron motion of electrons in a field with a finite component perpendicular to the plane. Naturally, this diamagnetic response is very anisotropic and only weakly temperature dependent. From this we can conclude that in nanographite ribbons with zigzag edges the susceptibility should consist mainly of these two competing contributions, $\chi_{\rm P}$ and $\chi_{\rm orb}$. Hence, a crossover occurs from a high-temperature diamagnetic to a low-temperature paramagnetic regime, where the characteristic temperature depends on the width of the ribbon and on the orientation of the external field. It is worth noting that the field direction is an important tool to distinguish the magnitude of the two components.

In this subsection, we briefly summarize the way how to calculate the orbital diamagnetic susceptibility $\chi_{\rm orb}$ of graphite ribbons. We use the tight-binding Hamiltonian for the calculation of orbital diamagnetic susceptibility $\chi_{\rm orb}$. The free energy $F(H,T)$ in the presence of magnetic field is given by,

$$F(H,T) = \mu N - \frac{1}{\beta\pi}\int_{\rm BZ} {\rm d}k \sum_n \ln\left(1+{\rm e}^{-\beta(\epsilon_{k,n}(H)-\mu)}\right) \tag{29}$$

where $\beta = 1/k_{\rm B}T$, μ is the chemical potential and $\epsilon_{k,n}(H)$ (n is the band index) is the tight-binding energy spectrum of the graphite ribbons in the magnetic field. Then magnetic moment $M(H)$ and magnetic susceptibility $\chi(H)$ per site are given by the 1st and 2nd derivative of free energy with respect to H, respectively. Thus we have

$$M(H) = -\frac{1}{N_e}\frac{\partial F}{\partial H}, \quad \text{and} \quad \chi(H) = \frac{1}{N_e}\frac{\partial M}{\partial H}\,. \tag{30}$$

Before showing the results for the graphite ribbons, let us first briefly review the large orbital diamagnetism in graphite. It was first shown by McClure [31] that the origin of large diamagnetism in graphite is due to the appearance of the Landau level at $E = 0$. When the energy dispersion is linear in k near the $\boldsymbol{K}$ or $\boldsymbol{K}'$ points, the Landau levels E_n in the 1st BZ of the graphite sheet can be expressed as,

$$E_n = \pm\frac{\sqrt{3}}{2}ta\sqrt{2nS} \quad (n = 0, 1, 2, \cdots)\,, \tag{31}$$

where n is an index of the Landau levels and S is given by $S = eH/\hbar c$. Here t is the nearest neighbor transfer integral, and considered to be about 3.0 eV. Each Landau level has a characteristic feature which is very different

from the behavior of the Landau levels of ordinary 3D free electron gas (see Fig. 13). It may be noted that the zero-th Landau level is always located in the zero energy gap at the $\boldsymbol{K}$ point. The zero-th Landau level does not shift by magnetic field. We can easily show that all the Landau levels with zero and negative Landau indices which are occupied by the valence electrons are responsible for the increase of free energy when we apply magnetic field and thus the orbital diamagnetism appears. When the Fermi energy is located in the zero energy gap, the oscillation of the free energy, which is known to cause de Haas-van Alphen effect at low temperatures, has a cusp at $E_{\rm F} = 0$ and yields large orbital diamagnetism as shown in Fig. 14(a). In Fig. 14(b), we plot free energy as a function of $E_{\rm F}$ for the 3D electron gas. In this case there are no sharp peaks in the oscillation in contrast to the results presented in Fig. 14(a) for graphite sheet. The expression for orbital susceptibility of graphite sheet at finite temperatures has been derived by McClure [31] based on the $\boldsymbol{k} \cdot \boldsymbol{p}$ approximation and it is given by,

$$\chi_{\rm MC} = -0.044 \left(\frac{4}{\pi C_{\rm L}}\right)^{3/2} (ta)^2 \left(\frac{e}{\hbar c}\right)^2 (k_{\rm B}T\rho)\, {\rm sech}^2 \left(\frac{\mu}{2k_{\rm B}T}\right) , \qquad (32)$$

where μ is the Fermi energy and ρ is the density of carbon atoms in the unit volume of graphite. The typical value of $\chi_{\rm MC}$ for graphite sheet at room temperature is 21.0×10^{-6} emu/g.

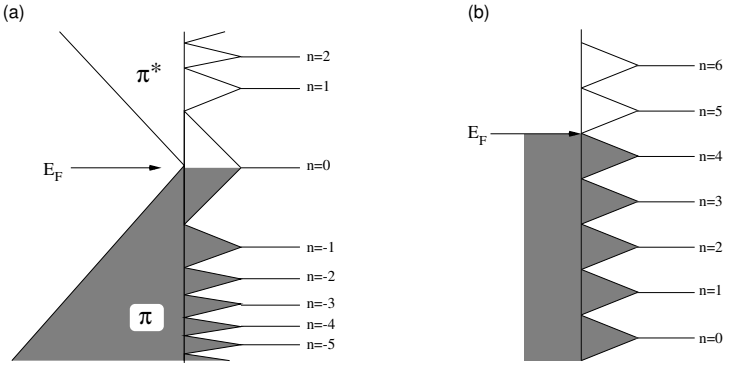

Fig. 13. (a) Landau levels of graphite sheet near $\boldsymbol{K}$ point **(b)** Landau levels of 3D metal

The orbital magnetism is influenced by the system geometry and size. Here we show the Fermi energy dependence of $\chi_{\rm orb}$. Actually in real graphite materials, a small change in the carrier density from the half-filling is possible and can even be controlled by the substrate properties. The calculated Fermi energy dependence is shown in Fig. 15(a), where it is found that $\chi_{\rm orb}/W$ is a universal function of μW. We normalize $\chi_{\rm orb}$ by dividing it by W, since it

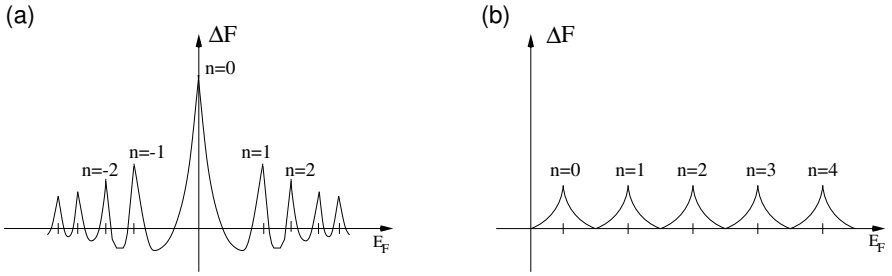

Fig. 14. (a) Oscillation of free energy of graphite sheet as a function of $E_{\rm F}$. When the Fermi energy is located at zero, this oscillation has sharp peak, however **(b)** the simple 3D metal does not show such sharp peaks. The existence of this peak is the origin of large diamagnetism in graphites

is proportional to W. Furthermore we multiply $E_{\rm F}$ by W, because the direct gap is proportional to $1/W$ at $k = 0$ for armchair ribbons and at $k = 2\pi/3$ for zigzag ribbons with large W.

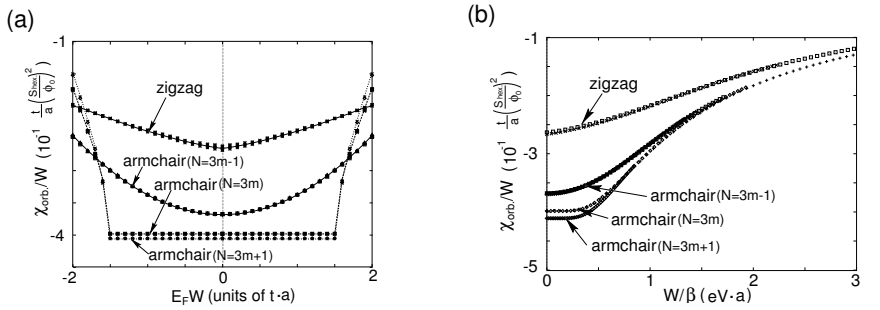

Fig. 15. (a) The Fermi energy dependence of the orbital magnetic moments $\chi_{\rm orb}$ of graphite ribbons at $T = 0$. **(b)** The temperature dependence of $\chi_{\rm orb}$, where $\chi_{\rm orb}$ is scaled by $1/W$ and β is scaled by W

In Fig. 15(b), the temperature dependence of $\chi_{\rm orb}$ is shown. It is important from the viewpoint of experiments on nanographites. In all the cases magnitude of $\chi_{\rm orb}$ decreases with increasing temperature. It is also found that the temperature dependence of $\chi_{\rm orb}/W$ scales as a function of βW, because the energy gap is proportional to $1/W$. Our calculation also demonstrates that the edge effect becomes more significant at lower temperature. Similar scaling properties can also be found in carbon nanotubes [36, 37].

3.2 Pauli Paramagnetism

In the previous subsection, we have seen that the orbital diamagnetic susceptibility depends on the edge shape of nanographite ribbons. Here we discuss about another important component of magnetic susceptibility, the Pauli paramagnetic susceptibility $\chi_{\rm P}$ as the zigzag ribbons have a sharp peak in the DOS at the Fermi level. The width of this peak is of the order of meV, which is comparable to the temperature scale at room temperature. Therefore, we expect that the Pauli susceptibility of zigzag ribbons is sensitive to temperature, although the Pauli susceptibility of usual metals is temperature independent. On the other hand, since the DOS of armchair ribbons at $\epsilon = 0$ is zero or very tiny, we can neglect the effect of Pauli paramagnetism in armchair ribbons.

The magnetic moment due to the Zeeman effect is $M = \mu_{\rm B}\left(n_\uparrow - n_\downarrow\right)$, where $\mu_{\rm B}$ is the Bohr magneton and $n_\uparrow$ $(n_\downarrow)$ denotes the electron density with up-spin (down-spin). The electron density at arbitrary temperature for each spin is given by

$$n_\sigma = \frac{1}{\pi}\int_{\rm 1st\,BZ} {\rm d}k \sum_n \frac{1}{1+{\rm e}^{\beta(\epsilon_{n,k}-\sigma\mu_{\rm B}H)}}\ , \tag{33}$$

where $\sigma(=\uparrow,\downarrow)$ denotes the spin index. Therefore, the Pauli susceptibility $\chi_{\rm P}$ per site is given by

$$\chi_{\rm P} = \lim_{T\to 0}\frac{\partial M}{\partial H} = \frac{\beta\mu_{\rm B}^2}{\pi N_e}\sum_n \int {\rm d}k \frac{1}{\cosh\left(\beta\epsilon_{n,k}\right)}\ . \tag{34}$$

We calculate numerically the finite temperature Pauli susceptibility of graphite ribbons using this equation up to room temperature.

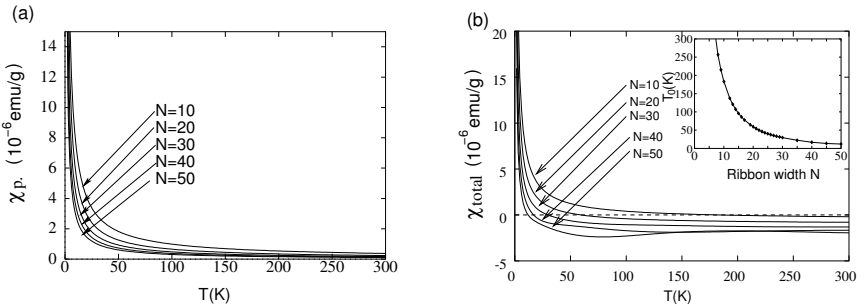

Fig. 16. (a) The temperature dependence of $\chi_{\rm P}$ for $N = 10, 20, \ldots, 50$ up to room temperature. **(b)** The temperature dependence of total susceptibility $\chi = \chi_{\rm orb} + \chi_{\rm P}$ is shown for $N = 10, 20, \ldots, 50$. The inset shows the width dependence of the crossing temperature, T_0, where $\chi = 0$

It is possible to separate the contribution of the edge states to $\chi_{\rm P}$. As we have seen in Sect. 2 of this article, the DOS due to the edge states is given

by (5). After substitution of (5) into (34), we replace the k-integration by the energy integration. Then we obtain the $\chi_{\rm P}$ contribution due to the edge states as,

$$\chi_{\rm P} = \frac{1}{N_e N \beta^{\alpha}} \int {\rm d}x \frac{x^{\alpha}}{\cosh x + 1} \sim \frac{1}{N} T^{\alpha} \ , \tag{35}$$

where $x = \beta\epsilon_k$ and $\alpha = 1/N - 1$. Interestingly, $\chi_{\rm P}$ has the Curie-like temperature dependence. The exponent of $\chi_{\rm P}$ depends on the ribbon width through α. When N becomes infinite, the exponent α approaches -1 and $\chi_{\rm P}$ shows the Curie-law. However, in this limit, the contribution of $\chi_{\rm P}$ is diminished by a factor $1/N$ in (35). Numerical results for the Pauli susceptibility $\chi_{\rm P}$ of zigzag ribbons up to room temperature are shown in Fig. 16(a) for various values of N. As expected, because of the edge states, $\chi_{\rm P}$ shows Curie-like temperature dependence.

The observed susceptibility χ is essentially the sum of the orbital diamagnetic susceptibility $\chi_{\rm orb}$ and the Pauli susceptibility $\chi_{\rm P}$. The temperature dependence of the total susceptibility χ is shown in Fig. 16(b). The total susceptibility χ shows the diamagnetic behavior in the high temperature regime and paramagnetic behavior in the low temperature regime. In the inset, the width dependence of the crossing temperature, i.e., where $\chi = 0$, is plotted, which is approximately proportional to $1/N$.

Here we should remind that both aromatic molecules and bulk graphite show diamagnetic behavior, however, nanographite with zigzag edges have a remarkable paramagnetic behavior because of the edge states. If this paramagnetic behavior is experimentally detected, it will be an indirect evidence of the existence of the edge states.

3.3 Magnetic Instability

The presence of the sharp peak in the density of states should induce lattice distortion due to electron-phonon interaction and/or magnetic polarization due to electron-electron interaction. Because of the non-bonding character of the edge states, the lattice distortion in the vicinity of zigzag edges is unlikely with the realistic strength of the electron-phonon coupling [38]. The absence of lattice distortion has also been confirmed in terms of the density functional approach [39]. We examine the effect of electron-electron interaction by using the Hubbard model with unrestricted Hartree-Fock approximation. We find the possibility of a spontaneous magnetic ordering near the edge, peculiar to the nanometer scale fragments of graphite.

In order to study the magnetic instability of nano-graphite ribbons, we use Hubbard model in the unrestricted Hartree-Fock (HF) approximation. The Hamiltonian is written as

$$H_{HF} = -t \sum_{\langle i,j \rangle, s} c^{\dagger}_{i,s} c_{j,s} + U \sum_{i,s} \left(\langle n_{i,-s} \rangle - \frac{1}{2} \right) n_{i,s} \ , \tag{36}$$

where the operator $c^{\dagger}_{i,s}$ creates an electron with spin s on the site i and $n_{i,s} = c^{\dagger}_{i,s}c_{i,s}$. The indices of the sites in graphite ribbons are shown in Fig. 1(b) and Fig. 1(c). The parameters t and U are the nearest-neighbor transfer integral and the on-site Coulomb repulsion, respectively. The symbol $\langle \cdots \rangle$ denotes the expectation value in the HF state. We solve the unrestricted Hartree-Fock (HF) Hamiltonian with the self-consistence conditions, i.e., $m_i = \sum_{i,s} s\langle n_{i,s} \rangle$, where m_i represents the magnetic moment at site i in the unit of Bohr magneton.

In Fig. 17, the U dependence of magnetization m for the zigzag ribbons with (a) $N = 2$, (b) $N = 3$ and (c) $N = 10$ are shown. The dashed lines are HF solutions for a 2D graphite sheet. The results for the armchair ribbons with (a) $N = 3$, (b) $N = 4$ and (c) $N = 5$ are shown in Fig. 18. A peculiar feature has been observed in the zigzag ribbons, the appearance of large magnetic moments on the edge carbons even for weak U. It can be explained as follows: Since the 2D graphite is a zero-gap semiconductor whose DOS is zero at the Fermi level, the broken line stands up at a finite value of $U(=U_C)$. This is consistent with the fact that graphite is non-magnetic when U is much smaller than t. On the other hand, the zigzag ribbon has a large density of states at the Fermi level originating from the edge states. Thus, non-zero magnetic solutions can emerge for infinitesimally small U as has been observed in the present mean-field result. However, special emphasis should be put on the behavior of the magnetization at the edge site 1A. As shown in Fig. 17, the magnetization at the site $1A$ rapidly rises up and reaches about 0.2 even at small $U(\approx 0.1)$ when the width of ribbon is increased. We note that the armchair ribbon does not show such singular magnetic behavior as shown in Figs. 18(a)-(c).

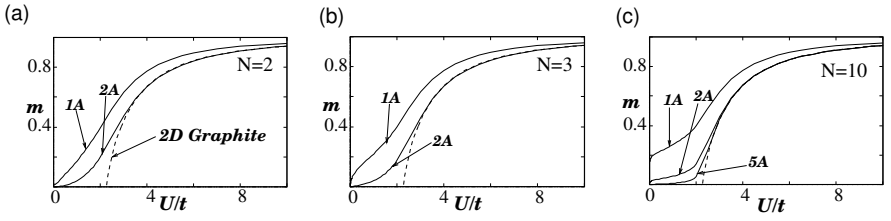

Fig. 17. The U dependence of the magnetization m for the zigzag ribbons of $N =$ **(a)** 2, **(b)** 3 and **(c)** 10. The dashed lines mean the mean-field solutions for the graphite sheet

Next, we should also stress the local ferrimagnetic structure for the zigzag ribbons. We exhibit the magnetic structure of the ribbon with $N = 10$ at $U/t = 0.1$ in Fig. 19(a), where spin alignment is visible at both edge sites. The origin of this structure can be explained from the nature of the edge states, which are responsible for magnetization. Since the amplitude of the

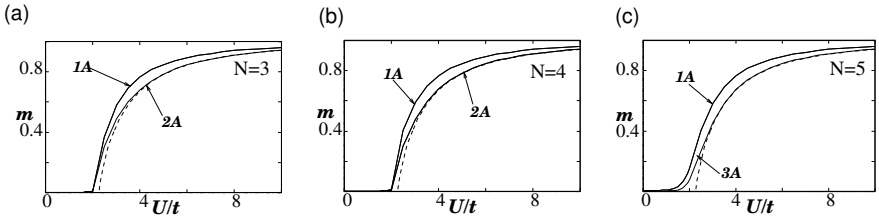

Fig. 18. The U dependence of the magnetization m for the the armchair ribbons of $N =$ **(a)** 3, **(b)** 4 and **(c)** 5. The dashed lines mean the mean-field solutions for the graphite sheet

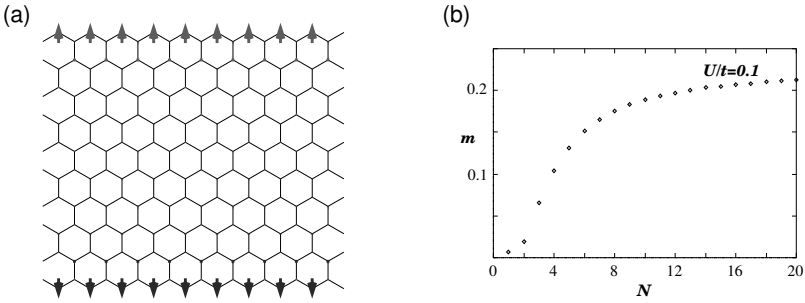

Fig. 19. (a) The schematic magnetic structure for the zigzag ribbon of $N = 10$ at $U/t = 0.1$. **(b)** The ribbon width, N, dependence of the magnetization at the most outer site (1A). The magnetization at site NB has same magnitude but with opposite sign

edge state is non-zero only on one of the two sublattices at an edge and damps inwards, the magnetic moments grows selectively on this sublattice forming local ferrimagnetic spin configuration, and the amplitude of these moments diminishes rapidly on the inner sites. As the sites on the opposite edges belong to different sublattices, the total magnetization for the zigzag graphite ribbon is zero, and this vanishing of total magnetic moment in the ground state is consistent with the exact statement for the half-filled Hubbard model [40]. Figure 19(b) shows the ribbon width N dependence of magnetization of the outer most site (1A or NB). The magnitude of magnetization increases rapidly as we increase ribbon width and it saturates around $N = 10$.

The ferrimagnetic spin polarizations along the zigzag edges are interesting in view of the magnetic properties of nanographites. Nevertheless, the long-range order derived from the mean-field calculation is spurious, because no finite-momentum long-range spin order is expected in an one-dimensional system with full spin-rotation symmetry [41]. Even we may argue that quasi-long-range order, similar to the spin-1/2 Heisenberg chain, is not possible in zigzag ribbons of any finite width for the following reason: The unit cell of the

ribbons contains an even number of sites such that Haldane's conjecture applies, i.e., the system should exhibit a spin gap [42]. This is very analogous to the case of the ladder systems with even number of legs, which display a resonating valence bond (RVB) ground state, i.e., a short-range correlated spin liquid state. With increasing the ribbon width, however, the analysis by the random phase approximation (RPA) shows that the spin gap Δ_s decreases exponentially due to the diminishing overlap between two edge states [43]. This fact means that the zigzag edges favor spin polarization with ferromagnetic alignment. The systematic analysis of topological network in nanographites gives a good indication for designing new magnetic carbon materials [44–47].

4 Electronic Transport Properties

In this section, we discuss the electron transport properties of nanographite ribbon junctions, and show that the edge states play a very important role on the electron transport of nanographite systems. Because of the non-bonding character of edge states, a single edge state cannot contribute to the electron transport. However, in zigzag ribbons, the edge states can provide a single-channel for electron conduction in the low-energy region. This is due to the bonding and anti-bonding interaction between the two edge states arising from both the edges that overlap. In order to analyze the electronic transport properties of the edge states, we consider the nanographite ribbon junction systems which connect two zigzag ribbons of different or same width. The electrical conductance of such junction is calculated using the Landauer-Büttiker formula based on the simple tight-binding model, in which conductance is written in terms of the transmission coefficient [1, 48–50]. We calculate the transmission coefficients of the junctions by of the recursive Green's function method which provides high efficiency and accuracy for the numerical calculations [51, 52].

Conductance of a nanographite ribbon junction as a function of Fermi energy shows rich structures in the energy region corresponding to single-channel transport [15–17]. The remarkable feature in the behavior of conductance is the appearance of sharp zero-conductance dip structures corresponding to total reflection resonances. These conductance zeros are associated with the presence of resonant localized states within the junction which resemble "flux states" in the sense that they show a strong current-current correlation with a Kekulé-like vortex pattern. The resonant state is a standing wave resulting from the superposition of the two flux states which are the time-reversed of each other. We also show that zero-conductance dips can be well interpreted in terms of the result of an asymmetric Aharonov-Bohm(AB) ring connected to current leads. Since these resonances are intimately connected with the time-reversal symmetry of the system, the application of a magnetic field removes the conductance zeros, yielding a pronounced negative magnetoresistance. In this article, we emphasize that the edge states in nanographite ribbons lead to

electronic transport properties distinctively different from those of the usual quantum wires or carbon nanotubes.

4.1 Landauer Formula

We use the multi-channel Landauer formula (MCLF) [48–50] in order to evaluate the conductance of nanographite junctions. This formula was originally developed in the context of Anderson localization problem. Nowadays it is widely used for studying mesoscopic and nanoscopic systems. The MCLF derives conductance from the scattering matrix. Through the scattering matrix, $\mathbf{S}$, the amplitudes of the scattered waves $\mathbf{b}$ are related to the incident wave amplitudes $\mathbf{a}$:

$$\begin{bmatrix} \mathbf{b}_{\mathrm{L}} \\ \mathbf{b}_{\mathrm{R}} \end{bmatrix} = \mathbf{S} \begin{bmatrix} \mathbf{a}_{\mathrm{L}} \\ \mathbf{a}_{\mathrm{R}} \end{bmatrix} = \begin{bmatrix} \mathbf{r} & \mathbf{t}' \\ \mathbf{t} & \mathbf{r}' \end{bmatrix} \begin{bmatrix} \mathbf{a}_{\mathrm{L}} \\ \mathbf{a}_{\mathrm{R}} \end{bmatrix} , \tag{37}$$

where $\mathbf{t}$, $\mathbf{t}'$ are the transmission matrices, $\mathbf{r}$, $\mathbf{r}'$ are the reflection matrices and subscripts L and R denote the left and right lead lines. Conductance per spin is defined in units of $e^2/\pi\hbar$ by

$$g(E) = \mathrm{tr}\left(\mathbf{t}^\dagger \mathbf{t}\right) = \sum_{\mu,\nu} |t_{\mu\nu}(E)|^2 , \tag{38}$$

where $t_{\mu\nu}(E)$ is the transmission coefficient from ν^{th} to μ^{th} channel at energy E. The scattering matrix can be written in terms of the lattice Green function (GF). The lattice Green functions and the transmission coefficients can be calculated from the tight-binding model. In order to calculate the lattice Green functions efficiently, we adopt the recursive Green function technique [51, 52].

An example of the nanographite ribbon junctions is depicted in Fig. 20(a). The shaded central region corresponds to the scattering region, and the zigzag ribbons of the left and right parts correspond to the lead lines. We need to know the Green function of zigzag ribbons in order to apply the recursive GF method (see next subsection).

4.2 Zigzag Ribbon as a Lead Line

Let us consider a zigzag ribbon of width N. We define the translational invariant direction as the y-axis and, the x-axis lies perpendicular to y-axis as shown in Fig. 20(b). In zigzag ribbons, there are two kinds of columns in a unit cell, which we call the α- and β-column, respectively. The equations of motion for the electrons on the zigzag ribbon in the $j-$th cell are given by

$$\begin{cases} (E\mathbf{I} - \mathbf{H}_0^\alpha)\mathbf{C}_j^\alpha - \mathbf{V}^\dagger \mathbf{C}_j^\beta - \mathbf{V}\mathbf{C}_{j-1}^\beta = \mathbf{0} \\ (E\mathbf{I} - \mathbf{H}_0^\beta)\mathbf{C}_j^\beta - \mathbf{V}^\dagger \mathbf{C}_{j+1}^\alpha - \mathbf{V}\mathbf{C}_j^\alpha = \mathbf{0} \end{cases} , \tag{39}$$

where $\mathbf{C}_j^\alpha$ ($\mathbf{C}_j^\beta$) is a vector ($N \times 1$ matrix) which describes the wave function amplitudes of the $\alpha(\beta)$-column of j-th layer and $\mathbf{H}_0^\alpha$ ($\mathbf{H}_0^\beta$) denotes the column Hamiltonian of α (β)-column. The $\mathbf{V}$ ($\mathbf{V}^\dagger$) represents the inter-column

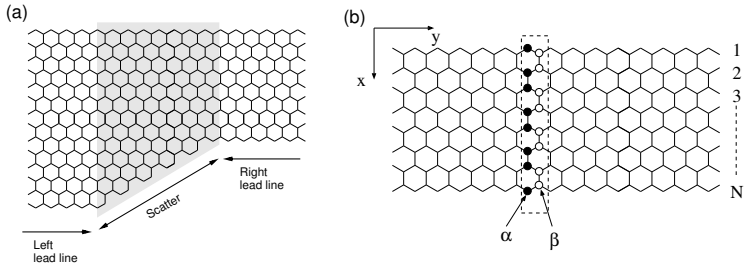

Fig. 20. (a) An example configuration of nanographite ribbon junctions. The shaded central region is the scatter. Lead lines (zigzag ribbons) are attached on the both sides of the scatter. **(b)** The structure of zigzag ribbon. We classify two columns called α and β in the unit cell. The rectangle with dashed lines is the unit cell

Hamiltonian ($N \times N$ matrix) describing hopping. From these two equations, we can obtain the following system of equations

$$\begin{bmatrix} E\mathbf{I}-\mathbf{H}_0^\beta, & 0, & -\mathbf{V}^\dagger, & 0, & -\mathbf{V} \\ 0, & E\mathbf{I}-\mathbf{H}_0^\beta, & 0, & -\mathbf{V}, & -\mathbf{V}^\dagger \\ -\mathbf{V}, & -\mathbf{V}^\dagger & 0, & 0, & E\mathbf{I}-\mathbf{H}_0^\alpha \end{bmatrix} \begin{bmatrix} \mathbf{C}_j^\beta \\ \mathbf{C}_{j-1}^\beta \\ \mathbf{C}_{j+1}^\alpha \\ \mathbf{C}_{j-1}^\alpha \\ \mathbf{C}_j^\alpha \end{bmatrix} = 0 \,. \tag{40}$$

Using either numerical or analytical method, we eliminate the two variables of the β-column, $\mathbf{C}_j^\beta$ and $\mathbf{C}_{j-1}^\beta$, to obtain the following Harper's equation,

$$\mathbf{C}_{j+1}^\alpha + \mathbf{u}\mathbf{C}_{j-1}^\alpha + \mathbf{v}\mathbf{C}_j^\alpha = 0 \,. \tag{41}$$

Assuming the Bloch form,

$$\mathbf{C}_{j+1}^\alpha = \lambda \mathbf{C}_j^\alpha \,, \tag{42}$$

we obtain the following eigenvalue problem,

$$\lambda \begin{pmatrix} \mathbf{C}_j^\alpha \\ \mathbf{C}_{j-1}^\alpha \end{pmatrix} = \begin{pmatrix} -\mathbf{v} & -\mathbf{u} \\ 1 & 0 \end{pmatrix} \begin{pmatrix} \mathbf{C}_j^\alpha \\ \mathbf{C}_\beta^\beta \\ \mathbf{C}_{j-1}^\beta \end{pmatrix} \,. \tag{43}$$

This equation has $2N$ eigenvalues and $2N$ eigenvectors, which are classified as N right- and N left-going waves. The N right-going solutions consist of traveling waves having positive velocities along the y direction and evanescent waves decaying exponentially in the positive y direction. Similarly, the N left-going solutions consist of traveling waves having negative velocities in the y direction and evanescent waves decaying exponentially in the negative y direction.

Let $\boldsymbol{u}_1(-), \cdots, \boldsymbol{u}_N(-)$ be the left-going solutions of $\mathbf{C}_0^\alpha$ corresponding to $\lambda_1(-), \cdots, \lambda_N(-)$ and $\boldsymbol{u}_1(+), \cdots, \boldsymbol{u}_N(+)$ be the right-going solutions of $\mathbf{C}_0^\alpha$ corresponding to $\lambda_1(+), \cdots, \lambda_N(+)$. We then define two kinds of matrices,

$$\mathbf{U}(\pm) = (\boldsymbol{u}_1(\pm), \cdots, \boldsymbol{u}_N(\pm)) \tag{44}$$

and

$$\Lambda(\pm) = \begin{bmatrix} \lambda_1(\pm) & & \\ & \ddots & \\ & & \lambda_N(\pm) \end{bmatrix} . \tag{45}$$

Any left- and right-going wave can be written, for example at j=0, as

$$\mathbf{C}_0^\alpha(\pm) = \mathbf{U}(\pm)\mathbf{C}^\alpha(\pm) , \tag{46}$$

where $\mathbf{C}(\pm)$ is an appropriate vector consisting of the expansion coefficients. For general j, we have

$$\mathbf{C}_j^\alpha(\pm) = \mathbf{U}(\pm)\Lambda(\pm)^j\mathbf{C}^\alpha(\pm) , \tag{47}$$

which leads to the relation

$$\mathbf{C}_j^\alpha(\pm) = \mathbf{F}(\pm)^{j-j'}\mathbf{C}_{j'}^\alpha(\pm) , \tag{48}$$

with

$$\mathbf{F}(\pm) = \mathbf{U}(\pm)\Lambda(\pm)\mathbf{U}^{-1}(\pm) . \tag{49}$$

Note that $\mathbf{U}(\pm)$ is not a unitary matrix in general.

4.3 Scattering Matrix

Now let us consider the scattering problem. We assume that the scattering object has N_s columns as shown in Fig. 21, starting from column 1 to column N_s. The 0-th column corresponds to the end of the left lead line, and the $N_s + 1$ column corresponds to the beginning of the right lead line. Infinite zigzag ribbons are attached to each side of this scattering object as the leads. First, we separate the amplitude $\mathbf{C}_0^\alpha$ at cell 0 into the right-going (incident) and left-going (reflected) solutions:

$$\mathbf{C}_0^\alpha = \mathbf{C}_0^\alpha(+) + \mathbf{C}_0^\alpha(-) . \tag{50}$$

By using (48), the $\mathbf{C}_{-1}^\alpha$ can be rewritten as follows,

$$\begin{aligned} \mathbf{C}_{-1}^\alpha &= \mathbf{C}_{-1}^\alpha(+) + \mathbf{C}_{-1}^\alpha(-) \\ &= \mathbf{F}^{-1}(+)\mathbf{C}_0^\alpha(+) + \mathbf{F}^{-1}(-)\mathbf{C}_0^\alpha(-) \\ &= \mathbf{F}^{-1}(-)\mathbf{C}_0^\alpha + \left\{\mathbf{F}^{-1}(+) - \mathbf{F}^{-1}(-)\right\}\mathbf{C}_0^\alpha(+) . \end{aligned} \tag{51}$$

Let us consider the case where a zigzag ribbon is attached to the scattering object at 0-th column. In order to do this, we have to derive an equation which relates $\mathbf{C}_0^\alpha$ to $\mathbf{C}_0^\beta$ and $\mathbf{C}_{-1}^\alpha$ in terms of the following equation,

$$\begin{bmatrix} E\mathbf{I}-\mathbf{H}_0^{\beta}, & -\mathbf{V}, & 0, & -\mathbf{V}^{\dagger} \\ -\mathbf{V}^{\dagger}, & E\mathbf{I}-\mathbf{H}_0^{\alpha}, & -\mathbf{V}, & 0 \end{bmatrix} \begin{bmatrix} \mathbf{C}_{-1}^{\beta} \\ \mathbf{C}_0^{\alpha} \\ \mathbf{C}_0^{\beta} \\ \mathbf{C}_{-1}^{\alpha} \end{bmatrix} = 0 \,. \tag{52}$$

The elimination of $\mathbf{C}_{-1}^{\beta}$ gives

$$(E\mathbf{I}-\mathbf{H}_{\mathrm{L}})\mathbf{C}_0^{\alpha} - \mathbf{V}_{\mathrm{L}}\mathbf{C}_0^{\beta} - \mathbf{V}_{\mathrm{L}}^{\dagger}\mathbf{C}_{-1}^{\alpha} = 0 \,. \tag{53}$$

Now we substitute (51) into the above equation in order to eliminate $\mathbf{C}_{-1}^{\alpha}$, and obtain obtain

$$(E\mathbf{I}-\mathbf{H}_{\mathrm{L}}-\mathbf{V}_{\mathrm{L}}^{\dagger}\mathbf{F}^{-1}(-))\mathbf{C}_0^{\alpha} - \mathbf{V}_{\mathrm{L}}\mathbf{C}_0^{\beta} = \mathbf{V}_{\mathrm{L}}^{\dagger}\left(\mathbf{F}(+)^{-1}-\mathbf{F}(-)^{-1}\right)\mathbf{C}_0^{\alpha}(+) \,. \tag{54}$$

Defining $\tilde{\mathbf{H}}_{\mathrm{L}}$ as

$$\tilde{\mathbf{H}}_{\mathrm{L}} = \mathbf{H}_{\mathrm{L}} + \mathbf{V}_{\mathrm{L}}^{\dagger}\mathbf{F}(-)^{-1} \,, \tag{55}$$

we obtain

$$(\mathbf{H}-\tilde{\mathbf{H}}_{\mathrm{L}})\mathbf{C}_0^{\alpha} - \mathbf{V}_{\mathrm{L}}\mathbf{C}_0^{\beta} = \mathbf{V}_{\mathrm{L}}^{\dagger}\left(\mathbf{F}(+)^{-1}-\mathbf{F}(-)^{-1}\right)\mathbf{C}_0^{\alpha}(+) \,. \tag{56}$$

On the other hand, in $\mathrm{N_s}+1$-th cell, only right-going waves exist, i.e.,

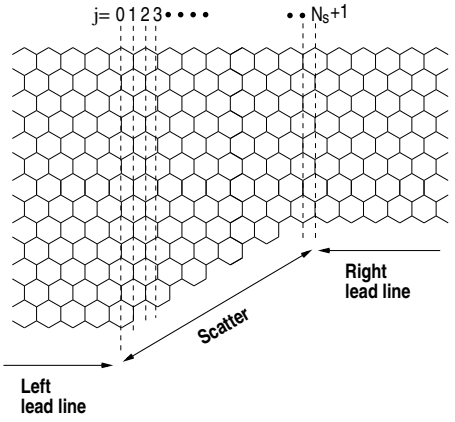

Fig. 21. The columns in the scattering region are denoted by j=0,1,2$\cdots$,N_s,N_s+1

$$\mathbf{C}_{\mathrm{N_s}+2}^{\alpha} = \mathbf{F}(+)\mathbf{C}_{\mathrm{N_s}+1}^{\alpha} \,, \tag{57}$$

and

$$\begin{bmatrix} E\mathbf{I}-\mathbf{H}_0^{\beta}, & -\mathbf{V}^{\dagger}, & 0, & -\mathbf{V} \\ -\mathbf{V}, & E\mathbf{I}-\mathbf{H}_0^{\alpha}, & -\mathbf{V}^{\dagger}, & 0 \end{bmatrix} \begin{bmatrix} \mathbf{C}_{\mathrm{N_s}+1}^{\beta} \\ \mathbf{C}_{\mathrm{N_s}+1}^{\alpha} \\ \mathbf{C}_{\mathrm{N_s}}^{\beta} \\ \mathbf{C}_{\mathrm{N_s}+2}^{\alpha} \end{bmatrix} = \mathbf{0} \,. \tag{58}$$

Eliminating $\mathbf{C}^{\beta}_{\mathrm{N_s}+1}$ we obtain

$$(E\mathbf{I} - \mathbf{H}_{\mathrm{R}})\mathbf{C}^{\alpha}_{\mathrm{N_s}+1} - \mathbf{V}^{\dagger}_{\mathrm{R}}\mathbf{C}^{\beta}_{\mathrm{N_s}} - \mathbf{V}_{\mathrm{R}}\mathbf{C}^{\alpha}_{\mathrm{N_s}+2} = 0\ . \tag{59}$$

Inserting (57) to eliminate $\mathbf{C}^{\alpha}_{\mathrm{N_s}+1}$, we find

$$(E\mathbf{I} - \tilde{\mathbf{H}}_{\mathrm{R}})\mathbf{C}^{\alpha}_{\mathrm{N_s}+1} - \mathbf{V}^{\dagger}_{\mathrm{R}}\mathbf{C}^{\beta}_{\mathrm{N_s}} = 0\ , \tag{60}$$

where

$$\tilde{\mathbf{H}}_{\mathrm{R}} = \mathbf{H}_{\mathrm{R}} + \mathbf{V}_{\mathrm{R}}\mathbf{F}(+)\ . \tag{61}$$

Thus, the Schrödinger equation for the scattering problem has the form

$$(E\mathbf{I} - \tilde{\mathbf{H}})\mathbf{C} = \mathbf{D}\ , \tag{62}$$

where $\mathbf{C}$, $\mathbf{D}$ and $\tilde{\mathbf{H}}$ are given by the following matrices,

$$\mathbf{C} = \begin{bmatrix} \mathbf{C}^{\alpha}_{0} \\ \mathbf{C}_{1} \\ \mathbf{C}_{2} \\ \mathbf{C}_{3} \\ \vdots \\ \mathbf{C}_{\mathrm{N_s}-1} \\ \mathbf{C}_{\mathrm{N_s}} \\ \mathbf{C}^{\alpha}_{\mathrm{N_s}+1} \end{bmatrix}, \qquad \mathbf{D} = \begin{bmatrix} \mathbf{V}^{\dagger}_{\mathrm{L}}(\mathbf{F}(+)^{-1} - \mathbf{F}(-)^{-1})\mathbf{C}^{\alpha}_{0}(+) \\ \vdots \\ \vdots \\ 0 \end{bmatrix} \tag{63}$$

and

$$\tilde{\mathbf{H}} = \begin{bmatrix} \tilde{\mathbf{H}}_{\mathrm{L}} & \mathbf{V}_{\mathrm{L}} & & & & & & & \\ \mathbf{V}_{1,0} & \tilde{\mathbf{H}}_{1} & \mathbf{V}_{1,2} & & & & & & \\ & \mathbf{V}_{2,1} & \tilde{\mathbf{H}}_{2} & \mathbf{V}_{2,3} & & & & & \\ & & \mathbf{V}_{3,2} & \tilde{\mathbf{H}}_{3} & \mathbf{V}_{3,4} & & & & \\ & & & \mathbf{V}_{4,3} & \tilde{\mathbf{H}}_{4} & \mathbf{V}_{4,5} & & & \\ & & & & \ddots & \ddots & \ddots & & \\ & & & & & \ddots & \ddots & \ddots & \\ & & & & & & \ddots & \ddots & \ddots \\ & & & & & & & \mathbf{V}_{\mathrm{N_s},\mathrm{N_s}-1} & \tilde{\mathbf{H}}_{\mathrm{N_s}} & \mathbf{V}_{\mathrm{N_s},\mathrm{N_s}+1} \\ & & & & & & & & \mathbf{V}^{\dagger}_{\mathrm{R}} & \tilde{\mathbf{H}}_{\mathbf{R}} \end{bmatrix} \tag{64}$$

respectively. Here $\mathbf{V}_{j,j+1}$ is the inter-column matrix from (j+1)-th column to j-th column.

Let us now define the Green function $\mathbf{G}$ as

$$\mathbf{G} = \left(E\mathbf{I} - \tilde{\mathbf{H}}\right)^{-1}\ . \tag{65}$$

Then we can derive the relations

$$\mathbf{C}^{\alpha}_{\mathrm{N_s}+1}(+) = \mathbf{C}^{\alpha}_{\mathrm{N_s}+1} = (\mathrm{N_s}+1|\mathbf{G}|0)V_L^{\dagger}(\mathbf{F}(+)^{-1} - \mathbf{F}(-)^{-1})\mathbf{C}^{\alpha}_0(+)$$

and

$$\mathbf{C}^{\alpha}_0(-) = \mathbf{C}^{\alpha}_0 - \mathbf{C}^{\alpha}_0(+) = [(0|\mathbf{G}|0)\mathbf{V}^{\dagger}_{\mathrm{L}}(\mathbf{F}(+)^{-1} - \mathbf{F}(-)^{-1}) - 1]\mathbf{C}^{\alpha}_0(+)\ . \quad (66)$$

From these equations, we can obtain the transmission coefficient $t_{\mu\nu}$ for the incident channel ν with velocity v_ν and out-going channel μ with velocity v_μ as

$$t_{\mu\nu} = \left[\frac{v_\mu}{v_\nu}\right]^{1/2} \left\{\mathbf{U}(+)^{-1}(N_s+1|\mathbf{G}|0)\mathbf{V}^{\dagger}_{\mathrm{L}}(\mathbf{F}(+)^{-1} - \mathbf{F}(-)^{-1})\mathbf{U}(+)\right\}_{\mu\nu} \quad (67)$$

and the reflection coefficient $r_{\mu\nu}$ for incident channel ν and out-going channel μ as

$$r_{\mu\nu} = \left[\frac{v_\mu}{v_\nu}\right]^{1/2} \left\{\mathbf{U}(-)^{-1}\left\{(0|\mathbf{G}|0)\mathbf{V}^{\dagger}_{\mathrm{L}}(\mathbf{F}(+)^{-1} - \mathbf{F}(-)^{-1}) - 1\right\}\mathbf{U}(+)\right\}_{\mu\nu} \quad (68)$$

Once we obtain the transmission coefficient $t_{\mu\nu}$, we can evaluate conductance by the Landauer formula,

$$G(E) = \frac{e^2}{\pi\hbar}\sum_{\mu,\nu}|t_{\mu\nu}(E)|^2\ . \quad (69)$$

Thus conductance can be calculated from the lattice GF.

4.4 Recursive Relation of the Green Function

In order to evaluate the transmission and reflection coefficients, we have to calculate the Green functions. Most powerful strategy is the recursive Green function method. Let us define the following four Green functions by,

$$(j|\mathbf{G}|j) = [j|(E\mathbf{I} - \mathbf{H}^{(j)})^{-1}|j] \quad (70)$$
$$(j|\mathbf{G}|0) = [j|(E\mathbf{I} - \mathbf{H}^{(j)})^{-1}|0] \quad (71)$$
$$(0|\mathbf{G}|j) = [0|(E\mathbf{I} - \mathbf{H}^{(j)})^{-1}|j] \quad (72)$$
$$(0|\mathbf{G}|0) = [0|(E\mathbf{I} - \mathbf{H}^{(j)})^{-1}|0] \quad (73)$$

where $\mathbf{H}^{(j)}$ is the total Hamiltonian for the strip comprising the 0 to j cells excluding the intercell Hamiltonian $\tilde{\mathbf{H}}_{j,j+1}$ and $\tilde{\mathbf{H}}_{j+1,j}$. Then the Green function for strips of any length can be obtained by a set of recursive formulas,

$$(j+1|\mathbf{G}^{(j+1)}|j+1)^{-1} = E\mathbf{I} - \tilde{\mathbf{H}}_{j+1} - \tilde{\mathbf{H}}_{j+1,j}(j|\mathbf{G}^{(j)}|j)\tilde{\mathbf{H}}_{j,j+1}$$

$$(j+1|\mathbf{G}^{(j+1)}|l) = (j+1|\mathbf{G}^{(j+1)}|j+1)\tilde{\mathbf{H}}_{j+1,j}(j|\mathbf{G}^{(j)}|l)$$

$$(l|\mathbf{G}^{(j+1)}|j+1) = (l|\mathbf{G}^{(j)}|j)\tilde{\mathbf{H}}_{j,j+1}(j+1|\mathbf{G}^{(j+1)}|j+1)$$

$$(l|\mathbf{G}^{(j+1)}|l) = (l|\mathbf{G}^{(j)}|l) + (l|\mathbf{G}^{(j)}|j)\tilde{\mathbf{H}}_{j,j+1}(j+1|\mathbf{G}^{(j+1)}|l)\ ,$$

where the suffix l must be less than $j+1$. In actual numerical calculations, we use the following Green function as a starting point for our recursive calculations.

$$(0|\mathbf{G}^{(0)}|0) = \left(E\mathbf{I} - \tilde{\mathbf{H}}_{\mathrm{L}}\right)^{-1} . \tag{74}$$

In terms of the above set of recursive relations, we can also calculate the Green function at arbitrary positions. By the analytic continuation, we can obtain the (local or total) density of states in the scattering region in terms of the following relation,

$$\rho(E) = -\frac{1}{\pi}\Im\mathrm{m}\mathbf{G}(E+i\eta) . \tag{75}$$

In addition, we can also study the behavior of the incident wave and current flow in the scattering region by this method.

4.5 Design of Single-Barrier Nanographite Ribbon Junctions

Before we discuss the design of the junctions and their conductance properties, let us summarize a few important facts concerning the low-energy states for zigzag and bearded ribbons.

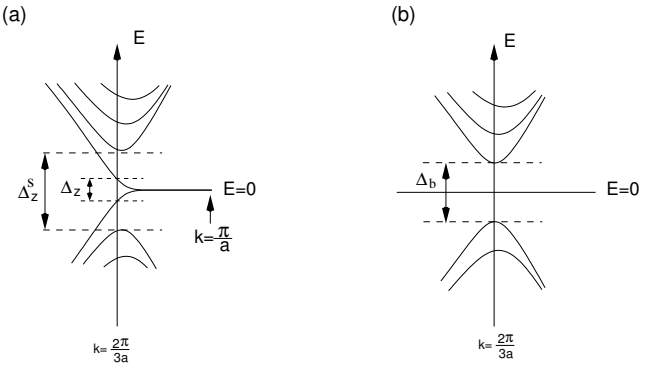

Fig. 22. The band structure near $E = 0$ of **(a)** zigzag ribbons and **(b)** bearded ribbons

Zigzag ribbon: The zigzag ribbons are metallic for arbitrary ribbon width with an energy dispersion near $E = 0$ as shown in Fig. 22(a). The partly flat band appears at $E = 0$ due to the edge states. Each edge state has a non-vanishing amplitude only on one of the two sublattices, i.e., non-bonding in character. However, in a zigzag ribbon of finite width, two edge states coming from both sides have a finite overlap. Because they are located on different sublattices, they mix into a bonding and anti-bonding configuration. In this way the partly flat bands acquire a dispersion. Note that the overlap increases

as k deviates from π/a, because then penetration depth of the edge states increases and it diverges at $k = 2\pi/3a$. The dispersion depends on the ribbon width N (number of zigzag lines from one side to the other), and has the approximate form

$$E_k = \pm 2tND_k^{N-1}\left[1 - \cos\left(\frac{ka}{2}\right)\right] , \tag{76}$$

where $D_k = 2\cos(ka/2)$. Thus, although the edge states on each side separately have non-bonding character, but taking together through their overlap they provide one conducting channel except at exactly $E = 0$. The energy region of single-channel transport is restricted by the energy gap Δ_z^s (see Fig. 22(a))

$$\Delta_z^s = 4t\cos\left[\frac{N-1}{2N+1}\pi\right] . \tag{77}$$

Note that Δ_z^s is different from Δ_z defined in Sect. 2, which is the direct gap of zigzag ribbons at $k = 2\pi/3$.

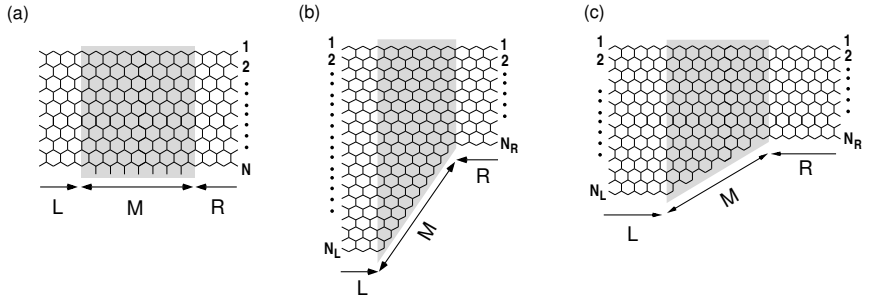

Fig. 23. The structure of junction **(a)** I, **(b)** II and **(c)** III

Bearded ribbon: The bearded ribbon has one zigzag edge and one edge which has additional bonds (beard) to the zigzag edge. Here the edge states of both edges exist on the same sublattice. Consequently, there is no mixing between these edge states that would remove the non-bonding character, and we obtain a completely flat band at $E = 0$ for any width N. The absence of dispersion leads to the insulating behavior for the edge state channel. The gap to the first conducting channel depends on N and is given by

$$\Delta_b = 4t\cos\left[\frac{N}{2N+2}\pi\right] . \tag{78}$$

Now let us turn to the design of junctions. In this subsection we study three types of junctions sandwiched by two zigzag ribbons denoted by L and

R. The middle region will be denoted by M. The junction I is depicted in Fig. 23(a). The M-region between the two zigzag ribbons of width N is a bearded ribbon which has l attached bonds. This is an example to illustrate the peculiar features of transport properties in zigzag ribbon junctions, regardless of the question as to whether bearded ribbons could be realized in nature. It should be noted that this model represents a metal-insulator-metal junction. In Fig. 23(b) and Fig. 23(c), the junctions II and III are depicted. These junctions connect zigzag ribbon leads of different width. The M-region contains a tilted zigzag edge for junction II and a tilted armchair edge for junction III. These junctions are more realistic than junction I. In the junctions II and III, the length of the M-region is proportional to $N_{\mathrm{L}} - N_{\mathrm{R}}$.

4.6 Fermi Energy Dependence of Conductance

We now discuss the energy dependence of conductance $G(E)$ of the junctions I - III. The energy may be considered as the chemical potential which could be adjusted by a gate voltage. The energy $E = 0$ corresponds to the undoped system which is half-filled.

Figure 24(a) shows the Fermi energy dependence of the ballistic conductance of junction I with $N = 10$ for the whole energy region, where the number of the attached bonds is 0, 1 or 3. Since the system with $l = 0$ is a perfect conductor, the ballistic conductance is proportional to the number of conducting channels at the Fermi energy, i.e., the number of subbands at the Fermi energy. The conductance has a clear step-like feature as a function of Fermi energy. With increasing number of attached bonds conductance decreases due to backward scattering. In the multi-channel energy region the structure of the function $G(E)$ is rather smooth, but in the single-channel region near $E = 0$ conductance has strong features for $l \neq 0$. Since in realistic systems the low-energy region close to $E = 0$ is most important, we shall focus our attention on the behavior of conductance in the low-energy region. The $G(E)$ of junction I in the single-channel region, $|E| < \Delta_b/2$, is shown in Fig. 24(b) and Fig. 24(c) as a function of Fermi energy. The characteristic feature is the appearance of zero-conductance dips at the specific values of E where the system shows complete back-scattering. The number of dips increases and the maximum height of the conductance decreases as we increase the number of attached bonds. Interestingly, even one-attached bond makes a zero-conductance dip. We should note that near the bottom of the valence bands (the top of the conduction bands), which is also a single-channel region, the dip structures of zero-conductance does not appear. Since the bottom of the valence bands (the top of the conduction bands) has the character of ordinary free electrons, the appearance of the zero-conductance dips near $E = 0$ is related to the topology of the lattice and the electron spectrum around $E = 0$. We should remark that the conductance of the junctions connecting two nanotubes of different circumference does not show this type of dip structures [53].

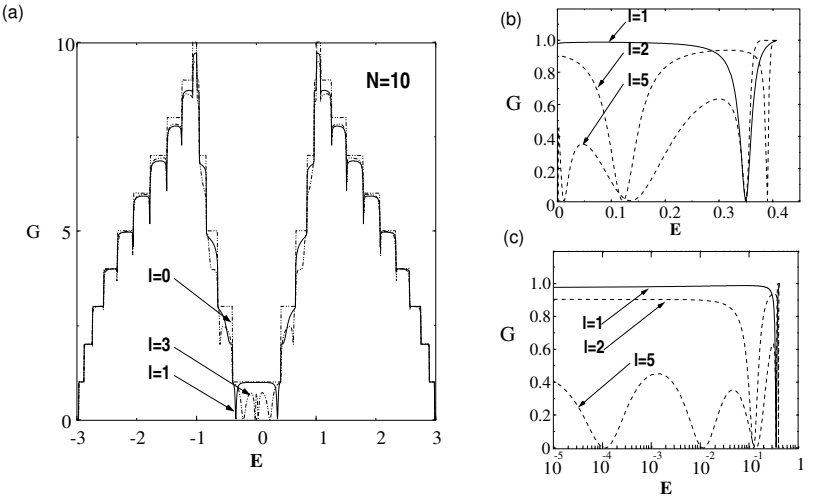

Fig. 24. The Fermi energy dependence of the transmission probability of junction I for **(a)** the whole energy region, **(b)** the single conducting channel energy region in linear-scale, and **(c)** the single conducting channel energy region in log-scale

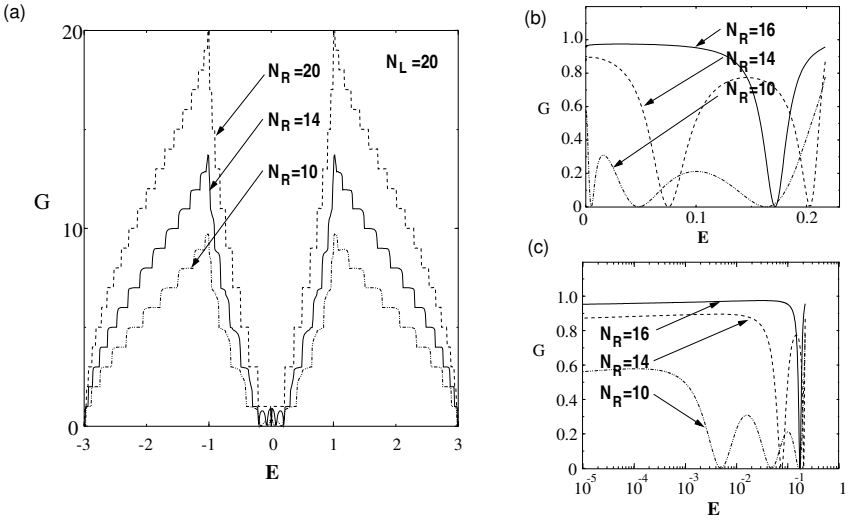

Fig. 25. The Fermi energy dependence of the transmission probability of junction II for **(a)** the whole energy region and **(b)** the single conducting channel energy region in linear-scale, and **(c)** the single conducting channel energy region in log-scale

In Fig. 25, the Fermi energy dependence of the conductance of junction II for (a) the whole energy range, (b) the single conducting channel region and (c) single conducting channel region with log-scale are displayed. Here we define the single-channel region as $|E| < \Delta_b(N_{\rm L})/2$, where both leads have a single conducting channel ($N_{\rm L} > N_{\rm R}$ implies $\Delta_b(N_{\rm L}) < \Delta_b(N_{\rm R})$). Similarly, Fig. 26 is for the junction III. For both the junctions, we fix the width of the left zigzag ribbon at $N_{\rm L} = 20$, while the width of the right zigzag ribbon is varied as $N_{\rm R} = 16$, 14, 10. Since the maximum number of conducting channels is equal to the width of the lead lines, conductance decreases with the decrease of width of the right zigzag ribbon in the multi-channel region. The behavior of conductance of the junctions II and III are qualitatively same in the multi-channel region. However, in the single-channel region near $E = 0$, the behavior of conductance of the junctions II and III are quite different. The conductance of junction III is smooth even in the single-channel region, but for junction II dip structures appear analogous to those of junction I.

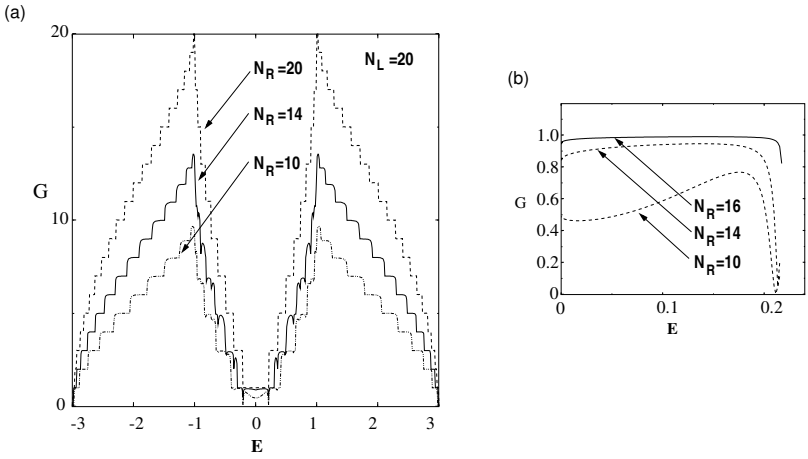

Fig. 26. The Fermi energy dependence of the transmission probability of junction III for **(a)** the whole energy region and **(b)** the single conducting channel energy region

In junction II, with the increase of length of the M-region, the number of the zero-conductance dips increases and conductance decreases in magnitude. The tilted edge in junction II supports an edge state which, similar to the bearded ribbon, lies on the same sublattice as the edge state on the other side of the ribbon.

4.7 Large Induced Current Vortex

The appearance of the zero-conductance resonances can be understood by the formation of standing waves due to the interference effects between the incident and scattered electron waves [15]. This interference effects produce additional structures in the spatial distribution of electric currents over the scattering region of nanographite ribbon junctions. In this subsection, we draw attention on the spatial distribution of electric currents close to the energies of the zero-conductance dips. For energies close to a zero-conductance resonance, a Kekulé-like vortex pattern appears over the scattering region. The nearly regular pattern reminds of a flux state. We will see that the resonant state responsible for the zero-conductance dip may also be considered as a standing wave due to the superposition of such a "flux state" and its time-reversed state.

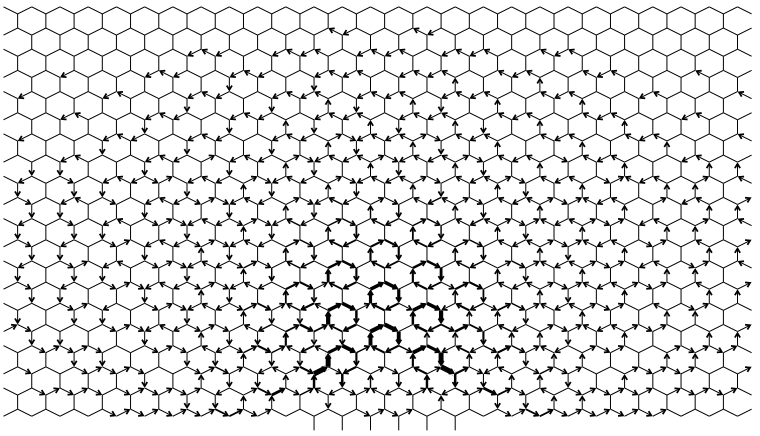

Fig. 27. The distribution of currents in the M-region of junction I with $N = 20$ and $l = 6$, at $E = 0.049565t$ (immediately below E_2 in the Fig. 28)

In order to visualize the spatial distribution of electric currents, we study here bond current distribution. The bond current flowing from site-j to site-i is defined by

$$J_{i,j} = \mathrm{i}\frac{et}{\hbar}\left(\mathrm{e}^{\mathrm{i}2\pi\theta_{i,j}}|i\rangle\langle j| + \mathrm{h.}c\right) \ . \tag{79}$$

In Fig. 27, we show the distribution of currents for junction I with $N = 20$ and $l = 6$, at $E = 0.049565$. This energy lies immediately below E_2, a zero-conductance resonance shown in Fig. 28(a). The overall current gives rise to a clockwise vorticity. In the junction region we observe a clear triangular Kekulé pattern of the current vortex. Interestingly, the center of the junction region

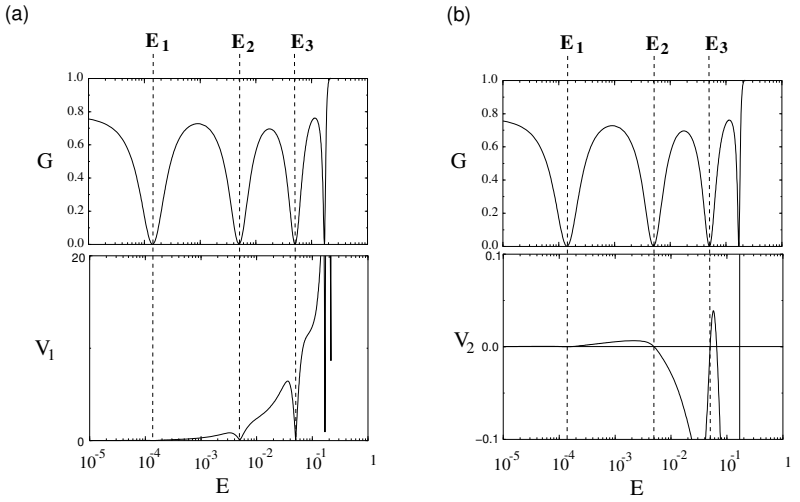

Fig. 28. The Fermi energy dependence of **(a)** V_1 and **(b)** V_2, together with the conductance, for junction I with $N = 20$ and $l = 6$

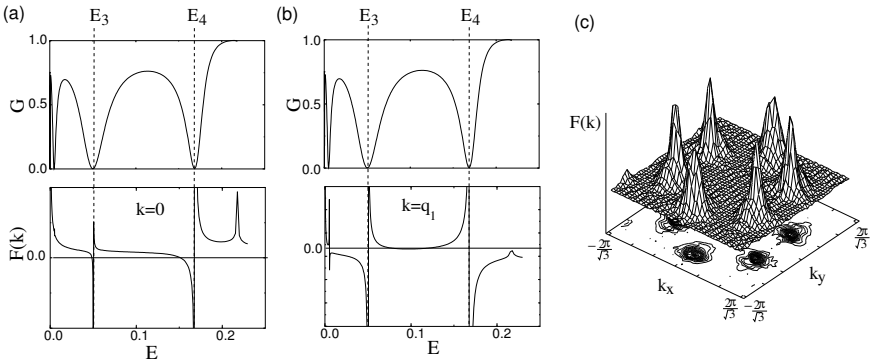

Fig. 29. The Fermi energy dependence of the Fourier transform of vorticity V_p at **(a)** $k = 0$ and **(b)** $k = q_1$ for junction I with $N = 20$ and $l = 6$. **(c)** The 3D plot of the correlation of circular current pattern in the M-region close to $E = E_2$

supports large circulating currents while the net current passing through the junction is rather small. We also see that the current pattern rapidly vanishes when we leave the junction region. When we increase the energy slightly above E_2, we find a very similar current distribution pattern in the junction region. While the net current is flowing in the same direction as in the previous case, the vortex pattern shows opposite orientation and the overall vorticity is also reversed.

In order to analyze the features of this current vortex pattern in more detail, we introduce the local *vorticity* which is defined on the dual (triangular) lattice as the clockwise sum of the currents flowing on the bonds of each hexagonal ring. The local vorticity on the $p^{\rm th}$ hexagonal ring is given by

$$V_p = \sum_{i=1}^{6} I_{i,p} \ , \tag{80}$$

where $I_{i,p}$ means the current on the $i^{\rm th}$ bond of $p^{\rm th}$ hexagonal ring. In order to quantify the total of all circulating currents flowing in the junction, we take an average of $|V_p|$. This quantity, V_1, is given by

$$V_1 = \langle |V_p| \rangle = \frac{\sum_p |V_p|}{\sum_p 1} \ . \tag{81}$$

Similarly, we also define the total vorticity of the junction, V_2, as

$$V_2 = \langle V_p \rangle = \frac{\sum_p V_p}{\sum_p 1} \ . \tag{82}$$

This represents the direction of the total circulating current component of the junction. It should be noted that the summation over plaquettes is taken over rings in the junction region M and slightly beyond including several columns of rings in the leads because there is a *proximity effect* of the current vortex pattern, i.e., the components of circulating currents penetrate into both the left and the right lead lines. In Fig. 28, we show the Fermi energy dependence of V_1 and V_2 for the junction I with $N = 20$ and $l = 6$ in the single conducting channel region. Both V_1 and V_2 vanish at each energy of zero-conductance, which means that not only the total vorticity V_2 disappears, but also no circular currents can be found in each individual plaquette ($V_1 = 0$). Moreover, the vorticity V_2 changes its sign at each zero-conductance energy point. Note that both V_1 and V_2 exhibit a linear $(E - E_n)$ dependence close to the zero-conductance point E_n. The fact that both vorticities go to zero at each zero-conductance energy E_n verifies the claim that the resonant state in the junction region may also be interpreted as a standing wave formed by the superposition of two flux phase-like states that are connected with each other by means of time-reversal operation. This also suggests that the resonance should disappear once time-reversal symmetry is explicitly violated, as for example, by an external field.

The study of V_p allows us also to observe the formation of the Kekulé-like vortex pattern easily, if we transform V_p into momentum space,

$$F(\boldsymbol{k}) = \sum_p V_p \cos(\boldsymbol{k} \cdot \boldsymbol{r}_{\rm p}), \tag{83}$$

where $\boldsymbol{r}_{\rm p}$ is the coordinate of the hexagonal ring center, and $k_x(k_y)$ is the wave number along (perpendicular to) zigzag lines in the junction. The sum runs

again over all rings in the junction and a few columns beyond. In Fig. 29(a) and Fig. 29(b), we show $F(\boldsymbol{k})$ again for junction I with $N = 20$ and $l = 6$ close to the zero-conductance point E_2. We see the Bragg peaks at $\boldsymbol{q}_1 = \frac{2\pi}{a}(\frac{1}{\sqrt{3}}, \frac{1}{3})$ (or $\boldsymbol{q}_2 = \frac{2\pi}{a}(0, \frac{2}{3})$) and $\boldsymbol{q}_3 = 0$, corresponding to a triangular correlation of the flux state-like current vortex pattern depicted in Fig. 27. In Fig. 29(c), we show the 3D-plot of the $F(\boldsymbol{k})$, where the pronounced triangular symmetry can be observed. We would like to mention here that we have observed analogous current pattern and properties of V_1 and V_2 for junction II.

Finally we would like to draw attention to a remarkable property of the vorticity V_2 (and V_1) close to each zero-conductance point. The net current passing through the junction is defined as

$$J_{\rm net} = (1 - |r|^2) J_{\rm lead} = |t|^2 J_{\rm lead} \ , \tag{84}$$

where $J_{\rm lead}$ is the incoming component of the current on the source lead line from one of the reservoirs. Close to each zero-conductance point E_n we find that $J_{\rm net} \propto (E - E_n)^2$, a quadratic dependence. With the linear dependence of $V_2 \propto (E - E_n)$, the total vorticity of the system considered as a "response" to a current through the junction ($J_{\rm net}$) diverges as the energy approaches E_n, i.e., $V_2(E)/J_{\rm net}(E) \propto (E - E_n)^{-1}$. Consequently, for energies E close to a zero-conductance point, even a small net current may generate a large vorticity, within the linear response regime. This is not a real linear response, since the external source corresponds to the lead current $J_{\rm lead}$, and the actually measured current $J_{\rm net}$ includes all the scattering renormalizations. Nevertheless, the relation between vorticity and transmitted current may be experimentally verified.

4.8 Connection with Asymmetric Aharonov-Bohm Ring

In this section we consider a simple model which has analogous electronic transport properties as the nanographite ribbon junctions and we study it on the basis of the scattering matrix theory developed by Büttiker and coworkers [54]. The system that we study is the single-channel asymmetric Aharonov-Bohm (AB) ring connected to current leads, as shown in Fig. 30. We will adopt the notation of [54]. This system reproduces well the qualitative properties of our ribbon junctions. If the two branches in the AB-ring have different length or different transmission probabilities, zero-transmission resonances appear as a function of energy. A current through the device generates circulating currents in the ring which change sign at each zero-conductance resonance. The application of an external field leading to a finite magnetic flux through the ring yield a negative magnetoresistance at the zero-conductance resonance. We also find that at the zero-conductance resonance, the two branches of the ring possess electron wave functions of opposite parity. All these features are common with the nanographite ribbon junctions.

Following the theory developed by Büttiker, we consider the single-channel electron transport through the AB-ring as shown in Fig. 30, where the upper-

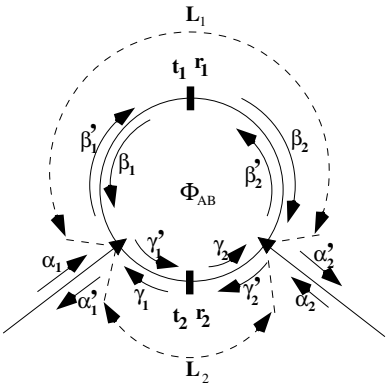

Fig. 30. Schematic figure of Aharonov-Bohm ring, where the definition of the amplitudes of wave functions are written. The length of the upper(lower) branch of the ring is $L_1(L_2)$

and lower-branch have different lengths, L_1 and L_2, respectively. The circumference is $L = L_1 + L_2 = (1+R)L_1$. The notations for the amplitudes of wave functions on each branch are given in Fig. 30. We assume that each branch has only one conduction channel. The upper(lower) branch has a scatterer expressed by the transfer matrix $\underline{t}_1$ ($\underline{t}_2$), which relates the amplitudes to the left with the amplitudes to the right of the scatterer. These transfer matrices are defined as,

$$\begin{bmatrix} \beta_2 \\ \beta_2' \end{bmatrix} = \underline{t}_1 \begin{bmatrix} \beta_1' \\ \beta_1 \end{bmatrix} = \begin{bmatrix} 1/t_1^* & -r_1^*/t_1^* \\ -r_1/t_1 & 1/t_1 \end{bmatrix} \begin{bmatrix} \beta_1' \\ \beta_1 \end{bmatrix} , \tag{85}$$

and

$$\begin{bmatrix} \gamma_1 \\ \gamma_1' \end{bmatrix} = \underline{t}_2 \begin{bmatrix} \gamma_2' \\ \gamma_2 \end{bmatrix} = \begin{bmatrix} 1/t_2^* & -r_2^*/t_2^* \\ -r_2/t_2 & 1/t_2 \end{bmatrix} \begin{bmatrix} \gamma_2' \\ \gamma_2 \end{bmatrix} , \tag{86}$$

where $t_i = T_i^{1/2} \mathrm{e}^{\mathrm{i}\phi_i}$ $(i = 1, 2)$ is the transmission amplitude of the scatterer, T_i the transmission probability, and ϕ_i the phase shift of the transmitted wave ($i = 1$ and 2 represents upper and lower branchs, respectively). Here r_i (r_i') is the reflection amplitude. It is sufficient for our purpose to consider the case where $r_i = 0$, i.e., we have perfect transmission ($T_i = 1$).

Now let us consider the junction between lead and ring. The amplitudes of the three outgoing waves are connected with the three incoming ones via a 3×3 scattering matrix (which depends on three parameters only [54]),

$$\begin{bmatrix} \alpha' \\ \beta' \\ \gamma' \end{bmatrix} = \underline{\mathrm{S}} \begin{bmatrix} \alpha \\ \beta \\ \gamma \end{bmatrix} = \begin{bmatrix} -(a+b) & \epsilon^{1/2} & \epsilon^{1/2} \\ \epsilon^{1/2} & a & b \\ \epsilon^{1/2} & b & a \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \\ \gamma \end{bmatrix} , \tag{87}$$

where ϵ ($0 \leq \epsilon \leq 1/2$) is the key parameter determining the coupling between the ring and a lead. Note that the matrix $\underline{\mathrm{S}}$ is unitary because of the current conservation and symmetric because of the time-reversal invariance. The probability (current) conservation requires that

$$(a+b)^2 + 2\epsilon = 1 \,, \tag{88}$$

$$a^2 + b^2 + \epsilon = 1 \,. \tag{89}$$

Thus we can rewrite a and b as the functions of ϵ: $a_\pm = \pm\frac{1}{2}\left(\sqrt{1-2\epsilon}-1\right)$ and $b_\pm = \pm\frac{1}{2}\left(\sqrt{1-2\epsilon}+1\right)$.

Now we determine the transmission amplitude and the circulating currents in this system for some arbitrary value of ϵ. The boundary conditions are $\alpha_1 = 1$ and $\alpha_2 = 0$, i.e., an incident wave from the left lead line. It is straightforward to calculate the amplitude of the transmitted wave:

$$\alpha_2' = -\mathrm{e}^{-\mathrm{i}\theta_1} \frac{\epsilon h}{b^2 \det(\underline{\Pi})} \,, \tag{90}$$

where

$$h = \det(\underline{\Pi})[b-a, 1]\underline{t}_1 \underline{\Pi}^{-1} \begin{bmatrix} b-a \\ -1 \end{bmatrix} \,, \tag{91}$$

with $\underline{\Pi} = \underline{t}_l \mathrm{e}^{-\mathrm{i}\theta_2} \underline{t}_2' \underline{t}_l \mathrm{e}^{-\mathrm{i}\theta_1} \underline{t}_1 - \underline{1}$. We have generalized the problem here including magnetic flux ϕ through the loop with $\theta = \theta_1 + \theta_2 = 2\pi\phi/\phi_0$. The phase shifts due to the gauge field in the upper and lower branch are $\theta_1 = \theta/(1+R)$ and $\theta_2 = R\theta/(1+R)$, respectively. The link matrix $\underline{t}_l$ entering in $\underline{\Pi}$ is defined as

$$\begin{bmatrix} \gamma_2' \\ \gamma_2 \end{bmatrix} = \underline{t}_l \begin{bmatrix} \beta_2 \\ \beta_2' \end{bmatrix} = \frac{1}{b} \begin{bmatrix} b^2 - a^2 & a \\ -a & 1 \end{bmatrix} \begin{bmatrix} \beta_2 \\ \beta_2' \end{bmatrix} \,. \tag{92}$$

It is also straightforward to obtain the amplitudes in the two branches as follows,

$$\begin{bmatrix} \beta_1' \\ \beta_1 \end{bmatrix} = -\frac{\sqrt{\epsilon}}{b} \underline{\Pi}^{-1} \begin{bmatrix} b-a \\ -1 \end{bmatrix} \,, \tag{93}$$

$$\begin{bmatrix} \gamma_1 \\ \gamma_1' \end{bmatrix} = \frac{\sqrt{\epsilon}}{b} \tilde{\Pi}^{-1} t_l^{-1} \begin{bmatrix} b-a \\ -1 \end{bmatrix} \,, \tag{94}$$

with $\underline{\tilde{\Pi}} = \underline{t}_l^{-1} \mathrm{e}^{\mathrm{i}\theta_1} \underline{t}_1^{-1} \underline{t}_l^{-1} \mathrm{e}^{\mathrm{i}\theta_2} \underline{t}_2'^{-1} - \underline{1}$.

We can then express the transmission amplitude including magnetic flux as

$$\alpha_2'(\phi, \epsilon) = \frac{\mathrm{i}\epsilon(\sin\phi_1 + \mathrm{e}^{\mathrm{i}\theta}\sin\phi_2)}{a^2\cos\gamma\phi + b^2\cos\theta - (1-\epsilon)\cos\phi + \mathrm{i}\epsilon\sin\phi} \,, \tag{95}$$

where $\phi = \phi_1 + \phi_2$ and $\gamma = (1-R)/(1+R)$. It should be noted that ϕ_1 and ϕ_2 can be written as

$$\phi_1 = \frac{1}{1+R}\phi \quad \text{and} \quad \phi_2 = \frac{R}{1+R}\phi \,. \tag{96}$$

The transmission probability $T(\phi, \epsilon) = |\alpha_2'(\phi, \epsilon)|^2$, proportional to the conductance, now shows zero-conductance dips in the absence of magnetic flux ($\theta = 0$). The transmission zeros appear for $\phi = \phi_0$ given by

$$\phi_{0,1} = 2m\pi \qquad \text{or} \qquad \phi_{0,1} = (2m+1)\pi/\gamma, \tag{97}$$

where m is an integer.

The resonant behavior of transmission probability is determined by the poles of the transmission amplitude (95). In order to determine the poles of (95), we rewrite ϕ as $\phi = \phi_r + \mathrm{i}\phi_i$. The real and imaginary parts of the phase ϕ are determined by the following two equations,

$$2\sin\gamma\phi_r \sinh\gamma\phi_i - \sin\phi_r(\mathrm{e}^{\phi_i} - (1-2\epsilon)\mathrm{e}^{-\phi_i}) = 0 \,, \tag{98}$$

$$a^2\cos\gamma\phi_r\cosh\gamma\phi_i + b^2\cos\theta - \frac{1}{2}\cos\phi_r(\mathrm{e}^{\phi_i} + (1-2\epsilon)\mathrm{e}^{-\phi_i}) = 0 \,. \tag{99}$$

We find that a resonance solution is associated with two poles in complex ϕ-plane.

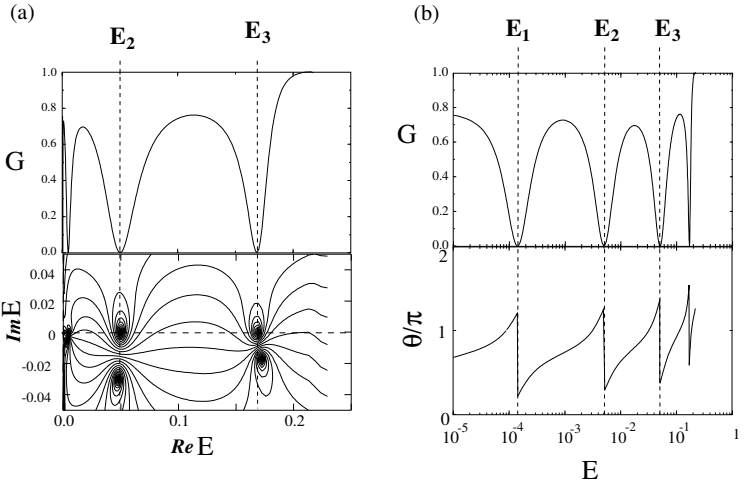

Fig. 31. Numerical results for junction I with $N = 20$ and $l = 6$: **(a)** The contour plot of the absolute value of the transmission amplitude in the complex energy plane. The zero-pole pairs appear at each zero-transmission energy. **(b)** The energy dependence of the phase of transmission

In the asymmetric case, i.e., $\gamma \neq 0$ ($R \neq 1$), it is not easy to deal with (98) and (99) in a simple analytic way. The numerical analysis shows two types of

solutions of zero-transmission, $\phi_{0,1}$ and $\phi_{0,2}$, which have the following different characteristic features.
i) $\phi = \phi_{0,1}$ solution: This solution satisfies (95) for arbitrary γ, and gives a zero-transmission resonance with in general one zero point and two poles in complex ϕ-plane. In the strong coupling limit of $\epsilon = 1/2$ with $\gamma \neq 1$, one of the two poles goes to infinity, so that the resonance can be characterized by a zero-pole pair in the complex ϕ-plane, resulting in the zero-conductance resonance. On the other hand, when the ring is symmetric, i.e., $\gamma = 0$ ($R = 1$) for arbitrary coupling, the zero point and one of two poles are canceled so that only one pole is left on the complex ϕ-plane. Then the feature of the resonance is usual resonant transmission due to one pole. Thus the $\phi_{0,1}$ solution gives the condition of anti-resonance for the asymmetric AB-ring system.
ii) $\phi = \phi_{0,2}$ solution: This solution appears when the ring is asymmetric, i.e., $\gamma \neq 0$. Although this solution also gives a zero point and poles, the zero point and poles are always degenerate on the real axis for arbitrary γ. Therefore the zero-transmission resonance of this solution has different character than that of the $\phi_{0,1}$ solution. The degeneracy of poles and zero point is not lifted by the variation of the coupling parameter ϵ.

In Fig. 31(a), we show the contour plot of the absolute value of transmission amplitude in the complex energy plane for junction I with $N = 20$ and $l = 6$. We clearly see the zero-pole pair at each zero-conductance energy points. Thus we can consider that the nanographite ribbon junction systems correspond to the asymmetric AB-ring systems in the strong coupling limit ($\epsilon \to 1/2$). In the strong coupling limit, we can rewrite the transmission probability as

$$T = |\alpha_2'(\epsilon \to \frac{1}{2})|^2 = \frac{4|\sin\phi_1 + \mathrm{e}^{\mathrm{i}\theta}\sin\phi_2|^2}{(\cos\gamma\phi + \cos\theta - 2\cos\phi)^2 + 4\sin\phi} \,. \tag{100}$$

Since the zero-conductance resonances can be characterized by the zero-pole pair on the complex energy plane, one zero-conductance resonance behavior can be written by the following Brite-Wigner form,

$$t_{\mathrm{BW}}(E) = \tilde{t}(E)\frac{E - E_0}{E - (E_p - \mathrm{i}\Gamma)} \tag{101}$$

where Γ is the width of the resonance. Here E_0 and E_p respectively gives the position of the zero and the pole. In general, E_0 is not equal to E_p. When $E_0 = E_p$, the resonance have the symmetric Lorentzian form. It is instructive to consider the behavior of the phase θ_{BW} of transmission coefficient when $|t_{\mathrm{BW}}|^2$ passes through zero. The phase is defined as

$$\theta_{\mathrm{BW}} = \tan^{-1}\left[\frac{\Im\mathrm{m}(t_{\mathrm{BW}})}{\Re\mathrm{e}(t_{\mathrm{BW}})}\right] \,. \tag{102}$$

It is easy to confirm that the phase of the transmission amplitude with the form of (101) jumps at $E = E_0$ (not at $E = E_p$) by π.

In Fig. 31(b), the phase of the transmission is depicted, where clear π phase jumps are observed at each zero-conductance energy points. It should be noted that recently the behavior of the phase of transmission coefficient has attracted much interests [55–58] in the context of the experiments of Yacoby et al. [59] and Schuster et al. [60]. It may be mentioned that the appearance of zero-transmission resonances accompanied with the zero-pole pair can be seen not only in the asymmetric AB-ring system, but also in the quantum wire system with an attached resonator [61–68]. Thus electron transport through the nanographite ribbon junctions is deeply connected to both of them.

Next we study the circulating current that flows in the AB-ring. The electric current in the upper(lower) branch J_1 (J_2) is given by $|\beta_1|^2 - |\beta_1'|^2$ ($|\gamma_1'|^2 - |\gamma_1|^2$), so that we define the vorticity ($V_{\rm AB}$) of electric currents through the AB-ring as $J_1 + J_2$. After some simple mathematical manipulations, we obtain the vorticity $V_{\rm AB}$ as

$$V_{\rm AB} = J_1 + J_2 = \frac{2\epsilon b \sin\phi[\sin\gamma\phi + (a+b)\sin\theta]}{\left[a^2\cos\gamma\phi + b^2\cos\theta - (1-\epsilon)\cos\phi\right]^2 + \epsilon^2\sin^2\phi} \ . \quad (103)$$

In the strong coupling limit ($\epsilon \to 1/2$), we can rewrite the above equation as follows,

$$V_{\rm AB}\left(\epsilon \to \frac{1}{2}\right) = \frac{8\sin\phi\sin\gamma\phi}{\left[\cos\gamma\phi + \cos\theta - 2\cos\phi\right]^2 + 4\sin^2\phi} \ . \quad (104)$$

We find that the vorticity $V_{\rm AB}$ changes its sign at the energies of zero-transmission resonances, i.e., at $\phi = \phi_0$. At the energies of zero-transmission, the vorticity $V_{\rm AB}$ becomes zero as no circulating currents can flow in the system. Since the expression for vorticity $V_{\rm AB}$ also has the resonance features, the vorticity becomes stronger around the energies of zero-transmission resonances. These behaviors of vorticity are consistent with the results of our numerical calculations on the nanographite junction systems.

The zero-conductance resonance is a consequence of the destructive time-reversal interference. The application of magnetic field can destroy these zero-transmission resonances, resulting in the negative magnetoresistances. In the weak magnetic field limits ($\theta \ll 1$), we can easily derive the magnetic field dependence of the transmission probability as follows.

$$T(\theta) = \left[\frac{\sin(m\gamma\pi)}{a^2\cos(2m\gamma\pi) + b^2 + \epsilon - 1}\right]\theta^2 + O(\theta^3) \qquad \text{at} \quad \phi_0 = 2m\pi \ , \quad (105)$$

$$T(\theta) = \left[\frac{\cos^2((2m+1)\pi/2\gamma)}{\left[a^2 + b^2 - (1-\epsilon)\cos((2m+1)\pi/\gamma)\right]^2 + \epsilon^2\sin^2((2m+1)\pi/\gamma)}\right]\theta^2 + O(\theta^3) \qquad \text{at} \quad \phi_0 = (2m+1)\pi/\gamma \ . \quad (106)$$

In both the cases, the zero-transmission gets removed in the form of θ^2.

In the summary of this section, we have established a connection between electron transport through nanographite ribbons and electron transmission

through the asymmetric Aharonov-Bohm ring. We observe that the zero-conductance resonances are the interference effect of two transmission paths that split in the scattering region of the nanographite ribbon junctions. The interference effects can be visualized by the formation of the standing waves whose parity is different on the two edges of the junctions, and also by the formation of circulating currents with Kékule-like vortex pattern. Recently, we have also studied the effect of the Rashba spin-orbit coupling on the zero-conductance resonances within the framework of the scattering matrix theory, where the lifting of conductance zeros due to spin-orbit coupling is related to the breaking of the spin reversal symmetry [69].

4.9 Summary

In this article, we have investigated the electronic, magnetic and transport properties of nanographite systems. It is found that the electronic states are strongly influenced by the existence of the graphite edges and also by their shapes. Zigzag edges produce localized edge states, which give a sharp peak in the density of states at the Fermi level.

The orbital magnetic susceptibility $\chi_{\rm orb}$ of nanographites has the intermediate values between those of aromatic molecules and bulk graphites. The $\chi_{\rm orb}$ can be scaled as a function of Fermi energy, temperature and ribbon width. Zigzag ribbons show large Pauli paramagnetic response at low-temperature due to the existence of sharp peak in the density of states at the Fermi level, while in the armchair ribbons Pauli paramagnetic susceptibility $\chi_{\rm P}$ is negligible. Since the width of the peak is of the order of meV, the paramagnetic susceptibility $\chi_{\rm P}$ is sensitive to temperature resulting the Curie-like behavior. In nanographite systems, this paramagnetic response competes with the diamagnetic responses, and we have crossover from high-temperature diamagnetic response to low-temperature paramagnetic response.

The edge states are not stable even for small onsite Coulomb interaction, and it yields localized spins of about $0.2\,\mu_{\rm B}$ at the edge sites. In the zigzag ribbons spins align ferrimagnetically due to the non-bonding nature of the edge states. The two magnetic states which are polarized at the zigzag edges couple antiferromagnetically resulting in zero total magnetization and this result is consistent with the Lieb's argument. Since the zigzag ribbons have even sites in the unit cell, the ground state of the system can be considered as a resonating valence bond state, which induce a gap for the spin excitation. However, since the coupling between two edges decreases rapidly with the increase of width, the spin gap is negligibly small when the ribbon width has a nanometer-scale.

The methods that we discuss here are also applicable to other edge shapes that support non-bonding edge states (e.g., the bearded and cove edges). On the basis of our analysis in the previous sections, the non-bonding edge states are expected at the edges which are not parallel to the armchair edge. Thus, there is a class of nanographites that show a crossover from the paramagnetic

phase at low temperature to the diamagnetic phase at high temperature, which can be assigned as the characteristic feature of this class of nanographite systems.

In the activated carbon fiber (ACF) and graphitized nanodiamonds, the dangling bonds are terminated by other elements. Nevertheless, their magnetic susceptibility shows the Curie-Weiss behavior originating from the existence of localized spins [20, 21]. Since in the ACF and graphitized nanodiamond systems dangling bonds are terminated, the presence of the dangling bond spins are ruled out. Our results show that the non-bonding edge states give strong Pauli paramagnetic response even in the absence of electron-electron interaction. Thus, we conclude that the localized spin originating from the non-bonding edge states is one of the strong candidates to explain the behavior of magnetic response of nanographite systems.

In general on observes that the graphite-related materials are quite silent in respect of their magnetic response. However, our results suggest that in nanographite systems paramagnetic response due to the non-bonding edge states are very important, and the nanographites are quite different from the bulk graphites and aromatic molecules from the viewpoint of their magnetic properties.

We have also studied the electronic transport properties of nanographite ribbons by the Landauer-Büttiker approach within the tight-binding framework, sandwiching the sample between two zigzag ribbons of same or different widths. The zigzag shape of a graphite edge provides a non-bonding edge localized state at $E = 0$. A single edge state cannot contribute to electron transport due to this non-bonding character, however, in the zigzag ribbons the bonding and anti-bonding configuration between two edge states can provide a single conducting channel. Our numerical study shows that the electrical conductance of nanographite ribbon junctions crucially depends on the morphology of the edge shapes.

In the single-channel conducting region, the Fermi energy dependence of electrical conductance of nanographite ribbon junctions exhibit rich structures with sharp zero-conductance dips. We have analyzed the origin of the zero-conductance dips from the behavior of the electronic wave functions and electric currents. Each zero-conductance resonance can be associated with a quasi-bound state in the scattering region of the junctions with the formation of standing waves. It is also found that in the scattering region electron waves split into two edge localized electron waves which have opposite parity.

Furthermore, when the energies of the incident electron waves are close to the energy of zero-conductance resonances, the electric currents form circulating currents with Kekulé-like vortex pattern in the scattering region. This behavior of electric current resembles the flux states in the sense that they have strong current-current correlation. The flux states that appear in the scattering region decrease rapidly outside this region. This circulating currents change their directions as we cross the energies of the zero-conductance

dips. Therefore, the zero-conductances are caused by the superposition of two flux states with opposite chirality, resulting in the formation of standing waves.

We have presented not only the numerical analysis, but also the phenomenological theory for the zero-conductance resonances associated with the electric current vortex. It has been pointed out that the single-channel transport through the nanographite ribbon junctions can be connected to the physics of the asymmetric Aharonov-Bohm ring system. The zero-conductance resonance is a consequence of the destructive interference effects between two conducting-channels states in the scattering region which arise from the splitting of the single-channel conducting edge states. The analysis in the complex energy plane shows the Brite-Wigner form of (101). This means that the zero-conductance resonance can be viewed as the so-called Fano resonances [70], which are known to occur when two scattering channels are available, one corresponding to a continuum of states and the other to a discrete quasi-bound state. It is also known that the Fano resonances can occur in quantum wire with a t-stub resonator [63, 64]. Thus, the single-channel electron transport through nanographite ribbon junctions has the similarities not only with the asymmetric Aharonov-Bohm ring, but also with the quantum wire with a t-stub resonator.

The zero-conductance resonances are the consequence of the time-reversal symmetry of the system. The application of a magnetic field removes these zero-conductance dips yielding a pronounced negative magnetoresistance. Since carbon nanotubes show large positive magnetoresistance, the negative magnetoresistance in nanographite systems is in sharp contrast.

References

1. Y. Imry: *Introduction to Mesoscopic Physics* (Oxford University Press 1997); S. Datta: *Electronic Transport in Mesoscopic Systems* (Cambridge University Press 1995)
2. S. Iijima: Nature **354**, 56 (1991)
3. M. S. Dresselhaus, G. Dresselhaus, P. C. Eklund: *Science of Fullerenes and Carbon Nanotubes* (Academic Press, San Diego 1996)
4. R. Saito, G. Dresselhaus, M. S. Dresselhaus: *Physical Properties of Carbon Nanotubes* (Imperial College Press, London 1998)
5. M. Fujita, K. Wakabayashi, K. Nakada, K. Kusakabe: J. Phys. Soc. Jpn. **65**, 1920 (1996)
6. L. G. Cançado, M. A. Pimenta, B. R. A. Neves, G. Medeiros-Ribeiro, T. Enoki, Y. Kobayashi, K. Takai, K. Fukui, M. S. Dresselhaus, R. Saito, A. Jorio: Phys. Rev. Lett. **93**, 047403 (2004)
7. E. Dujardin, T. Thio, H. Lezec, T. Ebbesen: Appl Phys. Lett. **79**, 2474 (2001)
8. R. Saito, M. Fujita, G. Dresselhaus, M. S. Dresselhaus: Appl. Phys. Lett. **60**, 2204 (1992)
9. R. Saito, M. Fujita, G. Dresselhaus, M. S. Dresselhaus: Phys. Rev. B **46**, 1804 (1992)
10. N. Hamada, S. Sawada, A. Oshiyama: Phys. Rev. Lett. **68**, 1579 (1992)

11. J. W. G. Wildöer, L. C. Venema, A. G. Rinzler, R. E. Smalley, C. Dekker: Nature **391**, 59 (1998)
12. T. W. Odom, J. Huang, P. Kim, C. M. Lieber: Nature **391**, 62 (1998)
13. K. Nakada, M. Fujita, G. Dresselhaus, M. S. Dresselhaus: Phys. Rev. B **54**, 17954 (1996)
14. K. Wakabayashi, M. Fujita, H. Ajiki, M. Sigrist: Phys. Rev. B **59**, 8271 (1999)
15. K. Wakabayashi: Ph.D Thesis, Tsukuba University (2000)
16. K. Wakabayashi, M. Sigrist: Phys. Rev. Lett. **84**, 3390 (2000)
17. K. Wakabayashi: Phys. Rev. B **64**, 125428 (2001)
18. K. Wakabayashi: Electronic and Magnetic Properties of Nanographite. In: *Carbon-based Magnetism—An Overview of the Magnetism of Metal Free Carbon-based Compounds and Materials*, ed by T. Makarova, F. Palacio (Elsevier 2006) pp 279–304
19. Y. Niimi, T. Matsui, H. Kambara, K. Tagami, M. Tsukada, H. Fukuyama: cond-mat/0404069 (2004)
20. Y. Shibayama, H. Sato, T. Enoki, M. Endo: Phys. Rev. Lett. **84**, 1744 (2000)
21. O. E. Andersson, B. L. V. Prasad, H. Sato, T. Enoki, Y. Hishiyama, Y. Kaburagi, M. Yoshikawa, S. Bandow: Phys. Rev. B **58**, 16387 (1998)
22. K. Han, D. Spemann, P. Esquinazi, R. Höhne, V. Riede, T. Butz: Adv. Mater. **15**, 1719 (2003)
23. P. Esquinazi, D. Spemann, R. Höhne, A. Setzer, K.-H. Han, T. Butz: Phys. Rev. Lett. **91**, 227201 (2003)
24. P. R. Wallace: Phys. Rev. **71**, 622 (1947)
25. M. S. Dresselhaus, G. Dresselhaus, K. Sugihara, I. L. Spain, H. A. Goldberg: *Graphite Fibers and Filaments* (Springer-Verlag 1988)
26. K. Tanaka, S. Yamashita, H. Yamabe, T. Yamabe: Synth. Met. **17**, 143 (1987)
27. K. Kobayashi: Phys. Rev. B **47**, 1757 (1993)
28. J. C. Slonczewski, P. R. Weiss: Phys. Rev. **109**, 272 (1958)
29. J. M. Luttinger: Phys. Rev. **84**, 814 (1951)
30. W. Kohn: Phys. Rev. **115**, 1460 (1959)
31. J. W. McClure: Phys. Rev. **104**, 666 (1956)
32. H. Ajiki, T. Ando: J. Phys. Soc. Jpn. **62**, 1255 (1993)
33. H. Ajiki, T. Ando: J. Phys. Soc. Jpn. **65**, 505 (1996)
34. K. S. Novoselov, A. K. Geim, S. V. Morozov, D. Jiang, M. I. Katsnelson, I. V. Grigorieva, S. V. Dubonos, A. A. Firsov: Nature **438**, 197 (2005)
35. D. J. Klein: Chem. Phys. Lett. **217**, 261 (1994)
36. J. P. Lu: Phys. Rev. Lett. **74**, 1123 (1995)
37. H. Ajiki, T. Ando: J. Phys. Soc. Jpn. **62**, 2470 (1993)
38. M. Fujita, M. Igami, K. Nakada: J. Phys. Soc. Jpn. **66**, 1864 (1997)
39. Y. Miyamoto, K. Nakada, M. Fujita: Phys. Rev. B **59**, 9858 (1999) [Errata: **60**, 16211 (1999)]
40. E. H. Lieb: Phys. Rev. Lett. **62**, 1201 (1989)
41. L. Pitaevskii, S. Stringari: J. Low. Temp. Phys. **85**, 377 (1991)
42. E. Dagotto, T. M. Rice: Science **271**, 618 (1996)
43. K. Wakabayashi, M. Sigrist, M. Fujita: J. Phys. Soc. Jpn. **67**, 2089 (1998); see also H. Yoshioka: J. Phys. Soc. Jpn. **72**, 2145 (2003)
44. K. Kusakabe, M. Maruyama: Phys. Rev. B **67**, 092406 (2003)
45. T. L. Makarova: Semiconductors **38**, 615 (2004)
46. K. Wakabayashi, K. Harigaya: J. Phys. Soc. Jpn. **72**, 998 (2003)

47. S. Okada, A. Oshiyama: Phys. Rev. Lett. **87**, 146803 (2001)
48. P. W. Anderson, D. J. Thouless, E. Abrahams, D. S. Fisher: Phys. Rev. B **22**, 3519 (1980)
49. M. Büttiker, Y. Imry, R. Landauer, S. Pinhas: Phys. Rev. B **31**, 6207 (1985)
50. R. Landauer: Z. Phys. B **68**, 217 (1987)
51. A. MacKinnon: Z. Phys. B **59**, 385 (1985)
52. T. Ando: Phys. Rev. B **44**, 8017 (1991)
53. R. Tamura, M. Tsukada: Phys. Rev. B **58**, 8120 (1998)
54. M. Büttiker, Y. Imry, M. Ya. Azbel: Phys. Rev. A **30**, 1982 (1984)
55. P. S. Deo, A. M. Jayannavar: Mod. Phys. Lett. B **10**, 787 (1996)
56. C.-M. Ryu, S. Y. Cho, Phys. Rev. B **58**, 3572 (1998)
57. H.-W. Lee: Phys. Rev. Lett. **82**, 2358 (1999)
58. T. Taniguchi, M. Büttiker: Phys. Rev. B **60**, 13814 (1999)
59. A. Yacoby, M. Heiblum, D. Mahalu, H. Shtrikman: Phys. Rev. Lett. **74**, 4047 (1995)
60. R. Schuster, E. Buks, M. Heiblum, D. Mahalu, V. Umansky, H. Shtrikman: Nature **385**, 417 (1997)
61. P. S. Deo, A. M. Jayannavar: Phys. Rev. B **50**, 11629 (1994)
62. R. Sordan, K. Nikolić: Phys. Rev. B **52**, 9007 (1995)
63. Z. Shao, W. Porod, C. S. Lent: Phys. Rev. B **49**, 7453 (1994)
64. W. Porod, Z. Shao, C. S. Lent: Phys. Rev. B **48**, 8495 (1993); W. Porod, Z. Shao, C. S. Lent: Appl. Phys. Lett. **61**, 1350 (1992)
65. J. Wang, Y. Wang, H. Guo: Appl. Phys. Lett. **65**, 1793 (1994)
66. F. Sols, M. Macucci, U. Ravaioli, K. Hess: J. Appl. Phys. **66**, 3892 (1989)
67. H. Xu, W. Sheng: Superlattices and Microstructures **25**, 79 (1999)
68. H. Xu, W. Sheng: Phys. Rev. B **57**, 11903 (1998)
69. U. Aeberhard, K. Wakabayashi, M. Sigrist: Phys. Rev. B **72**, 075328 (2005)
70. U. Fano: Phys. Rev. **124**, 1866 (1961)

Thermoelectric Phenomena from Macro-Systems to Nano-Systems

K. A. Chao[1] and Magnus Larsson[2]

[1] Department of Physics, Lund University, Sölvegatan 14A, S-223 62 Lund, Sweden. chao@teorfys.lu.se
[2] Nanofreeze Technologies Lund AB, Ole Rmers Väg 12, S-223 70 Lund, Sweden. magnus.larsson@nanofreeze.se

1 Introduction

The simplest thermoelectric system is a closed loop made with two different metals connected in the form of junctions at both ends. In 1822–1823 Seebeck discovered an electric current flowing through the loop when the junctions are kept at different temperatures, the so-called Seebeck effect. This system then works as a thermoelectric power generator. Since the electric current flows through both metals of the closed loop system, in each metallic branch there exists a voltage difference V between its two ends. Let us consider the simpler case of one conductor in which a relevant electric field $\boldsymbol{E}$ associated to V is created by a gradient ∇T of the temperature T. The above mentioned experiment suggests a relation

$$\boldsymbol{E} = S\,\nabla T\ , \tag{1}$$

where S is the Seebeck coefficient or the thermopower.

Twelve years later, in 1834 Peltier observed the temperature change around a junction between two different metals when an electric current passed through, which is now called Peltier effect. This observation must be related to the heat absorbed or released at the junction, depending on the direction of the electric current j. One can simply imagine that when the electric current flows in each metallic branch, an amount of heat Q is generated, which is connected to the current j as

$$Q = \Pi\, j \tag{2}$$

by the Peltier coefficient Π. The Peltier effect is closely related to the Seebeck effect. We will show later that $\Pi{=}ST$.

Soon after the discovery of Peltier effect, in 1838 Heinreich Lenz performed a simple experiment with enormous impact to the development of

thermoelectricity. Lenz put a water droplet at each of the two junctions of a bismuth-antimony closed loop and derived an electric current through the system. One water droplet then freezes into ice while the other droplet remains in the form of water. By reversing the direction of the electric current, the ice at one junction melts into a water, but the water droplet at the other junction freezes into ice. This is a clear indication that at one junction the external electric current supplies heat to the loop system, and at the other junction the electric current extracts heat from the loop system. Consequently, not only thermoelectric power generator can work as a thermoelectric refrigerator, one can also control the thermodynamic phenomenon at the junction of two different conductors as power generation or refrigeration with an external voltage source.

Thermoelectricity is a typical problem of energy conversion. The development of thermodynamics in the next thirty years after Seebeck's discovery brought all types of energy conversion to the interest of physicists. When an electric current j flows through a homogeneous conductor in which there is a temperature gradient $\partial T/\partial x$, say along the x-axis, Thomson discovered that an amount of heat

$$Q = \Gamma j \frac{\partial T}{\partial x} \tag{3}$$

is generated or absorbed, where Γ is the Thomson coefficient. Γ is again related to both S and Π.

For a long time the practical use of thermoelectricity is measuring temperature with thermocouples. Its use in power application emerged much later, and now has become one important approach to solve the energy problem. However, the efficiency of thermoelectric generator and refrigerator was investigated in 1885 by Rayleigh, and was correctly calculated by Altenkirch in 1909. Since the efficiency is the most important problem in power application, we will present its complete derivation in Sect. 2. Only after going through this derivation, we can understand the physical significance of the fundamental thermoelectric quantity *Figure of Merit*.

The Figure of Merit of a thermoelectric material is defined as [1] ZT with $Z \equiv S^2\sigma/\kappa$, where κ is the (combined electron and phonon) thermal conductivity, σ the electric conductivity, and the temperature T is measured in Kelvin. For technological applications one would like to have a bulk thermoelectric material with its ZT value as high as possible. We will show in Sect. 2 that if $ZT>1$ the efficiency of thermoelectric power generation becomes larger than 20% of the Carnot engine efficiency.

Before the wide use of semiconductors in modern industries, metals were the only conductors technically known. For all metals the values of ZT are much less than unity at all temperatures. Using the free electron gas as a model system, for which the electron density n is the only varying parameter, Ioffe has calculated Z and found [2] a broad maximum in the curve Z vs n around the electron density $10^{19}\,\mathrm{cm}^{-3}$. Since this range of n can be easily reached by doping semiconductors with impurities, bulk semiconductors are among the

best thermoelectric materials and have been much studied. Although many thermoelectric materials with $ZT \simeq 1$ were discovered in the past, they are still not good enough to meet the requirement for commercial power generator and/or refrigerator.

Looking for new thermoelectric material systems often requires the guidance of fundamental theoretical analysis. The semiclassical theory for thermoelectricity is based on the Boltzmann transport equation, the accurate solutions of which can only be obtained numerically such as through the Monte Carlo method. The commonly accepted analytical solutions are within the scope of *relaxation time approximation*. Within this approximation the fundamental thermoelectric theory will be summarized in Sect. 3. Since the dominating physical process in Boltzmann equation is scattering, the resulting diffusive transport is rather slow and therefore cannot produce high value of ZT. To overcome this problem one needs to find new thermoelectric mechanisms. One favorable process which is much studied in recent year is the *thermionic transport*.

Thermionic emission of electrons from a hot surface is a well-studied physical process. If the surface is used as a cathode, and if all emitted electrons are collected, the emitted current density is called the Richardson current which depends on the cathode temperature T and the cathode work function. The simplest thermionic device, the so-called *electron engine*, consists of two parallel metal plates separated by a very narrow vacuum gap. Keeping the two metal plates at different temperatures, there is a net thermally emitted electron current from the hot plate to the cold plate. In principle, large thermionic current can be achieved if one can reduce the work function to sufficiently low and the vacuum gap sufficiently narrow. With the advancement of material fabrication technology to produce high quality layer materials, there has been much progress in thermionics. In Sect. 4 we will discuss this new research field.

The modern material fabrication technology can also produce one-dimensional systems such as quantum wires. Using the formula given in Sect. 3, it is easy to calculate the figure of merit of quantum wires, and the calculated values of ZT can be very high. However, for power applications, we will show in Sect. 5 that quantum wires are not good candidates.

Regardless the transport mechanism, the definition of ZT suggests that in the desired systems σ and S should be enhanced, but κ suppressed. The thermal conductivity κ consists of an electronic part $\kappa_{\rm el}$ and a phonon part $\kappa_{\rm ph}$. While it is relatively easy to control $\kappa_{\rm el}$ by quantum confinement, to manipulate $\kappa_{\rm ph}$ and to confine phonons spatially is a difficult task [3]. The phonon heat flow from the hot electrode to the cold electrode will reduce the temperature gradient, and thus becomes an increasingly important problem in the nano-scale devices that already have important commercial applications. There have been much studies on $\kappa_{\rm ph}$ in terms of the particle motion of phonons, but not in terms of their wave motion. The wave motion picture of phonon heat transfer will be analyzed in Sect. 6.

It is well-known that quantum effects become increasingly important when the sample size decreases. Concerning transport properties, the two most important processes are ballistic transport and tunneling transport. While ballistic transport is connected to the over-the-barrier emission in thermionics, resonant tunneling usually occurs within a small energy window. This window can be used as a filter to select electrons with specific energy, for example, cold electrons with low kinetic energy. In this respect, the tunneling related thermoelectric phenomena will be studied in Sect. 7.

The recent advancement of materials sciences and manipulation techniques on nanometer scale have doubtlessly great impact on thermoelectricity. Along with this advancement, novel ideas also appear from time to time. Some of these ideas have found their applications in thermionics. In Sect. 8 we will outline the consequence of some new concepts.

In the area of thermoelectricity, it happens often that new technologies revives old research topics. In the final Sect. 9 we will give one such example where the modern nanotechnology revives the old research work on *electron engine*.

2 Thermodynamic Efficiency

We will use thermodynamic laws to derive the thermoelectric efficiency of power generation and refrigeration. With the electric current j in the proper direction, the system is schematically shown in Fig. 1 for a power generator. There are two heat reservoirs, one with a hot temperature $T_{\rm h}$ and the other with a cold temperature $T_{\rm c}$. The reservoirs are connected by a thermoelectric conductor. To avoid the unnecessary complication in mathematical expressions, we assume that the thermoelectric conductor has a unit length and a unit cross-section area. Since our results are consequence of the thermodynamic laws, they are valid for any material systems.

2.1 Thermoelectric Power Generation

As we have discussed in the previous section, a voltage V will be built up between the two ends of the thermoelectric conductor. Now, we will connect these two ends to an external load via two metallic wires, which are extremely short and have extremely low thermal conductivity. Hence, we only have to consider electric current j but not heat current flowing through the external part of the closed circuit. Let R be the resistance of the external load. Then, the amount of work that can be done by this thermoelectric power generator is

$$W = jV = V^2/R\,. \tag{4}$$

This thermoelectric power generator is schematically shown in Fig. 1.

Using (2) with $\Pi{=}ST$, the heat current flowing from the hot reservoir into the thermoelectric conductor is

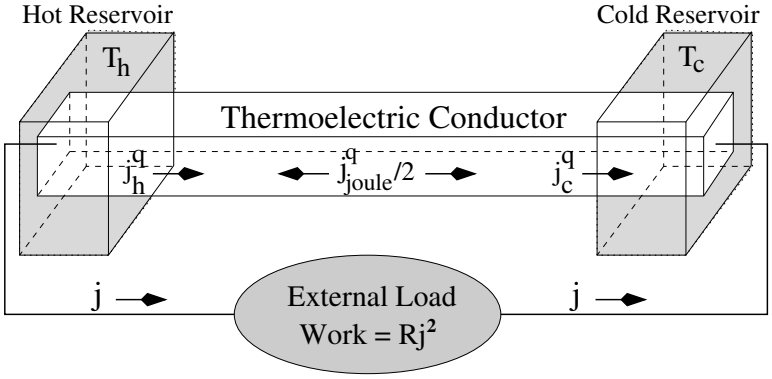

Fig. 1. Schematic description of a thermoelectric power generator

$$j_{\mathrm{h}}^{q} = ST_{\mathrm{h}}j + \kappa(T_{\mathrm{h}} - T_{\mathrm{c}}) \,, \tag{5}$$

where $\kappa{=}\kappa_{\mathrm{el}}{+}\kappa_{\mathrm{ph}}$ is the total thermal conductivity. Similarly, the heat current flowing from the thermoelectric conductor into the cold reservoir is

$$j_{\mathrm{c}}^{q} = ST_{\mathrm{c}}j + \kappa(T_{\mathrm{h}} - T_{\mathrm{c}}) \,. \tag{6}$$

When the electric current j flows through the thermoelectric conductor with an electric resistivity ρ, a Joule heat

$$j_{\mathrm{joule}}^{q} = j^2\rho \tag{7}$$

is generated. Half of this heat, $j^2\rho/2$ moves into the hot reservoir and the other half into the cold reservoir.

The efficiency of the generator, $\eta(R)$ is the ratio of the useful work W to the net heat extracted from the hot reservoir:

$$\eta(R) = \frac{W}{j_{\mathrm{h}}^{q} - j_{\mathrm{joule}}^{q}/2} = \frac{V^2R}{ST_{\mathrm{h}}VR + \kappa(T_{\mathrm{h}} - T_{\mathrm{c}})R^2 - V^2\rho/2} \,. \tag{8}$$

For a given structure of the thermoelectric power generator, this efficiency is a function of the load resistance R. To determine the maximum efficiency $\eta(R_0)$ for the optimum load resistance R_0, we solve the equation

$$\left.\frac{\partial\eta(R)}{\partial R}\right|_{R=R_0} = 0 \tag{9}$$

to obtain

$$R_0 = \rho\sqrt{1 + \left(\frac{S^2}{\rho\kappa}\right)\left(\frac{T_{\mathrm{h}} + T_{\mathrm{c}}}{2}\right)} \,. \tag{10}$$

It is more convenient to introduce the dimensionless quantity $r{=}R_0/\rho$. If we define a mean temperature $T_{\mathrm{m}}{=}(T_{\mathrm{h}}{+}T_{\mathrm{c}})/2$ and the thermoelectric parameter

$$Z = S^2/\rho\kappa > 0 \,, \tag{11}$$

so that we can rewrite

$$r = \sqrt{1 + ZT_{\rm m}} > 1 \tag{12}$$

to obtain the final form

$$\eta_{\max} \equiv \eta(R_0) = \left(\frac{T_{\rm h} - T_{\rm c}}{T_{\rm h}}\right)\left(\frac{r-1}{r + T_{\rm c}/T_{\rm h}}\right) \,. \tag{13}$$

The parameter $ZT_{\rm m}$ is called the figure of merit as we introduced previously, which plays a crucial role in the field of thermoelectricity. It depends on both the mean temperature $T_{\rm m}$ and the transport features of the thermoelectric conductor. In (13), on the right hand side, the first ratio $(T_{\rm h}–T_{\rm c})/T_{\rm h}$ is the Carnot efficiency for a reversible heat engine. Hence, by increasing the value of r (or the figure of merit $ZT_{\rm m}$), the efficiency of the thermoelectric power generator approaches the ideal Carnot efficiency. The efficiency $\eta_{\max}$ can also be increased with higher value of $T_{\rm h}$ as expected, which enhances both factors on the right hand side of (13).

2.2 Thermoelectric Refrigeration

The processes in a thermoelectric refrigeration are shown in Fig. 2, where an external power is supplied to drive the electric current j in the reversed direction. We see from (5) and (6) that in a thermoelectric generator, there is a net heat flow

$$\Delta j^q = j^q_{\rm h} - j^q_{\rm c} = S(T_{\rm h} - T_{\rm c})j \tag{14}$$

from the hot reservoir to the cold reservoir. This continuous heat flow will eventually remove the temperature difference $T_{\rm h} - T_{\rm c}$. To maintain the cold reservoir to be cold, it is necessary to balance this heat flow Δj^q. Besides, we also need to compensate the Joule heating in the thermoelectric conductor. This is the reason that we must supply an external power to keep the reservoir temperatures at $T_{\rm h}$ and $T_{\rm c}$. The amount power supplied from a power source with a voltage output V is

$$P(V) = j^q_{\rm joule} + \Delta j^q = jV + S(T_{\rm h} - T_{\rm c})j \,. \tag{15}$$

It is clear that the operating process of thermoelectric refrigeration is opposite to that of thermoelectric power generation.

The efficiency of thermoelectric refrigeration is measured by the coefficient of performance $K(V)$, which is defined as the ratio of the heat extracted from the cold reservoir, $j^q_{\rm c} - j^q_{\rm joule}/2$, to the external power supplied $P(V)$:

$$K(V) = \frac{j^q_{\rm c} - j^q_{\rm joule}/2}{P(V)} = \frac{ST_{\rm c}j - jV/2 - \kappa(T_{\rm h} - T_{\rm c})}{jV + S(T_{\rm h} - T_{\rm c})j} \,. \tag{16}$$

Using the relation $V{=}j\rho$ the above equation reads as

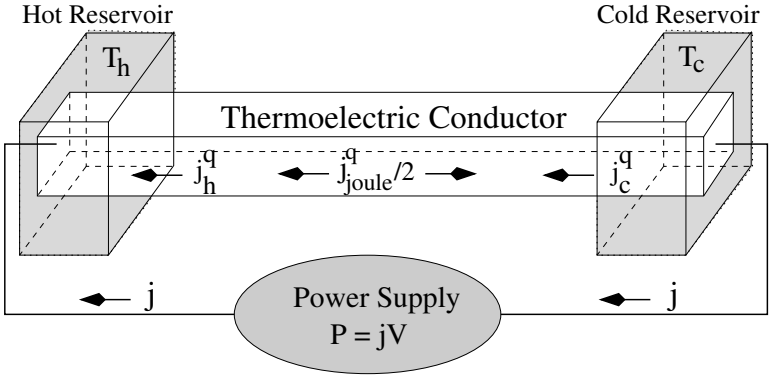

Fig. 2. Schematic description of a thermoelectric refrigerator

$$K(V) = \frac{ST_{\mathrm{c}}V - V^2/2 - \kappa(T_{\mathrm{h}} - T_{\mathrm{c}})\rho}{V^2 + S(T_{\mathrm{h}} - T_{\mathrm{c}})V} \,, \tag{17}$$

and from the condition

$$\left.\frac{\partial K(V)}{\partial V}\right|_{V=V_0} = 0 \tag{18}$$

we derive the maximal value of the coefficient of performance

$$K_{\max} \equiv K(V_0) = \left(\frac{T_{\mathrm{c}}}{T_{\mathrm{h}} - T_{\mathrm{c}}}\right)\left(\frac{r - T_{\mathrm{h}}/T_{\mathrm{c}}}{r+1}\right) \,, \tag{19}$$

where r is given by (12). As in the case for thermoelectric power generation, a high value of figure of merit also yields a high value of the coefficient of performance. This is not a surprising conclusion, since thermoelectric refrigeration is just the reversal of thermoelectric power generation.

It is instructive to give some measured values of the figure of merit, from which we can estimate the efficiency of the corresponding thermoelectric devices. Here we will follow the convention to express the figure of merit as ZT instead of ZT_{m}. So far the bulk material, crystals or alloys, with the highest figure of merit at room temperature is Bi_2Te_3, as shown in Fig. 3. Using such *good* thermoelectric materials to make devices, the highest efficiency is about 20% of the Carnot efficiency. Lately the main focus of device research has shifted from traditional bulk compounds to layer materials. We will return to this topic in the later sections.

3 Fundamental Thermoelectricity

We need to understand thoroughly the electric conductivity σ, the thermal conductivities κ_{el} and κ_{ph}, as well as the thermopower S in order to be able to fully analyze the figure of merit. A more general aspect of the problem

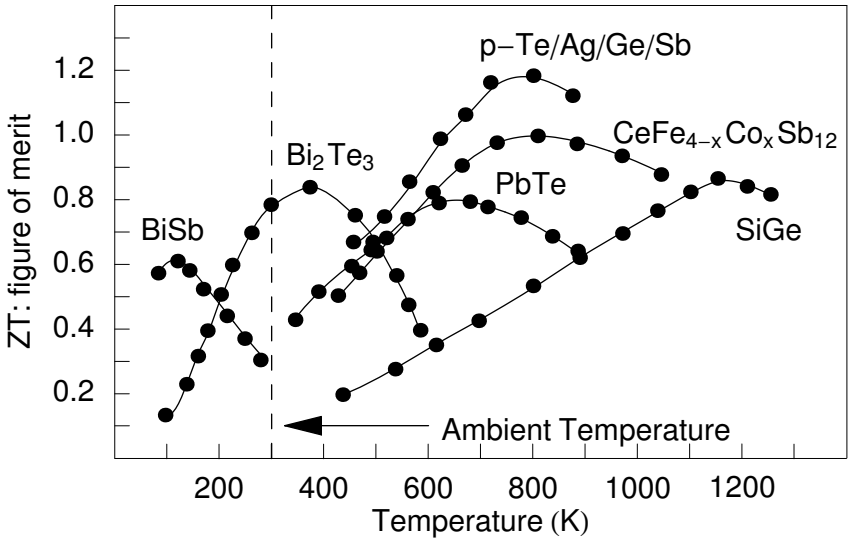

Fig. 3. Measured ZT of several *good* bulk thermoelectric materials

is carrier and heat transport under the influence of an external electric field and a temperature gradient. The starting point is the Boltzmann transport equation, the accurate solution of which can be obtained numerically only in almost all cases. The most commonly used analytical approach is semi-classical and is based on the *relaxation time approximation*, which we will follow here.

3.1 Relaxation Time Approximation

When an electron moves through a material it experiences various types of collisions, and stronger collisions result in higher resistance. An electron can be scattered by lattice ions (or phonons), other electrons, impurities, defects, surfaces and interfaces, etc. When electrons collide with lattice ions, their kinetic energy can be transferred to the lattice system and it heat up the material. This is known as Joule heating. When a current j passes through a conductor of resistance R, in a time interval t, the amount of heat dissipated in the conductor is

$$Q = j^2 R\, t\ , \tag{20}$$

which is known as Joule's law.

Electronic conduction in conductors has been well-studied with the relaxation time approximation using a non-equilibrium distribution function $g(\boldsymbol{r}, \boldsymbol{k}, t)$, where $\boldsymbol{r}$ is the position of the carrier and $\boldsymbol{k}$ is its wave vector. This distribution function is so defined that at the time t the number of electrons in the volume $\mathrm{d}\boldsymbol{r}\mathrm{d}\boldsymbol{k}$ in phase space is $g(\boldsymbol{r}, \boldsymbol{k}, t)\mathrm{d}\boldsymbol{r}\mathrm{d}\boldsymbol{k}/4\pi^3$. When the system is under equilibrium, the distribution function reduces to the Fermi function

$$g^0(\boldsymbol{r},\boldsymbol{k},t) = f[\epsilon(\boldsymbol{k})] = \frac{1}{\mathrm{e}^{[\epsilon(\boldsymbol{k})-\mu]/k_\mathrm{B}T}+1}\ , \tag{21}$$

where $\epsilon(\boldsymbol{k})$ is the energy of the electron and μ the chemical potential.

In the presence of an uniform external electric field $\boldsymbol{E}$ and a homogeneous temperature gradient ∇T, using the semiclassical equations of motion

$$\boldsymbol{v}(\boldsymbol{k}) = \frac{\partial \boldsymbol{r}}{\partial t} = \frac{1}{\hbar}\frac{\partial \epsilon(\boldsymbol{k})}{\partial \boldsymbol{k}}\ , \tag{22}$$

$$\hbar\frac{\partial \boldsymbol{k}}{\partial t} = -e\boldsymbol{E}\ , \tag{23}$$

the distribution function is spatially independent and steady in time, and can be written as [4]

$$g(\boldsymbol{k}) = f[\epsilon(\boldsymbol{k})] + \tau[\epsilon(\boldsymbol{k})]\Big(-\frac{\partial f(\epsilon)}{\partial \epsilon}\Big)\boldsymbol{v}(\boldsymbol{k})\left[-e\boldsymbol{E} + \frac{\epsilon(\boldsymbol{k})-\mu}{T}(-\nabla T)\right]\ , \tag{24}$$

where $-e$ is the electron charge. The relaxation time $\tau[\epsilon(\boldsymbol{k})]$ is the average time between two collisions of an electron. In the above equation we have neglected the change of chemical potential with temperature, which is of the order of $(k_\mathrm{B}T)^2/\epsilon_F$ and is small with respect to the Fermi energy ϵ_F.

When electrons flow through a conductor, they generate both an electric current density $\boldsymbol{j}$ and a heat current density $\boldsymbol{j}^q$. These current densities can be calculated from

$$\boldsymbol{j} = -e\int \frac{\mathrm{d}\boldsymbol{k}}{4\pi^3}\boldsymbol{v}(\boldsymbol{k})g(\boldsymbol{k})\ , \tag{25}$$

$$\boldsymbol{j}^q = \int \frac{\mathrm{d}\boldsymbol{k}}{4\pi^3}(\epsilon(\boldsymbol{k})-\mu)\boldsymbol{v}(\boldsymbol{k})g(\boldsymbol{k})\ . \tag{26}$$

We should mention that equation (26) does not include the Joule heat j^2R which is generated in the conductor due to electron-phonon collisions. This Joule heating is a second order effect in applied electric field, and is therefore much smaller than the heat current defined in (26). Nevertheless, Joule heating is still important since it is an energy dissipation process. Since it has the form j^2R, Joule heat dissipation has no specific direction. When a large current is passed through an electric circuit, if the removal of Joule heat is not sufficiently fast, the electric circuit can break down. This is an important aspect to be considered in nano-electronics.

3.2 Thermoelectric Coefficients

Using the distribution function given by (24), we can rewrite the equations (25) and (26) as

$$\boldsymbol{j} = \mathbf{L}^{11}\boldsymbol{E} + \mathbf{L}^{12}(-\nabla T)\ , \tag{27}$$

$$\boldsymbol{j}^q = \mathbf{L}^{21}\boldsymbol{E} + \mathbf{L}^{22}(-\nabla T)\ , \tag{28}$$

where $\mathbf{L}^{ij}$ are known as *thermoelectric coefficients*. In terms of the compact expression

$$\mathcal{L}^{\alpha} = e^2 \int \frac{\mathrm{d}\boldsymbol{k}}{4\pi^3} \Big(-\frac{\partial f(\epsilon)}{\partial \epsilon} \Big) \tau[\epsilon(\boldsymbol{k})] \boldsymbol{v}(\boldsymbol{k}) \boldsymbol{v}(\boldsymbol{k}) [\epsilon(\boldsymbol{k}) - \mu]^{\,\alpha} \,, \tag{29}$$

the four thermoelectric coefficients $\mathbf{L}^{ij}$ are defined as

$$\mathbf{L}^{11} = \mathcal{L}^0 \,, \tag{30}$$

$$\mathbf{L}^{21} = T\mathbf{L}^{12} = -\frac{1}{e}\mathcal{L}^1 \,, \tag{31}$$

$$\mathbf{L}^{22} = \frac{1}{e^2 T}\mathcal{L}^2 \,. \tag{32}$$

In the absence of a temperature gradient, equation (27) reduces to the Ohm's law $\boldsymbol{j}{=}\sigma\boldsymbol{E}$. Hence the electric conductivity is identified as

$$\sigma = \mathbf{L}^{11} \,. \tag{33}$$

The electronic part of the thermal conductivity is defined as $\kappa_{\mathrm{el}}{=}\boldsymbol{j}^q/(-\nabla T)$ when there is no electric current flowing through the system. Setting $\boldsymbol{j}{=}0$ in (27), from (28) we obtain

$$\kappa_{\mathrm{el}} = \mathbf{L}^{22} - \frac{\mathbf{L}^{21}\mathbf{L}^{12}}{\mathbf{L}^{11}} \,. \tag{34}$$

By introducing the Seebeck coefficient

$$S = \frac{\mathbf{L}^{12}}{\mathbf{L}^{11}} \,, \tag{35}$$

we can thus rewrite (27) and (28) as

$$\boldsymbol{j} = \sigma \left[\boldsymbol{E} + S(-\nabla T) \right] , \tag{36}$$

$$\boldsymbol{j}^q = S\,T\,\boldsymbol{j} + \kappa_{\mathrm{el}}\,(-\nabla T) \,, \tag{37}$$

which are the commonly used forms in literatures.

For metals, the second term in (30) is smaller than the first term by a factor of the order of $(k_{\mathrm{B}}T/\epsilon_F)^2$, and can be neglected. Furthermore, $\mathbf{L}^{22}$ has the explicit expression

$$\mathbf{L}^{22} = \frac{\pi^2}{3}\frac{k_{\mathrm{B}}^2 T}{e^2}\sigma + O((k_{\mathrm{B}}T/\epsilon_F)^2) \,. \tag{38}$$

Inserting this in (34) and neglecting the terms of order $(\epsilon_F/k_{\mathrm{B}}T)^2$, we obtain the relation

$$\kappa_{\mathrm{el}} = \frac{\pi^2}{3}\frac{k_{\mathrm{B}}^2}{e^2} T\sigma = LT\sigma \,, \tag{39}$$

which is the Wiedemann-Franz law, and $L{=}\pi^2 k_{\mathrm{B}}^2/(3e^2)$ is called the Lorentz number.

Besides the heat transport by electrons and the Joule heating due to electron-phonon collisions, heat can also be transported by phonons. This heat conduction is proportional to the temperature gradient in the material:

$$\boldsymbol{j}^q_{\mathrm{ph}} = -\kappa_{\mathrm{ph}}\,\nabla T\,. \tag{40}$$

Since phonons are neutral quasi-particles, the phonon heat conduction will not be affected by an electric field.

3.3 Seebeck Effect and Peltier Effect

We have introduced in Sect. 1 the Seebeck effect and the Peltier effect, which have been known and been measured for more than 160 years. Here we will explain these two observed phenomena with the semi-classical theory.

It was observed by Thomas Seebeck in 1823 that when a metallic rod is heated up at one end, an electric field in the material is established, which is proportional to the temperature gradient in the rod:

$$\boldsymbol{E} = S\nabla T\,. \tag{41}$$

This field stops the electron motion driven by the temperature gradient, resulting in no net electron flow in the rod. The Seebeck coefficient S is also called the thermopower. The above relation is a direct consequence of (36) with $\boldsymbol{j} = 0$, since the metallic rod is an open circuit in which no electric current can flow.

Measuring of the Seebeck coefficient is not as straightforward as connecting a voltmeter between the ends of the metallic rod, because such a connection introduces a temperature gradient in the voltmeter as well. Therefore, the thermopower of the voltmeter contributes to the measured Seebeck coefficient. If one cannot reduce this influence to a required accuracy, only relative Seebeck coefficients between two thermoelectric materials can be obtained. The experimental setup is to keep one end of each rod at the same temperature, and then to connect the other end of each rod across a voltmeter. Since the voltmeter is under constant temperature, it measures the difference between the thermopowers of the two materials.

When an electric current $\boldsymbol{j}$ flows through a metallic rod which is kept at uniform temperature, there will be an accompanying heat current $\boldsymbol{j}^q$. The two currents are proportional to each other as

$$\boldsymbol{j}^q = \Pi\,\boldsymbol{j}\,. \tag{42}$$

This is the Peltier effect and Π is the Peltier coefficient. From (31), (35) and (37), we obtain

$$\Pi = TS = \frac{\mathbf{L}^{21}}{\mathbf{L}^{11}}\,. \tag{43}$$

It is important to note that when passing a current through a metallic rod, a Joule heat $j^q_{\rm joule}$ will be produced according to the Joule's law. This heat transports uniformly in all directions, and in a metallic rod half of $j^q_{\rm joule}$ goes along one direction of the rod and the other half in the opposite direction. However, this is a second order effect, and to a first order approximation the heat current through the rod is given by (42).

4 Thermionics

Although the analysis of thermoelectric processes based on the relaxation time approximation is not accurate quantitatively in some cases, the scattering of carriers is the most important feature of the entire problem. It make the carrier transport diffusive, and therefore can hardly yield high values of ZT. To increase the value of ZT, one needs to find new transport mechanisms and/or new sample structures. To minimize the scattering effect, thermionic transport has been extensively studied in recent years. We will first describe the basic theoretical aspects of thermionics and then use the so-derived knowledge to investigate the thermoelectric phenomena in semiconductor superlattices.

4.1 Classical Theory

Thermal emission of electrons from a metal surface into vacuum is a well-known phenomenon and has found its use in old-fashion electronic devices such as vacuum tubes. Let $e\phi$ be the work function of the metal emitter. In classical description, as soon as an electron inside the metal is thermally excited above the work function level (the top of the potential barrier), it escapes from the metal surface and contributes to the thermionic current. At temperature T, the thermionic current density emitted from a metal surface is given by the Richardson's equation

$$J(\phi, T) = \frac{emk_{\rm B}^2 T^2}{2\pi^2\hbar^3} {\rm e}^{-e\phi/k_{\rm B}T} \ . \tag{44}$$

It should be noticed that while the exponential part of $J(\phi, T)$ is independent of the dimensionality of the electronic system, the power of T in the prefactor of $J(\phi, T)$ does depend on the system dimensionality. One has to be careful when calculating the Richardson current in one-dimensional quantum wires.

To demonstrate the thermionic power generation and refrigeration, we will consider a simple 3-layers metal-insulator-metal structure, the potential profile of which is illustrated in Fig. 4. We assume that the two metal plates are identical with the same work function $e\phi$, but have different temperatures $T_{\rm L}$ and $T_{\rm R}$. Our theoretical analysis also remains valid if the two metal plates are different. With an applied bias V to raise the Fermi energy ε_f of the metal at the left with respect to that of the metal at the right, as shown in

Fig. 4, the effective potential barrier for the electrons to be thermally excited from the right to the left is increased from the work function $e\phi$ to $e\phi{+}eV$. The thermionic current across the potential barrier is the sum of two Richard currents

$$J(\phi, T_{\mathrm{L}}, T_{\mathrm{R}}, V) = \frac{emk_{\mathrm{B}}^2}{2\pi^2\hbar^3}\left[T_{\mathrm{L}}^2 \mathrm{e}^{-e\phi/k_{\mathrm{B}}T_{\mathrm{L}}} - T_{\mathrm{R}}^2 \mathrm{e}^{-e(\phi+V)/k_{\mathrm{B}}T_{\mathrm{R}}}\right] . \qquad (45)$$

Let us assume that $T_{\mathrm{R}}{>}T_{\mathrm{L}}$. At zero bias $V{=}0$, the thermionic current is negative (from right to left against the arrow shown in Fig. 4) with a net electron flow from the high temperature metal to the low temperature metal. This system is a power generator. With increasing bias V, the thermionic current will decrease to zero, and then reverse its direction. Then the thermionic current flows from the cold metal with temperature T_{L} to the hot metal with temperature T_{R}. In this case the system is a refrigerator.

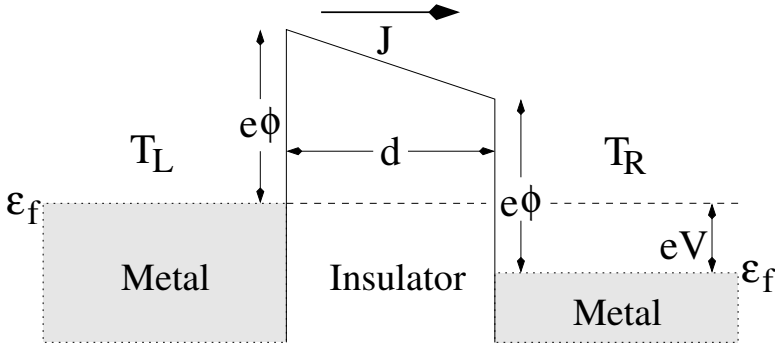

Fig. 4. A simple 3-layers metal-insulator-metal thermionic system

Form (45) it is clear that a low work function is required for generating a large thermionic current, but ordinary metals have rather large work functions. Even for metals with work function $e\phi \simeq 1\,\mathrm{eV}$, the practical operating temperature for such a device is in the range from 600 K to 1000 K, far above the room temperature. Equation (45) does not take into account the effect of image force which reduces the potential barrier height from the ideal work function. This image force is important if the separation d of the metallic layers becomes less than $1\,\mu m$. It has been reported that a separation of $d = 60\,\text{Å}$ can reduce the potential barrier height by 63% for Ag-O-Cs system [5]. However constructing and maintaining this small separation in a real device can be a major problem.

Another important aspect which is ignored in the above analysis is the space charge accumulated in the vacuum between the two metal plates. This space charge can change drastically the simple potential profile shown in Fig. 4. We will return to this topic after we present the quantum mechanical description of thermionics.

4.2 Quantum Theory

One would like to have a weak potential barrier in Fig. 4 in order to produce large thermionic current. It is well-known that in this case quantum mechanical tunneling through the barrier becomes important if the kinetic energy of an electron lies in certain energy range. When this occurs, the Richardson formula (44) is no longer valid. To analyze the quantum processes, we will follow the concept *conductance from transmission* [6] of Landauer. Instead of Fig. 4, we will describe the system and the relevant physical processes with the help of Fig. 5. The two metal plates in Fig. 4 are replaced by two perfect conducting leads at the left (labeled by L) and at the right (labeled by R), which have different temperatures ($T_{\rm L}$ and $T_{\rm R}$) and different chemical potentials ($\mu_{\rm L}$ and $\mu_{\rm R}$). These two leads serve as two electron reservoirs, and therefore their chemical potentials are not affected by the flowing current. The region of potential barrier in Fig. 4 is replaced by a ballistic conductor because scattering plays no role in the calculation of tunneling probability.

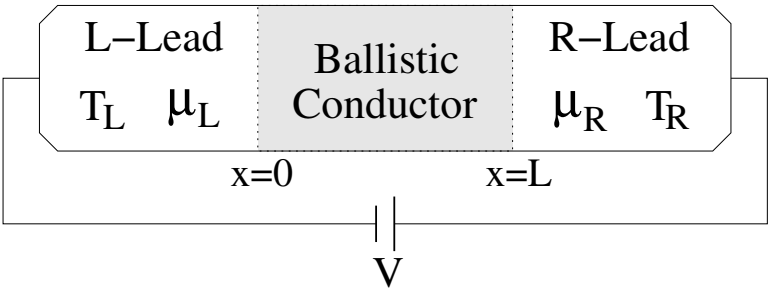

Fig. 5. A model system for the quantum description of thermionics

To simplify the writing of mathematical formula, let us consider a potential energy $U(x)$ which depends on x only. This is indeed realistic when we come to treat the thermionic processes in semiconductor superlattices. Let us first consider the dynamics of electron in the x-direction. Ignoring the kinetic energy of the electron in the yz-plane, the electron energy then has the simple form

$$\epsilon(k) = \frac{\hbar^2 k^2}{2m^*(x)} + U(x) \; , \tag{46}$$

where $m^*(x)$ is the effective mass of the electron. The electrons in each reservoir are under equilibrium obeying the Fermi distribution

$$f_{\rm L,R}(\epsilon) = \left(\mathrm{e}^{[\epsilon-\mu_{\rm L,R}-U(x)]/k_{\rm B}T_{\rm L,R}} + 1\right)^{-1} \; . \tag{47}$$

Since the conductor is ballistic, electrons with $k{>}0$ in the conductor must have originated from the left lead, while electrons with $k{<}0$ must have come

from the right lead. Let $T(\epsilon)$ be the energy-dependent transmission coefficient through the conductor. Then, all electrons in $k>0$ states contribute a current

$$j^+ = \frac{-e}{L}\sum_k \frac{1}{\hbar}\frac{\partial \epsilon}{\partial k} f_{\rm L}(\epsilon)\left[1-f_{\rm R}(\epsilon)\right] T(\epsilon) \qquad (48)$$
$$= \frac{-e}{\pi\hbar}\int_0^\infty {\rm d}\epsilon\, f_{\rm L}(\epsilon)\left[1-f_{\rm R}(\epsilon)\right] T(\epsilon)\ .$$

Assuming that the transmission function is the same for both the $k>0$ and the $k<0$ states, the current produced by all electrons in $k<0$ states has the similar form

$$j^- = \frac{e}{L}\sum_k \frac{1}{\hbar}\frac{\partial \epsilon}{\partial k} f_{\rm R}(\epsilon)\left[1-f_{\rm L}(\epsilon)\right] T(\epsilon) \qquad (49)$$
$$= \frac{e}{\pi\hbar}\int_0^\infty {\rm d}\epsilon\, f_{\rm R}(\epsilon)\left[1-f_{\rm L}(\epsilon)\right] T(\epsilon)\ .$$

The net current flowing through the conductor is then

$$j = j^+ - j^- = \frac{-e}{\pi\hbar}\int_0^\infty {\rm d}\epsilon\, T(\epsilon)\left[f_{\rm L}(\epsilon)-f_{\rm R}(\epsilon)\right]\ . \qquad (50)$$

For two identical leads, the net current is zero if there is no temperature gradient ($T_{\rm L}$=$T_{\rm R}$) and no bias (V=0). At zero temperature $T_{\rm L}$=$T_{\rm R}$=0, for two identical leads with a finite bias V=[$U(0)$–$U(L)$]/e, the current

$$j = \frac{e^2 V}{\pi\hbar} \qquad (51)$$

is linear in applied bias. This means that the conductance

$$G = \frac{j}{V} = \frac{e^2}{\pi\hbar} \qquad (52)$$

is a constant.

For a semiconductor superlattice grown along the x–axis, each (k_y,k_z) state is treated as an independent conducting channel. Hence, the total current is simply proportional to the cross-section area A of the yz–plane. When A becomes sufficiently small such that quantum confinement occurs, the eigen-energies of these conducting channels, $\epsilon_{m,n}$ with integer quantum numbers m and n, become discrete. This is the case of quantum wires and the electron energy is then modified as

$$\epsilon_n(k) = \epsilon_{m,n} + \frac{\hbar^2 k^2}{2m^*(x)} + U(x)\ . \qquad (53)$$

An electron with energy ϵ can propagate in the $\epsilon_{m,n}$ channel as long as $\epsilon_{m,n}\leq\epsilon$. Let M be the total number of channels in the energy range

$U(x{=}L){<}\epsilon{<}U(x{=}0)$. Then, at zero temperature the current j and the conductance G are scaled as $j{=}M(e^2V/\pi\hbar)$ and $G{=}M(e^2/\pi\hbar)$. With a split-gate technique, one can change the value of M one by one. Whenever a channel is opened or closed, the conductance makes a discrete jump of $e^2/\pi\hbar$. This conductance quantization was observed experimentally [7, 8]. Further details can be found in the book by Datta [9].

The heat current is derived in the same manner as the electric current:

$$j^q = \frac{1}{\pi\hbar}\int_0^\infty \mathrm{d}\epsilon\,(\epsilon-\mu)\,T(\epsilon)\big[f_{\mathrm{L}}(\epsilon)-f_{\mathrm{R}}(\epsilon)\big]\ . \tag{54}$$

Using the total current given by (50), the Joule heat can be calculated as

$$\begin{aligned} j^q_{\mathrm{joule}} &= j\,V = j\,\frac{U(0)-U(L)}{e} \\ &= \frac{U(0)-U(L)}{\pi\hbar}\int_0^\infty \mathrm{d}\epsilon\,T(\epsilon)\big[f_{\mathrm{L}}(\epsilon)-f_{\mathrm{R}}(\epsilon)\big]\ . \end{aligned} \tag{55}$$

In terms of these j^q and j^q_{joule}, for the case that $\mu(L){\simeq}\mu(R)$, the coefficient of performance of a thermionic refrigerator can be calculated from

$$K = \frac{j^q - j^q_{\mathrm{joule}}/2}{j^q_{\mathrm{joule}}}\ . \tag{56}$$

Clearly, one should reduce the Joule heating in order to achieve a higher coefficient of performance.

Assuming a small temperature difference ΔT between the two leads, and a weak applied bias $\Delta U/e{=}[U(0){-}U(L)]/e$, we can expand the integrand of (50) in terms of ΔT and $\Delta U/e$. Keeping only terms linear in ΔT and ΔU we arrive at

$$I = \frac{-e^2}{\pi\hbar}\int_0^\infty \mathrm{d}\epsilon\,T(\epsilon)\frac{\mathrm{d}f(\epsilon)}{\mathrm{d}\epsilon}\Big[\frac{\epsilon-\mu}{T}\Delta T+\Delta U\Big]\ . \tag{57}$$

From this expression we can calculate the Seebeck coefficient

$$S = -\frac{\Delta U}{e\Delta T}\ . \tag{58}$$

This expression of Seebeck coefficient was used in the study of a quantum point contact [10].

4.3 Semiconductor Superlattices

In 1958 Hatsopoulos and Kaye of the Department of Mechanical Engineering, Massachusetts Institute of Technology, reported at a conference: *The literature on thermionic devices contains suggestions that a fairly simple tube, the diode, might be used to convert heat directly into useful electrical work. On the basis of a research program started four years ago in the Research Laboratory of*

Heat Transfer in Electronics at M.I.T., we have succeeded in obtaining both an efficient and a practical engineering method to convert heat directly into electricity, with no moving parts. The configuration of this thermo-electron engine diode consists two parallel metal plates separated by a narrow vacuum gap. However, this engine operates well only if the temperature of the hot plate reaches as high as 1100^o C. To improve the efficiency of such an electron engine, one needs to reduce the work function of the electrode.

Many experiments were carried out using electrodes made from low work function materials. However, accumulated data indicated that efficiency of such type of device depends on the temperatures and the width of the vacuum gap, but not on specific electrode material. In other words, referring to Fig. 4, there is an effective potential barrier regardless what material is used for the electrodes. In fact, the problem of this effective potential barrier for thermal emission of electrons was studied in 1923 by Langmuir [11]. When the thermally emitted electrons move from the cathode to the anode, a space charge builds up in the vacuum gap. The problem has to be solved self-consistently with the Poisson equation taken into account. The result is that the space charge in the vacuum region creates an additional repulsive potential which depends on the gap width. Mahan has performed a detailed numerical analysis [12] to confirm the theoretical prediction of Langmuir, and suggested a metal-semiconductor multi-layer structure, in which there is no vacuum gap and therefore the space charge will be removed.

The characteristic feature of thermionics is Richardson current over potential barrier. Therefore, a variety of systems such as single-barrier, multiple-barriers, and superlattice have been investigated by many authors both experimentally and theoretically. Here we will not discuss individual works even selectively. Instead, we will outline only the key issues of this physics.

It is easy to show that room-temperature refrigeration is efficient if the potential barrier height in Fig. 4 is lower than 0.3 eV. Such a low potential barrier can be achieved in semiconductor superlattices with doped wells. In semiconductor superlattices the well width L_{W}, the barrier width L_{B}, the barrier height, and the Fermi level can be tailor made. In such heterostructures the doped wells are considered as metallic so that the carriers relax very fast. Therefore, local thermal equilibrium is assumed in every well. Such assumption implies a carrier mean free path l shorter than the superlattice constant $L{=}L_{\mathrm{W}}{+}L_{\mathrm{B}}$. When a carrier in one well is thermally excited over the barrier, it cannot move beyond the adjacent well in either direction perpendicular to interfaces.

However, in many semiconductors with normal doping level, l is of the order of several hundred Å at room temperature. For high mobility semiconductors such as GaAs or Ge, l is about 1000 Å at room temperature, and much longer at lower temperatures. On the other hand, the typical lattice constant in practical superlattices, which are used for thermoelectric measurements, is about 100 Å. Hence, there is a large probability for a thermally excited carrier to pass over several barriers before it emits phonons and relaxes in a well. As

a consequence of such multi-barrier hopping, in each quantum well the carrier distribution is perturbed from equilibrium when non-thermalized carriers arrive from other wells.

This problem was overcome conceptually and a way was found to solve it [13]. To clarify the important role of multi-barrier hopping when l is longer than L, let us rewrite (29) in a more suitable form:

$$\mathcal{L}_{ij} = (-e)^{2-i-j} \int \mathrm{d}\epsilon (\epsilon - \mu)^{i+j-2} G(\epsilon) \left[-\frac{\mathrm{d}}{\mathrm{d}\epsilon} f(\epsilon) \right] . \tag{59}$$

Then, the linear response transport parameters electric conductivity σ, thermopower S, and electron thermal conductivity κ_{el} can be expressed as

$$\sigma = \mathcal{L}_{11} , \tag{60}$$

$$S = \frac{\mathcal{L}_{12}}{T\mathcal{L}_{11}} , \tag{61}$$

$$\kappa_{\mathrm{el}} = \frac{(\mathcal{L}_{11}\mathcal{L}_{22} - \mathcal{L}_{21}\mathcal{L}_{12})}{T\mathcal{L}_{11}} . \tag{62}$$

For a homogeneous bulk material, in relaxation time approximation, $G(\epsilon)$ can be calculated as [4]

$$G(\epsilon) = \frac{2}{3} e^2 N(\epsilon) v l , \tag{63}$$

where $N(\epsilon)$ is the carrier density of states, and v the carrier velocity.

One can express σ, S, and κ in terms of $\mathcal{L}_{ij}$, as given above, not only for homogeneous bulk systems, but also for systems in which the electric field and the temperature gradient vary slowly in space. The key issue is to find a correct $G(\epsilon)$. We will demonstrate this feature for a semiconductor superlattice. It has been shown in [14] that if the voltage and the temperature variation across L are small, its transport parameters can be expressed in terms of certain effective bulk thermoelectric parameters which are derived as averaged thermionic transport processes over the barriers of a semiconductor superlattice. For the Richardson current $J(\phi, T)$ given by (44), if $L_{\mathrm{B}} \gg L_{\mathrm{W}}$, these effective bulk parameters are [14]

$$\sigma = (eJ/k_{\mathrm{B}}T)\, L_{\mathrm{B}} , \tag{64}$$

$$\kappa_{\mathrm{el}} = 2(k_{\mathrm{B}}/e)\, J(\phi, T)\, L_{\mathrm{B}} , \tag{65}$$

$$S = (k_{\mathrm{B}}/e)[\, (e\phi/k_{\mathrm{B}}T) + 2\,] . \tag{66}$$

It is easy to show that, when the electron distribution in the energy region above the barrier follows the Boltzmann statistics, the above expressions for σ, S, and κ_{el} can be reproduced from (59)–(62) with

$$G(\epsilon) = (1/2)e^2 N(\epsilon) v L_{\mathrm{B}} . \tag{67}$$

From (59) and (67), we conclude that thermoelectric transport in a semiconductor superlattice is equivalent to that in a homogeneous bulk semiconductor

in which the carrier mean free path is $3L_{\rm B}/4$ and the chemical potential is $e\phi$ below the band edge. The fact that $L_{\rm B}$ plays the role of the carrier mean free path can be easily understood because in [14] carrier transport over a barrier is *ballistic*, followed with an immediate thermalization upon entering the adjacent well.

It is important to point out that although the potential barrier is made of undoped semiconductor but not vacuum, charge still accumulates in the barrier regions. Based on theoretical understanding given above, thermionic transport phenomena in semiconductor superlattice have been studied taking into account both the charge accumulation in the barrier region and the multi-barrier hopping [13]. The key work in [13] is to construct a correct effective $G(\epsilon)$. One interesting outcome from this study is that most efficient thermionic transport is achieved if the energy difference between the top of the effective barrier and the chemical potential is 1–2 times of the thermal energy $k_{\rm B}T$. The condition is satisfied in the experiment [15], and a very high value ZT=2,4 is obtained in p-type Bi_2Te_3/Sb_2Te_3 superlattice.

While almost all works on thermionics in semiconductor superlattices study transport perpendicular to interfaces because of the use of Richardson current, we should mention that the transport parallel to interfaces was investigated for both heat current and electric current [16, 17].

5 One-Dimensional Systems

The definition of the figure of merit suggests that the desired sample structures should enhance σ and S, but suppress κ, regardless what mechanism controls the transport processes in the sample. Before various samples are made and characterized experimentally, some relevant theoretical information is often useful. Since the control of the phonon part $\kappa_{\rm ph}$ of thermal conductivity κ is a difficult task [3], it is relatively easy to calculate the transport properties of the electron subsystem, especially for the one-dimensional (1D) conductor using the independent particle model with an effective mass m^*.

Using the semi-classical approach with the relaxation time approximation given in Sect. 3, for a wire with a cross-section a^2, the electric conductivity σ, the thermopower S, and the electron thermal conductivity $\kappa_{\rm el}$ have been calculated as [18]

$$\sigma = \frac{e\mu_l}{\pi a^2}\sqrt{2m^* k_{\rm B}T/\hbar^2}\,F_{-1/2}\,, \tag{68}$$

$$S = -\frac{k_{\rm B}}{e}\left(\frac{3F_{1/2}}{F_{-1/2}} - \frac{\mu}{k_{\rm B}T}\right), \tag{69}$$

$$\kappa_{\rm el} = \frac{2k_{\rm B}T\tau}{\pi a^2}\sqrt{2k_{\rm B}T/\hbar^2 m^*}\left(\frac{5}{2}F_{3/2} - \frac{9F_{1/2}^2}{2F_{-1/2}}\right), \tag{70}$$

where μ is the chemical potential and μ_l the mobility along the wire. The function F_n is defined as

$$F_n = \int_0^\infty \mathrm{d}x\, x^n [\exp(x - \mu/k_\mathrm{B}T) + 1]^{-1} . \tag{71}$$

Assuming 1D phonon transport and using the typical values of the material parameters of semiconductor heterostructures, at room temperature a large value of ZT higher than 5 was obtained if the cross-section is reduced to $a < 1\,\mathrm{nm}$. Recently a detail numerical study was carried to investigate the thermoelectric properties of superlattice nanowires made of various lead salts [19]. It was again predicted that at the temperature 77 K, the figure of merit is higher than 5 if the wire diameter becomes less than 5 nm.

Experiments on various types of 1D materials have been performed in the last few years to measure their relevant thermoelectric properties. Thermopower of conducting polymers [20], organic conductors [21], and individual carbon nanotubes [22,23] were systematically studied. The thermoelectric properties were investigated for several whiskers systems such as Pb- and Ca-doped $(Bi_2Sr_2O_4)_xCoO_2$ [24], $Bi_2Sr_2Co_2O_9$ [25], and Na_xCo_2 [26]. The figure of merit of $Bi_2Sr_2Co_2O_y$ whiskers was found [27] to be 1.1 at 973 K, and the corresponding conversion efficiency is about 10%. The very best measurements on an individual carbon nanowire [28] show a small value of its thermopower.

While it is important to understand the thermoelectric physics in 1D materials, the raise in the value of ZT alone does not mean that 1D materialis are useful in relevant technology. The most important application of thermoelectric materials is in power generation and refrigeration. Besides knowing the value ZT of a thermoelectric material, it is important to know the energy efficiency of the devices made from this thermoelectric material. To illustrate this point, we will refer to Fig. 1 and Fig. 2 in Sect. 2, and let A be the lateral cross-section of the thermoelectric conductor. The various types of energies appeared in the following discussion are then normalized by A into energy densities.

We will first consider the thermoelectric power generation. Let $W_\mathrm{out}(Z)A$ be the output of useful work and $Q_\mathrm{h}(Z)A$ the net heat extracted from the hot reservoir. For the convenience of our analysis, we rewrite (13) as

$$\eta(Z) \equiv \frac{W_\mathrm{out}(Z)A}{Q_\mathrm{h}(Z)A} = \eta_\mathrm{r} \times F_\eta(Z) , \tag{72}$$

where η_r=(T_h–T_c)/T_h is the Carnot efficiency of a reversible heat engine, and

$$F_\eta(Z) = \frac{r(Z) - 1}{r(Z) + T_\mathrm{c}/T_\mathrm{h}} \tag{73}$$

is called a *fraction factor*. Using the room temperature 300 K for T_h, the calculated $F_\eta(Z)$ is plotted in Fig. 6 as solid curves for $T_\mathrm{c}/T_\mathrm{h}$=0.9, 0.7, and 0.5.

Similarly, we let $Q_\mathrm{c}(Z)A$ be the heat extracted from the cold reservoir and $W_\mathrm{in}(Z)A$ the work supplied externally. We then rewrite the coefficient of performance of a thermoelectric refrigerator given by (17) as

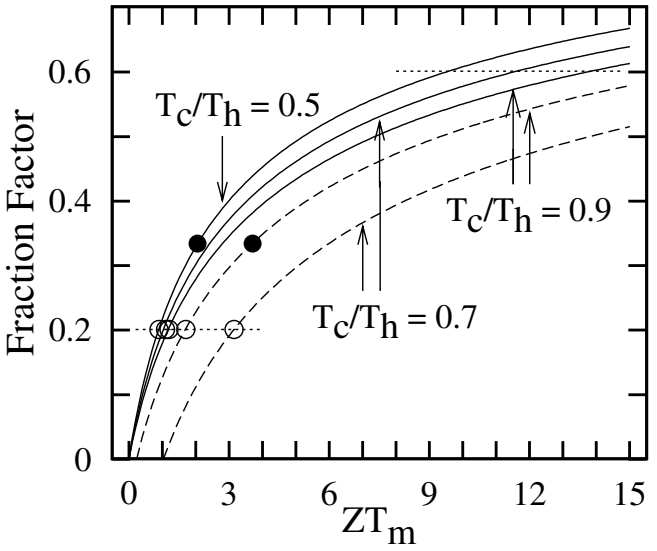

Fig. 6. The fraction factors as functions of the figure of merit $ZT_{\rm m}$ for the power generation efficiency F_η (solid curves) and for the refrigeration coefficient of performance F_K (dashed curves), calculated with $T_{\rm h} = 300\,{\rm K}$. The value of $T_{\rm c}/T_{\rm h}$ used to calculate each curve is marked correspondingly. For other features, see the text

$$K(Z) \equiv \frac{Q_{\rm c}(Z)A}{W_{\rm in}(Z)A} = K_{\rm r} \times F_K(Z)\,, \tag{74}$$

where $K_{\rm r} = T_{\rm c}/(T_{\rm h} - T_{\rm c})$ is the Carnot coefficient of performance, and the corresponding fraction factor F_K is

$$F_K(Z) = \frac{r(Z) - T_{\rm h}/T_{\rm c}}{r(Z)+1}\,. \tag{75}$$

The calculated $F_K(Z)$ is also shown in Fig. 6 as the dashed curves. We notice that with increasing $ZT_{\rm m}$ each curve in Fig. 6 exhibits a rapid initial increase and then levels off, approaching 1 as $ZT_{\rm m}$ approaching infinity. Under practical working condition, the value of $T_{\rm c}/T_{\rm h}$ is between 0.7 and 0.5 for power generation, and is around 0.9 for refrigeration.

Let us express $Z_{3d}T$ for the figure of merit of a bulk thermoelectric material (3D-TEM) with a macro-size cross-section A. Using as a power generator, $W^{3d}_{\rm out}(Z_{3d})A = Q_{\rm h}\eta(Z_{3d})A = Q_{\rm h}\eta_{\rm r}AF_\eta(Z_{3d})$ is the output of work. Under the same working conditions, if we shrink A to a nano-size area S such that the quantum confinement changes the 3D-TEM into a 1D thermoelectric material (1D-TEM), the figure of merit is also modified from $Z_{3d}T$ into $Z_{1d}T$. The output of power is then reduced drastically to $W^{1d}_{\rm out}(Z_{1d})S = Q_{\rm h}\eta_{\rm r}SF_\eta(Z_{1d})$. We can pack N of such 1D-TEM parallel to each other to form a macro-size

cable with the same cross-section A. The power output of this cable becomes $W^{\mathrm{C}}_{\mathrm{out}}(Z_{1d},N)S = NW^{1d}_{\mathrm{out}}(Z_{1d})S = Q_{\mathrm{h}}\eta_{\mathrm{r}}A\gamma F_{\eta}(Z_{1d})$, where $\gamma{=}NS/A$. In order to maintain their 1D properties, within the cable every 1D-TEM must be separated from all other 1D-TEMs both electrically and thermally. Therefore, NS is less than A and so γ is less than 1. It is reasonable to expect that within the cable the distance between the nearest surfaces of two adjacent 1D-TEM is about equal to or a little less than $\sqrt{S}$. Hence, γ is about 1/4 to 1/3.

For a given 1D-TEM with $Z_{1d}T$, there is a corresponding 3D-TEM with $Z_{3d}T$ such that $W^{3d}_{\mathrm{out}}(Z_{3d})A{=}W^{\mathrm{C}}_{\mathrm{out}}(Z_{1d},N)S$. That is, under the same operation condition, the 3D-TEM system and the 1D-TEM system have the same amount of power output. The above equality reduces simply to $F_{\eta}(Z_{3d}){=}\gamma F_{\eta}(Z_{1d})$. Since $\gamma{<}1$ and $F_{\eta}(Z)$ is a monotonic function of Z, we have $Z_{3d}{<}Z_{1d}$. Consequently, for thermoelectric power generation, one can use a 3D-TEM of lower figure of merit to produce the same total power output as that from an equivalent-size cable of 1D-TEM of higher figure of merit. The same conclusion can be reached for thermoelectric refrigeration because (72)–(73) and (74)–(75) have similar mathematical structure.

To illustrate this important conclusion, we set $\gamma{=}1/3$ and $T_{\mathrm{h}} = 300\,\mathrm{K}$ to find out the pair $(Z_{1d};Z_{3d})$ such that the condition $F_{\eta}(Z_{3d}){=}F_{\eta}(Z_{1d})/3$ is satisfied for power generation or $F_K(Z_{3d}){=}F_K(Z_{1d})/3$ is satisfied for refrigeration. One example of calculated result is shown in Fig. 6 as the two dotted lines: the upper dotted line marks the regime of the values of Z_{1d} for $F_{\eta}(Z_{1d}){=}F_K(Z_{1d}){=}0.6$, while the lower dotted line with circles marks the corresponding region of the values of Z_{3d} for $F_{\eta}(Z_{3d}){=}F_K(Z_{3d}){=}0.2$. More explicitly, for power generation, a 3D-TEM with $Z_{3d}{\simeq}1$ has the same efficiency as a 1D-TEM with $Z_{1d}{\simeq}10$. Furthermore, for a 1D-TEM with $Z_{1d}{\to}\infty$, the corresponding values of Z_{3d} remain in the region around the solid dot at $Z_{3d}{\simeq}2$ for power generation, and in the region around the solid dot at $Z_{3d}{\simeq}4$ for refrigeration.

We have already discussed in Sect. 4 that semiconductor superlattices and alloys, which are 3D-TEMs, have been extensively studied both theoretically and experimentally. Figure of merit higher than one at room temperature have been found in many of such 3D-TEMs. In particular, a very large value of $ZT{\simeq}2.4$ was discovered [15] in p-type Bi_2Te_3/Sb_2Te_3 superlattice devices. Judging from our numerical results shown in Fig. 6, it would be extremely difficult to find a corresponding 1D-TEM which can have better performance in power generation and/or refrigeration. We should mention that the numerical study on GaAs and PbTc quantum well and wire superlattice systems also indicates that 1D-TEM is not good for power applications [17]. But it is important to notice that here our proof is based on the fundamental thermodynamics which is independent of material systems.

6 Heat Transport by Phonon Wave Motion

Although the phonon thermal conductivity $\kappa_{\rm ph}$ appears in the expression of the figure of merit, so far we have not touched this important problem yet. Joule heating and thermal insulation are among the most crucial issues in modern electronic and optoelectronic devices. Following the advancement of nanotechnology, to understand the thermal transport in semiconductor nanostructures has emerged as an urgent task. In some devices which generate heat, such as integrated circuits and semiconductor lasers, one needs materials of high thermal conductivity to remove the excess heat as fast as possible. On the other hand, to operate a thermoelectric cooler efficiently, one needs materials of low thermal conductivity to be used as thermal insulating walls. This is not only a problem for technology, but also a challenge for fundamental science.

Traditionally phonon transport has been treated in terms of diffusive process, and heat conduction is governed by the corresponding Fourier's Law. Most theoretical analysis were based on the Boltzmann transport equation [29,30], treating phonons as particles and ignoring their wave-like properties. The key quantity in this equation, the scattering mechanism varies from system to system. One important scattering process which appears in almost all systems is the scattering of phonons by interfaces between different materials, since an interface breaks the regular lattice structure. This phenomena is referred to as Kapitza resistance after his first experimental observation [31]. The first theory to explain the interface resistance is the acoustic-mismatch model [32], in which the phonon transmission amplitude is expressed in terms of the acoustic impedances of the materials at opposite sides of the interface. Later the interface resistance was also investigated with the diffusive mismatch model [33] and the molecular-dynamics simulation method [34]. With these methods the scattering terms in the Boltzmann equation are determined within the framework of particle motion.

While nano-scale thermal transport has been reviewed recently [35], the corresponding theory remains to be established. Since the typical length scale in a nanostructure is of the order of phonons mean free path, the particle-type diffusive description of thermal transport is no longer appropriate. In a semiconductor superlattice, the relevant length scale is the superlattice period. Whether the thermal transport is dominated by the particle motion or of the wave propagation depends on the ratio of the superlattice period to the phonon mean free path. The crossover between these two characteristic regimes was studied recently in terms of a simple model [3].

Instead of using the Boltzmann transport approach to investigate the particle nature of phonon diffusive motion, we will study the thermal transport in the framework of phonon wave propagation. To discuss the phonon heat transport in thermoelectric systems, we will focus our attention on semiconductor superlattices, since proper samples can be fabricated to ensure that phonon wave propagation is the dominating process of thermal transport.

In this respect, we will mention briefly the interesting relevant experimental observations. The lattice thermal conductivity of Si/Ge superlattice was found [36] increasing with temperature but decreasing with the lattice period. In a GaAs/AlAs superlattice the thermal conductivity along the growth direction is three to ten times smaller than that in a GaAlAs alloy with the same mass ratio [37]. The characteristic feature of phonon dispersion curve in a superlattice, namely, the formation of mini-gaps due to the zone folding effect, was detected in a selective phonons transmission measurement [38]. This experiment proved unambiguously the interference of phonon wave motion in a $GaAs/Ga_{0.5}Al_{0.5}As$ superlattice. It was conjectured that the formation of mini-gaps is the origin of the large reduction of thermal conductivity in semiconductor superlattices.

If a sample is not perfectly ordered, such as most semiconductor superlattices containing repeated parts of random alloys, the phonon wave will be scattered when propagating in the sample. If one neglects the phonon-phonon scattering, the phonon propagation can be studied conveniently with the transfer matrix approach. While this method is very general and has been much used to study the electron transport in disordered systems, to present the key results here with simple mathematical analysis, we will consider the $GaAs/Ga_{1-x}Al_xAs$ superlattice grown along the [001] direction.

Let us start with the propagation of a lattice wave along the [001] direction of a bulk III-V semiconductor crystal. It has been shown [39] that the transverse vibrations and the longitudinal vibrations belong to different symmetry representations and therefore are decoupled. For a longitudinal mode all atoms in a given (001) plane vibrate as a whole. By projecting the positions of all atoms in the (001) plane on the [001] axis, we obtain an equivalent one-dimensional model to investigate the longitudinal acoustic branch propagating along the [001] direction. This 1D model is valid also for III-V semiconductor heterostructures, such as GaAs/AlAs superlattices. However, this theory is not correct for heterostructures containing alloys which destroy the crystal symmetry, for example, $GaAs/Ga_{1-x}Al_xAs$ superlattices. Nevertheless, if the distribution of aluminum atoms in $Ga_{1-x}Al_xAs$ layers is sufficiently uniform, the complete decouple of the longitudinal and the transverse modes for phonons propagating in [001] direction remains a good approximation, which will be adopted in our following discussion.

The eigen-modes of this one-dimensional model have been used to calculate relevant physical properties, for example, the Raman spectrum [40]. Here we will use this one-dimensional model to study the transmission coefficients of longitudinal acoustic phonons and the related phonon heat transport. The one-dimensional samples are generated with a computer. We define the z-axis along the [001] direction, and a the width of one monolayer (the distance between two nearest As atoms) in the 1D system. At every position $z{=}na$ with integer n, there is an As atom. To create the alloy $Ga_{1-x}Al_xAs$, with a probability according to the value of x, the random number generated by a computer will determine whether a Ga atom or an Al atom to be placed

between two adjacent As atoms. The crystals GaAs and AlAs are special cases with x=0 and x=1, respectively.

Let U_i be the displacement of the atom at position ia, and $K_{i,i+\frac{1}{2}}$ the force constant between the two adjacent atoms at ia and $(i+\frac{1}{2})a$, where i is an integer or a half-integer. Then, within the harmonic approximation, the equation of motion for the atom at position ia is

$$m_i \frac{\mathrm{d}^2 U_i}{\mathrm{d}t^2} = -K_{i-\frac{1}{2},i}(U_i - U_{i-\frac{1}{2}}) - K_{i,i+\frac{1}{2}}(U_i - U_{i+\frac{1}{2}}) \,, \tag{76}$$

where the mass m_i can be either m_{Ga} for a Ga atom, m_{As} for an As atom, or m_{Al} for an Al atom. Since GaAs and AlAs have almost the same values of elastic moduli, it is reasonable to use one value K for both the Ga–As pair and the Al–As pair.

For pure systems GaAs and AlAs, the solutions of (76) give the phonon dispersions and the corresponding eigen-modes. For a given frequency ω, the propagating wave vector is

$$k_\gamma(\omega) = \frac{2}{a} \sin^{-1} \sqrt{\frac{m_{\mathrm{Ga}} + m_{\mathrm{As}}}{2K}\,\omega^2 - \frac{m_{\mathrm{Ga}} m_{\mathrm{As}}}{4K^2}\,\omega^4} \tag{77}$$

in GaAs, and is

$$k_\alpha(\omega) = \frac{2}{a} \sin^{-1} \sqrt{\frac{m_{\mathrm{Al}} + m_{\mathrm{As}}}{2K}\,\omega^2 - \frac{m_{\mathrm{Al}} m_{\mathrm{As}}}{4K^2}\,\omega^4} \tag{78}$$

in AlAs. For a GaAs/AlAs heterostructure separated by a perfect interface, we can solve (76) to derive the transmission and reflection amplitudes of a propagating phonon, which has the form $\mathrm{e}^{\mathrm{i}kz}+r\mathrm{e}^{-\mathrm{i}kz}$ at the incident side of the interface and $t\mathrm{e}^{\mathrm{i}kx}$ at the transmitted side.

In a one-dimensional sample there are many interfaces. We will then use the general transfer-matrix technique to calculate the resulting transmission amplitude $T(\omega)$ of a propagating mode with frequency ω. Using the matching conditions that the amplitude and the phase of the atomic displacement at an interface remain the same when calculated from either side of the interface, the 2×2 scattering-matrix $\mathcal{T}(l;\omega)$ for the l-th interface in our one-dimensional model can be easily derived with the matrix elements as

$$\begin{aligned} \mathcal{T}(l;\omega)_{11} &= \mathcal{T}(l;\omega)_{22} = \frac{k_2(l;\omega) + k_1(l;\omega)}{2k_2(l;\omega)} \,, \\ \mathcal{T}(l;\omega)_{12} &= \mathcal{T}(l;\omega)_{21} = \frac{k_2(l;\omega) - k_1(l;\omega)}{2k_2(l;\omega)} \,. \end{aligned} \tag{79}$$

Depending on the materials at each side of the l-th interface, $k_1(l;\omega)$ and $k_1(l;\omega)$ can be either $k_\gamma(\omega)$ or $k_\alpha(\omega)$ given by (77) and (78). The total scattering-matrix of the entire one-dimensional sample is then

$$\mathcal{T}(\omega) = \prod_{l=1}^{N} \mathcal{T}(l;\omega) \,, \tag{80}$$

where N is the total number of interfaces. From this we obtain the transmission spectrum of the one-dimensional sample

$$T(\omega) = \left| \frac{1}{\mathcal{T}(\omega)_{22}} \right|^2 \,. \tag{81}$$

Using the transmission coefficients to calculate the phonon heat conduction, we connect the sample to a phonon injector at z=0 with temperature $\theta_{\rm I}$ and a phonon collector at z=Na with temperature $\theta_{\rm C}$. Let $n(\omega,\theta)$ be the Bose-Einstein distribution function, and $\boldsymbol{v}(\boldsymbol{k})$ the phonon group velocity. The thermal current density flowing through the sample from the injector to the collector can be expressed in a Landuaer-type formula

$$J_z = \sum_{\lambda} \int \frac{\mathrm{d}\boldsymbol{k}}{(2\pi)^3} \, \hbar\omega_\lambda(\boldsymbol{k}) v_{\lambda,z}(\boldsymbol{k}) \left[\, n(\omega_\lambda,\theta_{\rm I}) - n(\omega_\lambda,\theta_{\rm C}) \, \right] T_\lambda(\omega_\lambda;k_z) \,, \tag{82}$$

with λ is the phonon branch index. Here we have expressed the transmission coefficient as $T_\lambda(\omega_\lambda;k_z)$ to emphasize explicitly that it does not depend on the component of $\boldsymbol{k}$ parallel to interfaces.

The total thermal energy moving into the sample from either the injector or the collector is

$$E_\lambda(k_z,\theta_\nu) = \int \frac{\mathrm{d}k_x \, \mathrm{d}k_y}{(2\pi)^2} \, \hbar\omega_\lambda(\boldsymbol{k}) v_{\lambda,z}(\boldsymbol{k}) \, n(\omega_\lambda,\theta_\nu) \,, \tag{83}$$

where ν = I for injector and ν = C for collector. In terms of these thermal energy flows, we can rewrite (82) as

$$J_z = \sum_{\lambda} \int_0^\infty \frac{\mathrm{d}k_x}{2\pi} \left[\, E_\lambda(k_z,\theta_{\rm I}) - E_\lambda(k_z,\theta_{\rm C}) \, \right] T_\lambda(k_z) \,. \tag{84}$$

Equation (84) is very similar to the one-dimensional formula used to analyze the quantization of thermal conductance in suspended insulating nanowires [41]

$$J_{\rm th} = \sum_{\lambda} \int_0^\infty \frac{\mathrm{d}k}{2\pi} \left[\, \mathcal{E}_\lambda(k_z,\theta_{\rm I}) - \mathcal{E}_\lambda(k_z,\theta_{\rm C}) \, \right] T_\lambda(k_z) \,, \tag{85}$$

where the thermal energy flows into the one-dimensional sample are

$$\mathcal{E}_\lambda(k_z,\theta_\nu) = \hbar\omega_\lambda(k_x) v_{\lambda,z}(k_z) \, n(\omega_\lambda,\theta_\nu) \,. \tag{86}$$

By comparing (83)–(84) and (85)–(86), it is reasonable to expect that the thermal transport by phonon wave motion along the [001] growth direction

of a GaAs/$Ga_{1-x}Al_xAs$ superlattice can be understood qualitatively in terms of our simple one-dimensional model. This conjecture is consistent with the detailed study [3] of phonon thermal transport perpendicular to the interfaces of a multilayer system, which shows the same qualitative features in one-, two- and three-dimensional calculations.

Within the framework of the one-dimensional model, in the limit of $|\theta_{\rm I} - \theta_{\rm C}| \ll \theta_{\rm I}$, the one-dimensional longitudinal acoustic phonon thermal conductance

$$G_{\rm th}(\theta_{\rm I}) = \frac{J_{\rm th}(\theta_{\rm I})}{|\theta_{\rm I} - \theta_{\rm C}|} \tag{87}$$

can be easily derived as

$$G_{\rm th}(\theta) = \frac{k_{\rm B}^2 \theta}{h} \int_0^\infty {\rm d}\nu \, \frac{\nu^2 {\rm e}^\nu}{({\rm e}^\nu - 1)^2} \, T(\frac{\nu k_{\rm B} T}{\hbar}) \, , \tag{88}$$

where $k_{\rm B}$ is the Boltzmann constant. In a pure GaAs or AlAs sample, the transmission coefficient $T(\omega)$=1 for all allowed phonon frequency. In this case we obtain the quantum of thermal conductance $g_{_0}(\theta)$:

$$G_{\rm th}(\theta) = \frac{\pi^2 k_{\rm B}^2}{3h} \, \theta \equiv g_{_0}(\theta) \, . \tag{89}$$

Using the above analytical expressions, a detailed numerical calculation was performed [42] to prove definitely that the formation of mini-gaps is not the origin of the large reduction of thermal conductivity in semiconductor superlattices. This calculation also suggests that within the harmonic approximation without taking into account of the phonon-phonon scattering, the phonon wave motion cannot explain the experimental observation that the heat conductivity in a superlattice is substantially lower than that in an alloy of same mass ratio. For an improvement, one must include in the model the anharmonic terms.

7 Tunneling Thermoelectricity in Nano-Systems

The physical quantity *temperature* is well-defined in a macroscopic system under thermodynamic equilibrium. The system reaches equilibrium through the dynamic motion of the ions, either arranged in lattice structure or in random positions. Such a dynamic motion has a characteristic length, for example, the phonon mean free path in a crystal. Our concept of temperature becomes unclear when the sample dimension decreases to a scale comparable to this characteristic length. This is a fundamental problem in nano-systems, which so far has not been explored at all, and is too difficult to be discussed here.

The other well-known phenomenon is the emergence of quantum effect when the sample dimension is reduced. The quantum process which is relevant

to our study is tunneling through potential barriers. The thermionic process drives a carrier out of a system only if the energy of the carrier is higher than the highest potential barrier. That is, the thermionic process removes hot electrons from the system. If there is another process which efficiently moves cold electrons into the system, then the system works as a refrigerator. The colder the incoming electrons, the higher the cooling efficiency. The tunneling process can serve as the required second process, and will be discussed here. We will first mention a very interesting fundamental phenomenon *current reversal*, and then introduce a cooler based on quantum mechanical tunneling.

7.1 Electric Current Reversal

When a temperature gradient is applied to a sample, in semi-classical theory where electrons move diffusively, the thermoelectric carrier current flows through the sample from the hot reservoir to the cold reservoir. However, with quantum mechanical tunneling transport, this current can reverse its direction [43].

To demonstrate the physical picture with a simple one-dimensional electron transport along the x-axis, we consider a sample of finite width between $x{=}0$ and $x{=}d$, attaching to two conducting leads at $x{=}0$ with a hot temperature $T_{\rm h}$, and at $x{=}d$ with a cold temperature $T_{\rm c}$. In the yz plane the potential is constant which we set to zero, and $k_\parallel$ labels the electron wave vector for its motion in the yz plane. Along the x-axis the potential $V(x)$ is zero in the two conducting leads $x{<}0$ and $x{>}d$, but can be any function of finite value within the sample region $0{<}x{<}d$.

Let $\mu_{\rm h}$ be the chemical potential in the hot reservoir at the left side, and $\mu_{\rm c}$ be the chemical potential in the cold reservoir at the right side. Consider an electron coming in from the left lead with the kinetic energy $\epsilon_\perp$ along the x-axis and the parallel wave vector $k_\parallel$ in the yz plane. Let $T(\epsilon_\perp)$ be the transmission probability through the sample, and $f_{\rm h}(k_\parallel\epsilon_\perp, T_{\rm h})$ [or $f_{\rm c}(k_\parallel\epsilon_\perp, T_{\rm c})$] the electron distribution function in the left [or right] lead. Then the thermoelectric current density flowing through the sample from the hot lead to the cold lead is given by

$$j = \frac{e}{\pi\hbar}\sum_{k_\parallel}\int_0^\infty {\rm d}\epsilon_\perp\, T(\epsilon_\perp)\,\left[f_{\rm h}(k_\parallel\epsilon_\perp, T_{\rm h}) - f_{\rm c}(k_\parallel\epsilon_\perp, T_{\rm c})\right]\ . \tag{90}$$

Let us assume that in a specific sample the energy dependent transmission probability $T(\epsilon_\perp)$ is zero except for a small energy region $\varepsilon{-}\Gamma{<}\epsilon_\perp{<}\varepsilon{+}\Gamma$ in which $T(\epsilon_\perp){\simeq}1$. In this case the current density simplifies to

$$\frac{j}{2e\Gamma} \propto T_{\rm h}\,\ln(1+{\rm e}^{(\mu_{\rm h}-\varepsilon)/k_{\rm B}T_{\rm h}}) - T_{\rm c}\,\ln(1+{\rm e}^{(\mu_{\rm c}-\varepsilon)/k_{\rm B}T_{\rm c}})\ . \tag{91}$$

Since the chemical potential decreases monotonically with increasing temperature, we have $\mu_{\rm c}{>}\mu_{\rm h}$.

We will consider the case that the energy window around ε for $T(\epsilon_\perp)\simeq 1$ is so located that both conditions $|\mu_{\rm h}-\varepsilon| \gg k_{\rm B}T_{\rm h}$ and $|\mu_{\rm c}-\varepsilon| \gg k_{\rm B}T_{\rm c}$ are satisfied. Then, if $\mu_{\rm h}<\varepsilon$ and $\mu_{\rm c}<\varepsilon$, equation (91) reduces to

$$\frac{j}{2e\Gamma} \propto T_{\rm h}\, {\rm e}^{(\mu_{\rm h}-\varepsilon)/k_{\rm B}T_{\rm h}} - T_{\rm c}\, {\rm e}^{(\mu_{\rm c}-\varepsilon)/k_{\rm B}T_{\rm c}} > 0 \,, \tag{92}$$

which is the normal thermoelectric current direction. On the other hand, if $\mu_{\rm h}>\varepsilon$ and $\mu_{\rm c}>\varepsilon$, from (91) we have

$$\frac{j}{2e\Gamma} \propto \mu_{\rm h} - \mu_{\rm c} < 0 \,, \tag{93}$$

which exhibits a thermoelectric current reversal, implying a thermopower anomaly. A numerical analysis based on (90) with small Γ confirms this conclusion.

An energy window with $T(\epsilon_\perp)\simeq 1$ can be realized in double-barrier (or multiple-barrier) tunneling structures, which has been extensively studied both theoretically and experimentally [44]. The energy position and the width of this window are controlled by the chemical composition and the geometrical structure of the tunneling system, and the chemical potential $\mu_{\rm c}$ (or $\mu_{\rm h}$) can be adjusted by varying the impurity concentration in cold (or hot) reservoir. A detailed numerical study can be found in [43]. It is important to point out that this *thermoelectric current reversal* does not violate the thermodynamic law with respect to the heat current. The heat current in the system is always directed from the hot reservoir to the cold reservoir. It can be calculated from (90) with the electron charge e replaced by $\epsilon_{\rm tot}-\mu_{\rm h,c}$, where $\epsilon_{\rm tot}$ is the total electron energy. When the electric current changes sign, both the factor $f_{\rm h}(\epsilon_{\rm tot},T_{\rm h})-f_{\rm c}(\epsilon_{\rm tot},T_{\rm c})$ and the factor $\epsilon_{\rm tot}-\mu_{\rm h,c}$ change sign, which implies the correct phenomenon that the heat current always flows from the hot to the cold reservoir. It is easy to show that the energy flow is connected to the Joule heating power [45] $jV_{\rm th}$, where $V_{\rm th}$ is the thermoelectric voltage.

Here we have used the temperature difference $T_{\rm h}-T_{\rm c}$ to control the difference $\mu_{\rm c}-\mu_{\rm h}$ of chemical potentials, which can also be controlled with an applied bias. It is interesting to investigate the combined influence of the temperature and the bias on thermopower anomaly. This problem will be more relevant to the device applications.

7.2 Tunneling Thermionic Refrigeration

We have demonstrated in the previous subsection that via the use of tunneling channel, we can inject electrons into a system with selected energy, especially low energy or cold electrons. Incorporating this process with the thermionic emission of hot electrons from the same system, a novel solid state cooler can be fabricated to work at room temperature.

We will use AlGaAs heterostructure to construct such tunneling-thermionic refrigeration (TTR). It is well-known that AlGaAs heterostructures are not good thermoelectric materials. However, these heterostructures

have been thoroughly studied and their material parameters were determined rather accurately. Hence, using AlGaAs heterostructures as an example to illustrate the room temperature TTR, the results can be presented with quantitative accuracy. Since room temperature TTRs can be made even with the bad thermoelectric material AlGaAs heterostructures, it is optimistic to expect that higher efficiency TTRs can be realized with better thermoelectric materials.

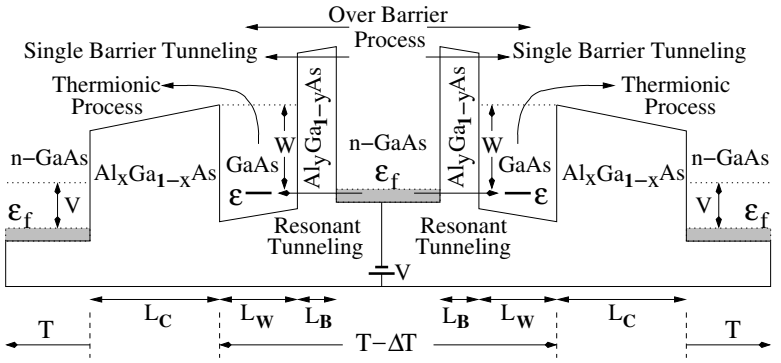

Fig. 7. The schematic illustration of a tunneling-thermionic refrigerator and the corresponding potential profile under a bias voltage V. When a current flows under a bias voltage V, the temperature in the region between the two thick $Al_xGa_{1-x}As$ alloy layers is lowered by an amount ΔT

Figure 7 shows the TTR structure and the corresponding potential profile under a bias voltage V. In the center of the TTR, sandwiched between two thin $Al_yGa_{1-y}As$ alloys which work as potential barriers, is an n-type GaAs. The materials beyond the potential barriers are thin layers of undoped GaAs, thick $Al_xGa_{1-x}As$ alloy layers, and n-type GaAs. Each undoped GaAs layer is a quantum well in which there is only one quasi-bound state with energy ε. The $Al_xGa_{1-x}As$ alloy layer is very thick such that when a bias voltage V is applied as shown in Fig. 7, the carriers which tunnel from the n-GaAs into the quantum well cannot tunnel through the $Al_xGa_{1-x}As$ layer. The Al concentration in $Al_yGa_{1-y}As$ is higher than that in $Al_xGa_{1-x}As$ ($y>x$). The sample structure can be asymmetric with respect to the central n-type GaAs region. However, for the convenience of drawing, here we have assumed a symmetric TTR structure.

Since the structure of the TTR is symmetric with respect to the center n-type GaAs, the physical processes in the right half of the system are the same as those in the left half of the system. Therefore, we only need to illustrate the three carrier transport processes in the right half of the device. In the n-doped GaAs, the Fermi energy is marked as ε_f. The over-barrier process has no thermoelectric effect. The single-barrier-tunneling process has a weak

refrigeration effect [46,47]. The third transport process, the main cooling process, is the resonant tunneling from the middle n-GaAs into the quasi-bound level ε of the GaAs quantum wells, followed by the thermionic escape over the thick $Al_xGa_{1-x}As$ barriers. In this process, each carrier removes an amount of thermal energy W from the lattice. Therefore, if the device is surrounded by an environment of temperature T, the temperature of the inner part (between the two $Al_xGa_{1-x}As$ alloy layers) will be lowered to T–ΔT. ΔT can be increased by reducing the heat back-flow through the two thick $Al_xGa_{1-x}As$ alloy layers, which have very low phonon thermoconductivity as discussed in Sect. 6. However, the alloy layer thickness L_C must be less than the carrier mean free path in order to maintain the ballistic transport of the thermionic process.

The amount of heat removed by the electric current J from the inner part to the outer environment has two contributions: $Q_{\rm s}$ due to the single-barrier tunneling process and $Q_{\rm rt}$ due to the resonant-tunneling-thermionic process. Let $Q_{\rm ph}$ be the heat back-flow through the two $Al_xGa_{1-x}As$ alloy layers due to the phonon heat conduction. Then, the *coefficient of performance* (COP) of the TTR is defined as

$$K = \frac{Q_{\rm s} + Q_{\rm rt} - Q_{\rm ph}}{JV} \ . \tag{94}$$

To operate the TTR, K must be positive. When K is positive, the inner part of the system will cool down to the temperature T–ΔT when the temperature of the outer environment is T. The TTR has been analyzed in details together with extensive numerical calculation [48]. Even for the *bad* thermoelectric material AlGaAs, the numerical results show that at room temperature the TTR can easily lower the temperature by 5-7 degrees. Such devices can be fabricated with the present semiconductor technology. Besides its use as a kitchen refrigerator, the TTR can efficiently cool micro-electronic devices.

8 New Concepts in Thermionic Applications

Thermoelectricity in bulk materials has been an old topic, and its applications have been limited by the lack of materials with high value of figure of merit. The modern trend towards miniaturization of electronic and optoelectronic devices has brought in another problem: to integrate bulk thermoelectric cooling with micro-electronics and future nano-electronics. Benefited by the development of thin film technology in the past, micro-thermoelectrics has emerged as the present focus of thermoelectric devices research. While we will discuss in the following two new concepts for future solid state coolers, the reader is referred to [49] for micro-thermoelectric applications.

8.1 Opto-Thermionic Refrigeration with Energy Recycle

We have introduced in previous sections the modulation doped semiconductor superlattices as potential candidate for thermionic devices. However, using the results in [14] and [50], for a typical value of semiconductor thermal conductivity, it is easy to see that the refrigeration efficiency is very low with a very small value of figure of merit. Only when the thermal conductivity in the barriers becomes as low as that of glasses, the thermoelectric figure of merit approaches the value of unity. One way to suppress the unwanted opposite heat flow is to increase the barrier width. Nevertheless, this may also lower the thermionic efficiency, and hence reduces the required in-flow of cold electrons over the barriers from the cold side to the hot side. This is an intrinsic dilemma.

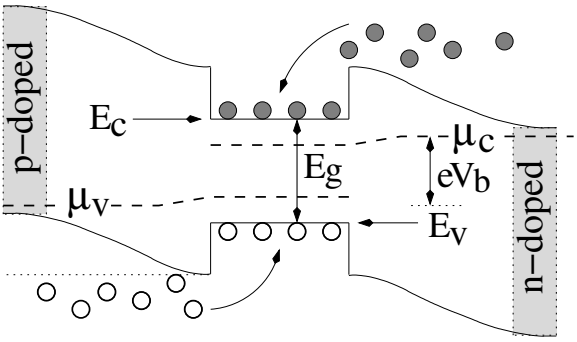

Fig. 8. The schematic illustration of the energy levels in a p-AlGaAs/GaAs/n-AlGaAs junction under a bias voltage V_b. Shaded areas are p-doped and n-doped regions

An entirely new approach of heat pumping, the so-called *opto-thermionic refrigeration* cooling process, was proposed [51] to overcome this problem. The system under an applied bias V_b, schematically illustrated by Fig. 8, is a modified pn-junction with a wide undoped quantum well embedded at the center part. The carriers in the thermionic processes are both electrons (solid circles in Fig. 8) in the n-doped semiconductor at one side and holes (open circles in Fig. 8) in the p-doped semiconductor at the other side. When thermally excited electrons and holes are driven into the quantum well by the applied bias voltage V_b, they recombine. The energy supplied to an electron-hole pair by the applied bias is eV_b, and if an electron-hole pair recombines radiatively, the so-created photon carries an amount energy E_g out of the system. If eV_b is less than E_g, then an amount of heat E_g-eV_b is released from the system by photon emission. If the recombination is non-radiative, the system will be heated up.

In many semiconductors the dominant non-radiative recombination channel is the Auger scattering, which can be largely suppressed in high-quality samples. Based on the theoretical analysis and prediction in [51], a detailed numerical calculation was performed [52] very recently on GaAlAs systems with impurity concentration and quantum well width as varying parameters. The numerical results show that the opto-thermionic cooler can extract an amount of thermal energy from the sample as much as 10 watts/cm^2. Since the fabrication and measurement technologies for GaAlAs heterostructures are well-developed, experimental works on opto-thermionic refrigeration are expected to be reported in the near future.

When using opto-thermionic process to cool portable electronic/optoelectronic equipments, an energy related problem arises because the refrigeration part also draw energy from battery. One can make a photo-cell with the same material as the material of the quantum well in the opto-thermionic cooler, and incorporate them together. Then, a large part of the photons emitted from the opto-thermionic refrigerator will be reabsorbed by the photo-cell to extend the life time of the battery.

8.2 Thermionic Couple

One very important application of the conventional thermoelectric materials is to measure temperature with thermocouple made from two metals. Although micro-thermocouple is used to map the temperature distribution in micro-scale, to reduce its size further to nano-thermocouple remains a difficult problem. A breakthrough in concept is need in order to establish a working theory.

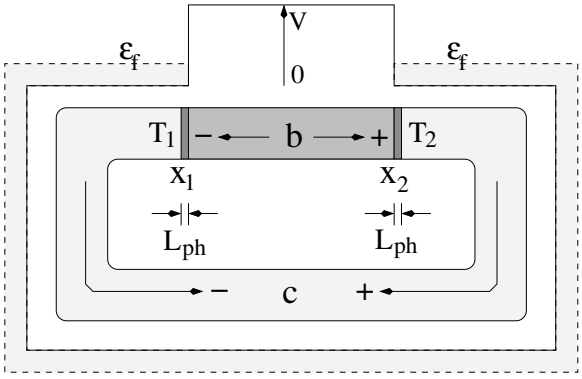

Fig. 9. A proposed thermionic couple with a conductor marked by **c** connected at both ends (positions x_1 and x_2) with an insulator marked by **b**. The two local phonon sources of narrow width $L_{\rm ph}$ are set at the junction positions x_1 and x_2 with respective temperatures T_1 and T_2. The potential profile around the ring is shown by the outer loop with the Fermi energy ϵ_F

Thermionic processes occur in nano-scale systems, and a nano-scale *thermionic couple* can be constructed accordingly [53] at least in principle. A proposed thermionic couple is shown in Fig. 9, with the inner part for the geometric structure of the system and the outer part for the corresponding potential profile. The system is a close loop, and the x-axis is defined as along the loop in clockwise direction. The loop system consists of a conducting wire (light shaded region marked with **c**) and an insulating wire (dark shaded region marked with **b**). The length of the conducting wire, L is shorter than the carrier mean free path, and the carrier transport in the wire is ballistic. The single-barrier potential profile around the loop is plotted as the outer part in Fig. 9, with the Fermi energy marked as ϵ_F. Around positions x_1 and x_2 we introduce two very narrow regions of width $L_{\rm ph}$ for local heating. In these two narrow regions of phonon sources, the phonon temperatures T_1 and T_2 are well-defined. The electron-phonon interaction is then restricted to these two point contacts at x_1 and x_2, which can be considered as heat reservoirs. There are several existing techniques [54–59] which may be used here for local heating, but it is not sure which one works best.

Since the system is not connected to any particle reservoir, in [53] the non-equilibrium electron distribution function was solved using the Boltzmann equation approach with electron-phonon interaction at the two junctions. It was found that there is always a heat current flowing from the hot junction to the cold one. The electric current over the potential barrier increases monotonically as the Fermi energy in the metallic wire approaches the potential barrier height from below. Using a modulation doped semiconductor ring for numerical calculation, the thermionic current was found to be almost linear with respect to the temperature difference between the two junctions. This discovery provides the first theoretical guideline for the fabrication of nanometer scale thermionic couple.

9 A Remark

Thermionics has been developed rapidly in the last 15 years as a result of close joint research efforts by theory, experiments, materials fabrications, and device designs. The very original idea of *electron engine*, using two parallel metallic plates, did not work practically at room temperatures because of the space-charge-induced additional potential barrier in the vacuum gap between the two metallic plates. In the very original studies on this topic, the image potential was ignored because the vacuum gap is not sufficiently narrow. The width of this vacuum gap can be reduced significantly due to the recent advancement of technologies. When the image potential is taken into account in a sufficiently narrow vacuum gap, the resultant potential barrier is lowered to the extent that very large thermionic current can be achieved. The remaining problem is to maintain a steady narrow gap in vacuum, which is essential to stop heat back-flow from the cold metal plate to the hot metal plate. For the details of

new development in this area, the reader is referred to one of the co-authors (M.L), who is one of the leading researchers in this field.

References

1. H. J. Goldsmid: *Thermoelectric Refrigeration* (Plenum, New York 1964)
2. A. F. Ioffe: *Semiconductor Thermoelements and Thermoelectric Cooling* (Infosearch Limited, London 1957)
3. M. V. Simkin, G. D. Mahan: Phys. Rev. Lett. **84**, 927 (2000)
4. N. W. Ashcroft, N. D. Mermin: *Solid State Physics* (Saunders College, Philadelphia 1976) Chap. 13
5. Y. Hishinuma, T. H. Geballe, B. Y. Moyzhes, T. W. Kenny: Appl. Phys. Lett. **78**, 2572 (2001)
6. R. Landauer: Physica Scripta **T42**, 110 (1992)
7. B. J. van Wees, H. van Houten, C. W. J. Beenakker, J. G. Willamson, L. P. Kouwenhoven, D. van der Marel, C. T. Foxon: Phys. Rev. Lett. **60**, 848 (1988)
8. D. A. Wharam, T. J. Thornton, R. Newbury, M. Pepper, H. Ahmed, J. E. F. Frost, D. G. Hasko, D. C. Peacock, D. A. Ritchie, G. A. C. Jones: J. Phys. C **21**, L209 (1988)
9. S. Datta: *Electron Transport in Mesoscopic Systems* (Cambridge University Press, New York 1995)
10. P. Streda: J. Phys. Cond. Matt. **1**, 1025 (1989)
11. I. Langmuir: Phys. Rev. **21**, 419 (1923)
12. G. D. Mahan: J. Appl. Phys. **76**, 4362 (1994)
13. A. G. Mal'shukov, Zhongshui Ma, V. B. Antonyuk, K. A. Chao: Solid State Commun. **119**, 563 (2001)
14. G. D. Mahan, L. M. Woods: Phys. Rev. Lett. **80**, 4016 (1998)
15. R. Venkatasubramanian, E. Siivola, T. Colpitts, B. O'Quinn: Nature (London) **413**, 597 (2001)
16. V. B. Antonyuk, A. G. Mal'shukov, Zhongshui Ma, K. A. Chao: Appl. Phys. Lett. **79**, 3791 (2001)
17. D. A. Broido, T. L. Reinecke: Phys. Rev. B **64**, 045324 (2001)
18. L. D. Hicks, M. S. Dresselhaus: Phys. Rev. B **47**, 16631 (1998)
19. Yu-Ming Lin, M. S. Dresselhaus: Phys. Rev. B **68**, 075304 (2003)
20. N. Mateeva, H. Niculescu, J. Schlenoff, L. R. Testardi: J. Appl. Phys. **83**, 3117 (1998)
21. Harukazu Yoshino, Keizo Murata: Phys. Rev. B **67**, 035111 (2003)
22. L. Grigorian, G. U. Sumanasekera, A. L. Loper, S. L. Fang, J. L. Allen, P. E. Eklund: Phys. Rev. B **60**, R11309 (1999)
23. P. Kim, L. Shi, A. Majumdar, P. L. McEuen: Phys. Rev. Lett. **87**, 215502 (2001)
24. Ryoji Funahashi, Ichiro Matsubara: Appl. Phys. Lett. **79**, 362 (2001)
25. F. Chen, K. L. Stokes: Appl. Phys. Lett. **81**, 2379 (2002)
26. G. Peleckis, T. Motohashi, M. Karppinen, H. Yamauchi: Appl. Phys. Lett. **83**, 5418 (2003)
27. Ryoji Funahashi, Masahiro Shikano: Appl. Phys. Lett. **81**, 1459 (2002)
28. Joshua P. Small, Kerstin M. Perez, Philip Kim: Phys. Rev. Lett. **91**, 256801 (2003)
29. G. Chen, M. Neagu: Appl. Phys. Lett. **71**, 2761 (1997)

30. G. Chen: Phys. Rev. B **57**, 14958 (1998)
31. P. L. Kapitza: J. Phys. (Moscow) **4**, 181 (1941)
32. I. M. Khalatnikov: Sov. Phys. JETP **22**, 687 (1952)
33. E. T. Swartz, R. O. Pohl: Rev. Mod. Phys. **61**, 605 (1989)
34. P. K. Schelling, S. R. Phillpot, P. Keblinski: Appl. Phys. Lett. **80**, 2484 (2002)
35. D. G. Cahill, W. K. Ford, K. E. Goodson, G. D. Mahan, A. Majumdar, H. J. Maris, R. Merlin, S. R. Phillpot, J. Appl. Phys. **93**, 793 (2003)
36. T. Borca-Tasciuc, D. Achimov, W. L. Liu, G. Chen, H.-W. Ren, C.-H. Chen, K. L. Wang, M. Goorsky, T. Redetic, R. Gronsky, T. Koga, M. S. Dresselhaus: Superlattices Microstruct. **28**, 199 (2000)
37. W. S. Capinski, H. J. Maris, T. Ruf, M. Cardona, K. Ploog, D. S. Katzer: Phys. Rev. B **59**, 8105 (1999)
38. V. Narayanamurti, H. L. Stromer, M. A. Chin, A. C. Gossard, W. Wiegmann: Phys. Rev. Lett. **43**, 2012 (1979)
39. J. L. Yarnell, J. L. Warren, R. G. Wenzel, P. J. Dena: *Neutron Inelastic Scattering*, vol 1 (IAEA, Vienna 1968) p 301
40. B. Zhu, K. A. Chao: Phys. Rev. B **36**, 4906 (1987)
41. K. Schwab, E. A. Henriksen, J. M. Worlock, M. L. Roukes: Nature **404**, 974 (2000)
42. Vadim B. Antonyuk, Magnus Larsson, A. G. Mal'shukov, K. A. Chao: Semicond. Sci. Technol. **20**, 347 (2005)
43. Magnus Larsson, Vadim B. Antonyuk, A. G. Mal'shukov, K. A. Chao: Phys. Rev. B **68**, 233302 (2003)
44. Hiroshi Mizuta, Tomonori Tanoue: *The Physics and Applications of Resonant Tunnelling Diodes* (Cambridge University Press, USA 1995)
45. T. E. Humphrey, R. Newbury, R. P. Taylor, H. Linke: Phys. Rev. Lett. **89**, 116801 (2002)
46. K. Fujimaru, R. Sasajima, H. Matsumura: J. Appl. Phys. **85**, 6912 (1999)
47. Y. Hishinuma, T. H. Geballe, B. Y. Moyzhes, T. W. Kenny: Appl. Phys. Lett. **78**, 2572 (2001)
48. K. A. Chao, Magnus Larsson, A. G. Mal'shukov: Appl. Phys. Lett. **87**, 022103 (2005)
49. Christopher LaBounty: Heterostructure integrated thermionic cooling of optoelectronic devices. Ph.D. Dissertation, Department of Electrical and Computer Engineering, University of California, Santa Barbara (2001)
50. G. D. Mahan, J. O. Sofo, M. Bartkowiak: J. Appl. Phys. **83**, 4683 (1999)
51. A. G. Mal'shukov, K. A. Chao: Phys. Rev. Lett. **86**, 5570 (2001)
52. Peng Han, Kui-juan Jin, Yueliang Zhou, Xu Wang, Zhongshui Ma, Shang-Fen Ren, A. G. Mal'shukov, K. A. Chao: J. Appl. Phys. **99**, 074504 (2006)
53. Magnus Larsson, A. G. Mal'shukov, K. A. Chao: Phys. Rev. B **70**, 235339 (2004)
54. K. Schwab, E. A. Henriksen, J. M. Worlock, M. L. Roukes: Nature **404**, 974 (2000)
55. Z. Chen, R. S. Sorbello: Phys. Rev. B **47**, 13527 (1993)
56. K. S. Ralls, D. C. Ralph, R. A. Buhrman: Phys. Rev. B **40**, 11561 (1989)
57. F. Dietzel, W. Dietsche: Phys. Rev. B **48**, 4713 (1993)
58. L. W. Molenkamp, H. van Houten, C. W. J. Beenakker, C. T. Foxon, R. Eppenga: Phys. Rev. Lett. **65**, 1052 (1990)
59. L. W. Molenkamp, Th. Gravier, H. van Houten, O. J. A. Buijk, C. T. Foxon, M. A. A. Mabesoone: Phys. Rev. Lett. **68**, 3765 (1992)

Coherence and Interactions in Diffusive Systems

Gilles Montambaux

Laboratoire de Physique des Solides, associé au CNRS, Université Paris–Sud, 91405 Orsay, France. montambaux@lps.u-psud.fr

1 Introduction : Phase Coherence and Disorder

Although this volume mainly concerns nanoscopic systems, this article is devoted to an intermediate range, between the nanoscopic and macroscopic scales, the so-called mesoscopic regime [1]. In this regime, the system to be considered may be large compared to the mean free path of the electrons. Disorder plays then a very important role and, in the so-called diffusive regime, the interplay between disorder and quantum interference effects is crucial. This is the main subject of this review article. Here, electronic interactions will be treated as a perturbation, in contrast with the topics discussed in the other chapters where the electronic correlations may play the most important role. I will try to present some personal points of view in order to describe these well-known signatures of phase coherence like weak localization or universal conductance fluctuations. The goal here is to avoid technicalities as most as possible. The last part concerns the effect of electron-electron interactions.

To describe interference effects in electronics, it is useful to compare with simple facts known in optics. The simplest experiment with light is the two-slit Young experiment and phase coherent effects considered in electronics are nothing but some more sophisticated versions of the Young experiment. This two-slit experiment can be also performed with electrons in vacuum [2] but here we shall consider metallic wires, that are complex disordered media.

In vacuum, an electron beam is split in two parts and the intensity is measured on a screen. The topological equivalent in a metal consists in a loop pierced by a magnetic flux and we measure the current resulting from the interferences between the two paths, see Fig. 1. In optics the way to probe the interference pattern on the screen is to change the optical path between the two trajectories, by changing the nature of the medium, that is its optical index. For electrons, the charge is coupled to the vector potential $\boldsymbol{A}$ and these interference pattern can be modified with a magnetic field.

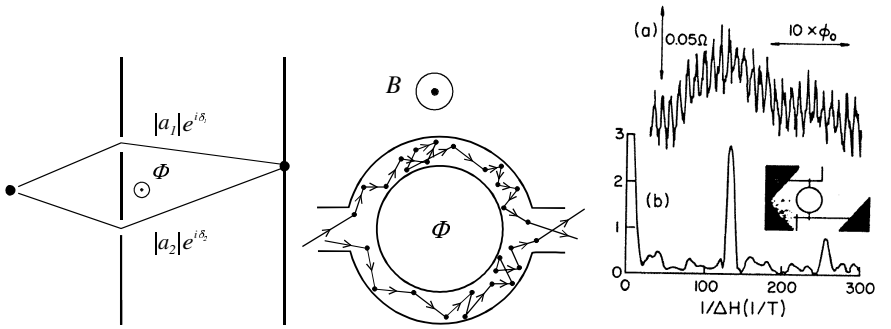

Fig. 1. Left : schematic representation of the Aharonov-Bohm effect. A flux tube of flux ϕ is placed behind the two slits. Center : Schematic equivalent of the Aharonov-Bohm experiment in a metallic ring. Right : Magnetoresistance oscillations of a Au ring and its Fourier transform [3]

If we try to transpose what is known from optics to electronics, we measure a current intensity which is proportional to the probability for the electrons to traverse the loop. To calculate this probability in quantum mechanics, we have to add the contributions of two quantum amplitudes corresponding to the two sides of the loop, and the current (a probability) is proportional to the square of this quantum amplitude. Each quantum amplitude ψ_i has a phase φ_i :

$$\psi_1 = \psi\, \mathrm{e}^{\mathrm{i}\varphi_1} \qquad , \qquad \psi_2 = \psi\, \mathrm{e}^{\mathrm{i}\varphi_2} \; . \tag{1}$$

We have to sum the amplitudes and take the modulus square. For the current, we get a classical term plus an interference term :

$$I \propto |\psi_1 + \psi_2|^2 = 2\psi^2 \left[1 + \cos(\varphi_1 - \varphi_2)\right] . \tag{2}$$

Classically, the current should be the sum of the two currents, this corresponds to Ohm's law, but quantum mechanically there is some additional phase effect. And if we apply a magnetic field (this is the well-known Aharonov-Bohm effect [5]), there is an additional phase along each of the two trajectories

$$\delta\varphi_1 = \frac{e}{\hbar}\int_1 \boldsymbol{A}\cdot \mathrm{d}\boldsymbol{l} \qquad , \qquad \delta\varphi_2 = \frac{e}{\hbar}\int_2 \boldsymbol{A}\cdot \mathrm{d}\boldsymbol{l} \; , \tag{3}$$

so that the phase difference is modulated by the circulation of the vector potential along the closed loop formed by the two trajectories,

$$\Delta\varphi = \delta\varphi_1 - \delta\varphi_2 = \frac{e}{\hbar}\oint \boldsymbol{A}\cdot \mathrm{d}\boldsymbol{l} = 2\pi\frac{\phi}{\phi_0} \; , \tag{4}$$

and is proportional to the magnetic flux ϕ through the loop. Here $\phi_0 = h/e$ is the flux quantum. This tells us that the wave functions, the energy levels, thermodynamic and transport properties must be function of this flux with

period ϕ_0. Of course this does not tell us anything about the amplitude of these oscillations. Unlike the case of the Young experiment with light or for electrons in vacuum, the problem here is much more complicated due to disorder. The question now is : *do interference effect survive in the presence of disorder?* Do oscillations persist?

Indeed, some oscillations remain in the presence of disorder. The pioneer experiment founding the field of mesoscopic physics was performed by Webb et al. in 1985 [3]. They measured the resistance of a ring of micronic size, in the presence of a magnetic field (Fig. 1). They found that the resistance oscillates with the field, proving the existence of an interference effect, even in the presence of disorder. Two interesting features must be noted : firstly, the period of the oscillations is the flux quantum ϕ_0 as expected. Secondly, the typical amplitude of the oscillations of the conductance is $\Delta G = \Delta R/R^2 \simeq e^2/h$.

The necessary condition for these oscillations to exist is that phase coherence is preserved, a condition which is obeyed in vacuum. But in a metal, this phase coherence is broken because electrons interact with other degrees of freedom (e.g., phonons, other electrons, magnetic impurities). Because of this coupling, each electron can lose the memory of its phase. This happens on a typical length, called the *phase coherence length*, denoted by L_ϕ, which depends on the coupling to these degrees of freedom. Typically at 1 K, it is of order of 1 µ. This is the *mesoscopic range*. Of course, such oscillations do not exist for a macroscopic ring, because phase coherence is lost at the macroscopic level.

There is another important scale, due to static disorder, over which electrons experience collisions with the impurities. This scale l_e is named the *elastic mean free path*. It is much smaller than the phase coherence length. Since elastic collisions do not break phase coherence, electrons may experience many collisions without losing the memory of their phase. Each electron keeps the memory of its phase typically on a scale $L_\phi \gg l_e$. The physics we are going to discuss corresponds to length scales which are much larger than the elastic mean free path l_e, but smaller than L_ϕ, so that the sample can be considered as quantum mechanically coherent. This regime is called the *mesoscopic regime*.

We want to construct a theory for these oscillations. How to explain their amplitude? How can they survive disorder? Another important key experiment was performed by Sharvin and Sharvin on a cylinder [4], before the Webb's experiment on a single ring. In this case, there are also oscillations, but instead of being ϕ_0-periodic, they have a period $\phi_0/2$. The fundamental ϕ_0 disappears and the experiment reveals the second harmonics of these oscillations.

There is a simple way to understand this frequency doubling : assume that the cylinder can be viewed as a pile of independent rings, it realizes an average of the oscillations of several independent rings (similar effect for a network of rings [6]). For a given ring, the oscillations have a given phase which

depends on the disorder configuration (This phase is 0 or π for a two-terminal geometry [7]). For another ring, the oscillations have a different phase. When averaging over several rings (as is done in a cylinder), because of this random phase, the oscillations vanish.

So, disorder is destructive for quantum interferences. However this simple argument would tell us that no oscillations should survive disorder averaging. This is not the case since the cylinder experiment shows oscillations with period $\phi_0/2$. In average, there are still oscillations, but with period $h/2e$. This simple and very important fact tells us that *some contributions survive disorder averaging*. So the question is : how is it possible that some robust contribution survive disorder averaging ?

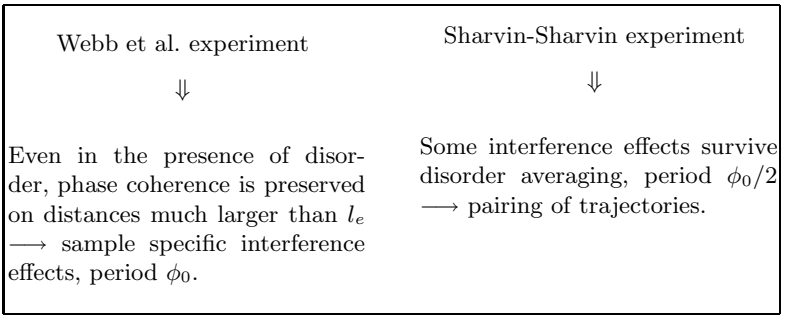

Fig. 2. Conclusions that can be drawn from the two pioneering experiments in mesoscopic physics

2 Important Scales

We shall consider weakly disordered metals, such that the average distance between two collision events is much larger than the Fermi wave length : $l_e \gg \lambda_F$. This condition allows for a semiclassical description of electronic waves. Moreover we assume that the typical size L of the system is much larger than the mean free path l_e so that the electronic motion in the sample is diffusive : electrons collide elastically many times while traversing the system. Finally we assume that the system is completely phase coherent, that is $L \ll L_\phi$. To summarize, the system is weakly disordered, diffusive and mesoscopic :

$$\lambda_F \ll l_e \ll L \ll L_\phi \ . \tag{5}$$

We shall see that phase coherent effects may not disappear but are simply reduced in the macroscopic limit $L_\phi < L$, so that we shall also consider the case :

$$\lambda_{\rm F} \ll l_e \ll L_\phi \ll L \ . \tag{6}$$

In strong disorder, when the mean free path becomes of order of the Fermi wave length $k_{\rm F} l_e \simeq 1$, interference effects are strong and lead to localization of the electronic waves. This is the domain where the electronic states are *exponentially localized* in space, with Anderson localization from extended to localized waves. This topic will not be covered here. In the opposite regime of very weak disorder, the mean free path becomes so large (or the system is so small) that the mean free path becomes larger than the system size. Collisions with impurities are rare, and occur mainly on the boundaries of the system. This the so-called *ballistic regime*, where the physics is mainly driven by the structure of the boundaries, i.e., the shape of the system. For most shapes, the trajectories are chaotic. A common method to describe this regime is the so-called Random Matrix Theory also used in other fields of physics like nuclear physics. There are quite interesting common features between some aspects of transport in chaotic dots and nuclear physics. Let us also emphasize that the Random Matrix Theory of scattering or transmission matrices can also be used to describe diffusive systems [8].

Since the typical size L of the system is much larger than the mean free path, the electronic motion is diffusive. The average distance between collision events, the mean free path is related to the collision time τ_e, $l_e = v_{\rm F}\tau_e$, $v_{\rm F}$ being the Fermi velocity, since the electronic motion between two collisions is ballistic. For times much longer than the collision time τ_e, the motion is diffusive and the typical distance an electron can reach after a time t scales like

$$r^2 = Dt \ , \tag{7}$$

where D is the diffusion coefficient given by $D = v_{\rm F} l_e/d$, d being the space dimensionality. This relation tells us that for a finite system of size L, a very important scale appears : this is the time for which an electron typically sees the boundaries of the system. It is called the traversal time, or *Thouless time*. It is the time for an electron to "realize" that the system is finite. It is given by

$$\tau_D = \frac{L^2}{D} \ . \tag{8}$$

To this characteristic time, is associated a characteristic energy, the *Thouless energy* E_c :

$$E_c = \frac{\hbar}{\tau_D} = \frac{\hbar D}{L^2} \ . \tag{9}$$

This energy scale plays a major role in the description of thermodynamic and transport properties of mesoscopic diffusive systems. For time scales smaller than τ_D, the electron propagates like in infinite space. The diffusive motion depends on the space dimensionality of the system. On the other hand, in the long time scale, the electronic motion explores the entire system, this is the so-called ergodic regime.

Using (7), we can associate to the phase coherence length L_ϕ a characteristic time, the phase coherence time τ_ϕ :

$$\tau_\phi = \frac{L_\phi^2}{D} \,. \tag{10}$$

This is the time during which an electron keeps the memory of its phase.

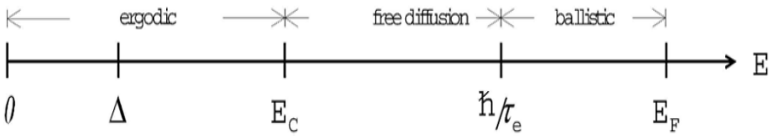

Fig. 3. Characteristic energy scales defining the different regimes studied in coherent multiple scattering. Explanations are given in the text

Figure 3 presents a scale of characteristic energies (or inverse characteristic times). At short time scales, the motion is ballistic. For times larger than τ_e, the motion is diffusive in free space. Above τ_D, the motion is bounded, this is the ergodic regime. Then τ_ϕ separates the mesoscopic regime and the classical regime. If $\tau_\phi > \tau_D$ ($L_\phi > L$), the system is mesoscopic.

In the diffusive regime, the space dimensionality d plays an important role. Moreover, the one-dimensional case is somehow special. In strictly one dimension, it is known that there is no diffusive regime since, in the presence of disorder, all states are exponentially localized. So we shall not consider this case but rather *quasi*-one-dimensional systems, with a transverse width a, so that the real motion is three-dimensional (the proper Schrödinger equation to be solved would be three-dimensional with a quantization of the transverse component of the wave vector), but the diffusion is one-dimensional. Instead of having one transverse propagation channel, there are many transverse channels (the transverse size is much larger than λ_F). At short time (smaller than a "transverse" Thouless time a^2/D), diffusion is three-dimensional, but at larger times, there is a one-dimensional propagation of the diffusion cloud. When considering the transport through a wire ($L \gg a$), we shall be interested in time scales necessary to traverse the wire, that is times of order of τ_D, which is much larger than the transverse time, so that we can consider that at this time scale the diffusion is one-dimensional.

3 Classical Probability and Diffusion Equation

The aim of this article is to propose a qualitative description of physical phenomena, trying to avoid sophisticated tools and keeping in mind that we are essentially concerned by the calculation of disordered averaged quantities.

An average quantity like the conductance basically measures the probability for electrons to cross the system. What is the nature of this probability? Let us first spend some time to describe the probability $P(\boldsymbol{r}, \boldsymbol{r}')$ which describes the propagation of a particle from a point $\boldsymbol{r}$ to a point $\boldsymbol{r}'$. In quantum mechanics, this propagation is described by a probability *amplitude*. This amplitude is called a Green's function $G(\boldsymbol{r}, \boldsymbol{r}')$. We do not aim to develop the theory of Green's functions. For our purpose here, it is sufficient to note that there are many possible scattering trajectories from $\boldsymbol{r}$ to $\boldsymbol{r}'$. Thus a Green's function has the following structure : it is the sum of all the probability amplitudes corresponding to various multiple scattering trajectories from $\boldsymbol{r}$ to $\boldsymbol{r}'$, each trajectory being characterized by an amplitude and a phase proportional to its action, that is its length [9] :

$$G(\boldsymbol{r}, \boldsymbol{r}') = \sum_j A_j(\boldsymbol{r}, \boldsymbol{r}') \ . \tag{11}$$

Now, we want to know the probability to find a particle at point $\boldsymbol{r}'$ if it has been injected at point $\boldsymbol{r}$. The probability to go from $\boldsymbol{r}$ to $\boldsymbol{r}'$ is the modulus square of the amplitude. From (11), we see that this probability is the sum of amplitude squared terms, plus interference terms which pair different trajectories j and j' :

$$|G(\boldsymbol{r}, \boldsymbol{r}')|^2 = \sum_{j,j'} A_j(\boldsymbol{r}, \boldsymbol{r}') A_{j'}^*(\boldsymbol{r}, \boldsymbol{r}') = \sum_j |A_j(\boldsymbol{r}, \boldsymbol{r}')|^2 + \sum_{j' \neq j} A_j(\boldsymbol{r}, \boldsymbol{r}') A_{j'}^*(\boldsymbol{r}, \boldsymbol{r}') \tag{12}$$

which an obvious generalization of (2) for the two-slit configuration. Since we know that, in quantum mechanics, one must add amplitudes instead of intensities, the interference term (the second term in (12)) cannot be *a priori* neglected. This second term describes interferences between *different* trajectories j and j'. Each contribution in this sum has a random phase which depends on the detail of the impurity configuration. Since the phases are uncorrelated, at the first level of approximation, we may expect that the contribution of the interference term cancels upon disorder averaging. So quantum effects seem not to be so important because of the vanishing of this contribution. We shall see however that this is not exactly the case. Within this approximation, the second term cancels and the probability is essentially given by the sum of intensities :

$$\overline{|G(\boldsymbol{r}, \boldsymbol{r}')|^2} = \overline{\sum_j |A_j(\boldsymbol{r}, \boldsymbol{r}')|^2} \ . \tag{13}$$

We see that the phases have disappeared. So the remaining term is completely classical. Indeed, let us assume that some event changes the phase of the amplitude A_j. The complex amplitude A_j^* gets the opposite phase, leaving the probability unchanged.

To have a simple picture of (13), we represent on Fig. 4 a quantum amplitude as a line (it is rather a sort of Brownian trajectory). Its complex

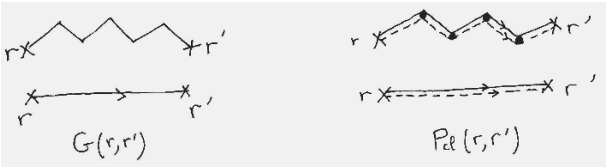

Fig. 4. Schematic representations of a Green's function $G(\boldsymbol{r},\boldsymbol{r}')$ and of the classical probability $P_{\rm cl}(\boldsymbol{r},\boldsymbol{r}') \propto \overline{\sum_j |A_j(\boldsymbol{r},\boldsymbol{r}')|^2}$. The upper diagrams exhibit a few collision events, which are not represented on the lower diagrams

conjugate is represented as a dashed line. The first term in (13) corresponds to the pairing of a trajectory with its complex conjugate, and we see immediately why the phase disappears. The quantity $\sum_j \overline{|A_j(\boldsymbol{r},\boldsymbol{r}')|^2}$ resembles the classical probability. We call it a "*Diffuson*". To be more precise, but without any proof, we define the probability $P(\boldsymbol{r},\boldsymbol{r}',\omega)$ as

$$P(\boldsymbol{r},\boldsymbol{r}',\omega) = \frac{1}{2\pi\rho_0}\overline{G_\epsilon(\boldsymbol{r},\boldsymbol{r}')G^*_{\epsilon-\omega}(\boldsymbol{r}',\boldsymbol{r})}\ . \tag{14}$$

The Green's function and its complex conjugate are taken at different energies (or frequencies) ϵ and $\epsilon-\omega$. One can check that this probability is correctly normalized, that is $\int_0^\infty P(\boldsymbol{r},\boldsymbol{r}',t)\mathrm{d}\boldsymbol{r}' = 1$, where $P(\boldsymbol{r},\boldsymbol{r}',t)$ is the Fourier transform of $P(\boldsymbol{r},\boldsymbol{r}',\omega)$. Starting from the Schrödinger equation in a random potential and after disorder averaging, it is possible to show that in the limit $k_{\rm F}l_e \gg 1$, the probability $P(\boldsymbol{r},\boldsymbol{r}',\omega)$ defined by (14) reduces to the Diffuson $P_{\rm cl}(\boldsymbol{r},\boldsymbol{r}') \propto \overline{\sum_j |A_j(\boldsymbol{r},\boldsymbol{r}')|^2}$. For slow spatial variations, $P_{\rm cl}(\boldsymbol{r},\boldsymbol{r}',\omega)$ is the solution of a classical diffusion equation :

$$(-\mathrm{i}\omega - D\Delta)\, P_{\rm cl}(\boldsymbol{r},\boldsymbol{r}',\omega) = \delta(\boldsymbol{r}-\boldsymbol{r}')\ , \tag{15}$$

where D is the diffusion coefficient. Doing this, we have only considered classical contributions to the average (14). We shall study later the corrections to this classical probability.

Among the solutions of this diffusion equation, one very important is the return probability which enters in many physical quantities. It is the probability $P(\boldsymbol{r},\boldsymbol{r},t)$ for an electron to return to its original position after time t. I will also consider the space integrated return probability :

$$P(t) = \int P(\boldsymbol{r},\boldsymbol{r},t)\, d\boldsymbol{r}\ . \tag{16}$$

In free space, the solutions of (15) are simply obtained from the Fourier transform

$$\left(\frac{\partial}{\partial t} + Dq^2\right) P(\boldsymbol{q},t) = \delta(t) \tag{17}$$

whose solution $P(\boldsymbol{q},t)$ is simply

$$P(\boldsymbol{q},t) = \mathrm{e}^{-Dq^2 t} \ . \tag{18}$$

Fourier transforming back, we find easily

$$P(\boldsymbol{r},\boldsymbol{r}',t) = \frac{1}{(4\pi Dt)^{d/2}} \mathrm{e}^{-|\boldsymbol{r}-\boldsymbol{r}'|^2/4Dt} \ , \tag{19}$$

so that the return probability is given by

$$P(\boldsymbol{r},\boldsymbol{r},t) = \frac{1}{(4\pi Dt)^{d/2}} \qquad \text{and} \qquad P(t) = \frac{\Omega}{(4\pi Dt)^{d/2}} \ , \tag{20}$$

where Ω is the volume of the system. The dependence on the dimensionality d of the return probability is crucial since it will explain why dimensionality plays a so important role in mesoscopic physics of diffusive systems.

4 Conductance

Now I wish to come to very simple and qualitative considerations about the conductance of a disordered system, which will be useful in the rest of this chapter.

4.1 Classical Conductance as the Ratio of Two Volumes

Consider the conductance G. Since it has the dimensions of e^2/h, we can introduce a dimensionless conductance g as

$$g = G/(se^2/h) \ . \tag{21}$$

Since this quantity is dimensionless, it may be written as the ratio of two physical quantities. For example, by simple manipulations, it can be written as the ratio of two energies : $g \propto E_c/\Delta$, the Thouless energy E_c and the average level spacing Δ. Here I would like to write it as the *ratio of two volumes*. Let us start with the classical Drude conductivity σ_0. From Einstein relation, it is given by

$$\sigma_0 = se^2 D\rho_0 \ , \tag{22}$$

where D is the diffusion coefficient and ρ_0 is the density of states at the Fermi level for one spin direction. The factor $s = 2$ accounts for spin degeneracy. By Ohm's law, the conductance G for a three-dimensional system is given by $G = \sigma_0 S/L$, S being the section and L the length of the sample. More generally for a hypercube of typical size L in d dimensions, it is given by $G = \sigma_0 L^{d-2}$. Introducing the Thouless time τ_D defined by (8), let us rewrite the conductance as

$$G = se^2 \rho_0 L^d/\tau_D \ . \tag{23}$$

Density of states at the Fermi level ρ_0 can be written as $\rho_0 = dA_d/2\pi\lambda_F^{d-1}\hbar v_F$, where λ_F is the Fermi wavelength, v_F is the Fermi velocity, and A_d is the volume of the unit sphere ($A_3 = 4\pi/3$, $A_2 = \pi$, $A_1 = 2$). An easy way to recover immediately this result is to say that the total number of states is $(k_F L)^d$, so that by derivation with respect to the energy, we have necessarily $\rho_0 \propto k_F^d/\epsilon_F \simeq k_F^{d-1}/\hbar v_F \propto 1/\hbar v_F \lambda_F^{d-1}$. As a result, the dimensionless conductance g can be written as

$$g = dA_d \, \frac{\Omega}{\lambda_F^{d-1} v_F \tau_D} \, , \tag{24}$$

where $\Omega = L^d$ is the volume of the system. The dimensionless conductance quantity appears as the *ratio of two volumes*, the volume Ω of the system and the volume of a tube of length $v_F\tau_D$ and of section λ_F^{d-1}. We shall see later that this formulation will be quite useful to measure the importance of interference effects.

4.2 Conductance and Transmission

Our starting point to describe electric transport is the Landauer formalism. Even staying at a very qualitative level, this formalism is quite natural since it expresses the *conductance as a transmission coefficient* through the disordered sample.

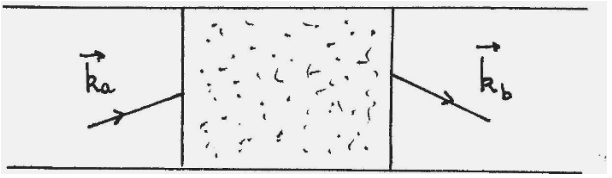

Fig. 5. In the Landauer formalism, the conductance is related to the transmission coefficient between different incoming and outgoing channels

Consider a disordered conductor of length L and section $S = W^{d-1}$. It is connected to perfect conductors (Fig. 5) which can be considered as wave guides where free electronic waves propagate. In this geometry, the transverse wave vectors of the eigenmodes (also called channels) are quantized by transverse boundary conditions. One can define a transmission coefficient T_{ab} from an incoming channel a (ingoing wave vector $\boldsymbol{k}_a$) to an outgoing channel b (outgoing wave vector $\boldsymbol{k}_b$). The Landauer formula reads :

$$G = s\frac{e^2}{h}\sum_{a,b} T_{ab} \, . \tag{25}$$

To calculate the number of transverse channels, one considers that electrons are injected at the Fermi energy, i.e., such that $|\boldsymbol{k}_a| = |\boldsymbol{k}_b| = k_{\rm F}$. The transverse component is quantized in units of $2\pi/W$. This quantization imposes the number of channels. In $d = 2$ and $d = 3$, their number is

$$M_2 = \frac{2\pi k_{\rm F}}{2\pi/W} = k_{\rm F} W \qquad , \qquad M_3 = \frac{\pi k_{\rm F}^2}{4\pi^2/W^2} = \frac{k_{\rm F}^2 S}{4\pi} \; . \tag{26}$$

Let us consider now the structure of the transmission coefficient T_{ab}. It is the square of an amplitude and it has, with minor differences, the same structure as the probability $P(\boldsymbol{r}, \boldsymbol{r}', \omega)$. The main difference is the following : instead of injecting a particle at a point $\boldsymbol{r}$ inside the sample, we inject a plane wave $\boldsymbol{k}_a$ from outside the sample. In particular, the boundary conditions have to be treated properly. But, without entering into details, we may easily understand that, after disorder averaging, the average transmission coefficient and consequently the conductance can be related to the probability to cross the sample. More precisely for a $3d$ sample, one can show that the dimensionless conductance is [1]

$$g = \frac{4}{9} M v_{\rm F} P(0, L) \tag{27}$$

where $P(0, L)$ is the solution of the diffusion equation (15) with appropriate boundary conditions. It is given by $P(0, L) = l_e^2/DL$ so that

$$g = \frac{4}{3} M \frac{l_e}{L} \; . \tag{28}$$

To obtain these results quantitatively, there are some technicalities that we do not describe here [1]. What should be remembered is the message of Fig. 6 : the conductance is proportional to the classical probability to cross the sample. This statement is sufficient to understand how coherence effects appear.

Fig. 6. The conductance is proportional to the classical probability to transmit channel a to channel b (summed over channels). The object which represents this probability is the sum of contributions of paired trajectories as introduced in Fig. 4. We call it a "*Diffuson*"

5 Quantum Corrections and Quantum Crossings

The Diffuson is a classical object. It does not depend on the phases of the complex amplitudes. In the diffusive regime, it is solution of a diffusion equation. However, we may have to check whether we have not left aside additional

effects when throwing out all the interference terms in relation (12). It turns out that some of these terms have quite interesting consequences.

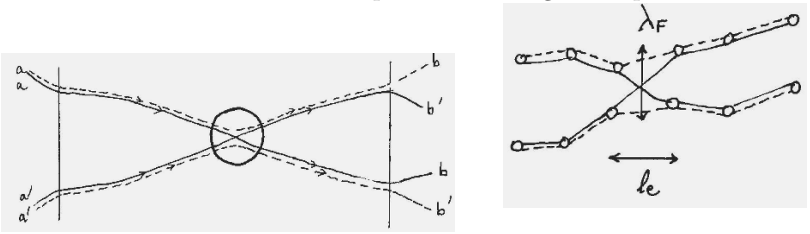

Fig. 7. (a) Crossing of two Diffusons. **(b)** Detail : the volume of the intersection region is proportional to $\lambda_F^{d-1} l_e$

Indeed, quantum effects can appear when two Diffusons cross, or when a Diffuson crosses with itself. The notion of *quantum crossing* is extremely important because it is the source of quantum effects. The Diffuson being a classical object, coherence effects can only appear because of these quantum crossings. They are at the origin of the weak localization correction and of universal conductance fluctuations. Let us try to get some intuition about these crossing events.

Figure 7a shows that a crossing mixes four complex amplitudes which belong to two incoming Diffusons and pair them differently. The two emerging Diffusons are built with amplitudes A_j and $A_{j'}$ coming respectively from each of the incoming Diffusons. They have the same phase since they follow the same path. The quantum crossing, also often called a Hikami box in a more technical context, is thus an object whose role is to permute the quantum amplitudes. It is necessarily short-ranged, because trajectories have to be as close as possible to each other to avoid dephasing (Fig. 7b). Since it appears between two successive collisions on impurities, and since the phase mismatch between trajectories has to be smaller that 2π, one sees that the volume of this object is of order $\lambda_F^{d-1} l_e$.

It is important to evaluate the probability of occurrence of such quantum crossings because it will be shown to be a measure of the importance of quantum effects. Since the volume of a quantum crossing is of order $\lambda_F^{d-1} l_e$, a Diffuson propagating during a time t can be seen as an effective object of length $\mathcal{L} = v_F t$ and of cross-section λ_F^{d-1}. Thus, it has a finite volume $v_F \lambda_F^{d-1} t$. The probability $\mathrm{d}p_\times(t)$ of crossing of two Diffusons after a time $\mathrm{d}t$ in a volume $\Omega = L^d$ is thus proportional to the ratio between the volume of a Diffuson and the volume of the system :

$$\mathrm{d}p_\times(t) = \frac{\lambda_F^{d-1} v_F \mathrm{d}t}{\Omega} \propto \frac{1}{g} \frac{\mathrm{d}t}{\tau_D} \tag{29}$$

where we have used (24) to introduce the dimensionless conductance g. Consider now an open system coupled to reservoirs. The time needed to travel

throughout the sample is the Thouless time $\tau_D = L^2/D$. The probability of quantum crossing during this time is given by

$$p_\times(\tau_D) = \int_0^{\tau_D} \mathrm{d}p_\times(t) = \frac{\lambda_\mathrm{F}^{d-1} v_\mathrm{F} \tau_D}{\Omega} \simeq \frac{1}{g} \,. \tag{30}$$

This is exactly the inverse conductance ! I believe that this is the most important message to understand phase coherence effects in disordered systems. All these effects can be simply understood in terms of quantum crossings and the probability of such crossings which measures the importance of quantum mechanical effects is simply given by the inverse of the dimensionless conductance g.

In a good metal, the conductance g is large, the volume of the tube is small, electrons do not spend much time in the system and quantum effects are very small. In the opposite limit, when g becomes of order 1, the volume of this tube is of the order of the volume of the system. It is so big that the probability of quantum crossing is of order 1. This corresponds to the Anderson regime where electronic waves are localized by strong disorder. Here we shall not consider this regime but only the small disorder regime where quantum effects remain small. The approach to Anderson localization can be viewed as the proliferation of quantum crossings.

As a first qualitative but important conclusion of our discussion, we see that classical transport is described by a conductance $G = sge^2/h$. Quantum corrections are smaller than classical terms by a ratio $1/g$. This tells us immediately that the quantum corrections are of order G/g, that is e^2/h !

6 Weak Localization

6.1 Weak Localization and Quantum Crossings

We have seen that the classical probability and the conductance can be expressed as a sum of contributions of pairs of complex conjugated trajectories. Since trajectories can have quantum crossings, they can form closed loops (Fig. 8). It turns out that in such a loop (whose contribution is not included in (13)), the trajectories are *time-reversed*. One trajectory j and its time-reversed j^T go in opposite directions. If there is time-reversal symmetry, they have the same action and thus they have exactly the same phase. This phase can be quite complicated because it depends on the disorder configuration but it is the same for both trajectories. So the contribution of these loops does not cancel on average. If the end points are far away like in Fig. 8, the contribution of these new trajectories is small, of order $1/g$, but it leads to an experimentally observable effect : the weak localization correction to the conductance. This is a phase coherent effect because only trajectories of size smaller than the phase coherence length L_ϕ contribute to this additional contribution.

At this point, I want to stress that many presentations of weak localization correction emphasize the existence of a loop of opposite trajectories, but do not insist on the structure of the quantum crossing. This is rather quite important, because this is what explains the amplitude $1/g$ of the correction. This is where phase coherence is lost.

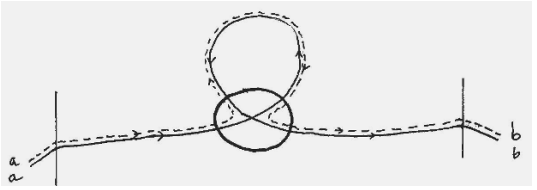

Fig. 8. Trajectory with a quantum crossing and a loop. In the loop, the two propagations are time-reversed

Using the same type of argument as in the previous section, let us evaluate the probability to have a loop for a trajectory which travels through the sample. Since there is a quantum crossing, the probability is small, of order $1/g$. Moreover, it depends on the distribution of loops in the disordered system. Let us call it $P_{\rm int}(t)$. For the probability of traversing the sample with a loop, we have :

$$p_o(\tau_D) = \int_0^{\tau_D} P_{\rm int}(t)\ {\rm d}p_\times(t) = \frac{1}{g}\int_0^{\tau_D} P_{\rm int}(t)\frac{{\rm d}t}{\tau_D}\ . \tag{31}$$

We have also to remember that, because of decoherence in the loops, only those with time t smaller than τ_ϕ contribute. The resulting probability to have trajectories with loops of time smaller than τ_ϕ is :

$$p_o(\tau_\phi) = \int_0^{\min(\tau_D,\tau_\phi)} P_{\rm int}(t)\ {\rm d}p_\times(t) = \frac{1}{g}\int_0^{\min(\tau_D,\tau_\phi)} P_{\rm int}(t)\frac{{\rm d}t}{\tau_D} \tag{32}$$

where $P_{\rm int}(t)$ is the probability to have loops of time t. This leads to a relative correction to the conductivity (or to the conductance) given by

$$\frac{\Delta\sigma}{\sigma_0} = \frac{\Delta G}{G_0} = -p_o(\tau_\phi)\ . \tag{33}$$

The sign of the correction is negative because the trajectories j and j^T have opposite momenta. This quantum correction to the classical Drude conductivity is called the *weak localization* correction [10–12]. The phase coherence is broken by the coupling of the electrons to other degrees of freedom or due to electron-electron interactions. Such coherence breakdown is temperature dependent and can be phenomenologically described by a temperature dependent phase coherence length $L_\phi(T) = \sqrt{D\tau_\phi(T)}$: trajectories larger than L_ϕ do not contribute to the weak localization correction.

As we have seen above, the amplitude of the correction is proportional to $P_{\rm int}(t)$, the distribution of loops. This quantity is nothing but the *return probability* after a time t. It is not exactly the classical return probability, which is the product of an amplitude with its complex conjugate (Diffuson). Here it corresponds to the product of an amplitude with the complex conjugate *time-reversed* amplitude. This object is named a "*Cooperon*". For closed trajectories, and with time-reversal symmetry, the two contributions, Cooperon and Diffuson, are equal. The return probability is thus doubled due to quantum coherence.

Equations (32) and (33) have a meaning only in the diffusive regime for which $t > \tau_e$ (otherwise a loop cannot be formed). The contribution of the return probability has thus to be integrated between τ_e, the smallest time for diffusion, and the phase coherence time τ_ϕ. Replacing the bounds by exponential cutoffs, the weak localization correction can be cast in the form [13] :

$$\Delta G = -2s\frac{e^2}{\hbar}\int_0^\infty P_{\rm int}(t)\ (\mathrm{e}^{-t/\tau_\phi} - \mathrm{e}^{-t/\tau_e})\frac{\mathrm{d}t}{\tau_D}\ . \tag{34}$$

In order to evaluate ΔG in various situations, we now study the diffusion equation and its solutions.

6.2 How to Solve the Diffusion Equation

In order to perform practical calculations, we have to calculate this distribution of closed trajectories, that is, the return probability. We have to solve a diffusion equation for this interference part. It looks very much like a classical diffusion equation (15), but there is an important difference. To account for magnetic field effects, it should be replaced by (in time representation) :

$$\left[\frac{\partial}{\partial t} - D\left(\nabla + \frac{2\mathrm{i}e\boldsymbol{A}}{\hbar c}\right)^2\right] P(\boldsymbol{r},\boldsymbol{r}',t) = \delta(\boldsymbol{r}-\boldsymbol{r}')\delta(t)\ . \tag{35}$$

The effect of the field is described by a covariant derivative (like in a Schrödinger equation) with an effective charge $2e$ to account for the pairing of trajectories.

To solve this equation (35), let us notice that it is a Green's equation, whose solutions are

$$P(\boldsymbol{r},\boldsymbol{r}',t) = \theta(t)\sum_n \psi_n^*(\boldsymbol{r})\psi_n(\boldsymbol{r}')\mathrm{e}^{-E_n t}\ , \tag{36}$$

where $\theta(t)$ is the step function and $\{E_n, \psi_n\}$ are the eigenvalues and eigenfunctions of the eigenvalue equation associated to (35) :

$$-D\left(\nabla_{\boldsymbol{r}} + \frac{2\mathrm{i}e\boldsymbol{A}}{\hbar c}\right)^2 \psi_n(\boldsymbol{r}) = E_n\psi_n(\boldsymbol{r})\ . \tag{37}$$

From (36), we find that the integrated probability $P(t)$ has the simple form :

$$P(t) = \theta(t) \sum_n \mathrm{e}^{-E_n t} \, . \tag{38}$$

This important result tells us that in order to evaluate the weak localization correction in any geometry, we simply need the eigenvalues of the diffusion equation in the corresponding geometry. We consider now a few examples.

6.3 Dimension Dependence of the Weak Localization Correction

Consider an infinite system, or with size $L \gg L_\phi$. For free diffusion in infinite space, the eigenvalues E_n of the diffusion equation are Dq^2 and the return probability $P(t)$ is given by (20). Since $P(t)$ is dimension dependent, we see that this weak localization correction depends dramatically on the space dimensionality. Inserting (20) in (34), and writing $\tau_\phi = L_\phi^2/D$, we obtain the well-known results :

$$\Delta g = \begin{cases} -\dfrac{L_\phi}{L} & \mathrm{quasi}-1d \\ -\dfrac{1}{\pi} \ln \dfrac{L_\phi}{l_e} & d=2 \\ -\dfrac{1}{2\pi} \left(\dfrac{L}{l_e} - \dfrac{L}{L_\phi} \right) & d=3 \end{cases} \tag{39}$$

Since $L_\phi(T)$ varies as a power-law with temperature, we obtain in particular the famous $\ln T$ dependence of the weak localization correction in $2d$.

It should be noticed that these results are meaningful if the correction stays smaller than the classical conductance (which can be written in the general form $g = A_d (k_\mathrm{F} L/2\pi)^{d-1} \, l_e/L$). This defines a characteristic length ξ given by $\Delta g(\xi) \simeq g$, for which the weak-disorder perturbative regime breaks down. In $1d$ and $2d$, it is given by

$$\xi_{1d} = 2 l_e \qquad , \qquad \xi_{2d} = l_e \mathrm{e}^{\pi k_\mathrm{F} l_e/2} \, , \tag{40}$$

and for a quasi-$1d$ system $\xi_{q1d} \simeq M l_e$ where M is the number of channels. ξ is the localization length. For a review on the strong localization regime, see for example [14].

6.4 Finite Systems, Boundary Conditions

In a mesoscopic system, the cutoff time in (32) is provided by τ_D. In other words, the cutoff length in (39) is now the size L of the system instead of L_ϕ. Therefore, from (39), we see that in quasi-$1d$, the weak localization correction is universal in the sense that it is a number, independent of disorder strength

(l_e). In $2d$ and $3d$, the integral (34) diverges at small time and is cut off by τ_e, so that the correction cannot be universal [1].

In order to calculate quantitatively the weak localization correction and the return probability in a finite system, one must be careful to account properly for correct boundary conditions.

If the system is closed, electrons stay inside the system, so that $P(t) \underset{t\to\infty}{\longrightarrow} 1$. The correct boundary condition is that the probability current vanishes at the boundary (Neumann condition). Therefore $q = n\pi/L$ with $n = 0, 1, 2, 3, \cdots$. For $t \to \infty$, the contribution of the zero mode in equation (38) gives correctly $P(t) \to 1$.

If the sample is perfectly connected to leads, electrons can leave the sample and $P(t) \underset{t\to\infty}{\longrightarrow} 0$. The probability at the boundary has to vanish (Dirichlet boundary condition), because if it goes in the leads it never comes back in the same state. The zero mode is now excluded, $q = n\pi/L$ with $n = 1, 2, 3, \cdots$. $P(t) \to 0$ when $t \to \infty$ since the particle leaves the box at large time. Inserting the expression (38) of the return probability with $E_n = n^2 E_c$ into (34) gives immediately, in the limit $L_\phi \to \infty$:

$$\Delta g = -2 \sum_{n\neq 0} \frac{1}{\pi^2 n^2} = -\frac{1}{3} \quad . \tag{41}$$

This result is proper to the perfectly connected wire.

6.5 Magnetic Field Effects

Ring or Cylinder Geometry : Sharvin-Sharvin Oscillations

Consider first the geometry of a ring pierced by a Aharonov-Bohm flux. In the presence of the flux, each closed trajectory accumulates an Aharonov-Bohm phase $2\pi\phi/\phi_0$, where ϕ is the flux through the ring. The time-reversed trajectory accumulates an opposite phase $-2\pi\phi/\phi_0$, so that the relative phase shift between the two trajectories is $4\pi\phi/\phi_0$, The fact that this relative phase between the two time-reversed trajectories is twice the phase enclosed by one trajectory is the reason why average quantities oscillate with period $\phi_0/2 = h/2e$.

We need to calculate the return probability in this geometry. This can be done directly by solving (35). Here let us proceed by simple arguments. Remember that in a $1d$ infinite space, the probability to go from $\boldsymbol{r}$ to $\boldsymbol{r}'$ is given by (19) with $d = 1$. The return probability is obtained by writing $\boldsymbol{r} = \boldsymbol{r}'$. On a ring, this would be the return probability without making a loop, $1/\sqrt{4\pi D t}$. The return probability after one loop of perimeter L necessarily contains a term $\mathrm{e}^{-L^2/4Dt}$. The accumulated phase is $4\pi\phi/\phi_0$ so that the flux dependent contribution of trajectories making one loop is modulated by $\cos 4\pi\phi/\phi_0$. Adding together the contributions of m loops, we get simply the Fourier expansion of the flux dependent return probability :

$$P_{\rm int}(t,\phi) = \frac{L}{\sqrt{4\pi Dt}} \sum_{m=-\infty}^{+\infty} {\rm e}^{-m^2L^2/4Dt} \cos 4\pi m\phi/\phi_0 \ . \tag{42}$$

Each harmonics of this expansion represents the return probability after m loops around the ring. Inserting this expression into (34) and after time integration, we obtain easily

$$\Delta G(\phi) = -s\frac{e^2}{h}\frac{L_\phi}{L}\left(1 + 2\sum_{m=1}^{+\infty} {\rm e}^{-mL/L_\phi}\ \cos 4\pi m\phi/\phi_0\right) \tag{43}$$

which can be resumed to obtain

$$\Delta G(\phi) = -s\frac{e^2}{h}\frac{L_\phi}{L}\frac{\sinh L/L_\phi}{\cosh L/L_\phi - \cos 4\pi\phi/\phi_0} \ . \tag{44}$$

The harmonics decay exponentially with their order, since they correspond to longer and longer diffusive trajectories.

For a cylinder, there is a possibility for the electrons to diffuse along the z axis of the cylinder, so that (42) is simply multiplied by $L_z/\sqrt{4\pi Dt}$. Inserting this new probability into (34), we obtain

$$\Delta G(\phi) = -s\frac{e^2}{\pi h}\frac{L}{L_z}\left[\ln\frac{L_\phi}{l_e} + 2\sum_{m=1}^{+\infty} K_0(mL/L_\phi)\ \cos 4\pi m\phi/\phi_0\right] \tag{45}$$

where K_0 is a modified Bessel function [15]. The $m = 0$ term is the usual $2d$ result (39). The contributions of the harmonics decay as ${\rm e}^{-L/L_\phi}$. These oscillations, predicted by Altshuler, Aronov and Spivak, where observed by Sharvin and Sharvin in 1981 [4, 16].

2*d* Gas in a Magnetic Field

In the ring geometry, all pairs of diffusive trajectories would pick the same phase $4\pi\phi/\phi_0$. In a uniform magnetic field, small and large trajectories accumulate different fluxes $\phi(\mathcal{A}) = B\mathcal{A}$ depending on their area $\mathcal{A}$. So the return probability is balanced by the average $\langle\cos 4\pi\phi(\mathcal{A})/\phi_0\rangle_{\mathcal{A}}$ on the distribution of areas $\mathcal{A}$ formed by the time-reversed diffusive trajectories.

Let us start with a qualitative evaluation. Short trajectories accumulate a flux smaller than the flux quantum and their contribution survives. Large trajectories accumulate flux larger than ϕ_0 and their contribution vanishes. When the magnetic field increases, the contribution of smaller and smaller trajectories is progressively suppressed. Trajectories smaller than some field dependent characteristic length L_B corresponding to $BL_B^2 \simeq \phi_0$ will not contribute. To this length L_B corresponds a characteristic time $\tau_B = L_B^2/D \simeq \phi_0/BD$, so that we can expect

$$\langle\cos 4\pi\phi/\phi_0\rangle_{\mathcal{A}} \simeq {\rm e}^{-t/\tau_B} \ . \tag{46}$$

Trajectories which enclose more than one flux quantum do not contribute to the return probability. Because of this new cutoff time, we can expect a field dependence of the weak localization of the form

$$\Delta g = -\frac{1}{\pi} \ln \frac{\min(L_\phi, L_B)}{l_e} \tag{47}$$

instead of (39).

The exact calculation is straightforward starting from equation (38). The eigenvalues E_n are solutions of an effective Schrödinger equation for a free particle of mass $m = \hbar/2D$ and charge $-2e$ in a uniform field B. They are precisely the Landau levels, namely

$$E_n = \left(n + \frac{1}{2}\right) \frac{4eDB}{\hbar} , \tag{48}$$

where n is an integer. The degeneracy of these levels for an area S is $g_n = \frac{2eB}{h} S$. The integrated return probability $P_{\rm int}(t)$ is just given by the sum $\sum_n g_n \mathrm{e}^{-E_n t}$, that is :

$$P_{\rm int}(t, B) = \frac{BS/\phi_0}{\sinh(4\pi BDt/\phi_0)} \tag{49}$$

where $\phi_0 = h/e$ is the flux quantum. This expression is nothing but the partition function of the harmonic oscillator. In the limit $B \to 0$, one recovers the result for free diffusion : $S/(4\pi Dt)$. For large times, $P_{\rm int}(t, B)$ decreases exponentially with the characteristic time $\tau_B = \phi_0/4\pi BD$ introduced qualitatively in (46). It describes the dephasing of time-reversed trajectories. Inserting (49) in (34), and performing the integral, we get :

$$\Delta g(B) = -\frac{1}{2\pi} \left[\Psi\left(\frac{1}{2} + \frac{\hbar}{4eDB\tau_e}\right) - \Psi\left(\frac{1}{2} + \frac{\hbar}{4eDB\tau_\phi}\right) \right] \tag{50}$$

where Ψ is the digamma function. This expression corresponds to the approximation (47). The weak localization correction is negative and cancelled by the magnetic field. As a result, a negative magnetoresistance is a well-known signature of weak localization (Fig. 9). A magnetoresistance measurement is a very interesting and useful tool to estimate τ_ϕ. The correction cancels when $\tau_B \simeq \tau_\phi$, that is for a field B_ϕ corresponding to a flux quantum through an area $B_\phi L_\phi^2$. Doing the same measurement at different temperatures is the usual method to extract $\tau_\phi(T)$.

7 Conductance Fluctuations

7.1 Universality as a Signature of Quantum Coherence

At a scale $L < L_\phi$, a conductor is a quantum object. Its conductance depends on the interference pattern between all diffusive trajectories. This interference pattern can be modulated by external parameters, like a magnetic field

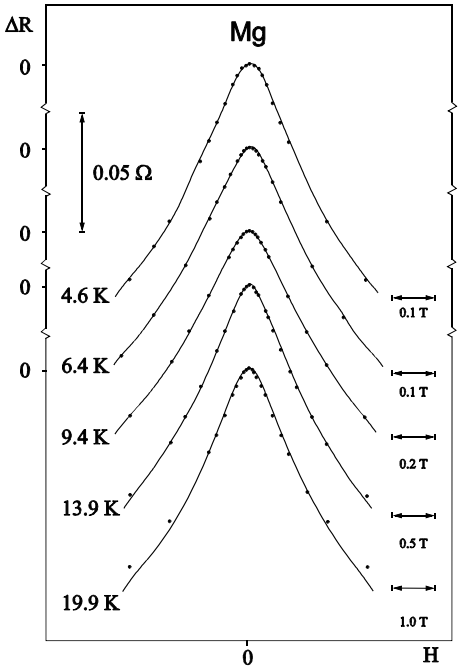

Fig. 9. Magnetic field dependence of the magnetoresistance of a Mg film, for different temperatures. The points are experimental results and the solid curves correspond to (50). The time $\tau_\phi(T)$ is a fitting parameter [11]

or a gate voltage. For example, Fig. 10a represents the variation of the conductance with the magnetic field, performed for 46 different samples (actually the same sample which has been annealed several times, so that the impurity configuration has changed and the interference pattern is different). It exhibits "fluctuations" which are reproducible for a given configuration of disorder. They are a "*fingerprint*" of this configuration. Figure 10b displays the average conductance, obtained by an average over the 46 samples. One clearly sees the weak localization correction, which is destroyed around some characteristic field B_ϕ. Interestingly, for the same characteristic field, the variance of the fluctuations, displayed on Fig. 10c is reduced by a factor 2.

Universality of the conductance fluctuations is a signature of quantum transport. Classically, one would expect the system to be considered as an addition of large number of incoherent elements. This number is of order $\mathcal{N} = (L/L_c)^d$, where L_c would be a correlation length, of the order of the mean free path. Then one would expect relative fluctuations of the conductance G of order of

$$\frac{\delta G}{G} \simeq \frac{1}{\sqrt{\mathcal{N}}} = \left(\frac{L_c}{L}\right)^{d/2} \tag{51}$$

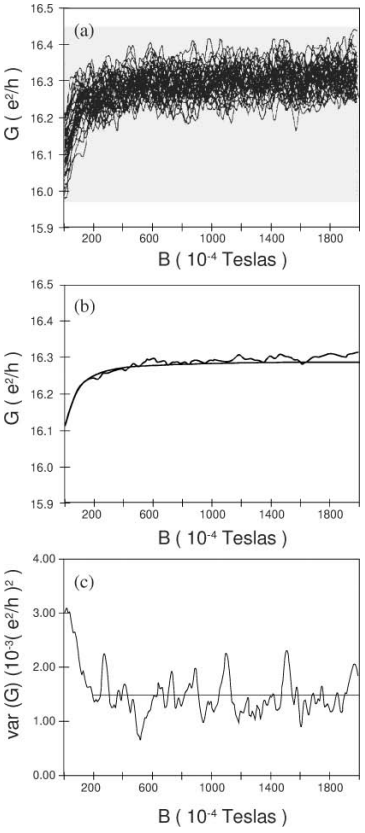

Fig. 10. Reproducible fluctuations of the magnetoconductance in units of e^2/h, at $T = 45\,\mathrm{mK}$ for Si doped GaAs. **(a)** Shows 46 plots as function of the magnetic field, for the same sample after successive annealing. Each plot corresponds to a disorder configuration and is called a *magnetofingerprint*. The amplitude of the fluctuations is smaller than e^2/h because $L > L_\phi$ (see (66)). **(b)** Presents the average conductance versus field. The weak localization correction disappears beyond a characteristic field. **(c)** Displays above the same field, the variance of the fluctuations divided by a factor 2, corresponding to the destruction of the Cooperon [17]

so that, since G varies as L^{d-2} :

$$\delta G \propto L^{\frac{d-4}{2}} \ , \tag{52}$$

and vanishes for large L. The system is said to be self-averaging. But the fact that the fluctuation δG stays actually finite means that there are strong correlations due to quantum coherence. Moreover, if one considers fully coherent ($L < L_\phi$) systems with quite different conductances, a good metal, a bad metal, or a semiconductor, one finds that the amplitude of the "oscillations" is always the same : it does not depend on the disorder. It is universal, of order e^2/h. *A priori*, we are not so surprised that these fluctuations are universal since our simple argument of Sect. 5 showed that all quantum effects have to be of order e^2/h.

7.2 Conductance Fluctuations and Speckle Correlations in Optics

Here it is quite useful to compare the physics of electronic transport with similar physics in optics where one measures the fluctuations of a transmission

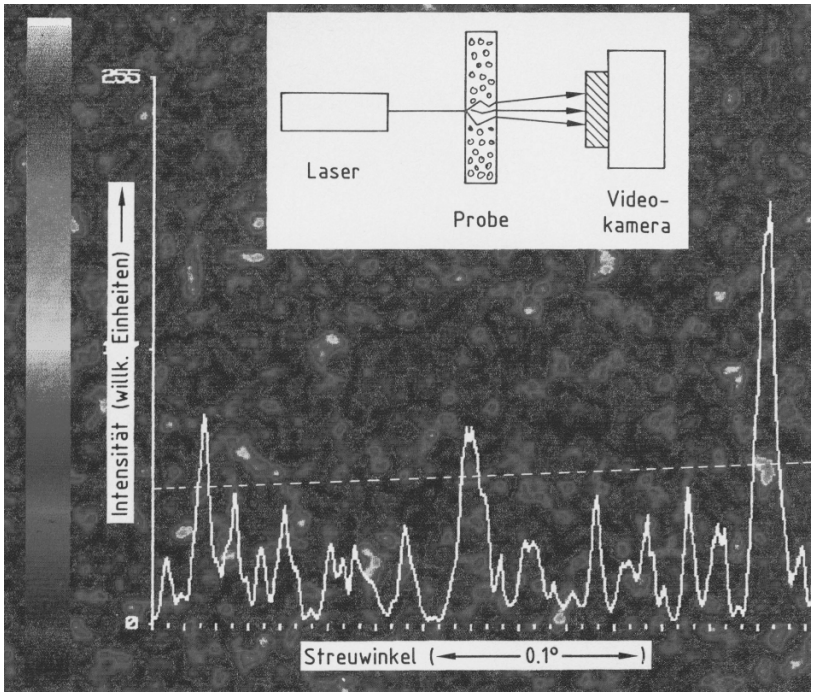

Fig. 11. A typical speckle pattern. The white and noisy curve represents the angular dependence of the light intensity along the cut represented by the dashed line. The relative fluctuations are of order unity (courtesy of G. Maret)

coefficient. In optics the light scattered by a diffusing medium forms a *speckle* pattern on a screen, and we want to describe the speckle fluctuations (This is exactly a generalization of Young experiments. Two slits produce well-defined fringes. Here the diffusing medium, like e.g. colloidal suspension, produces a complicated pattern called a speckle). A laser beam is sent on a diffusing medium along an incident direction a and the diffused intensity is measured along a direction b. The speckle pattern displayed on Fig. 11 represents the intensity measured along a direction b for a fixed incident direction a. So a given intensity on the screen represents the transmission coefficient T_{ab} from an incident direction a to an emergent direction b. We notice immediately that there are black spots, meaning that the relative fluctuations of this coefficient are of order 1. This is the Rayleigh law :

$$\overline{\delta T_{ab}^2} = \overline{T}_{ab}^2 \quad . \tag{53}$$

In electronics however, relative fluctuations of the conductance are very small, of order $1/g^2$. Here we want to understand why fluctuations are large in optics and small in electronics, namely to compare conductance fluctuations and fluctuations of transmission coefficient.

A very convenient way to link the two fields of optics and electronics is to use the Landauer formalism, which explicitly expresses the conductance as a transmission coefficient. This is formalized by Landauer formula $g = \sum_{ab} T_{ab}$ (25). The main difference between optics and electronics is that in optics, it is possible to measure each transmission coefficient T_{ab} while in electronics the conductance is related by (25) to a *sum* over all incoming and outgoing channels.

The average transmission is a probability. It is a sum of contributions of paired trajectories. Assuming that the angular (or channel) dependence of $\overline{T}_{ab}$ is negligible, all the channels contribute equally to the conductance, so that, from (25) :

$$\overline{G} = \frac{e^2}{h} M^2 \, \overline{T}_{ab} \tag{54}$$

and the average transmission coefficient is thus equal to

$$\overline{T}_{ab} \; = \frac{g}{M^2} \; . \tag{55}$$

Now we want to calculate the correlation between two transmission coefficients, that is the function $\overline{T_{ab} T_{a'b'}}$. This quantity is the product of two average transmission coefficients plus a correlation term :

$$\overline{T_{ab} T_{a'b'}} = \overline{T}_{ab} \;\; \overline{T}_{a'b'} \; + \overline{\delta T_{ab} \, \delta T_{a'b'}} \; . \tag{56}$$

The correlation term is constructed by pairing of trajectories corresponding to different transmission coefficients (Fig. 12).

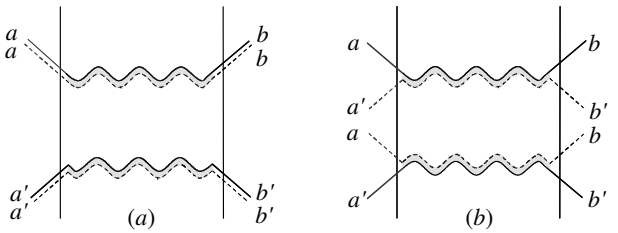

Fig. 12. Schematic representation of the two contributions to the product $\overline{T_{ab} T_{a'b'}}$. The first (a) corresponds to the product $\overline{T}_{ab} \;\; \overline{T}_{a'b'}$. The second (b) gives a contribution to the correlation function that we shall denote by $\overline{\delta T_{ab} \, \delta T_{a'b'}}^{C_1}$

One sees that there is no dephasing between the diffusive paired trajectories, expect outside the sample since a and a' (b and b') may correspond to different incoming (outgoing) channels. The second term is therefore of the form $\overline{T}_{ab}^{\,2} \; f(a, a', b, b')$ where f is a short-range function which vanishes rapidly as soon as $a \neq a'$ or $b \neq b'$. If a and b are angular directions, f is a rapidly decreasing function of the angles. In a wave guide geometry, where

the modes are quantized, $f = \delta_{aa'}\delta_{bb'}$. This contribution to the correlation function $\overline{\delta T_{ab}\,\delta T_{a'b'}}$ is called C_1 :

$$\overline{\delta T_{ab}\,\delta T_{a'b'}}^{C_1} = \overline{T}_{ab}^{\,2} \quad \delta_{aa'}\delta_{bb'} \; . \tag{57}$$

For $a = a'$ and $b = b'$, we obtain the Rayleigh law (53), that is

$$\overline{\delta T_{ab}^2} = 2\;\overline{T}_{ab}^{\,2} \quad . \tag{58}$$

The amplitude of the fluctuations is of the order of the average. This explains why there are black spots on Fig. 11.

In order to calculate the conductance fluctuations, we have to sum over all incoming and outgoing channels :

$$\overline{\delta g^2} = \sum_{aa'bb'} \overline{\delta T_{ab}\delta T_{a'b'}} = \sum_{aa'bb'} \overline{T}_{ab}^{\,2} \;\; \delta_{aa'}\delta_{bb'} = M^2\,\overline{T}_{ab}^{\,2} \;\; = \frac{g^2}{M^2} \ll 1 \; . \tag{59}$$

The sum is small since most of the terms are negligible. So our picture explains the important fluctuations of T_{ab} but *not* the amplitude of the conductance fluctuations. This means that additional contributions to the correlation function $\overline{\delta T_{ab}\,\delta T_{a'b'}}$ may have been forgotten.

A next contribution is obtained by pairing trajectories in a different way. One possibility is to exchange the quantum amplitudes, and to have one crossing as shown in Fig. 13a. This contribution is smaller by a factor $1/g$, so that its contribution to the correlation function $\overline{\delta T_{ab}\delta T_{a'b'}}$ is small. But it has an angular dependence different from the previous contribution. Figure 13a shows that there is a phase factor either for the incoming or the outgoing beam, so that instead of (57), we have for this second contribution, usually called C_2,

$$\overline{\delta T_{ab}\delta T_{a'b'}}^{C_2} = \frac{2}{3g}\overline{T}_{ab}^{\,2}\;(\delta_{aa'} + \delta_{bb'}) \; . \tag{60}$$

The factor $2/3$ results from an integration over the position of the quantum crossing [1]. A sum over all incoming and outgoing channels gives for this contribution :

$$\overline{\delta g^2} = \sum_{aa'bb'} \frac{2}{3g}\overline{T}_{ab}^{\,2}\;(\delta_{aa'} + \delta_{bb'}) = \frac{4g}{3M} \ll 1 \; . \tag{61}$$

Still this contribution cannot explain the amplitude of the observed conductance fluctuations, since it vanishes in the large M limit. Let us consider the next contribution shown on Fig. 13b, with two quantum crossings. We see that this so-called C_3 contribution is smaller than the first one by a factor $1/g^2$:

$$\overline{\delta T_{ab}\delta T_{a'b'}}^{C_3} = \frac{2}{15g^2}\,\overline{T}_{ab}^{\,2} \quad . \tag{62}$$

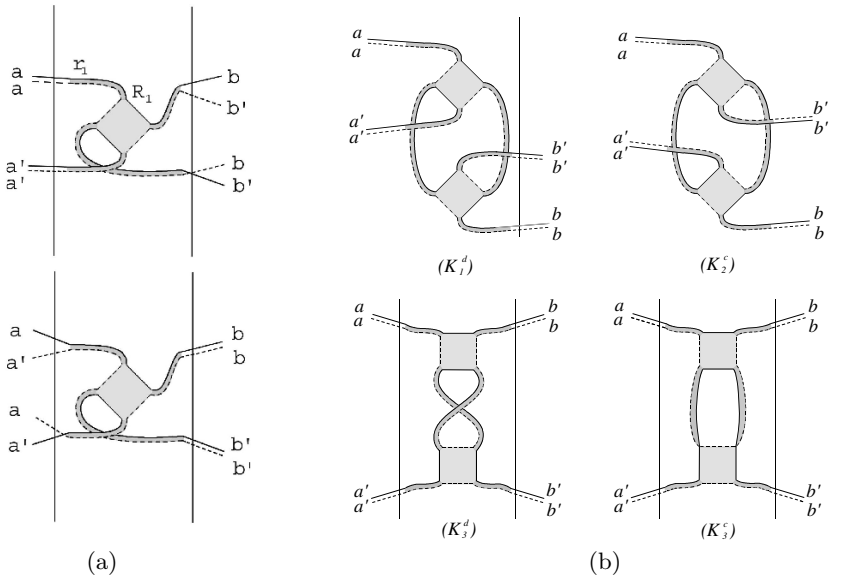

Fig. 13. Schematic representation of the C_2 and C_3 contributions to the correlation function $\overline{\delta T_{ab}\delta T_{a'b'}}$. κ_1^d and κ_3^d are Diffuson contributions. κ_2^c and κ_3^c are Cooperon contributions

The 2/15 factor comes from the integration over the positions of the two quantum crossings [1] (see next section). This term has *no angular dependence*, so that the summation over channels contains now M^4 terms and the contribution to the conductance fluctuations is

$$\overline{\delta g^2} = \frac{2}{15} \, . \tag{63}$$

The fluctuations are universal, independent on the strength of disorder. In summary, they are universal since the corresponding correlation function is constructed with two conductances and two quantum crossings, giving $g^2 \times 1/g^2 \simeq 1$. The contributions with no crossing or one crossing cancel because of angular dependences. The next terms with n crossings are negligible, of order $1/g^{n-2}$. In optics when one considers a speckle pattern, the first contribution is the most important, and the one crossing and two crossings contributions are very difficult to observe (They can be observed since, although very small, they have a different angular dependence, and also different temporal or frequency dependences [1]). In electronics, only the third contribution with two crossings is important after summation over incoming and outgoing channels. In summary, $\sum_{ab} T_{ab}$ has much smaller fluctuations than T_{ab}. We have now a simple recipe to evaluate average quantities or correlations functions : each quantum crossing gives a factor $1/g$.

7.3 Amplitude of the Conductance Fluctuations

In order to calculate quantitatively the conductance fluctuations, their dependence on geometry or external parameters, we must analyze more precisely the structure of the paired trajectories in Fig. 13b. In addition to the crossings, there is a loop. And we have to integrate on the distribution of loops, like for the weak localization correction. In contrast with the weak localization correction, this loop is formed by two crossings instead of one. So here for a given position of one crossing, we have to integrate on the position of the second crossing. Since this second crossing is necessarily along the loop formed by the two crossings, the integration over the position of the second crossing gives a volume element proportional to the length of the loop. For a trajectory of length $v_{\rm F}t$, the volume is $v_{\rm F}t\lambda_{\rm F}^{d-1}$. Moreover, $P(t)$ contains two-Diffusons and two-Cooperons contributions. A careful examination of the possible crossing and trajectories shows the only possible diagrams shown on Fig. 13. Taking into account their degeneracy [1], we obtain an expression which is as simple as the weak localization correction :

$$\overline{\delta G^2} = 6s^2 \left(\frac{e^2}{h}\right)^2 \int_0^\infty t\ [P_{\rm cl}(t) + P_{\rm int}(t)]{\rm e}^{-t/\tau_\phi} \frac{{\rm d}t}{\tau_D^2}\ . \tag{64}$$

There is an equal contribution of loops with Diffusons or Cooperons. In a magnetic field, the Cooperon contribution is suppressed so that the variance is reduced by a factor 2, as seen on Fig. 10c. This happens for the same magnetic field B_ϕ as the destruction of the weak localization correction, that is for a flux quantum through the system or through L_ϕ^2.

Incoherent processes not only destroy the Cooperon contribution but also the Diffuson contribution. This could appear surprising since we have seen that this Diffuson contribution corresponds to classical diffusion, and therefore should not be phase sensitive. However, this Diffuson contribution is not really the classical contribution, since it is constructed by pairing trajectories corresponding to *different* realizations of the system. If there is a phase breaking event, it affects equally one amplitude and its complex conjugate. But here, the phase breaking event may affect differently the two amplitudes since they correspond to different systems.

Like we have done above for the weak localization correction (34), we can now evaluate quite easily $\overline{\delta G^2}$ given by (64) for different geometries from the corresponding expression of the return probability $P(t)$. Let us do it for a quasi-1d mesoscopic wire. $P(t)$ is given by

$$P(\boldsymbol{q},t) = \sum {\rm e}^{-Dq^2t}$$

where $q = n\pi/L$ is quantized by the Dirichlet boundary conditions corresponding to a perfectly connected wire (no zero mode, see Sect. 6.4). Inserting $P(t)$ in (64), we get

$$\overline{\delta g^2} = 6 \sum_{n>0} \frac{1}{\pi^4 n^4} = \frac{2}{15} \, . \tag{65}$$

Let us remark that the choice of the boundary conditions is very important. If the system were closed or poorly connected, corresponding to Neumann boundary conditions, then the contribution of the zero mode would lead to a divergence, or at least non-universality of the fluctuations.

In the macroscopic limit, when $L_\phi \gg L$, we can treat the system as infinite, replace $P(t)$ by its dependence (20) for an infinite system $(\tau_D/4\pi t)^{d/2}$ and multiply by the exponential decay $\mathrm{e}^{-t/\tau_\phi}$ in the integral. Then we obtain :

$$\overline{\delta g^2} \propto \left(\frac{L_\phi}{L} \right)^{\frac{4-d}{2}} \, . \tag{66}$$

We are not surprised by this result. It is exactly the one anticipated from our simple argument (51) treating the fluctuations as due to incoherent contributions of correlated regions of size L_ϕ.

Finally, it is quite easy to notice, that correlations functions of n conductances imply $2n-2$ crossings. Therefore the n^{th} cumulant of the conductance distribution is of order $g^n/g^{2n-2} \propto 1/g^{n-2}$. It vanishes for $n > 2$ in the metallic limit $g \to \infty$, so that the conductance distribution is indeed Gaussian.

8 Diffusion on Graphs and Spectral Determinant

The calculation of the weak localization correction or of the conductance fluctuations, as well as of other thermodynamical quantities like orbital magnetic susceptibility [1, 18] can be extended to the case of any structure – called a network – made of quasi-one-dimensional diffusive wires. First, we note that the quantities of interest, like the weak localization correction (34) or the conductance fluctuations (64), have the same structure :

$$\int t^\alpha P(t) \mathrm{e}^{-\gamma t} \mathrm{d}t \, , \tag{67}$$

where $\gamma = 1/\tau_\phi$. From (38), the time integral of $P(t)$ can be straightforwardly written in terms of a quantity called the *spectral determinant* $S(\gamma)$:

$$\int_0^\infty \mathrm{d}t P(t) \mathrm{e}^{-\gamma t} = \sum_n \frac{1}{E_n + \gamma} = \frac{\partial}{\partial \gamma} \ln S(\gamma) \, , \tag{68}$$

where $S(\gamma)$ is, within a multiplicative constant independent of γ :

$$S(\gamma) = \prod_n (\gamma + E_n) \, , \tag{69}$$

E_n being the eigenvalues of the diffusion equation (37). Using standard properties of Laplace transforms, the above time integrals can be rewritten in

terms of the spectral determinant, so that the weak localization corrections to the conductivity and the conductivity fluctuations respectively read :

$$\Delta\sigma = -\,2s\frac{e^2}{h}\frac{D}{\Omega}\,\frac{\partial}{\partial\gamma}\ln S_{\rm int}(\gamma)\ , \tag{70}$$

$$\langle\delta\sigma^2\rangle = -\,6s^2\frac{e^4}{h^2}\frac{D^2}{\Omega^2}\,\frac{\partial^2}{\partial\gamma^2}\left[\ln S_{\rm cl}(\gamma)+\ln S_{\rm int}(\gamma)\right]\ , \tag{71}$$

where $S_{\rm cl}$ and $S_{\rm int}$ are the spectral determinants associated respectively to the diffusion equation for the Diffuson and the Cooperon. Here Ω is the total volume of the system. These expressions are quite general, strictly equivalent to expressions (34) and (64). Their interest is that, on a network, the spectral determinant takes a very simple form. By solving the diffusion equation (35) on each link, and then imposing Kirchoff type conditions on the nodes of the graph with N nodes, the problem can be reduced to the solution of a system of N linear equations relating the eigenvalues at the N nodes. Let us introduce the $N\times N$ matrix M :

$$M_{\alpha\alpha} = \sum_\beta \coth(\eta_{\alpha\beta})\quad ,\quad M_{\alpha\beta} = -\frac{{\rm e}^{{\rm i}\theta_{\alpha\beta}}}{\sinh\eta_{\alpha\beta}}\ . \tag{72}$$

The sum $\sum_\beta$ extends to all the nodes β connected to the node α, $l_{\alpha\beta}$ is the length of the link between α and β, and $\eta_{\alpha\beta}=l_{\alpha\beta}/L_\phi$. The off-diagonal coefficient $M_{\alpha\beta}$ is non-zero only if there is a link connecting the nodes α and β. Here $\theta_{\alpha\beta}=(4\pi/\phi_0)\int_\alpha^\beta \boldsymbol{A}.{\rm d}\boldsymbol{l}$ is the circulation of the vector potential between α and β. It can then be shown that the spectral determinant takes the very convenient form [1, 18] :

$$S = \left(\frac{L_\phi}{L_0}\right)^{N_B-N}\ \prod_{(\alpha\beta)}\sinh\eta_{\alpha\beta}\ \det M\ , \tag{73}$$

where L_0 is an arbitrary length independent of γ (or L_ϕ) and N_B is the number of links in the graph. We have thus transformed the spectral determinant which is an infinite product in a finite product related to $\det M$. Using (73), mesoscopic quantities (70–71) can be easily predicted for any geometry of diffusive networks [1, 18].

9 Interaction Effects

Until now we have not considered the role of electron-electron interactions. They turn out to give small corrections to transport quantities like the average conductivity, but they play an important role to understand thermodynamic properties like persistent currents [1]. Moreover, until now we have introduced by hand a phase coherence time τ_ϕ (or length L_ϕ). We wish now to understand

the microscopic origin for the loss of quantum coherence. This phase coherence is limited by the interactions with other degrees of freedom, in particular other electrons through their mutual interaction. We want to discuss now how e-e interactions break phase coherence.

On one hand, interaction effects can be considered as negligible. We know from Landau theory of Fermi liquids that in an interacting electron gas, free particles have simply to be replaced by *quasi-particles* which are dressed objects, screened by the cloud of other electrons. These quasi-particles have a long lifetime which diverges when approaching the Fermi level. From Landau, we know that

$$\frac{1}{\tau_{ee}(\epsilon)} \propto \epsilon^2 \tag{74}$$

where $\epsilon = E - E_F$ is the energy of the quasi-particle measured from the Fermi level. On the other hand, in a disordered metal, electrons move diffusively, that is very slowly. They spend long time close to each other. Qualitatively, we can expect that diffusion somehow enhance the effect of interactions. We may ask if and how expression (74) is changed because of the diffusion. Moreover, since the diffusive motion is dimensionality dependent, the modified lifetime should also depend on this dimensionality.

Interaction between electrons is expected to have two major effects :

- First, each electron is not only sensitive to the disordered potential but also to the fluctuations of the electronic density due to other electrons. This additional fluctuating potential modifies the position of the energy levels, especially near the Fermi level. So we may expect a modification in the repartition of the energy levels, that is a change of the density of states near the Fermi level. We shall show that the density of states exhibits a decrease, the so-called "Altshuler-Aronov" anomaly. This reduction of the density of states is accompanied by a reduction of the conductivity.
- Since the e-e interaction is a inelastic process, each quasi-particle has a finite lifetime which limits the phase coherent properties like weak localization, since the coherence between time-reversed trajectories are necessarily limited by this lifetime.

9.1 Screening

In order to describe interaction effects, let us start with a few reminders about the screening of interaction. The bare Coulomb interaction potential is $U_0(R) = e^2/R$, that is in $3d$, $U_0(q) = 4\pi e^2/q^2$. The screened Coulomb interaction $U(q)$ is given by

$$U(q) = \frac{U_0(q)}{1 + \chi_0(q) U_0(q)} \tag{75}$$

where, in the small q limit (we are interested in the diffusive regime where $ql_e \ll 1$), the susceptibility $\chi_0(q)$ is the Pauli susceptibility, that is the density of states $2\rho_0$. Therefore, the screened interaction is given by

$$U(q) = \frac{4\pi e^2}{q^2 + \kappa^2} \qquad , \qquad U(R) = \frac{e^2}{R} \mathrm{e}^{-\kappa R} \, , \tag{76}$$

where the Thomas-Fermi vector κ (inverse screening length) is

$$\kappa^2 = 8\pi e^2 \rho_0 \, . \tag{77}$$

In the diffusive limit $ql_e \ll 1$, the screened interaction can by approximated by

$$U(q) = \frac{4\pi e^2}{\kappa^2} = \frac{1}{\chi_0} = \frac{1}{2\rho_0} \qquad , \qquad U(R) = \frac{1}{2\rho_0} \delta(R) \, . \tag{78}$$

On the scale of diffusion, the screened interaction can be considered as a local interaction.

However, it turns out that screening is not instantaneous since electrons have to diffuse to screen a local charge. Therefore the interaction is actually frequency dependent. This dynamical screening is described by the frequency dependent susceptibility $\chi_0(q, \omega)$ which accounts for the dynamical charge reorganization :

$$\chi_0(q, \omega) = 2\rho_0 \frac{Dq^2}{-\mathrm{i}\omega + Dq^2} \, . \tag{79}$$

Therefore equation (78) becomes

$$U(q, \omega) = \frac{1}{\chi_0} = \frac{1}{2\rho_0} \frac{-\mathrm{i}\omega + Dq^2}{Dq^2} \, . \tag{80}$$

9.2 Density of States Anomaly

As we have done above, we shall avoid technicalities of diagrammatic theory, and try to get the important results from qualitative arguments. Although we have not elaborated on the theory of Green's functions, let me remind you at least that the Green's function is related to the density of states by

$$\rho(\epsilon) = -\frac{1}{\pi \Omega} \int \mathrm{Im} G_\epsilon(\boldsymbol{r}_0, \boldsymbol{r}_0) \mathrm{d}\boldsymbol{r}_0 \, . \tag{81}$$

As we have seen in Sect. 3, $G(\boldsymbol{r}_0, \boldsymbol{r}_0)$ is the sum of contributions from all closed trajectories from $\boldsymbol{r}_0$ to $\boldsymbol{r}_0$ (Fig. 14a). All these amplitudes have different and random phases and their contribution cancels in average. What remains is the contribution of short trajectories, giving an average density of states $\overline{\rho(\epsilon)} = 2\rho_0$.

How can the effects of diffusion + interactions appear on the density of states? The non-interacting density of states (or Green's function) is a single electron property, and therefore involve single trajectories. In the presence of electron-electron interaction, each electron trajectory can be paired with

the trajectory of a second electron, with which it interacts. Then by pairing these two trajectories, we can construct a Diffuson. More precisely, we pair an amplitude corresponding to one electron to the conjugate amplitude corresponding to another electron. Their interaction is represented by a wiggly line in Fig. 14b. There are actually two possible contributions, depending on the position of the interaction line. They are nothing but the Hartree and the exchange (Fock) contributions. We can conveniently separate the "diagrams" in three different parts :

- A short-range part close to the point $\boldsymbol{r}_0$. It resembles somehow to a quantum crossing, with a dephasing between the three trajectories,
- two long ranged Diffusons,
- an interaction region between $\boldsymbol{r}$ and $\boldsymbol{r}'$.

So, we can easily construct the structure of this additional contribution to the density of states :

$$\frac{\delta\rho}{\rho_0} \simeq -\frac{\lambda_\rho}{g}\int^{\hbar/\epsilon} P(t)\,\frac{\mathrm{d}t}{\tau_D}\,. \tag{82}$$

λ_ρ is a dimensionless parameter which describes the strength of the interaction. It contains both the Hartree and exchange contributions. The $1/g$ reduction comes from the quantum crossing. And the distribution of loops formed by the paired trajectories is given by the return probability $P(t)$. For an energy ϵ, the two amplitudes must be taken at different energies ω and $\omega - \epsilon$. Therefore the two trajectories can stay in phase only during a time $\Delta \simeq \hbar/\epsilon$, so that the upper cutoff in the integral is $\hbar/\epsilon$.

Here our aim is simply to present the structure of the result without entering into details. Another qualitative derivation is given in [19], a detailed discussion is proposed in [1], and the original calculation is done in [20]. We see that the amplitude and the structure of this correction to the density of states looks very similar to that of the weak localization correction (34), except that the upper cutoff is not τ_ϕ but $\hbar/\epsilon$. Unlike for the weak localization, this correction depends on the classical return probability, so that it is not suppressed by a magnetic field [21].

The form (82) is approximate. A more sophisticated calculation replaces the upper cutoff by a Fourier transform. Moreover since $g \propto \rho_0\Omega/\tau_D$, (82) becomes

$$\delta\rho(\epsilon) = -\frac{\lambda_\rho}{\pi\Omega}\int_0^\infty P(t)\cos\epsilon t\,\mathrm{d}t\,. \tag{83}$$

From the expression (20) of $P(t)$, we obtain the energy dependence of the density of states anomaly :

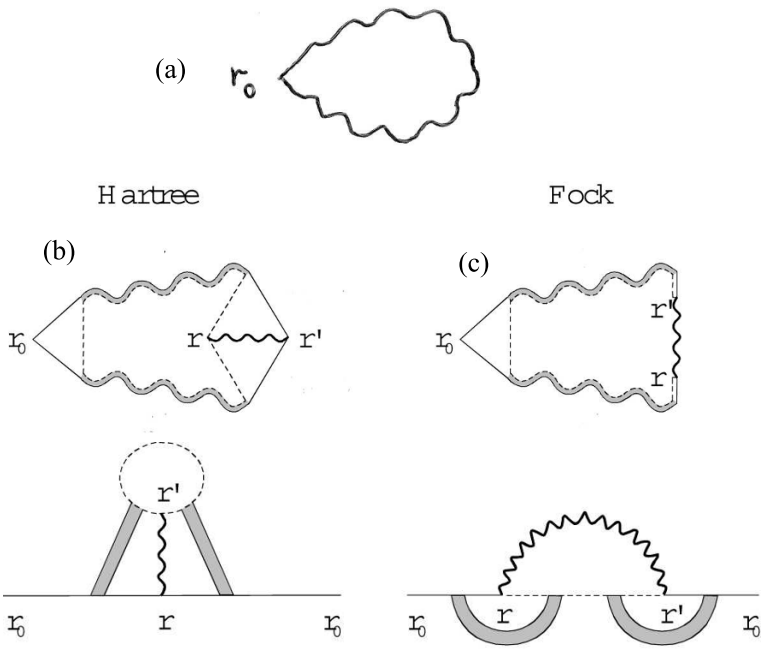

Fig. 14. (a) Diagram for the non-interacting density of states. After disorder averaging, the contribution of diffusive trajectories vanishes because of their random phase. **(b)** Hartree and **(c)** exchange (Fock) diagrams for the density of states anomaly. These contributions survive disorder averaging since they contain paired trajectories. Upper : schematic representation exhibiting clearly the three regions, a short-range region, diffusive trajectories and the interaction region. Bottom : usual diagrammatic representation. These two representations are equivalent

$$\delta\rho(\epsilon) \propto -\frac{\lambda_\rho}{D\Omega} \begin{cases} L_\epsilon & \text{quasi} - 1d \\ \ln \dfrac{L_\epsilon}{l_e} & d = 2 \\ \dfrac{1}{l_e} - \dfrac{1}{L_\epsilon} & d = 3 \end{cases} \tag{84}$$

where the characteristic length L_ϵ is $L_\epsilon = \sqrt{\hbar D/\epsilon}$. The structure of this correction related to $P(t)$ is similar to the weak localization correction (39), except that the characteristic length L_ϕ has been replaced by L_ϵ. This reduction of the density of states can be observed experimentally by tunnel measurements, since it reflects as an anomaly in the voltage dependence of

the tunnel conductance G_t. At zero temperature, the relative correction to the tunnel conductance is given by

$$\frac{\delta G_t}{G_t} = \frac{\delta\rho(\epsilon = eV)}{\rho_0} \ . \tag{85}$$

At finite temperature, it is not difficult to generalize the expression (83) of the density of states anomaly as

$$\delta\rho(\epsilon, T) = -\int f'(\epsilon - \omega)\delta\rho(\omega)\mathrm{d}\omega \ , \tag{86}$$

where f' is the derivative of the Fermi function. After a Fourier transform, we find

$$\delta\rho(\epsilon, T) = -\frac{\lambda_\rho}{2\pi\Omega}\int_0^\infty R_T(t)P(t)\cos\epsilon t\,\mathrm{d}t \ , \tag{87}$$

where the thermal function $R_T(t)$ is given by $R_T(t) = \pi Tt/\sinh\pi Tt$. The temperature dependence of the tunnel conductance anomaly, also called zero-bias anomaly is

$$\frac{\delta G_t(V,T)}{G_t} = -\frac{1}{\rho_0}\int \delta\rho(\epsilon, T)f'(\epsilon - eV)\mathrm{d}\epsilon \ , \tag{88}$$

or, after a Fourier transform :

$$\frac{\delta G_t(V,T)}{G_t} = -\frac{\lambda_\rho}{2\pi\rho_0\Omega}\int_0^\infty R_T^2(t)P(t)\cos eVt\,\mathrm{d}t \ , \tag{89}$$

This correction has been measured for various systems with different dimensionalities and the $1/\sqrt{V}$, $\ln V$, and $\sqrt{V}$ predicted by (84,89) respectively in 1, 2 and 3 dimensions have been observed [22].

9.3 Correction to the Conductivity

Taking into account the interaction between electrons leads also to a reduction of the conductivity. Without going into the details of the calculations, we can argue that this reduction is a consequence of the correction to the density of states. Both effects result from the scattering of an electron by the charge fluctuations induced by disorder. The temperature dependence of the conductivity $\sigma(T)$ is related to its energy dependence at $T = 0\,\mathrm{K}$ by $\sigma(T) = -\int f'(\epsilon)\sigma(\epsilon)\mathrm{d}\epsilon$, where $f'(\epsilon)$ is the derivative of the Fermi function. Since the conductivity is proportional to the density of states (Einstein relation), we expect that the density of states anomaly leads to a correction of the conductivity given by

$$\frac{\delta\sigma(T)}{\sigma_0} = \int \mathrm{d}\epsilon\left(-\frac{\partial f}{\partial\epsilon}\right)\frac{\delta\rho(\epsilon, T)}{\rho_0} \ , \tag{90}$$

where σ_0 is the Drude conductivity (22). For a static interaction, the density of states correction $\delta\rho(\epsilon, T)$ is given by (87). Upon Fourier transforming, we have

$$\delta\sigma(T) = -\lambda_\sigma \left(\frac{e^2 D}{\pi\Omega}\right) \int_0^\infty R_T^2(t) P(t) \mathrm{d}t \tag{91}$$

where λ_σ is a parameter which depends on the interaction [23]. Using the expression (20) of $P(t)$, we obtain the temperature dependence of the correction to the conductance

$$\delta g(T) \propto -\lambda_\sigma \begin{cases} \dfrac{L_T}{L} & \text{quasi} - 1d \\ \ln \dfrac{L_T}{l_e} & d = 2 \\ \dfrac{L}{l_e} - \dfrac{L}{L_T} & d = 3 \end{cases} \tag{92}$$

where the thermal length L_T is defined by $L_T = \sqrt{\hbar D/T}$. In $2d$, the temperature dependence is logarithmic like the weak localization correction. Unlike the weak localization correction, this correction to the conductivity is not sensitive to a magnetic field. Therefore both corrections can be separated experimentally by the application of a magnetic field.

9.4 Lifetime of Quasi-Particle

Consider a Fermi sea and inject a quasi-particle in a state $|\alpha\rangle$ with energy ϵ_α above the Fermi sea. It interacts with another particle $(|\gamma\rangle, \epsilon_\gamma)$ and the final state consists in two quasi-particles $(|\beta\rangle, \epsilon_\beta)$ and $(|\delta\rangle, \epsilon_\delta)$ above the Fermi sea (Fig. 15). Energy conservation implies $\epsilon_\alpha + \epsilon_\gamma = \epsilon_\beta + \epsilon_\delta$. The lifetime of the state $|\alpha\rangle$ is given by the Fermi golden rule and it is related to the matrix element of the interaction :

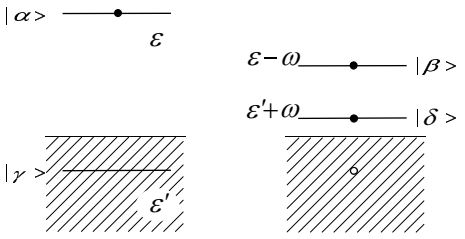

Fig. 15. A quasi-particle in a state $|\alpha\rangle$ of energy $\epsilon_\alpha = \epsilon$ interacts with another quasi-particle $|\gamma\rangle$ of energy $\epsilon_\gamma = \epsilon'$ in the Fermi sea. The final state is made of two quasi-particles above the Fermi sea and one hole

$$\frac{1}{\tau_\alpha} = 2\pi s \sum_{\beta\gamma\delta} |\langle\alpha\gamma|U|\beta\delta\rangle|^2 \delta(\epsilon_\alpha + \epsilon_\gamma - \epsilon_\beta - \epsilon_\delta) \ , \tag{93}$$

with the constraint that $\epsilon_\gamma < 0$, $\epsilon_\beta > 0$ and $\epsilon_\delta > 0$. If the matrix element does not depend on energies, we see immediately that the Landau ϵ^2 dependence comes simply from energy constraints : basically three final states can be chosen in a range of energy ϵ, with the constraint of energy conservation, whence the ϵ^2 dependence. Indeed the matrix element of the interaction is energy independent in the ballistic case, but this is not true anymore in the diffusive case. The goal of the following pages is to calculate the energy dependence of the litetime in the diffusive regime. Since we do not specify a given state $|\alpha\rangle$, (93) can be rewritten for the lifetime at a given energy ϵ

$$\begin{aligned}\frac{1}{\tau_{ee}(\epsilon)} &= \frac{1}{\nu_0} \sum_\alpha \frac{1}{\tau_\alpha} \delta(\epsilon - \epsilon_\alpha) \\ &= \frac{2\pi s}{\nu_0} \sum_{\alpha\beta\gamma\delta} |\langle\alpha\gamma|U|\beta\delta\rangle|^2 \delta(\epsilon_\alpha + \epsilon_\gamma - \epsilon_\beta - \epsilon_\delta)\delta(\epsilon - \epsilon_\alpha) \ ,\end{aligned} \tag{94}$$

where $\nu_0 = \rho_0 \Omega$ is the density of states. By introducing two energy integrals, it can be rewritten in the form

$$\begin{aligned}\frac{1}{\tau_{ee}(\epsilon)} = \frac{2\pi s}{\nu_0} \int_0^\epsilon \mathrm{d}\omega \int_{-\omega}^0 \mathrm{d}\epsilon' \sum_{\alpha\beta\gamma\delta} |\langle\alpha\gamma|U|\beta\delta\rangle|^2 \\ \times \delta(\epsilon - \epsilon_\alpha)\delta(\epsilon' - \epsilon_\gamma)\delta(\epsilon - \omega - \epsilon_\beta)\delta(\epsilon' + \omega - \epsilon_\delta) \ ,\end{aligned} \tag{95}$$

where $\epsilon > 0$, $\epsilon' < 0$ and $\epsilon - \omega > 0$, $\epsilon' + \omega > 0$ are respectively the two energies of the initial states and of the final states. If the matrix element is energy independent, we trivially recover the ϵ^2 dependence. Upon averaging over disorder, the lifetime has the form

$$\frac{1}{\tau_{ee}(\epsilon)} = 4\pi\nu_0^3 \int_0^\epsilon \omega W^2(\omega) \mathrm{d}\omega \tag{96}$$

with

$$W^2(\omega) = \frac{1}{\nu_0^4} \overline{\sum_{\alpha\beta\gamma\delta} |\langle\alpha\gamma|U|\beta\delta\rangle|^2 \delta(\epsilon - \epsilon_\alpha)\delta(\epsilon' - \epsilon_\gamma)\delta(\epsilon - \omega - \epsilon_\beta)\delta(\epsilon' + \omega - \epsilon_\delta)} \tag{97}$$

The characteristic matrix element $W(\omega)$ depends only on energy transfer ω, but neither on ϵ, nor on ϵ'.

The matrix element $\langle\alpha\gamma|U|\beta\delta\rangle$ can be evaluated on the basis of eigenfunction of the non-interacting particles and reads

$$\langle\alpha\gamma|U|\beta\delta\rangle = \int \mathrm{d}\boldsymbol{r}_1 \mathrm{d}\boldsymbol{r}_2 \phi_\alpha^*(\boldsymbol{r}_1)\phi_\gamma^*(\boldsymbol{r}_2)\phi_\beta(\boldsymbol{r}_1)\phi_\delta(\boldsymbol{r}_2) U_\omega(\boldsymbol{r}_1 - \boldsymbol{r}_2) \ , \tag{98}$$

where $U_\omega(\boldsymbol{r})$ is the dynamically screened potential. The combination of wave functions and δ function can be rewritten in terms of Green's functions

$$-\frac{1}{\pi}\mathrm{Im}G(\boldsymbol{r},\boldsymbol{r}') = \sum_\alpha \phi^*_\alpha(\boldsymbol{r})\phi_\alpha(\boldsymbol{r}')\delta(\epsilon-\epsilon_\alpha)\ , \tag{99}$$

so that $W^2(\omega)$ can be rewritten in the form

$$W^2(\omega) = \frac{1}{\nu_0^4\pi^4}\int \mathrm{d}\boldsymbol{r}_1\mathrm{d}\boldsymbol{r}_2\mathrm{d}\boldsymbol{r}'_1\mathrm{d}\boldsymbol{r}'_2 U_\omega(\boldsymbol{r}_1-\boldsymbol{r}_2)U_\omega(\boldsymbol{r}'_1-\boldsymbol{r}'_2) \times \overline{\mathrm{Im}G_\epsilon(\boldsymbol{r}_1,\boldsymbol{r}'_1)\mathrm{Im}G_{\epsilon-\omega}(\boldsymbol{r}'_1,\boldsymbol{r}_1)}\ \ \overline{\mathrm{Im}G_{\epsilon'}(\boldsymbol{r}_2,\boldsymbol{r}'_2)\mathrm{Im}G_{\epsilon'+\omega}(\boldsymbol{r}'_2,\boldsymbol{r}_2)} \tag{100}$$

where the average of the product of four Green's functions has been decoupled into the product of two average values. By pairing Green's functions, it is possible to show that the average product contains a long-range part related to the probability P (see (14) and [1])

$$\overline{\mathrm{Im}G_\epsilon(\boldsymbol{r},\boldsymbol{r}')\mathrm{Im}G_{\epsilon-\omega}(\boldsymbol{r}',\boldsymbol{r})} = \pi\rho_0\,\mathrm{Re}P(\boldsymbol{r},\boldsymbol{r}',\omega)\ . \tag{101}$$

We deduce

$$W^2(\omega) = \frac{1}{\pi^2\nu_0^2\Omega^2}\int \mathrm{d}\boldsymbol{r}_1\mathrm{d}\boldsymbol{r}_2\mathrm{d}\boldsymbol{r}'_1\mathrm{d}\boldsymbol{r}'_2 U_\omega(\boldsymbol{r}_1-\boldsymbol{r}_2)U_\omega(\boldsymbol{r}'_1-\boldsymbol{r}'_2) \times \mathrm{Re}P_d(\boldsymbol{r}_1,\boldsymbol{r}'_1,\omega)\mathrm{Re}P_d(\boldsymbol{r}_2,\boldsymbol{r}'_2,-\omega) \tag{102}$$

or, upon Fourier transformation :

$$W^2(\omega) = \frac{1}{\pi^2\nu_0^2\Omega^2}\sum_{\boldsymbol{q}\neq 0}|U(\boldsymbol{q},\omega)|^2[\mathrm{Re}P_d(\boldsymbol{q},\omega)]^2\ . \tag{103}$$

At the diffusion approximation, the dynamically screened potential is given by (80) so that

$$W^2(\omega) = \frac{1}{4\pi^2\nu_0^4}\sum_{\boldsymbol{q}\neq 0}\frac{1}{\omega^2+D^2q^4}\ , \tag{104}$$

which can be expressed in terms of the return probability $P(t)$

$$W^2(\omega) = \frac{1}{4\pi^2\nu_0^4}\frac{1}{\omega}\int_0^\infty P(t)\sin\omega t\ \mathrm{d}t\ . \tag{105}$$

Finally, the electronic lifetime (96) is given by

$$\frac{1}{\tau_{ee}(\epsilon)} = \frac{2}{\pi\nu_0}\int_0^\infty \frac{P(t)}{t}\sin^2\frac{\epsilon t}{2}\mathrm{d}t\ . \tag{106}$$

For a metal of volume Ω, we can identify two different regimes :

• $\epsilon \gg E_c$ where E_c is the Thouless energy. This corresponds to time scales $t \ll \tau_D$. In this case an electron described as a diffusive wave packet

is insensitive to the boundaries and behaves as in an infinite medium where, according to (20), $P(t) = \Omega/(4\pi Dt)^{d/2}$. We obtain [24] for the integral (105) :

$$W^2(\omega) = \frac{dc_d}{16}\frac{1}{\nu_0^4\omega^2}\left(\frac{\omega}{E_c}\right)^{d/2} , \tag{107}$$

so that the electronic lifetime is equal to

$$\frac{1}{\tau_{\rm ee}(\epsilon)} = \frac{\pi}{2}c_d\Delta\left(\frac{\epsilon}{E_c}\right)^{d/2} \qquad (\epsilon \gg E_c) \tag{108}$$

where $c_1 = \sqrt{2}/\pi^2$, $c_2 = 1/4\pi^2$, $c_3 = \sqrt{2}/6\pi^3$. Such a behavior has been indeed observed in silver wires ($d = 1$) for which $W^2(\omega) \propto \omega^{-3/2}$ and $1/\tau_{ee}(\epsilon) \propto \epsilon^{1/2}$, although the measured prefactor came out to be larger than the value predicted here [25].

• The limit $\epsilon \ll E_c$, that is $t \gg \tau_D$, corresponds to the ergodic regime in which the diffusive electronic wave packet explores all the accessible volume Ω. Thus we would expect $P(t)$ to be driven only by the zero mode. This is not so, because in expression (104) this mode has been removed in order to ensure electronic neutrality. The excitation energy ϵ is smaller than E_c and it is not possible to replace the sum (104) by an integral. In this limit, we obtain

$$W^2(\omega) = \frac{a_d}{4\pi^6}\frac{\Delta^4}{E_c^2} \propto \frac{\Delta^2}{g^2} , \tag{109}$$

where the coefficient a_d is defined by the series

$$a_d = \sum_{n_x,n_y,n_z} \frac{1}{(n_x^2+n_y^2+n_z^2)^2} . \tag{110}$$

The ratio E_c/Δ is the dimensionless conductance g. For $\omega \ll E_c$, the characteristic matrix element of the interaction is thus energy independent and of order Δ/g. The inverse lifetime in this case is [26]

$$\frac{1}{\tau_{\rm ee}(\epsilon)} = \frac{a_d}{2\pi^5}\Delta\left(\frac{\epsilon}{E_c}\right)^2 \qquad (\epsilon \ll E_c) . \tag{111}$$

9.5 Quasi-Particle Lifetime at Finite Temperature

In Landau theory it is well-known that the quasi-particle lifetime at zero energy $\epsilon = 0$ and finite temperature T is simply obtained by replacing ϵ by T,

so that it varies as T^2. The diffusive case is more subtle. It turns out that in this case we cannot simply substitute ϵ by T. This is wrong in low dimension. Let us see why.

What is changed at finite temperature? The Fermi golden rule is modified to account for Fermi factors. The condition of filled or empty states has to be replaced by Fermi factors and (95) generalizes as

$$\frac{1}{\tau_{\rm ee}(\epsilon,T)} = 4\pi\nu_0^3 \int_{-\infty}^{\infty} {\rm d}\omega \int_{-\infty}^{\infty} {\rm d}\epsilon' F(\epsilon,\epsilon',\omega) W^2(\omega) \tag{112}$$

where $F(\epsilon,\epsilon',\omega)$ is a combination of Fermi factors $f_\epsilon = 1/({\rm e}^{\beta\epsilon}+1)$:

$$F(\epsilon,\epsilon',\omega) = f_{\epsilon'}(1-f_{\epsilon-\omega})(1-f_{\epsilon'+\omega}) + (1-f_{\epsilon'})f_{\epsilon-\omega}f_{\epsilon'+\omega} \; . \tag{113}$$

The first term in this expression is larger when $\epsilon > 0$. It describes the decay of an electron-like state above the Fermi level. The second term dominates when $\epsilon < 0$ and describes the decay of a hole-like state into the Fermi sea. For $\epsilon = 0$, both terms are equal. Integrating upon ϵ', we obtain

$$\frac{1}{\tau_{\rm ee}(\epsilon,T)} = 4\pi\nu_0^3 \int_{-\infty}^{\infty} {\rm d}\omega \; \omega W^2(\omega) f_{\epsilon-\omega} \frac{{\rm e}^{\beta\epsilon}+1}{{\rm e}^{\beta\omega}-1} \; . \tag{114}$$

This lifetime can also be obtained from the imaginary part of the self-energy of a quasi-particle in the presence of a screened interaction [27]. At zero temperature, we recover the result (108).

9.6 Quasi-Particle Lifetime at the Fermi Level

We now consider more specifically the lifetime of a quasi-particle *at the Fermi level ($\epsilon = 0$) and at finite temperature*. Physical properties such as conductance are expressed in terms of *single-particle* states at the Fermi level. It is thus essential to understand the range of validity of the description in terms of independent quasi-particles. From relation (114), we have [28]

$$\frac{1}{\tau_{\rm ee}(T)} = 8\pi\nu_0^3 \int_0^{\infty} {\rm d}\omega W^2(\omega) \frac{\omega}{\sinh\beta\omega} \; . \tag{115}$$

For the diffusion in free space, the matrix element $W^2(\omega)$ is proportional to $\omega^{d/2}$ (relation (107)), so that

$$\frac{1}{\tau_{\rm ee}(T)} = \frac{\pi d c_d}{2\nu_0} \int_0^{\infty} \frac{{\rm d}\omega}{\omega\sinh\beta\omega} \left(\frac{\omega}{E_c}\right)^{d/2} \; . \tag{116}$$

Therefore, in three dimensions, we have

$$\frac{1}{\tau_{\rm ee}(T)} = \frac{\sqrt{2}}{4\pi^2\nu_0} \int_0^{\infty} \frac{{\rm d}\omega}{\omega\sinh\beta\omega} \left(\frac{\omega}{E_c}\right)^{3/2} \simeq \frac{T}{\nu_0} \int_0^{T} \frac{{\rm d}\omega}{\omega^2} \left(\frac{\omega}{E_c}\right)^{3/2} \; , \tag{117}$$

that is

$$\frac{1}{\tau_{ee}(T)} \simeq \Delta \left(\frac{T}{E_c}\right)^{3/2} \qquad (d=3) \tag{118}$$

up to a numerical factor. Note that the exponent of the power law is the same as the exponent for the energy dependence of the lifetime at zero temperature (108). This result follows at once if we notice that relevant processes in the quasi-particle relaxation described by $\omega W^2(\omega)$ are those for which the energy transfer ω is of order T.

It would be tempting to generalize this result to any dimension and to conclude that $1/\tau_{\rm ee}(T) \propto T^{d/2}$. This is not correct for $d \leq 2$. In this case, the contribution of e-e processes with *low energy transfer* $\omega \simeq 0$ dominates and leads to a divergence in the integral (116). In order to cure this divergence, it is worth noticing that $\tau_{\rm ee}(T)$ represents precisely the lifetime of an eigenstate, so that the energy transfer ω cannot be defined with an accuracy better than $1/\tau_{\rm ee}$. Consequently, there is no energy transfer smaller than $1/\tau_{\rm ee}(T)$, so that the integral (116) needs to be cut off self-consistently for ω smaller than $1/\tau_{\rm ee}(T)$. For $d \leq 2$, we thus obtain a self-consistent relation for $\tau_{\rm ee}$:

$$\frac{1}{\tau_{\rm ee}(T)} \simeq \frac{1}{\nu_0} \int_{1/\tau_{\rm ee}}^{\infty} \frac{{\rm d}\omega}{\omega \sinh \beta\omega} \left(\frac{\omega}{E_c}\right)^{d/2} \simeq \frac{T}{\nu_0} \int_{1/\tau_{\rm ee}}^{T} \frac{{\rm d}\omega}{\omega^2} \left(\frac{\omega}{E_c}\right)^{d/2} \tag{119}$$

where the thermal factor has been replaced by a cutoff at $\omega \sim T$. In two dimensions, $1/\tau_{\rm ee}(T)$ is proportional to the temperature (within logarithmic corrections) :

$$\frac{1}{\tau_{\rm ee}(T)} \simeq \Delta \frac{T}{E_c} \ln \frac{E_c}{\Delta} \qquad (d=2)\ . \tag{120}$$

In one dimension, and since $T\tau_{\rm ee} \gg 1$, the integral becomes proportional to $\sqrt{\tau_{\rm ee}}$ so that the self-consistent relation leads to

$$\frac{1}{\tau_{\rm ee}(T)} \simeq \Delta \left(\frac{E_c}{\Delta}\right)^{1/3} \left(\frac{T}{E_c}\right)^{2/3} \qquad (d=1)\ . \tag{121}$$

9.7 Phase Coherence

The time (115) has been defined as the lifetime of a quasi-particle, generalizing the notion introduced by Landau to the case of a diffusive system in d dimensions. We should now evaluate the phase coherence time $\tau_\phi(T)$ which limits coherent effects like the weak localization correction (34). This time can be interpreted as the lifetime of the Cooperon. Its derivation consists in calculating *directly* the dephasing $\langle {\rm e}^{{\rm i}\Phi(t)} \rangle$ resulting from electron-electron interaction and accumulated between time-reversed conjugated multiple scattering sequences. To that purpose, the interaction between electrons is replaced by an effective interaction which describes the coupling of a single electron to the electromagnetic field created by the other electrons. This calculation [12, 29] is not developed here; see [1] for a detailed derivation.

An alternative and qualitative approach is to consider that phase coherence is limited by the lifetime of quasi-particles. Since the multiple scattering trajectories that are paired in the Cooperon are defined for a given energy state, they cannot interfere for times larger than $\tau_{ee}(T)$. This results in an irreversible dephasing between the trajectories and thus a loss of phase coherence. It is therefore natural to assume that

$$\tau_\phi(T) = \tau_{ee}(T) \; . \tag{122}$$

Indeed, the temperature dependences predicted in equations (118), (120) and (121) have been confirmed experimentally, in all dimensions by weak localization measurements.

It turns out that not only these two characteristic times $\tau_{ee}(T)$ and $\tau_\phi(T)$ are equal (within a numerical factor), but also that the two processes, *quasi-particles relaxation* and *phase relaxation*, are very similar. Finally let us remark that the introduction of the low-energy cutoff in (119) may appear as a handwaving and artificial way to handle the low energy divergence. We have shown recently that the profound reason for this divergence is that, for $d \leq 2$, *relaxation of quasi-particles* as well as the *phase relaxation* are not exponential [30].

Acknowledgments

Many points of view presented here have been developed during a long collaboration with Eric Akkermans, and are detailed in [1].

References

1. For a comprehensive review, see E. Akkermans, G. Montambaux: *Mesoscopic Physics of Electrons and Photons* (Cambridge University Press 2007)
2. Y. Imry, R. A. Webb: Scientific American **260**, 36 (1989); M. Peshkin, A. Tonomura: The Aharonov-Bohm effect. In: *Lecture Notes in Physics*, vol 340 (Springer-Verlag, Heidelberg 1989)
3. R. A. Webb, S. Washburn, C. P. Umbach, R. P. Laibowitz: Phys. Rev. Lett. **54**, 2696 (1985)
4. D. Yu. Sharvin, Yu. V. Sharvin: JETP Lett. **34**, 272 (1981)
5. Y. Aharonov, D. Bohm: Phys. Rev. **115**, 485 (1959)
6. B. Pannetier, J. Chaussy, R. Rammal, P. Gandit: Phys. Rev. Lett. **53**, 718 (1984); B. Douçot, R. Rammal: Phys. Rev. Lett. **55**, 1148 (1985)
7. M. Büttiker: IBM J. Res. Develop. **32**, 317 (1988)
8. C. W. J. Beenakker: Rev. Mod. Phys. **69**, 731 (1997)
9. R. P. Feynman, A. R. Hibbs: *Quantum Mechanics and Path Integrals* (McGraw-Hill, New-York 1965); G. Roepstorff: *Path Integral Approach to Quantum Physics : An Introduction* (Springer-Verlag, New-York 1994)
10. D. E. Khmelnitskii: Physica **126 B**, 235 (1984)
11. G. Bergmann: Phys. Rep. **107**, 1 (1984)

12. S. Chakravarty, A. Schmid: Phys. Rep. **140**, 193 (1986)
13. Notice that this correct result differs from the qualitative evaluation (33) only by a factor 2
14. B. Kramer, A. MacKinnon: Rep. Prog. Phys. **56**, 1469 (1993)
15. M. Abramowitz, I. A. Stegun: *Handbook of Mathematical Functions* (Dover 1972)
16. B. L. Altshuler, A. G. Aronov, B. Z. Spivak: JETP Lett. **33**, 94 (1981)
17. D. Mailly, M. Sanquer: J. Physique I France **2**, 357 (1992)
18. M. Pascaud, G. Montambaux: Phys. Rev. Lett. **82**, 4512 (1999); E. Akkermans, A. Comtet, J. Desbois, G. Montambaux, C. Texier: Ann. Phys. **284**, 10 (2000)
19. G. Montambaux, A. Akkermans: Quantum transport in disordered conductors. In: *Nanotubes : Science and Applications*, ed by A. Loiseau, P. Launois, P. Petit, S. Roche, J. P. Salvetat (Springer Verlag, Frontiers in Physics 2006)
20. B. L. Altshuler, A. G. Aronov: Electron-electron interaction in disordered conductors. In: *Electron-electron Interactions in Disordered Systems*, ed by A. L. Efros, M. Pollak (Elsevier 1985) p 1
21. There is also a contribution in the Cooperon channel. However, this contribution is corrected by higher order diagrams and turns out to be small
22. Y. Imry, Z. Ovadyahu: Phys. Rev. Lett. **49**, 841 (1982); F. Pierre, H. Pothier, P. Joyez, N. O. Birge, D. Esteve, M. Devoret: Phys. Rev. Lett. **86**, 1590 (2001)
23. From our simple picture, the coefficients λ_σ and λ_ρ are equal. A better description shows that these parameters are slightly different, see [1, 20]
24. Alternatively, the sum (104) on wave vectors can be replaced by an integral and we recover (107)
25. F. Pierre, H. Pothier, D. Esteve, M. H. Devoret: J. Low Temp. Phys. **118**, 437 (2000); F. Pierre: Ann. Phys. **26**, vol.4 (2001)
26. U. Sivan, Y. Imry, A. G. Aronov: Europhys. Lett. **28**, 115 (1994)
27. E. Abrahams, P. W. Anderson, P. A. Lee, T. V. Ramakrishnan: Phys. Rev. B **24**, 6783 (1981)
28. W. Eiler: J. Low Temp. Phys. **56**, 481 (1984)
29. B. L. Altshuler, A. G. Aronov, D. E. Khmelnitskii: J. Phys. C **15**, 7367 (1982)
30. G. Montambaux, E. Akkermans: Phys. Rev. Lett. **95**, 016403 (2005); C. Texier, G. Montambaux: *Theory of quantum transport in multiterminal networks of diffusive wires*, in preparation

Transport and Persistent Currents in Mesoscopic Rings: Interplay Between Electron-Electron Interaction and Disorder

Georges Bouzerar

Laboratoire Louis Néel, CNRS, 25 avenue des Martyrs BP 166, F-38042 Grenoble Cedex 9, France. bouzerar@ill.fr

1 Introduction

The aim of this article is to provide a very general and pedestrian introduction to the notion of persistent currents in mesoscopic systems. Thus, this review is not exhaustive and mainly addressed to non-experts. Step by step, it will be shown that a single-particle picture is insufficient to explain the magnitude and sign of the measured persistent currents. The main idea would be to analyze whether the interplay between disorder and electron-electron interaction could eventually explain this discrepancy. This review gives the opportunity to emphasize the challenges raised by the experiments. Indeed, it will be shown that the amplitude and sign of the measured currents remain until now an open issue.

1.1 What is a Mesoscopic System?

Before giving more details one can say that mesoscopic physics is the missing bridge between macroscopic physics and atomic physics [1–4]. One can also imagine that starting with a macroscopic system we reduce its volume further and further: the mesoscopic regime starts when the system is small enough that its physical properties cannot anymore be described by standard classical physics, the quantum description becomes the relevant one. The typical length L of a mesoscopic system is between microscopic (Å) and macroscopic (10^{-3} m) lengthscale, it is typically of the order of a micron. To be more precise the typical length L should be much smaller than the coherence length L_ϕ. One can naively say that this important lengthscale separates the "classical world" from the "quantum world". L_ϕ is a typical lengthscale beyond which the coherent propagation of the electrons in the material is lost. An electron

propagates coherently when it keeps the memory of the phase of its wave function, the scattering with the defects are *elastic*. During elastic scatterings, the energy is conserved and the dephasing induced into the electronic wave function are well-defined and deterministic (calculable). These scattering are for example due to *static* defects in the material (dislocation, non-magnetic impurity, etc.). We define l_e as the average distance between two elastic processes. On the other hand, the processes during which the energy is not conserved and that suppress phase coherence are called *inelastic processes*. As a consequence the energy level of the electrons will have a finite lifetime. These inelastic processes introduce uncontrollable and unpredictable dephasing into the electronic wave function. The inelastic scatterings are the source of dissipation in the system. The inelastic processes which destroy the quantum interferences are for example (i) the electron-phonon scatterings, (ii) the electron-electron interaction, or (iii) the scattering between an electron with an impurity which possess an internal dynamical degree of freedom (e.g., the spin in the case of a magnetic impurity). This implies that mesoscopic physics is essentially a very low temperature physics. To get rid of the effects of inelastic processes the solution is to work at sufficiently low temperature. Typically below 1 Kelvin the phonons are "frozen out" and the electron-electron scattering is the dominant decoherence mechanism if no other source of inelastic processes is present. Also note that the electronic dephasing time τ_ϕ increases by lowering temperature. For quasi-1D systems $\tau_\phi \propto T^{-p}$, where $p \simeq 2/3$ [2]. Typically most of the experiments are performed usually below 100 mK.

1.2 Ballistic, Diffusive and Localized Regimes

Because of the finite size of the samples and the finite concentration of elastic scatterers, which we denote as disorder, we will see that the electron motion can be of different nature. We have already defined the typical sample size L and the elastic mean free path l_e (L_ϕ is here assumed to be infinite). The third relevant typical lengthscale is the Fermi wave length $\lambda_{\rm F}$. The strength of disorder is defined by comparing these three lengthscales L, l_e and $\lambda_{\rm F}$.

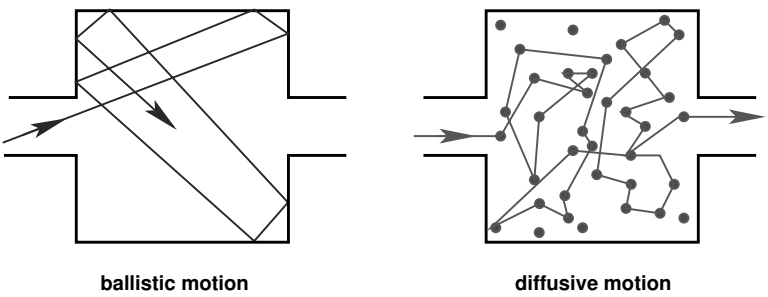

Fig. 1. Diffusive and ballistic motion of an electron inside a cavity

- Weak disorder:
  In this regime $\lambda_F \ll l_e$, the elastic collision between electrons and impurities are well-separated. The notion of trajectory is well-defined and a semi-classical approach is meaningful. This is the case for normal metal as gold and copper used in the experiments that we will discuss later on. For gold and copper $\lambda_F \sim 1\,\text{Å}$ and $l_e \sim 100\,\text{Å}$, and for semiconductors $\lambda_F \sim 4\,\text{Å}$ and $l_e \sim 10\,\mu\text{m}$. In the weak disorder regime one can distinguish two different situations (see Fig. 1):
  - *The ballistic regime*: This regime is characterized by the inequality $\lambda_F \leq L \leq l_e$. This is the very weak disorder regime (very low density of scatterers). The trajectories of the electrons are a succession of reflections on the sample boundaries.
  - *The diffusive regime*: It corresponds to $\lambda_F \leq l_e \leq L$. The electron undergo a large number of collisions with the defects during its trajectory across the sample. In this case the trajectories can be assimilated to classical diffusive random walks. A new timescale appears naturally, the diffusing time through the sample $\tau_D = L^2/D$ where the diffusion constant $D = v_F l_e/d$, v_F being the Fermi velocity and d the dimensionality of the system.
- Strong disorder:
  In this regime $\lambda_F \geq l_e$, the notion of trajectory looses its meaning, and the semi-classical approach becomes impossible. Thus we have to manage with a purely quantum regime. The electron wave function decays exponentially with the distance and a novel characteristic lengthscale appears ξ, the *localization length*. In such Anderson insulator systems [5] electrons are confined in boxes of volume ξ^d.

1.3 Persistent Currents

In 1983 Buttiker, Imry and Landauer have suggested the existence, at equilibrium, of persistent currents flowing without dissipation in *isolated* metallic rings pierced by a magnetic flux [6–8]. Figure 2 gives a schematic view of a quasi-1D disordered mesoscopic ring. These currents result from the quantum interferences of the electron wave functions [9]. It is important to stress that these currents have nothing to do with the nature of the persistent currents in superconducting materials. However, they could be compared to the diamagnetic currents induced by a magnetic field in atoms or molecules.

We now show how the persistent current is related to the ground-state energy of an electronic systems. For simplicity we assume a single-band Hamiltonian of non-interacting fermions in presence of disorder. The Hamiltonian $\mathcal{H}(\phi)$ reads,

$$\mathcal{H}(\phi) = \frac{1}{2m}\left(\boldsymbol{p} - \frac{e\boldsymbol{A}}{c}\right)^2 + V(x) \tag{1}$$

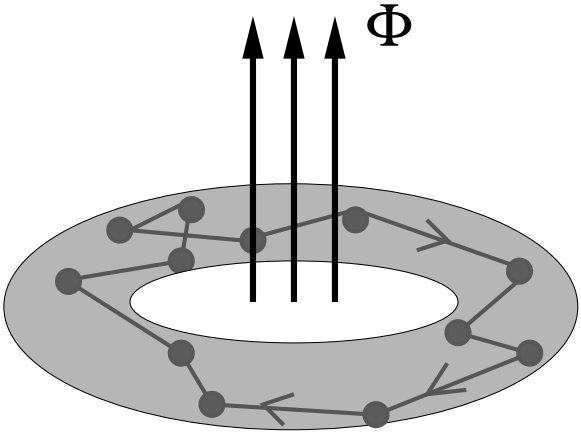

Fig. 2. Schematic diagram of a quasi-1D disordered mesoscopic ring threaded by magnetic flux. Closed trajectory shows that the electrons undergo multiple scatterings with the impurities

where $\boldsymbol{A}$ is the vector potential and its integral $\oint \boldsymbol{A} \cdot \boldsymbol{dl} = \phi$. Here $V(x)$ is the local potential which contains both the contribution due to the crystalline structure and impurities. With L as the ring perimeter, the periodicity of the potential reads,

$$V(x+L) = V(x) \tag{2}$$

and thus,

$$\langle x|\Psi_n\rangle = \langle x+L|\Psi_n\rangle \ . \tag{3}$$

$|\Psi_n\rangle$ is an eigenstate of the one-body Hamiltonian,

$$\mathcal{H}|\Psi_n\rangle = e_n|\Psi_n\rangle \ . \tag{4}$$

Both $|\Psi_n\rangle$ and e_n depend on ϕ. Let us define the function Φ_n as

$$\langle x|\Psi_n\rangle = \Phi_n(x)\mathrm{e}^{-2\mathrm{i}\pi \frac{\phi}{\phi_0}\frac{x}{L}} \tag{5}$$

where $\phi_0 = hc/e$ is the flux quantum. It immediately follows that

$$\left[\frac{1}{2m}p^2 + V(x)\right]\Phi_n = e_n\,\Phi_n \tag{6}$$

and the boundary conditions become,

$$\Phi_n(x+L) = \mathrm{e}^{2\mathrm{i}\pi \frac{\phi}{\phi_0}}\,\Phi_n(x) \ . \tag{7}$$

Thus with the simple gauge transformation, the Shrödinger equation does not depend on the magnetic flux ϕ but the boundary conditions for the eigenstates have changed. Equation (7) implies that the eigenvalues e_n are continuous functions of ϕ with ϕ_0 periodicity. On the other hand it also shows that the effect of the flux is to produce a shift of the momentum by $2\pi\phi/L\phi_0$. Thus the eigenvalues are,

$$e_n(\phi) = e_n(k_n + \frac{2\pi}{L}\frac{\phi}{\phi_0}) \,. \tag{8}$$

Let us now evaluate the current contribution of an energy level e_n. The velocity associated to this energy level is given (we set for simplicity $h = c = 1$) by

$$v_n = \frac{\partial e_n}{\partial k_n} = \frac{L\,\phi_0}{2\,\pi}\frac{\partial e_n}{\partial \phi} \,. \tag{9}$$

Thus the current of an electron in the energy level e_n is,

$$i_n = -\frac{ev_n}{L} = -\frac{1}{2\pi}\frac{\partial e_n}{\partial \phi} \,. \tag{10}$$

The total current in the ring is,

$$I(\phi) = -\frac{1}{2\pi}\sum_{n=1}^{N_e}\frac{\partial e_n}{\partial \phi} \tag{11}$$

where N_e is the total number of electrons in the ring. In terms of the total ground state energy $E_0(N_e) = \sum_{n=1}^{N_e} e_n$, the total persistent current can be re-expressed as

$$I(\phi) = -\frac{1}{2\pi}\frac{\partial E_0(N_e)}{\partial \phi} \,. \tag{12}$$

Note that this expression is in fact very general, valid in any dimension, in presence of disorder and electronic correlations. At finite temperature, the ground state energy $E_0(N_e)$ should be replaced by the total free energy $F(\phi)$.

2 Persistent Currents in Non-Interacting Systems

2.1 One-Dimensional Ring

In this subsection we consider the case of a pure one-dimensional ring threaded by a magnetic flux ϕ. We will discuss both the cases $T = 0\,\mathrm{K}$ for which $L_\phi = \infty$ and the finite temperature case for which L_ϕ becomes finite.

The Clean Case at $T = 0\,\mathrm{K}$

We have seen previously that the persistent current is ϕ_0 periodic. Let us now show that it also depends on the parity of the electron number N_e in the ring. The eigenvalues of $\mathcal{H}$ are

$$e_n(\phi) = \frac{2\pi^2}{mL^2}\left(n + \frac{\phi}{\phi_0}\right)^2 , \tag{13}$$

where $n = 0, 1, \ldots, L-1$. Thus the current associated to this level is

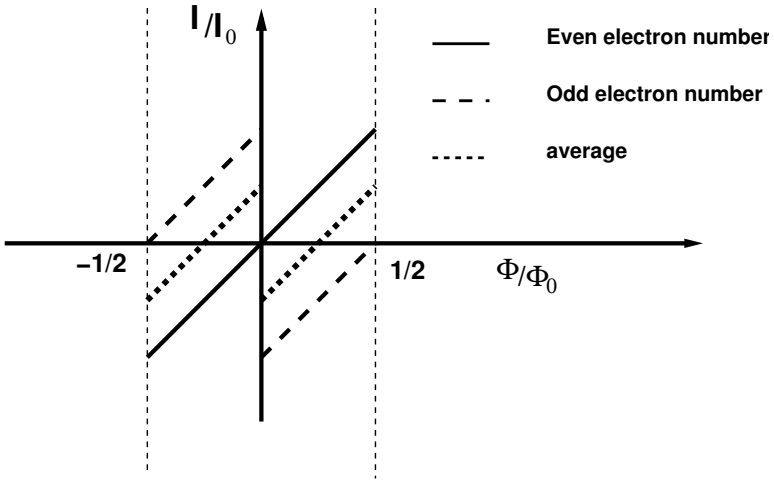

Fig. 3. Persistent current in 1D non-disordered mesoscopic ring for even, odd and averaged N_e

$$i_n(\phi) = -\frac{2\pi}{mL^2}\left(n + \frac{\phi}{\phi_0}\right) . \tag{14}$$

It is then straightforward to get the total current as

$$I(\phi) = -I_0\,\frac{2\phi}{\phi_0} \qquad \text{if } N_e \text{ even and } \phi \in [-\tfrac{1}{2}, \tfrac{1}{2}] , \tag{15}$$

$$I(\phi) = -I_0\left(\frac{2\phi}{\phi_0} - 1\right) \qquad \text{if } N_e \text{ odd and } \phi \in [0, 1] \ . \tag{16}$$

Here we have introduced,

$$I_0 = \frac{e}{\tau_{\mathrm{B}}} = \frac{e v_{\mathrm{F}}}{L} \tag{17}$$

where the Fermi velocity is $v_{\mathrm{F}} = k_{\mathrm{F}}/m$. The currents for even and odd number of electrons are shown in Fig. 3.

The decomposition of $I(\phi)$ into its harmonics leads to [10],

$$I(\phi) = \sum_{l}^{\infty} \frac{2I_0}{l\pi} \cos(lN_e\pi) \sin(2l\pi \frac{\phi}{\phi_0}) \; . \tag{18}$$

This expression shows that the sign of the harmonics depends on the parity of the electron number in the ring. The odd harmonics have opposite sign for odd and even N_e. Thus the average current over N_e leads to the suppression of all odd harmonics, the current becomes $\phi_0/2$ periodic. This is illustrated in Fig. 3. It is also interesting to remark that even harmonics are always of positive sign. We will come back to these points in the next sections.

Effects of Finite Temperature

The calculations are performed at fixed chemical potential μ and finite temperature T. The total persistent current is the sum of the contribution of each individual level weighted by the probability of occupation of the corresponding level,

$$I(T, \phi) = \sum_{n} p_n i_n(\phi) \tag{19}$$

where $p_n = 1/(\mathrm{e}^{\beta(e_n - \mu)} + 1)$.

In the case of metallic systems and in the limit $\mu \gg k_{\mathrm{B}}T$, the total current is given by [10],

$$I(T) = -4k_{\mathrm{B}}T \sum_{n}^{\infty} \sum_{l}^{\infty} \exp\left[-\frac{l(2n-1)T}{T^*}\right] \cos(lk_{\mathrm{F}}L) \sin(2l\pi\phi) \tag{20}$$

where a characteristic temperature scale T^* appears. It separates the high from the low temperature regime,

$$k_{\mathrm{B}}T^* = \frac{\Delta}{2\pi^2} = \frac{v_{\mathrm{F}}}{\pi L} \tag{21}$$

where $\Delta = 2\pi v_{\mathrm{F}}/L$ is the mean level spacing.

After performing the summation over n one finds,

$$I(T) = -\frac{4I_0T}{\pi\, T^*} \sum_{l}^{\infty} \frac{\exp(-lT/T^*)}{1 - \exp(-2lT/T^*)} \cos(lk_{\mathrm{F}}L) \sin(2l\pi\phi) \; . \tag{22}$$

This expression shows that the harmonics of the current are exponentially suppressed by the temperature,

$$I_l = -\frac{4I_0T}{\pi\, T^*} \frac{\exp(-lT/T^*)}{1 - \exp(-2lT/T^*)} \cos(lk_{\mathrm{F}}L) \; . \tag{23}$$

This expression also implies that the number of harmonics ΔN that contribute dominantly to the persistent current is typically of the order of $\Delta N \sim T^*/T$.

Effects of Static Disorder

As mentioned before static disorder implies that the scattering of the electrons with the defects is phase-coherent. To study the effects of disorder, it is particularly convenient to consider the Anderson model,

$$\mathcal{H} = -\frac{t}{2}\sum_i \left[\exp(2\mathrm{i}\pi\phi/L)\, c_i^\dagger\, c_{i+1} + \text{h.c.}\right] + \sum_i w_i\, n_i\ . \tag{24}$$

The first term describes the kinetic part (hopping of spinless fermions) and disorder is introduced in the on-site potentials w_i. The variable w_i are chosen randomly in the interval $[-W/2, W/2]$, where W measures the strength of the disorder. In 1D case, it is known that for uncorrelated disorder, all eigenstates are localized for an infinitesimal amount of disorder. The localization of the eigenstates results from the quantum interference of the electronic wave functions. The manifestation of quantum interferences in presence of disorder in small devices is also known as the weak localization phenomena [11–13]. For general references on the Anderson localization see for example [14–17], and also see the fundamental works of Thouless [18, 19] which is at the origin of the scaling theory of localization and of the majority of quantum transport theories in disordered systems.

Note that in the special case of a pure 1D system, there is no possibility of diffusive regime. Thus, the two possible regimes are respectively the ballistic and localized one. The ballistic regime corresponds to $L \leq \xi$ and the localized to $L \gg \xi$, ξ being the localization length. For the tight-binding model in the half-filled band case, the localization length has been calculated for long wires in the two limiting situations [20, 21]. These results are

- weak disorder regime, i.e., $W \ll t$: $\xi = 25t^2/W^2$,
- strong disorder regime, i.e., $W \gg t$: $\xi = 1/\ln(W/t)$.

In the weak disorder regime, Cheung et al. [10] have shown that the persistent current can be well approximated by the first harmonic only, the higher harmonics being much smaller. The amplitude of the current is found to decay exponentially with the system size,

$$\langle I(\phi)\rangle = \frac{I_0}{2}\exp\left(-\frac{L}{\xi}\right)\sin(2\pi\phi)\ . \tag{25}$$

In the case of very dirty ring ($W \gg t$) they found that

$$\langle I(\phi)\rangle = \frac{I_0}{2}\frac{W}{\pi\, t}\mathrm{e}^{-L\ln(W/t)}\sin(2\pi\phi)\ . \tag{26}$$

This shows that both for weak and strong disorder, the currents are always suppressed exponentially with increasing system size.

2.2 The Multi-Channel Case

The multi-channel case is more interesting since it is closer to real systems. Indeed the experiments which will be discussed in the following are performed with quasi-1D rings. Motivated by the experimental studies, a large amount of theoretical work was devoted to persistent currents in multi-channel mesoscopic rings, especially in the diffusive regime. In this paragraph, we will only present some results, those which appear to be the most relevant to the experiments both in the diffusive and ballistic regimes.

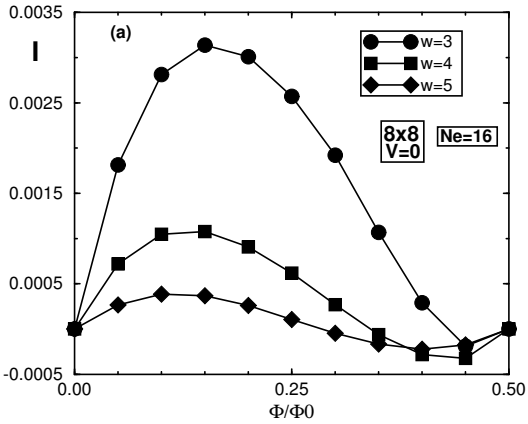

Fig. 4. $I(\phi)$ as a function of ϕ/ϕ_0 in the absence of interaction ($V = 0$) in a multi-channel ring of size 8×8 at quarter-filling $\langle n \rangle = 0.25$ for various values of W. For each value of W an average over 1000 configurations of disorder was done. Figure from [22]

First, it is interesting to estimate the typical value of the persistent current in the case of a multi-channel (quasi-1D) ring in the weak disorder regime (ballistic regime). If we assume naively that each individual channel provides a current of order I_0 with a random sign as the number of electron varies in the various channels, we would get for the amplitude of the typical current,

$$I_{\rm typ} = \sqrt{M} I_0 \tag{27}$$

where $I_{\rm typ} = \sqrt{\langle I^2 \rangle}$ and $M = k_{\rm F}^2 S/4\pi$ is the number of transverse channel, S being the cross-section of the ring.

On the other hand, in the diffusive regime and in absence of electron-electron correlation, an analytical expression was obtained within Green's functions approach by Cheung et al. [23]. The important conclusion of their study was that the typical current does not depend on the transverse dimensions and thus on the number of channel,

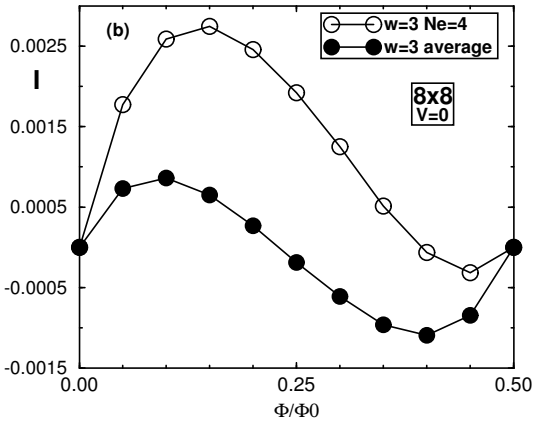

Fig. 5. $I(\phi)$ as a function of ϕ/ϕ_0 in the absence of interaction ($V = 0$) in a multichannel ring of size 8×8 for $W = 3t$. The open symbols are for $N_e = 4$ while the filled symbols correspond to an average over the rings with $N_e = 2, 3, ...6$. In both the cases an average over 1000 configurations of disorder was performed. Figure from [22]

$$I_{\rm typ} \propto \frac{e}{\tau_{\rm D}} = I_{\rm D} \ . \tag{28}$$

Note that this expression can be rewritten as $I_{\rm typ} = ev_D/L$, where $v_D = v_{\rm F} l_e/L$ is the diffusion velocity along the ring. On the other hand, the amplitude of the average persistent current $\langle I \rangle$ was shown to increase with the number of channels and decrease exponentially with the disorder strength,

$$\langle I \rangle \approx \left(\frac{M l_e}{L}\right)^{1/2} I_0 \exp\left(-\frac{L}{2l_e}\right) \sin\left(2\pi \frac{\phi}{\phi_0}\right) + \text{higher harmonics} \ . \tag{29}$$

Note that the average current is ϕ_0 periodic.

Later, a numerical study based on the exact digitalizations of the one particle Hamiltonian, performed by Montambaux et al. [24–26] had lead to quite different results. Let us summarize their results. In the zero disorder limit, it was found that the amplitude of the average current $\langle I \rangle \propto I_0$ and independent of the number of channels. It is to be noted that the symbol $\langle \cdots \rangle$ means average over many rings that have a given number of electrons that fluctuate from one ring to the another. As in the pure 1D case discussed previously, because of their random sign, the odd harmonics are suppressed after averaging. Thus the total average current is $\phi_0/2$ periodic. In the presence of disorder it was shown that $\langle \cdots \rangle_N = \langle \cdots \rangle_{N,D}$ where N denotes average over the electron number and D over the disorder configurations. Concerning the average current $\langle I \rangle$ it was found that, as in the clean case, the periodicity

remains $\phi_0/2$ while the amplitude decreases with the number of channels and the variation with disorder is a power law

$$\langle I \rangle = \sqrt{\frac{l_e}{ML}} I_0 \ . \tag{30}$$

This is in disagreement with the analytical results of Cheung et al., see (29), where the currents increase with the number of channels but are exponentially suppressed by disorder. The main and fundamental origin of this discrepancy is the following. The average persistent current is calculated at fixed chemical potential (grand canonical ensemble) in the analytical study whilst in the exact diagonalization approach the calculations were performed with a fixed number of electrons in each disordered ring (canonical ensemble). As already mentioned, we remind that in the exact diagonalization calculations the number of electrons fluctuates from one ring to another (see Fig. 4 and Fig. 5).

On the other hand, it was shown that the typical value of the current is ϕ_0 periodic and the typical values of the first and second harmonics appeared to be independent of the number of channels, in agreement with the analytical results of Cheung et al.,

$$I^p_{\rm typ} = C_p \frac{l_e}{L} I_0 \ . \tag{31}$$

The ratio $C_1/C_2 \approx 3.5 \pm 0.3$ was found to be independent of the characteristics of the ring.

Later, Von Oppen and Riedel performed the calculation of the averaged persistent current analytically in disordered mesoscopic ring with a fixed number of electrons [27]. They obtained for the current an expression which is independent of the disorder strength and inversely proportional to the number of channels,

$$\langle I^{2p} \rangle = I_0 / 2\pi M \ . \tag{32}$$

Unfortunately, this result is in disagreement with the exact diagonalization calculation mentioned above. An interesting work, which sheds light on the relation between the currents calculated in the grand canonical ensemble versus the canonical ensemble, is given in [28].

3 Experimental Results

To measure experimentally the persistent currents in quasi-1D disordered rings there are two kind of possible realizations. The first one consists in measuring the response of a single isolated mesoscopic ring. This require a very sensitive device (SQUID). The second approach consists in measuring simultaneously the total magnetic response of a very large number of rings

disconnected from each other. This setup requires much less sensitivity but raise some important questions related to averaging over disorder configurations. Indeed, one crucial question would be: What is the relation between the currents measured in a single isolated ring with respect to that measured in a multi-ring experiment where it is averaged over a large number of disorder configurations? This will be discussed in the following. These two different type of experiments were realized for the first time more than a decade ago. As we will see, even today, the measured current amplitudes are still a challenge to theory.

3.1 Single Ring Experiments

The measurement of the magnetization of a single isolated mesoscopic ring is an experimental challenge which has been accomplished respectively by Chandrasekhar et al. [29] in 1991 on a isolated golden ring and by Mailly et al. [30] in 1993 on a ring made from high mobility GaAs-GaAlAs heterojunction. In both experiments the magnetic moment was measured with a highly sensitive micro-SQUID. In both experiments, and as expected, the currents were found to be periodic of period ϕ_0.

- In the golden ring experiment the regime was diffusive and the characteristics of the ring were the following: $L \propto 3\,\mu\text{m}$, $L_\phi \geq L$, $l_e \approx 20\,\text{nm} \gg \lambda_\text{F}$, $M \approx 10^5$. The magnitude of the measured current I_typ was of the order of the current in a 1D clean ring. According to what was previously discussed in this regime we would expect a typical current independent of the number of channel, but of much smaller amplitude. Indeed, in the one particle framework we expect $I_\text{typ} \approx I_\text{D} = e/\tau_\text{D} = 10^{-2} I_0$ where the ratio $l_e/L \approx 0.01$. This disagreement with the theoretical predictions based on non-interacting electrons systems is still today an open issue. In the following we will discuss a mechanism which may explain this discrepancy, the electron-electron interaction.
- In the second experiment performed on semiconductors the magnitude of the measured signal was $I_\text{typ} \propto I_0$. This is in agreement with the predictions of one particle theory. This is in fact not really surprising since the disorder was very weak (i.e., the system was almost ballistic) and the number of channel M was relatively small. Indeed in the semiconductor ring used in this experiment $L \propto 8\,\mu\text{m}$ and $l_e \approx 5\,\mu\text{m}$ and $\sqrt{M} \approx 2$.

3.2 Many Rings Measurements

This kind experiment was first performed in 1990 by Levy et al. [31]. They have measured the total response of $N = 10^7$ disordered mesoscopic rings of copper. The distance between the rings was such that the rings could be considered as isolated from each other. The characteristics of the rings were the following: the perimeter $L \propto 2.2\,\mu\text{m}$, the cross-section $L_y\, L_z \propto 0.3\,(\mu\text{m})^2$,

the mean free path was estimated to be $l_e \propto 200\,\text{Å}$. The rings were in the diffusive regime $L > l_e > \lambda_F$. The temperature of the order of 400 mK was such that the coherence length was bigger than the size of the rings, i.e., $L < L_\phi$. In this experiment, the physical quantity which is measured is the induced total magnetization $\langle M \rangle$:

$$\langle M \rangle = \frac{1}{N} \sum_{k=1}^{N} \mu_k \ , \tag{33}$$

where $\mu_k = Si_k$, i_k is the persistent current flowing in the k-th ring and $S = L^2/4\pi$. The total average persistent current is,

$$\langle I \rangle = \frac{1}{NS} \sum_{k=1}^{N} \mu_k \ . \tag{34}$$

The measured current was $\phi_0/2$ periodic and its magnitude $\langle I \rangle_{\text{exp}} \simeq 3.10^{-3} I_0$. The explanation of $\phi_0/2$ periodicity was already given in the previous section. However, the theoretical estimate of the amplitude is, as in the single ring experiment, inconsistent with the measured value. By using the theoretical expression given in (30), valid in the diffusive regime, one finds a current $\langle I \rangle_{\text{th}} \simeq 10^{-4} I_0$. Thus the calculated amplitude is about two orders of magnitude small.

3.3 Recent Experiments and the Sign of Persistent Current

All the theoretical calculations which were done within the framework of one particle picture lead to the fact that the sign of the second harmonics is always positive (paramagnetic). After the experiment of Levy et al. where the sign of the current was not determined precisely, many efforts were made to, first confirm the validity of the measured current and also to determine the sign of the second harmonic. A first attempt was by Jariwala et al. [32] in 2001. They measured the response of a very small ensemble of only 30 diffusive mesoscopic gold rings. They found that both the first and second harmonics were negative (diamagnetic). The fact that the first harmonic was also negative is not so relevant. One would expect for a larger ensemble of rings the disappearance of the odd harmonics. One year later, Deblock et al. [33] measured the response of an ensemble of 10^5 GaAs disconnected rings. The measured amplitude was in agreement with previous experiment and the sign of the current was found to be negative. In the same year, another experiment was performed by the same people on a large ensemble of 10^5 silver rings in the diffusive regime [34]. They found that both the periodicity and amplitude of the current were in agreement with the pioneering experiment of Levy et al. They also clearly demonstrated the diamagnetic nature of the induced current.

4 Persistent Currents in Interacting Electron Systems

In the previous section we have seen that the one particle theory cannot explain neither the magnitude of the persistent currents in the diffusive regime nor the diamagnetic sign of the second harmonic. It is thus natural to wonder if the electron-electron correlations are responsible for this discrepancy. Is the effects of disorder are compensated by the effects of electron-electron correlations? To include in the theory the correlation effects there are in fact many possible ways. The possible theoretical approaches that were used are: the perturbation theory, the Green's functions methods, the diagrammatic approaches, the Renormalization Group (RG) methods, the exact diagonalization techniques, the unrestricted Hartree-Fock approach, the Density Matrix Renormalization Group (DMRG) method, etc. It is clear that the detailed discussion and comparison between so many possible different techniques would require more than a chapter. Thus, we will discuss only some results, and especially focus our attention on the calculations done in the framework of perturbation theory, perturbative RG approach, exact diagonalization method and Unrestricted Hartree-Fock approach.

4.1 The One-Dimensional Case

Persistent Current, Optical Conductivity and Charge Stiffness

To describe the motion of electrons in 1D disordered ring we consider the Anderson Hamiltonian to which the electron correlation part is added (the $t-V-W$ model). The many-body Hamiltonian reads,

$$\mathcal{H} = -\frac{t}{2}\sum_i \left[\exp\left(2\mathrm{i}\pi\phi/L\right) c_i^\dagger\, c_{i+1} + \text{h.c.}\right] + V\sum_i n_i\, n_{i+1} + \sum_i w_i\, n_i \ . \quad (35)$$

There are two free parameters, the strength of the interaction term V and the strength of disorder W. The hopping integral is set to $t = 1$.

We have seen in the non-interacting case that the total current is given by the first derivative of the ground state energy $E(\phi)$ with respect to flux. This expression is still valid in the many-body picture. This can be easily shown by starting with the definition of current

$$I(\phi) = \langle \Psi_0 | \hat{j}_x | \Psi_0 \rangle \ , \quad (36)$$

where the current operator is

$$\hat{j}_x = -\frac{1}{2\pi}\frac{\partial \mathcal{H}}{\partial \phi} \quad (37)$$

and $|\Psi_0\rangle$ is the many-body ground state. In our case the current operator is,

$$\hat{j}_x = -\frac{\mathrm{i}t}{2}\sum_i [\exp\left(2\mathrm{i}\pi\phi/L\right) c_i^\dagger\, c_{i+1} - \text{h.c.}] \ . \quad (38)$$

Another important physical quantity which can be calculated and directly compared to experimental measurement is the optical conductivity. The real part of conductivity can be calculated within the framework of linear response theory, and it can be shown that the real part of conductivity has the following form [35, 36],

$$\Re\mathrm{e}\,\sigma_{xx}(\omega) = \sigma_1(\omega) = \pi\, D\, \delta(\omega) + \sigma_{\mathrm{reg}}(\omega)\ . \tag{39}$$

The weight of the Dirac peak measures the contribution at zero frequency to the optical conductivity, this term is called the Drude weight peak. When the sample is connected to leads, the inverse of this term provides the resistivity value. In the thermodynamic limit, the Drude weight or stiffness is the order parameter which tells us about the nature of the system [37]:

- Metallic phase: $D \neq 0$,
- Insulating phase: $D = 0$.

This term is also a direct measure of the sensitivity of the energy spectrum to the change of the boundary conditions limit. The regular part is given by the Kubo formula,

$$\sigma_{\mathrm{reg}}(\omega) = \frac{\pi}{L} \sum_{m\neq 0} \frac{|\langle\Psi_0|\hat{j}_x(\Phi_m)|\Psi_m\rangle|^2}{E_m - E_0} \delta(\omega - (E_m - E_0)) \tag{40}$$

where $|\Psi_m\rangle$ is the mth excited states and E_m is the corresponding energy. Note that, depending on the parity of the number of electrons $\Phi_m = 0$ or 1. For this value of the flux the ground state energy is minimal.

There exists several sums rules and in particular one which relates the total Kinetic energy $\langle 0|\hat{T}(\Phi_m)|0\rangle$ to the integral of the optical conductivity,

$$\int_0^\infty \sigma_1(\omega)\mathrm{d}\omega = -\frac{\pi\, e^2}{2L}\, \langle 0|\hat{T}(\Phi_m)|0\rangle\ . \tag{41}$$

This expression is very important and useful. If we assume that the kinetic part is weakly affected by both disorder and correlation, then an important consequence emerges from the sum rule is that we can directly calculate the transfer of weight from the Drude peak to the regular part (finite frequency). Note also that this expression allows to calculate the Drude peak D,

$$\frac{D}{e^2} = -\frac{\langle\hat{T}\rangle}{L} + \frac{2}{L} \sum_{n\neq 0} \frac{|\langle\Psi_0|\hat{j}_x|\Psi_n\rangle|^2}{E_n - E_0}\ . \tag{42}$$

However to get the Drude weight directly from this term requires much computational efforts since it requires the knowledge of the complete excitation spectrum, eigenvalues and eigenvectors. We will see now that the Drude

weight can in fact be calculated directly from the ground-state energy. Let us expand the Hamiltonian to second order in ϕ,

$$\mathcal{H}(\phi) = \mathcal{H}(0) + \frac{2\pi\phi\,\hat{j}_x(0)}{L} - \frac{(2\pi\,\phi)^2\,\hat{T}(0)}{2L^2} + \cdots \,. \tag{43}$$

Thus, to second order, the ground state energy is,

$$E_0(\phi) = E_0(0) + \frac{2\pi\phi}{L} < 0|\hat{j}_x(0)\,|0> - \frac{(2\pi\,\phi)^2}{2L^2} < 0|\hat{T}(0)|0> + \frac{(2\pi\,\phi)^2}{L^2} \sum_m \frac{|\langle\psi_0|\hat{j}_x(0)|\psi_m\rangle|^2}{E_m(0) - E_0(0)} \tag{44}$$

where $|\psi_0\rangle$ is the ground state with energy $E_0(0)$ and $|\psi_m\rangle$ are the excited states with energy $E_m(0)$ in the absence of external flux. Then it follows,

$$\frac{L}{4\pi^2}\frac{\partial^2 E_0}{\partial\,\phi^2} = -\frac{1}{L} < 0|\hat{T}(0)|0> + \frac{2}{L}\sum_{m\neq 0} \frac{|<0|\hat{j}_x(0)|m>|^2}{E_m(0) - E_0(0)} \,. \tag{45}$$

After identification with the Kubo formula (42), we get,

$$\frac{D}{e^2} = \frac{L}{4\pi^2}\frac{\partial^2 E_0}{\partial\,\phi^2} \,. \tag{46}$$

Hence, the Drude weight is the second derivative of the ground-state energy with respect to flux. This also explains why Drude weight is also called stiffness.

4.2 Numerical and Analytical Results

The Case of Repulsive Interaction

A) The clean non-interacting case: $W = 0$ and $V = 0$

In the previous section we have already calculated the amplitude of the persistent currents. It is interesting now to estimate the optical conductivity. In the absence of disorder the Hamiltonian commute with the current operator,

$$[\hat{j}_x, \hat{H}] = 0 \,. \tag{47}$$

Thus, the conductivity reduces to the Drude peak (perfect metal),

$$D = -\frac{e^2\langle\hat{T}\rangle}{L} \,. \tag{48}$$

A straightforward calculation leads to the following expression for the Drude weight of a finite system,

$$\frac{D(L,\langle n\rangle)}{e^2\,t} = \frac{1}{\pi}\sin(\pi\langle n\rangle) + \frac{\pi}{3L^2}\sin(\pi\langle n\rangle) + \cdots \tag{49}$$

where $\langle n\rangle = N_e/L$ is the spinless fermion density. Thus in the thermodynamic limit

$$\frac{D_\infty(\langle n\rangle)}{e^2\,t} = \frac{1}{\pi}\sin(\pi\langle n\rangle)\ . \tag{50}$$

Note that in the limit of low density of carrier ($\lambda_{\rm F} \gg a$) we get

$$\int_0^\infty \Re\mathrm{e}\,(\sigma(\omega))\mathrm{d}\omega = \frac{\pi D}{2} = \frac{\pi\langle n\rangle e^2}{2m_{\rm eff}}\ , \tag{51}$$

where the effective mass is $m_{\rm eff} = 1/t$.

This expression is similar to that obtained with the classical Drude expression. Indeed according to the classical Drude model the real part of the classical conductivity $\sigma_{\rm cl}(\omega)$ is,

$$\Re\mathrm{e}\,(\sigma_{\rm cl}(\omega)) = \frac{ne^2\tau_0}{m(1+\omega^2\tau_0^2)} \tag{52}$$

which directly leads to,

$$\int_0^\infty \Re\mathrm{e}\,(\sigma_{\rm cl}(\omega))\mathrm{d}\omega = \frac{\pi ne^2}{2\,m}\ . \tag{53}$$

It may be noted that the integral does not depend on the scattering rate τ_0.

B) The clean interacting case: $W = 0$ and $V \neq 0$

In the absence of disorder $W = 0$, it appears that, much is known about the $t - V$ Hamiltonian. This concerns especially the half-filled band case. The method which appear to be the most powerful and appropriate to tackle the problem is the bosonization technique. For more details one can refer, for example to [38–41]. Note that the bosonization method allows to treat non-perturbatively the effect of the electron-electron correlation, but unfortunately it is suitable only for 1D systems. Let us now, remind some of the physical properties of the $t - V$ Hamiltonian. At half-filling, for repulsive interaction the system exhibits a phase transition from metal to insulator at the critical value $(V/t)_{c_1} = 1$. This transition is called Mott transition [42]. Above the critical value a gap Δ opens up in the charge excitation spectrum and the density-density correlation functions decays exponentially $\langle\rho(r)\rho(0)\rangle \propto \exp(-r/\xi)$, where $\xi \propto 1/\Delta$ is the correlation length which diverge when approaching $(V/t)_{c_1}$.

Note that the nature of Mott transition is completely different from the Anderson transition. In the Anderson transition the quantum interference of the wave functions are responsible for localization, thus the transition from metal

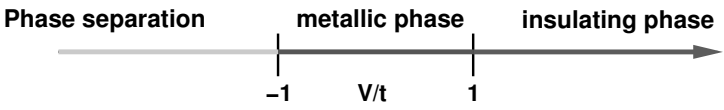

Fig. 6. Phase diagram of the $t-V$ model in the half-filled band case

to insulator occurs at any filling if the disorder strength is strong enough. On the other hand, the Mott transition results from the many-body character of the problem, the mechanisms responsible for the transition are particular backward scattering processes between the electrons involving the lattice. They are called the Umklapp processes. The Umklapp contribution is only relevant at half-filling. Thus away from half-filling, these contributions become irrelevant and thus no transition will be observed.

In the case of attractive interaction and at any filling, the system is phase separated below a second critical value $(V/t)_{c_2} = -1$. Below this critical value, the metallic phase becomes unstable where the holes and the fermions are separated from each other. In the metallic phase, or Luttinger liquid phase $(-1 \leq V/t \leq 1)$, the density-density correlation functions decays as a power law and we say that we have a quasi-long-range order. We should emphasize particularity on the Luttinger liquids: the low energy excitations are only collective modes, there are no individual well-defined quasi-particle states as in the usual Fermi liquids. In the Luttinger liquid phase exponents of the correlation functions vary continuously with V/t. Away from half-filling the system is always in the metallic phase. The phase diagram for the half-filled band case is shown in Fig. 6. It is also interesting to mention that the spinless fermion model can be mapped onto the spin-$\frac{1}{2}$ XXZ model. The transformation which allows this mapping is called the Wigner-Jordan transformation [43].

$$c_l = \exp(\mathrm{i}\Phi_l) S_l^- \ , \tag{54}$$

$$c_l^+ = \exp(-\mathrm{i}\Phi_l) S_l^+ \ , \tag{55}$$

where the phase factor is defined by $\Phi_l = \pi \sum_{k=1}^{l-1} (\frac{1}{2} + S_k^z)$ and $S_l^z = c_l^+ c_l - \frac{1}{2}$. Then the Hamiltonian reduces to

$$\mathcal{H}_{XXZ} = -\frac{t}{2} \sum_i \left[\exp\left(2\mathrm{i}\pi\phi/L\right) S_i^+ \, S_{i+1}^- + \mathrm{h.c.} \right]$$
$$+V \left(\frac{L}{4} + \sum_i S_i^z \, S_{i+1}^z \right) + \sum_i w_i \, \left(S_i^z - \frac{1}{2} \right) \ . \tag{56}$$

A site occupied by a spinless fermion corresponds to spin "up" and an unoccupied site to spin "down". Note that the on-site potential corresponds to random local magnetic fields. The metal insulator transition discussed above for spinless fermions corresponds now to a transition from XY phase to the Ising phase. The special point $V = t$ corresponds to the quantum spin-$\frac{1}{2}$ nearest-neighbor Heisenberg Hamiltonian. In the spin language, above the critical value V_{c_1} a gap opens up in the spin excitation spectrum. Note also that the Wigner-Jordan transformation is valid only in the 1D case.

C) Numerical results

In Fig. 7 we show $I(\phi)$ as a function of ϕ for a system with $L = 16$ at half- and quarter-filling for different values of V. We see that at half-filling the current is weakly sensitive to the amplitude of V when $V/t \leq 1$. However, for larger values, say $V/t = 2$, the amplitude decreases significantly. This is reminiscent of the metal insulator transition discussed above. On the other hand, away from half-filling the effects of interaction are negligible.

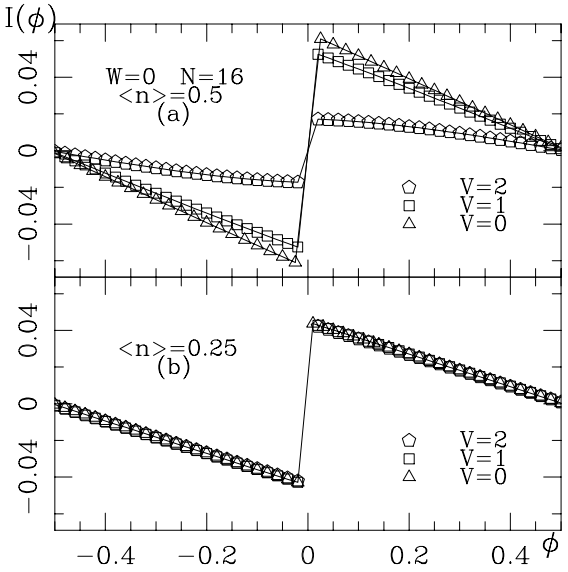

Fig. 7. $I(\phi)$ as a function of ϕ in units of ϕ_0 for a ring of size $L = 16$ at half- and quarter-filling for various values of V. Figure from [44]

In Fig. 8 and Fig. 9, we have plotted the amplitude of the Drude weight as a function of $1/L$ for the half-filled band and quarter-filled band cases, respectively. The advantage is that it is possible to visualize directly the

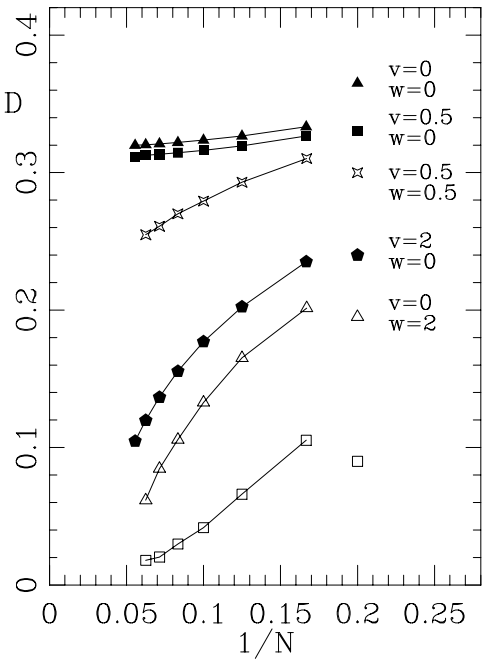

Fig. 8. Drude weight D as a function of the inverse of system size at half-filling ($\langle n \rangle = 0.5$) for various values of V and W. Figure from [44]

Mott transition. We observe the following in Fig. 8:

Half-filled case: $\langle n \rangle = 0.5$

- $V/t < 1$: In the thermodynamic limit D is finite and the system is perfectly metallic. D varies weakly with V, and we observe that in the thermodynamic limit $D_\infty(\langle n \rangle = 0.5, V/t < 1) \approx 1/\pi$ in agreement with the analytical result obtained for $V/t = 0$ in (50).
- $V/t > 1$: We observe clearly that the Drude weight is zero in the thermodynamic limit (Mott insulator).

It can be shown using bosonization techniques that the Drude weight is given by,

$$\pi D_\infty(V, W = 0) = \frac{1}{2} u_\rho K_\rho \tag{57}$$

where the quantities u_ρ is the renormalized velocity and K_ρ is the parameter which controls the critical exponent of the correlation functions. These quantities have been calculated by Luther and Peschel for the half-filled band case [45]. Later Haldane have extended numerically the calculations

away from the half-filled band case [46]. At half-filling we have the exponent $K_\rho = [2(1 - \frac{\delta}{\pi})]^{-1}$ and $u_\rho = \pi \sin(\delta)/\delta$ where the parameter δ is defined by the relation,

$$\cos(\delta) = \frac{V}{t} \,. \tag{58}$$

Thus, at half-filling, in the Luttinger liquid phase the stiffness is given by [47],

$$D_\infty(V, W = 0) = \frac{1}{4} \frac{\sin(\delta)}{\delta(1 - \frac{\delta}{\pi})} \,. \tag{59}$$

From the above equation we recover the previous results for $V = 0$, e.g., $D_\infty(V = 0, W = 0) = 1/\pi$ and we find $D_\infty(V = t, W = 0) = 1/4$ at the critical point $V/t = 1$.

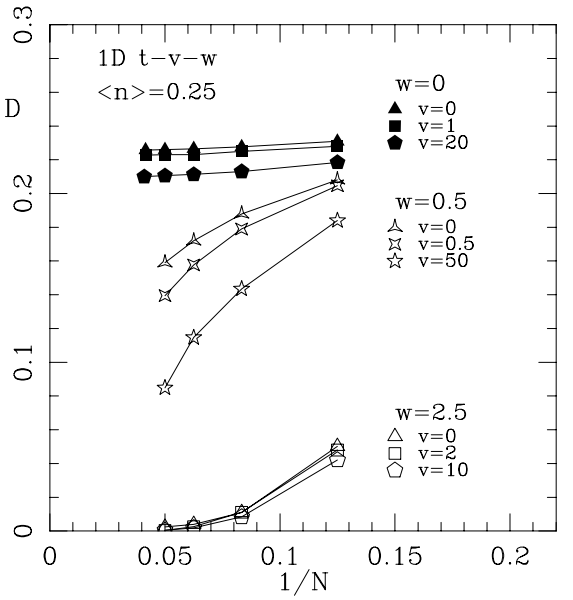

Fig. 9. Drude weight D as a function of the inverse of system size at quarter-filling ($\langle n \rangle = 0.25$) for various values of V and W. Figure from [44]

Non-half-filled case: $\langle n \rangle \neq 0.5$

In Fig. 9, we have plotted the Drude weight away from half-filling. We clearly observe that the effects of interaction are completely negligible since, in the clean case, for any values of V (say, $V/t = 0, 1, 20$), the

system remains metallic and D_∞ is almost insensitive to V. The numerical value is in agreement with the analytical expression (50) which gives $D_\infty(\langle n\rangle \neq 0.5, 0) = 0.225$.

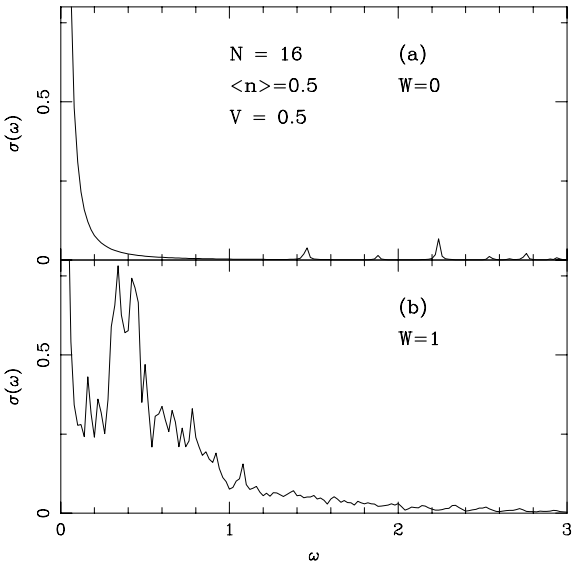

Fig. 10. Real part of the optical conductivity $\sigma(\omega)$ as a function ω (it includes the Drude part). Calculations were performed for $L = 16$ at half-filling **(a)** without and **(b)** with disorder. In the disordered case an average over 100 configurations was performed. Figure from [44]

Let us discuss the effect of interaction on the optical conductivity. In contrast to the case of non-interacting electrons the conductivity has also a contribution at finite frequency. In Fig. 10(a), we have plotted the conductivity for the half-filled band case, in absence of disorder, and for $V/t = 0.5$. We clearly see that the conductivity is concentrated in the Drude part. The contribution at finite frequency is less than 2%. This is consistent with the fact that at this value the Drude weight $\pi D \approx 0.98$ is very close to the maximum $\pi D = 1$ for the uncorrelated case.

D) The dirty interacting case: $W \neq 0$ and $V \neq 0$

In Fig. 11 we have plotted the persistent current $I(\phi)$ as a function of ϕ at half- and quarter-filling in the presence of both disorder and correlation. First we observe that the increase of the interaction strength reduces further the amplitude of the current. Thus in this case, the correlations do not

compensate the effect of Anderson localization due to disorder. The effect of correlation is to reduce further the localization length ξ. Note however that the effect of correlation is much stronger in the half-filled case which is a signature of the Mott transition discussed previously. In these calculations disorder strength is small and for this value $l_e \leq L$ so that we are in the diffusive regime.

In Fig. 8 and Fig. 9, we have plotted the average value of Drude weight in the presence of both interaction and disorder. The typical number of configurations were of the order of 150. First we observe that for any finite W, in both the cases $\langle n \rangle = 0.25$ or 0.5, the Drude weight goes to zero in the thermodynamic limit. In 1D, for an infinitesimal amount of disorder, the system is always in an insulating phase in the thermodynamic limit both in the presence or absence of repulsive interaction. Additionally we see that for a fixed value of W, the effect of V is much stronger at half-filling than away. The figures suggest that at this special filling factor, the effects of V on the localization length are of the same order of magnitude. However, at quarter-filling we clearly observe that at fixed W the effect of V is extremely weak and the localization length is essentially controlled by disorder.

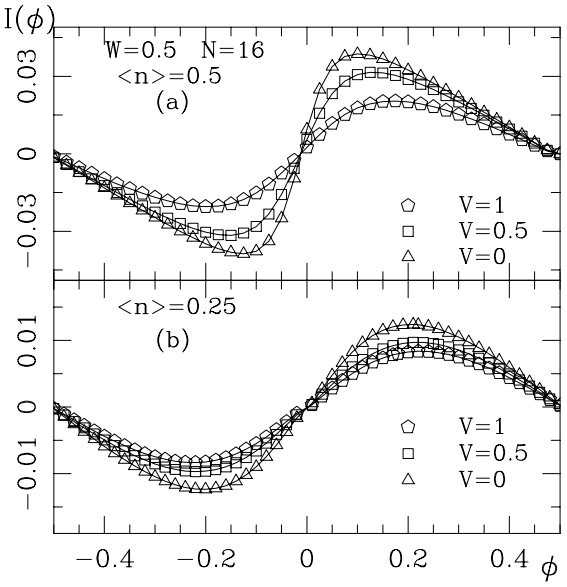

Fig. 11. $I(\phi)$ as a function of ϕ in units of ϕ_0 for a ring of size $L = 16$ at half- and quarter-filling for a fixed disorder strength $W = 0.5t$ with different values of the interaction strength V. Figure from [44]

In Fig. 10(a) and Fig. 10(b) we have plotted conductivity respectively for the ordered and the disordered systems. We observe that conductivity at finite frequencies are now significant. There is a strong reduction of the Drude weight: for $W = 0$ we have $\pi D \approx 1$ and for $W = t$ it is $\pi D \approx 0.52$. Since the kinetic energy variation is small, then according to the sum rules discussed previously, the transfer from the Drude weight to finite frequency is of the order of 50%. The broad absorption peak which appear at $\omega_{\rm abs} \approx 0.4t$ can be understood in the following way. The disorder localizes the electron in small finite regions of the sample. As a consequence the electron can oscillate in these disconnected regions of different sizes. By assuming that the typical size of these region is of order $\xi \propto l_e$, we would expect naively that $\omega_{\rm abs} \approx \pi v_{\rm F} l_e = \pi v_{\rm F} W^2/A$. It would be interesting to check this dependence numerically.

To conclude this part, in the case of spinless fermions, the screened short-range repulsive interactions do not compete with the localization due to disorder. In the contrary, their effects are added to increase the localization effects. At half-filling, the localization is both due to disorder and correlations, their contribution to the localization length is of the same order of magnitude. Away from half-filling, we can consider that the effects of the repulsive interaction are negligible. The localization is mainly due to Anderson localization. Thus the interaction reduces further the magnitude of the persistent currents.

In the case of long range Coulomb interaction, using a perturbative treatment, it was shown by Kato and Yoshioka [48] that for weak disorder the interaction does not increase the currents. On the other hand for very *strong* disorder it was shown by Abraham and Berkovits [49] that the persistent currents increase when the interaction are switched on.

The Case of Attractive Interaction

We have seen in the previous subsections that for the cases of 1D spinless fermionic rings, the repulsive interaction does not counteract the effects of localization due to disorder. In other words, the tendency to localization increases when the repulsive interactions are switched on, leading to further decrease of the persistent currents. It is interesting to ask oneself "What would happen in the case of attractive interaction?" Of course this does not mean that the electron-electron interaction is attractive in the rings used in experiments, but the problem of competition between disorder and attractive interaction is of interest from the theoretical point of view.

In this subsection we consider that the parameter V in the Hamiltonian (35) takes negative values. In Fig. 12 we have plotted the Drude weight calculated exactly as a function of the inverse of system size for various values of V and W. For weak but finite disorder amplitude, we observe that in the thermodynamic limit the stiffness D is finite and its value is insensitive to the disorder strength. Thus, the system remains metallic even in presence of disorder. This is in contrast to repulsive interaction case where for any arbitrary small value of W, we see that $D \to 0$ in the thermodynamic limit. For larger

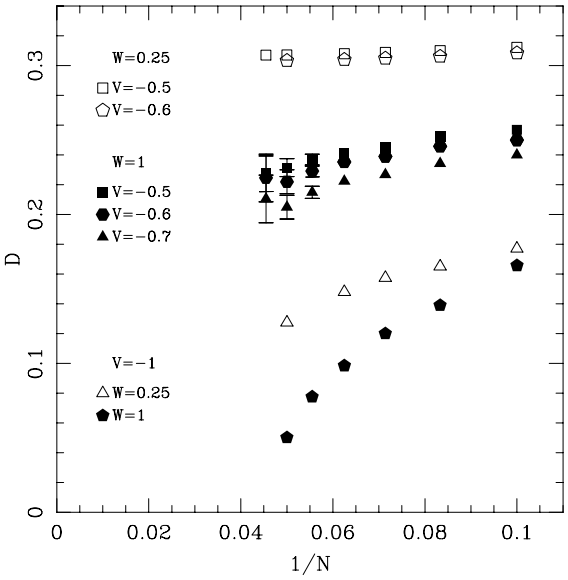

Fig. 12. Drude weight D of the attractive case as a function of $1/N$ at half-filling for different values of the parameters V and W. The disorder averaging was performed over 100 configurations in the case of weak disorder $W = 0.25t$ and over 300 configurations for the stronger disorder $W = t$ or $2t$ cases [22, 50]

values of the disorder strength (e.g., $W = t$), we observe that D decreases with the system size but it seems that it does not saturate. For the special case $V = -t$, we see that in all the cases Drude weight vanishes in the thermodynamic limit. This is consistent with the discussion that at this particular value, there is a phase separation in the system for any filling. The study of the attractive disordered systems show that in the (V,W) plane there is a region of stability of the metallic phase in the presence of disorder. A more systematic study should be performed to determine precisely the phase diagram. Unfortunately, this is not a simple task within exact diagonalization approach.

However, it is possible to calculate the phase diagram using perturbative renormalization group approach [51, 52]. Of course, the perturbative RG is not as accurate as the exact diagonalization calculations. In Fig. 13, we have plotted the phase diagram of the $t - V - W$ model calculated at half-filling within the perturbative RG method [50]. The equation of the critical line was found to be given by,

$$W_c(V) = \left[12\pi^3 \frac{\sin^2(\delta)}{\delta^2}\left(2 - \frac{3\delta}{\pi} - \ln[3(1-\delta\pi^{-1})]\right)\right]^{\frac{1}{2}} . \qquad (60)$$

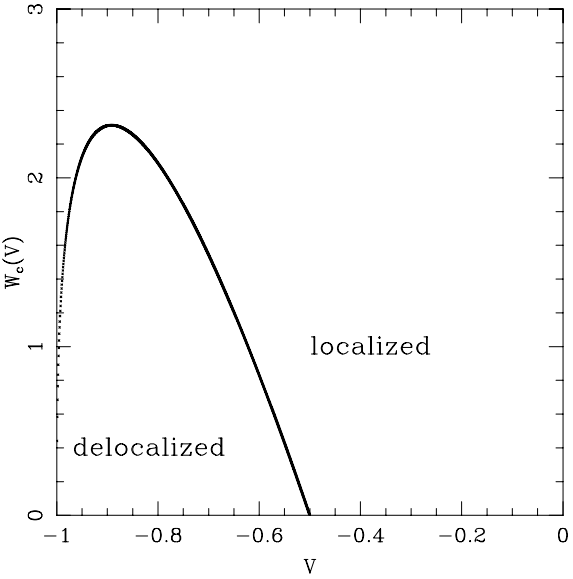

Fig. 13. Phase diagram that represents the Kosterlitz-Thouless critical line which separate the localized phase from the metallic phase in the attractive spinless fermion model at half-filling [22, 50]

The maximum value of the disorder which separates the metallic from the delocalized phase is $W_{\max} \sim 2.3\,t$. A comparison with the exact diagonalization results indicates that the region of stability of the metallic phase is overestimated by the RG calculations. It is also interesting to note that in the vicinity of the critical line, the RG equations are similar to those which describe the Kosterlitz-Thouless transition in the 2D XY model [53].

The Repulsive Hubbard Model: Role of Spin

We have seen in the previous subsections that the repulsive interaction in the case of 1D ring of spinless fermions does not lead to an increase of the persistent currents (stiffness). What would happen if we now take into account the additional degree of freedom of the electron, its spin?

The disordered Hubbard model reads,

$$H = -t \sum_{<i,j,\sigma>} c^{\dagger}_{i,\sigma} c_{j,\sigma} + U \sum_{i,\sigma} n_{i,\sigma} n_{i,-\sigma} + \sum_{i} w_i n_i \tag{61}$$

where σ denotes electron spin and U is the strength of the on-site interaction.

Exact diagonalization calculations on small rings have been performed by Deng et al. [54]. Their study of the amplitude of persistent current as a

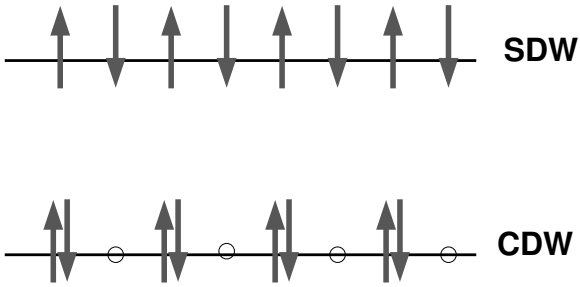

Fig. 14. Schematic representation of a charge density wave and a spin density wave

function of U had shown a peak at a characteristic interaction strength and a decrease beyond this value. Thus in the case of the Hubbard model interaction can increase persistent currents. Later, calculations based on the perturbative RG approach were performed by Giamarchi and Shastry [55]. They have also found that the currents are enhanced by the repulsive interaction. However, in contrast to the exact diagonalization calculations, no peak was observed in the RG calculations. Further RG calculations by Mori and Hamada [56] resolves the discrepancy. They have shown that by including in the RG equation the $4k_{\mathrm{F}}$ impurity scattering term, ignored in [55], the observed peak of the exact diagonalization calculations can be reproduced. The reason of the importance of this term is the following. In the spinless fermion model , the crucial term which is responsible for the suppression of the persistent currents is the $2k_{\mathrm{F}}$ scattering term. In the large U limit the Hubbard model approaches a spinless fermion model were k_{F} has to be replaced by $2k_{\mathrm{F}}$. Thus, it is then expected that $4k_{\mathrm{F}}$ term should become relevant for large U. A simple way to understand the effect of electron-electron interaction is the following: the on-site repulsion favors the formation of spin density wave in the system and thus tends to smoothen out the charge density (see Fig. 14). Since disorder couples directly to the charge, the repulsive Hubbard interaction reduces the pinning with the impurity and thus leads to an increase of the localization length or equivalently to larger persistent currents. On the other hand, the attractive interaction which favors the formation of pairs (superconductivity), does not, in contrast favor the naive expectation of delocalization [55]. In the case of large negative U, only pairs of particle can hop, the Hamiltonian can be mapped onto a hard-core boson (spinless fermions) with a hopping $t' = t^2/U$ in presence of disorder. It is possible to show more precisely that the Hamiltonian can be mapped onto a spinless fermion model with a narrow bandwidth and nearest neighbor interaction. Thus we expect a strong reduction of the persistent currents in the attractive Hubbard model. Note that in the case of spinless fermions the nearest neighbor repulsion V favors charge density wave in the system and thus leads to the decrease of the persistent currents.

4.3 The Multi-Channel Rings

In contrast to the 1D systems, the use of the exact diagonalization method becomes impossible in the multi-channel systems, because of its limitation to very small clusters. Other numerical approach like DMRG [57–60] cannot be used. Unfortunately, the bosonization methods which appeared to be powerful in the 1D case can neither be used in the multi-channel rings. Thus the number of efficient tools to tackle the problem of the interplay between interaction and disorder is drastically reduced. In the following section we will discuss two possible approaches, (i) the perturbative methods and (ii) the self-consistent Unrestricted Hartree Fock (UHF) approach.

The Perturbative Approach

The first attempt to evaluate the effects of electron-electron interaction on the amplitude of the persistent currents in disordered mesoscopic multi-channel rings was by Ambegaokar and Eckern [61]. Using diagrammatic methods, they have calculated the correction to persistent currents due to screened Coulomb interactions. More precisely they have calculated the Hartree and Fock corrections to the persistent currents. They have found that the correction to the second harmonic of the current is,

$$\langle I_{h2}\rangle = \frac{8\rho_0\langle U\rangle}{3\pi}\frac{l_e}{L}\frac{ev_{\rm F}}{L}\exp\left(-\frac{T}{3T_1}\right)\sin\left(4\pi\frac{\phi}{\phi_0}\right) \tag{62}$$

where $T_1 = \hbar D/L^2$.

They have used as input parameters those of the experiment performed by Levy et al. [31]: $l_e = 200\,\text{Å}$ and $L = 2.2\,\mu\text{m}$. Additionally, Ambegaokar and Eckern have estimated the value of the screened interaction parameter $\rho_0\langle U\rangle = 0.6$. This leads for the amplitude of the current to $\langle I_{h2}\rangle = 2.2\times 10^{-3} ev_{\rm F}/L$. We remind that the measured current was two times larger. This seems to indicate an improvement with respect to one particle approaches. Unfortunately, by including other relevant higher order diagrams, it was found that the currents were strongly reduced [62,63] and that the resulting current was about one order of magnitude smaller.

Similarly, Ramin et al. [64] have performed a numerical study and analyzed the effect of the Hartree correction (first order perturbation calculation) on the persistent current amplitude. They have obtained results in agreement with the diagrammatic approach of [61]. Their results indicate a significant enhancement of the second harmonic. Note, that this enhancement is valid both in the cases for the spinless fermion model and for the Hubbard model. One could also conclude from their study, that in contrast to the 1D case, the spin does not play an important/relevant role in the multi-channel case. Thus, on the basis of this promising results one could wonder if calculations beyond the Hartree picture could eventually lead to sufficiently large value of the persistent currents.

The Unrestricted Hartree Fock Approach

The advantage of performing the calculations within unrestricted Hartree Fock (UHF) approach is to allow to treat simultaneously and on equal footings the effects of disorder and correlations and thus restore the many-body physics. This is in contrast to the first order perturbative approaches discussed in the previous subsection for which the ground state is calculated in absence of interaction and the correction to the ground state energy is done at the lowest order.

A) The spinless fermion model case

In the Hartree-Fock decoupling picture the spinless fermion Hamiltonian reads,

$$\mathcal{H}_{\rm hf}(\phi) = -\sum_{i,j} t_{ij}^{\rm eff}\, c_i^\dagger\, c_j + \sum_i w_i^{\rm eff}\, n_i - \frac{1}{2}\sum_{i,j} U_{ij}(\langle n_i\rangle\langle n_j\rangle - |\langle c_j^\dagger\, c_i\rangle|^2) \ . (63)$$

Note that the site index corresponds to (x_i,y_i,z_i).

The effective hopping parameters $t_{ij}^{\rm eff}$ and on-site potentials $w_i^{\rm eff}$ contain the renormalization respectively due to Hartree and Fock terms,

$$w_i^{\rm eff} = w_i + \sum_{j\neq i} U_{ij}\langle n_j\rangle \ , \tag{64}$$

and

$$t_{ij}^{\rm eff} = t_{ij} + U_{ij}\langle c_j^\dagger\, c_i\rangle \ . \tag{65}$$

The terms $\langle n_i\rangle$ and $\langle c_j^\dagger\, c_i\rangle$ are determined self-consistently for a fixed disorder configuration and given filling factor.

Note that the perturbative calculations of Ramin et al. would correspond to $t_{ij}^{\rm eff} = t_{ij}$ and $w_i^{\rm eff} = w_i + \sum_{j\neq i} U_{ij}\langle n_j\rangle_0$, where $\langle\cdots\rangle_0$ denotes calculations with the unperturbed ground-state.

In Fig. 15 and Fig. 16 we have plotted the first two moment of the current distribution calculated at $\phi = \phi_0/4$ for $W = t$ and at fixed density $\langle n\rangle = 1/4$ as a function of V. In these figures we clearly see that the UHF calculations agrees very well with the exact diagonalization calculations. Note that, because of the small system size considered here, we can consider the system to be in a weak disorder regime. It is interesting to mention that exact diagonalization calculations performed by Berkovits et al. [66] for very strong disorder indicate a possible enhancement by the interaction. However the disorder strength was so strong that, in my view, the system could be already in the localized regime.

In Fig. 17 and Fig. 18 we compare the calculation of the second harmonic of the persistent current within the UHF approach to that obtained using a first order perturbation theory (HF) as done by Ramin et al. These figures

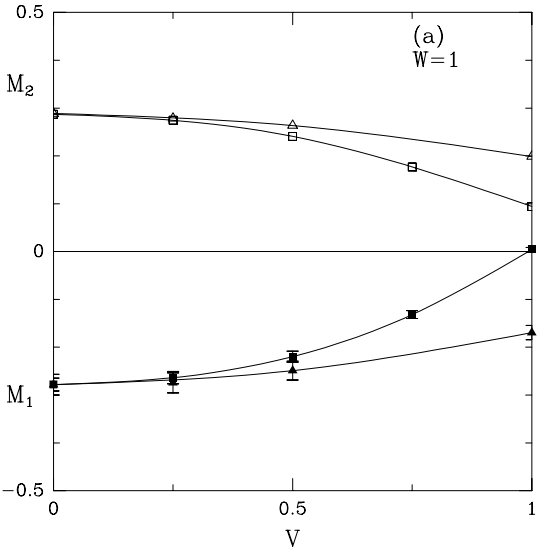

Fig. 15. First moment (M_1) and second moment(M_2) of the current distribution as a function of V in units of I_0, calculated at $\phi = \phi_0/4$ for $W = t$ with fixed density $\langle n \rangle = 1/4$. The system size is 4×4. The squares (triangles) correspond to the UHF (exact) calculations. Average over disorder was performed over at least 500 configurations [22, 65]

show that in the case of weak interaction the UHF and HF agrees as expected and lead to a small increase of the currents. However, for sufficiently strong electron-electron coupling UHF disagree with the HF predictions. It is seen that the currents are completely suppressed by the interaction. This also indicates that higher order diagrams become highly relevant which invalidate the first order perturbation calculations. Note also, that the reduction of the currents occurs for smaller values of the parameter V in the case of stronger disorder.

Thus, to conclude this part, in the case of spinless fermions, the screened Coulomb interaction reduces further the amplitude of the currents in disordered multi-channel rings as in the 1D case. This also indicate that the spin should play an important role and cannot be assimilated to an additional effective channel.

B) The Hubbard model case

In this subsection we analyze the case of the Hubbard model. We remind that, in contrast to the spinless fermion case, the repulsive interaction

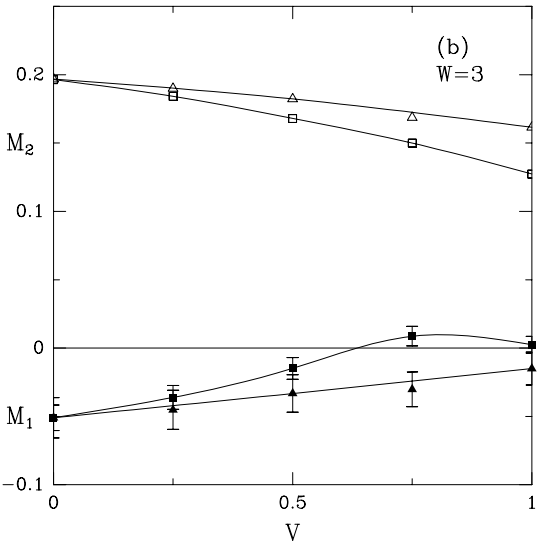

Fig. 16. The plot is the same as that of Fig. 15 but for $W = 3t$ [22,65]

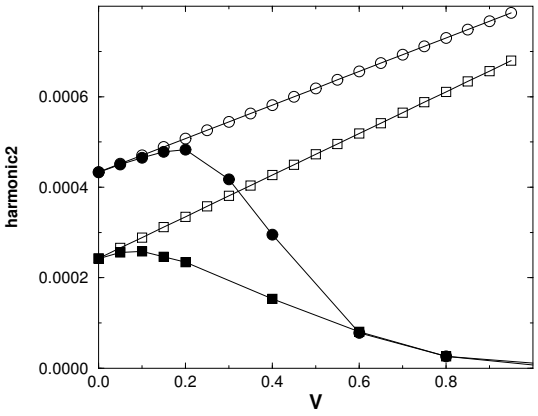

Fig. 17. $\langle I_{h2} \rangle$ as a function of V. The calculations were performed with a 8×8 system at half-filling. Average over disorder was performed over 1000 configurations. The circle correspond to $W = 3t$ and the square to $W = 4t$. The open symbols correspond to first order perturbation theory and the filled symbols to UHF results [22,65]

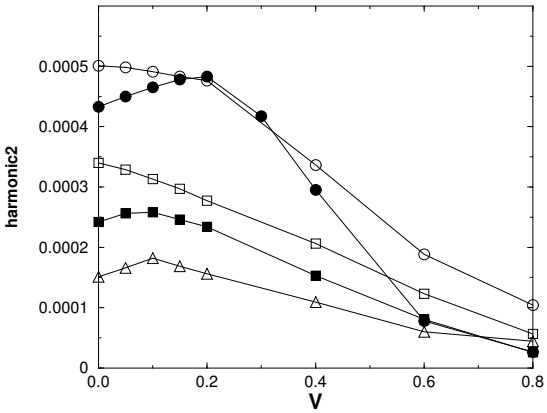

Fig. 18. $\langle I_{h2} \rangle$ as a function of V within the UHF approximation. The circles, squares and triangles correspond to $W = 3t$, $4t$ and $5t$, respectively. The open symbols correspond to the half-filled case and the filled symbols to the quarter-filled case [22, 65]

increases the amplitude of the current in 1D systems. Note that, in the Hubbard model, there is no Hartree term, thus no renormalization of the hopping integrals are needed. The calculations which will be presented in this subsection were obtained by diagonalization of the UHF Hamiltonian in the spin sector $\langle S^z \rangle = 0$ and assuming the absence of local magnetic moment, e.g., $\langle n_{i,\uparrow} \rangle = \langle n_{i,\downarrow} \rangle$.

In Fig. 19 we have plotted the variation of $\langle I_{h2} \rangle(U, W)/\langle I_{h2} \rangle(0, W)$ as a function of the Hubbard parameter U. This figure gives a comparison of the simple first order perturbation calculations in the HF and UHF treatments. We observe that in contrast to the results shown in Fig. 17 for the spinless fermion model, in the HF and UHF treatments current increases with the strength of the interaction parameter. However, we clearly see that the increase is smaller in the UHF picture with respect to HF results.

In Fig. 20, the relative variation of $\langle I_{h2} \rangle(U, W)/\langle I_{h2} \rangle(0, W)$ is plotted as a function of U for different system sizes. We clearly see that the increase of the second harmonic of the current with respect to the non-interacting value increases with the system size.

C) Difference between Hubbard model and spinless fermion model

The above results illustrate the crucial difference between spinless fermion model and Hubbard model, the spin appears to play an important role. The results obtained in the multi-channel case are similar to those obtained in the 1D case: in the spinless fermion model repulsion suppresses current whilst in

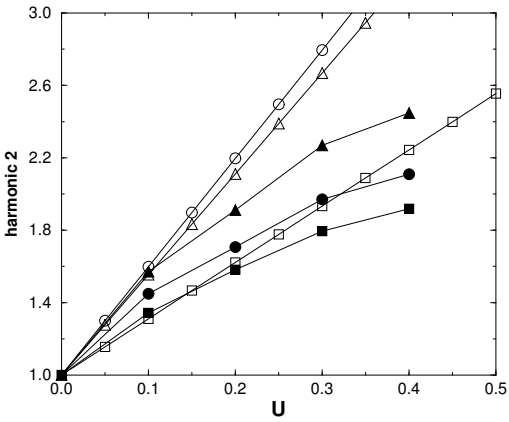

Fig. 19. Variation of $\langle I_{h2}\rangle(U,W)/\langle I_{h2}\rangle(0,W)$ as a function of U. Comparison between the first order calculation (HF) (open symbols) and the UHF results (filled symbols) with systems size 8×8. (i) circle ($W = 4t$ and $N_e = 16$), (ii) square ($W = 3t$ and $N_e = 32$) and (iii) triangle ($W = 4t$ and $N_e = 32$) [22, 65]

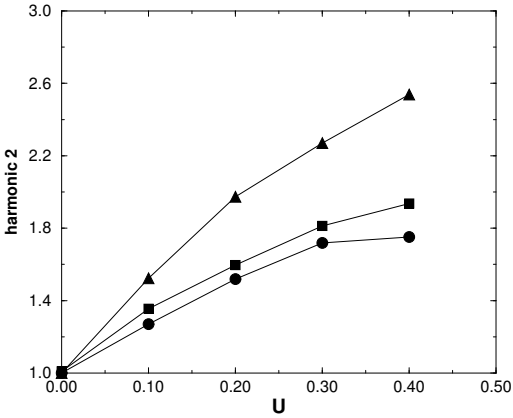

Fig. 20. $\langle I_{h2}\rangle(U,W)/\langle I_{h2}\rangle(0,W)$ as a function of U for $W = 3t$. The UHF calculations were performed at half-filling with system size 6×6 (circle), 8×8 (square), and 10×10 (triangles) [22, 65]

the Hubbard model current increases. However, as in the 1D Hubbard model, we would expect a maximum in $\langle I_{h2}\rangle(U,W)$ as a function of U. This would require a more detailed analysis. We have seen, in the 1D disordered rings that the difference between the spinless fermion case and the Hubbard model is that in the first case the correlation favors CDW fluctuations and in the other SDW. Let us show that this mechanism is in fact also relevant in the multi-channel case. The advantage to work in real space is to allow the direct observation of the charge density fluctuations in the system.

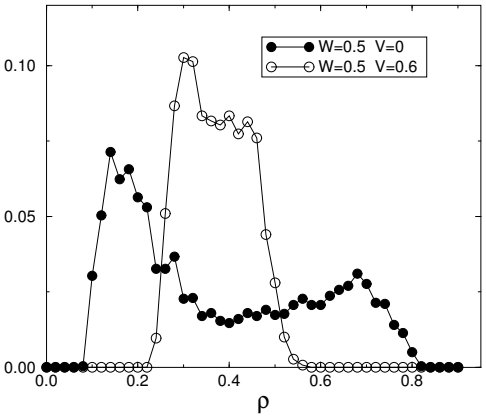

Fig. 21. Probability distribution of the local charge density of a 10×10 spinless fermion system with $N_e = 40$ and $V = 0$ or $0.6t$ for some fixed disorder strength. This distribution was obtained after averaging over 30 disorder configurations [22,65]

In Fig. 21 and Fig. 22 we have plotted the distribution of the charge density $P(\rho)$ in absence and presence of electron-electron correlations. We see in Fig. 21 a drastic effect of the repulsive interaction. In the case $V = 0$ we observe a broad peak at $\langle n\rangle = 0.4$. As the interactions are switched on, we see that charge fluctuations strongly increase and lead to two distinct peaks around $\rho = 0.7$ and $\rho = 0.2$ showing that the system become phase separated. Away from half-filling we observe a broadening of the charge density distribution but no tendency of phase separation. On the other hand, in Fig. 22 we see that the effect of U is to reduce the distribution width, thus U tends to uniformize the charge density.

4.4 Conclusion

To conclude, we have in this chapter presented how the electron-electron interaction could eventually compete with the effects of disorder and thus lead

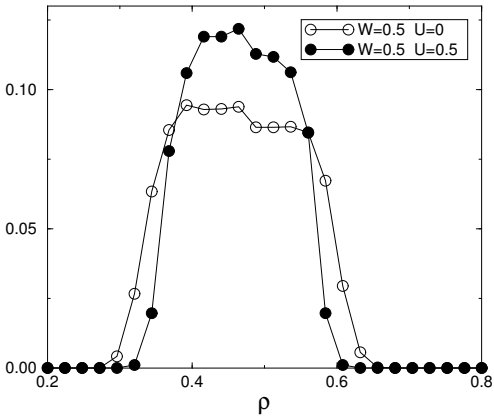

Fig. 22. Probability distribution of the local charge density of a 10×10 Hubbard system with $N_e = 2N_\uparrow = 2N_\downarrow = 100$ and $U = 0$ or $0.5t$ for some fixed disorder strength. This distribution was obtained after averaging over 30 disorder configurations [22, 65]

to an increase of the persistent current amplitude. We have seen that the nature of the interaction plays an important role. In the case of spinless fermions the short-range electron-electron repulsive interaction tends to suppress the persistent currents. The underlying mechanism is the tendency to favor CDW fluctuations which strongly couple to the disorder and thus enhances the tendency to localization. On the other hand, in the Hubbard case (spin taken into account), the repulsive electron-electron interaction favors the SDW fluctuations. Thus the pinning with the impurities is then reduced which leads to an increase of the current. Interestingly, it was shown that this physics is valid both in the cases of 1D rings and multi-channel systems. However, the presented results suggest that the increase is unfortunately not sufficient to explain the discrepancy between experiments and theory which concerns two orders of magnitude. Additionally, the averaged second harmonic, appears to be always of the same positive sign, in both the cases with or without electron-electron interactions. Thus both the amplitude of the persistent current and its sign remains an open issue and a challenge to theory. However, we should mention that calculations of the persistent currents in the presence of long range Coulomb interaction indicates a stronger enhancement of the current with respect to screened Coulomb interaction [67]. Note also that all the calculations are usually performed using only nearest neighbor hopping integrals, the influence of further couplings could also be interesting. Recent calculations performed for non-interacting electrons indicate a weaker effect of the disorder in the case of extended hopping integrals [68, 69].The influence of the spin-orbit coupling is probably the most promising ingredient and cer-

tainly one interesting problem which should be investigated in details. This would require a treatment of the disorder and of the spin-orbit coupling on equal footings.

References

1. A. G. Aronov, Yu. V. Sharvin: Rev. Mod. Phys. **59**, 744 (1987)
2. B. L. Altshuler, B. D. Simons: In: *Physique Quantique Mésoscopique, Les Houches 1994 Session LXI*, ed by E. Akkermans, G. Montambaux, J. L. Pichard, J. Zinn-Justin (North Holland)
3. D. Mailly, Sanquer: In: *Quantum Coherence in Mesoscopic Systems*, ed by B. Kramer (Pleneum Press, New York 1991)
4. *Mesoscopic Phenomena in Solids* , ed by B. L. Altshuler, P. A. Lee, R. A. Webb (North-Holland 1991)
5. P. W. Anderson: Phys. Rev. **109**, 1492 (1958)
6. M. Buttiker, Y. Imry, R. Landauer: Phys. Lett. **96 A**, 365 (1983)
7. N. Byers, C. N. Yang: Phys. Rev. Lett. **7**, 46 (1961)
8. F. Bloch: Phys. Rep. **137 A**, 787 (1965)
9. Y. Aharonov, D. Bohm: Phys. Rev. **115**, 485 (1959)
10. H. F. Cheung, Y. Gefen, E. K. Riedel, W. H. Shih: Phys. Rev. B **37**, 6050 (1988)
11. S. Chakravarty, A. Schmid: Phys. Rep. **140**, 193 (1986)
12. G. Bergmann: Phys. Rep. **107**, 1 (1984)
13. D. E. Khmelnetskii: Physica **126 B**, 235 (1984)
14. P. A. Lee, T. V. Ramakrishnan: Rev. Mod. Phys. **57**, 287 (1985)
15. E. Abrahams, P. W. Anderson, D. C. Licciardello, T. V. Ramakrishnan: Phys. Rev. Lett. **42**, 673 (1979)
16. *Anderson Localization*, ed by Y. Nagaoka, H. Fukuyama (Springer, Berlin 1982)
17. B. Kramer, A. Mac Kinnon: Rep. Prog. Phys. **56**, 1469 (1993)
18. D. J. Thouless: Physics Reports **132**, 93 (1974); Phys. Rev. Lett. **18**, 1167 (1977)
19. J. T. Edwards, D. J. Thouless: J. Phys. C **5**, 807 (1972)
20. P. D. Kirkman, J.B. Pendry: J. Phys. C **17**, 4327 (1984)
21. M. Kappus, F. J. Wegner: Z. Phys. B **45**, 15 (1981)
22. G. Bouzerar: PhD Thesis, University of Orsay, Paris (1996)
23. H. F. Cheung, Y. Gefen, E. K. Riedel: Phys. Rev. Lett. **62**, 587 (1989)
24. H. Bouchiat, G. Montambaux: J. Physique **50**, 2695 (1989)
25. G. Montambaux, H. Bouchiat, D. Sigeti, R. Friesner: Phys. Rev. B **42**, 7647 (1990)
26. H. Bouchiat, G. Montambaux, D. Sigeti: Phys. Rev. B **42**, 1682 (1990)
27. F. Von Oppen, E.K. Riedel: Phys. Rev. Lett. **66**, 84 (1991)
28. B. L. Altshuler, Y. Gefen, Y. Imry: Phys. Rev. lett. **66**, 88 (1991)
29. V. Chandrasekhar, R. A. Webb, M. J. Brady, M. B. Ketchen, W. J. Gallagher, A. Kleinsasser: Phys. Rev. Lett. **67**, 3578 (1991)
30. D. Mailly, C. Chapelier, A. Benoit: Phys. Rev. Lett. **70**, 2020 (1993)
31. L. P. Levy, G. Dolan, J. Dunsmuir, H. Bouchiat: Phys. Rev. Lett. **64**, 2074 (1990)
32. E. M. Q. Jariwala et al: Phys Rev. Lett. **86**, 1594 (2001)
33. R. Deblock et al: Phys Rev. B **65**, 075301 (2002)
34. R. Deblock et al: Phys Rev. Lett. **89**, 206803 (2002)

35. G. D. Mahan: *Many Particle Physics*, 3rd edn (Kluwer Academic/Plenum Publishers 2000)
36. D. J. Scalapino, S. R. White, S. Zhang: Phys. Rev. B **47**, 7995 (1993)
37. W. Kohn: Phys. Rev. **133**, A171 (1964)
38. R. Shankar: Int. J. Mod. Phys. **4**, 2371 (1990)
39. V. J. Emery: In: *Highly Conducting One-Dimensional Solids* (Plenum, New-York 1979)
40. J. Solyom: Adv. Phys. **28**, 209 (1979)
41. H. J. Shulz: In: *Correlated Electron Systems*, ed by V. J. Emery (World Scientific, Singapore 1993)
42. N. Mott: *Metal-Insulator Transitions* (Taylor & Francis, London 1974)
43. P. Jordan, E. Wigner: Z. Phys. **47**, 631 (1928)
44. G. Bouzerar, D. Poilblanc, G. Montambaux: Phys. Rev. B **49**, 8258 (1994)
45. A. Luther, I. Peschel: Phys. Rev. B **12**, 3908 (1975)
46. F. D. Haldane: Phys. Rev. Lett. **45**, 1358 (1980)
47. B. S. Shastry, B. Sutherland: Phys. Rev. Lett. **65**, 243 (1990)
48. H. Kato, D. Yoshioka: Phys. Rev. B **50**, 4943 (1994)
49. M. Abraham, R. Berkovits: Phys. Rev. Lett. **70**, 1509 (1993)
50. G. Bouzerar, D. Poilblanc: J. Phys. I **4**, 1699 (1994)
51. T. Giamarchi, H. Schulz: Phys. Rev. B **37**, 325 (1988)
52. A. Doty, D.S. Fisher: Phys. Rev. B **45**, 2167 (1992)
53. J. M. Kosterlitz, D. Thouless: J. Phys. C **6**, 1181 (1973)
54. W. Deng et al: Phys. Rev. B **50**, 7655 (1994)
55. T. Giamarchi, B. S. Shastry: Phys. Rev. B **51**, 10915 (1995)
56. H. Morii, M. Hamada: Phys Rev. B **53**, 4850 (1996)
57. D. Weinmann et al: Eur. Phys. J. B **19**, 139 (2001)
58. P. Schmitteckert et al: Phys. Rev. Lett. **81**, 2308 (1998)
59. E. Gambetti: Phys. Rev. B **72**, 165338 (2005)
60. F. Carvalho Dias et al: Phys. Rev. B **73**, 075109 (2006)
61. V. Ambegaokar, U. Eckern: Phys. Rev. Lett. **67**, 3192 (1991)
62. U. Eckernr: Z. Phys. B **82**, 393 (1991)
63. R. A. Smith, V. Ambegaokar: Europhys. Lett. **20**, 161 (1992)
64. M. Ramin, B. Reulet, H. Bouchiat: Phys. Rev. B **51**, 5582 (1995)
65. G. Bouzerar, D. Poilblanc:, Phys. Rev. B **52**, 10772 (1995)
66. R. Berkovits, Y. Avishai: Europhys. Lett. **29**, 475 (1995)
67. D. Yoshioka, H. Kato: Physica B **212**, 251 (1995)
68. S. K. Maiti, J. Chowdhury, S. N. Karmakar: cond-mat/0604011 (2006)
69. S. K. Maiti, J. Chowdhury, S. N. Karmakar: J. Phys.: Condens Matter **18**, 5349 (2006)

Electron Transport Through Mesoscopic Closed Loops and Molecular Bridges

Santanu K. Maiti[1] and S. N. Karmakar[2]

[1] Theoretical Condensed Matter Physics Division
Saha Institute of Nuclear Physics, 1/AF, Bidhannagar, Kolkata 700 064, India.
santanu.maiti@saha.ac.in
[2] Theoretical Condensed Matter Physics Division
Saha Institute of Nuclear Physics, 1/AF, Bidhannagar, Kolkata 700 064, India.
sachindranath.karmakar@saha.ac.in

1 Introduction

An emerging tendency of modern material science is to propose and investigate systems containing smaller and smaller structures. These smaller structures approach the so-called mesoscopic or nanoscopic regimes in which quantum effects become much more relevant for the behavior of these materials. This situates mesoscopic physics at the interface of statistical physics and quantum physics. The mesoscopic systems are very much smaller than the large-scale objects and they often have unusual physical and chemical properties. The study of mesoscopic systems provides a clear understanding of the behavior of a material as it goes from a few atoms to large visible and tangible objects.

1.1 Mesoscopic Regime

Mesoscopic regime refers to the length scale at which one can reasonably describe the properties of a material or a phenomenon without discussing the behavior of the individual atoms. Typically this length scale is of submicron order and involves averaging over a few thousand atoms or molecules. In this scale fluctuations of the physical quantities arising from the motion and behavior of individual particles can be reduced below some desirable threshold (often a few percent) and it must be established rigorously within the context of any particular problem. In the mesoscopic regime behavior of a system is highly influenced by quantum interference of the electronic wave functions. The quantum phase coherence, essential for the appearance of interference effects, is preserved only during a finite time τ_ϕ called the *phase coherence time*. In electronic conductors, finite phase coherence time corresponds to a *phase*

coherence length L_ϕ over which electrons can travel before their phase coherence get lost. The quantum effects appear only when the characteristic time or the characteristic length of the system are smaller than the phase coherence time or the phase coherence length, respectively. In many cases this means that the relevant system size L must be smaller than the phase coherence length L_ϕ. For an electron, the phase coherence time/length is limited by electron-electron and electron-phonon scattering processes. These processes are important at high temperatures, but both are suppressed at low temperatures implying that the phase coherence time/length is strongly material and temperature dependent. The mesoscopic regime is therefore characterized by small time and/or length scales and low temperatures. When temperature is lowered, the phase coherence time/length increases (by a factor T^{-1}) and the mesoscopic regime gets extended. At sub-Kelvin temperatures, the time and length scales in the semiconductor samples are of the order of picoseconds and micrometers, respectively.

1.2 Some Extraordinary Mesoscopic Phenomena

The samples like quantum dots, quantum wires, two-dimensional electron gases in the semiconductor heterostructures, etc. exhibit many exotic physical properties due to extreme spatial confinement of electrons at the submicron length scale. Here we briefly describe some spectacular effects that appear in such systems as a consequence of quantum phase coherence of the electronic wave functions.

Aharonov-Bohm Oscillations

One of the most remarkable consequences of quantum phase coherence is the observation of Aharonov-Bohm (AB) oscillations in the conductance of normal metal mesoscopic rings. At very low temperature, superposition of the electronic wave functions corresponding to the two arms of the ring becomes important. The pioneering experiment on the AB effect was on a ring-shaped resistor made from a 38 nm film of polycrystalline gold. The diameter of the ring was 820 nm and its thickness was 40 nm [1]. Conductance of this ring was found to oscillate as a function of magnetic flux enclosed within the ring with $h/|e|$ periodicity [2]:

$$g = g_0 + \hat{g} \cos\left[\frac{|e|BS}{\hbar} + \overline{\phi}\right] \tag{1}$$

where S is the area enclosed by the ring and B is the magnetic field perpendicular to the plane of the ring.

Integer and Fractional Quantum Hall Effects

An important discovery of 1980s was the phenomenon of integer quantum Hall effect [3] which also results from quantum phase coherence of the electronic wave functions of two-dimensional electron gas systems. In the Hall

measurement one drives a current along a two-dimensional conductor, and, measures the longitudinal voltage V_x and the transverse Hall voltage V_H as functions of the magnetic field B applied perpendicular to the plane of the conductor. According to the classical Drude formula, the Hall resistance R_H should be linearly proportional to B and the longitudinal resistance R_x should remain unaffected by magnetic field. This behavior holds true only when magnetic field is very weak. In the strong magnetic field and at low temperature, one gets completely new behavior which cannot be explained by the classical Drude model. The longitudinal resistance exhibits a oscillatory behavior while the Hall resistance shows sharp step-like structures with plateaus. The values of R_H on these plateaus are given by h/ne^2, where n is an integer with values $1, 2, 3, \cdots$. It turns out that these values of R_H are very robust and highly reproducible with great precision, and this is the reason why they are often used as the standard of resistance. The integer quantum Hall effect is a purely quantum mechanical phenomenon due to the formation of the Landau levels and we refer to [4–6] for its details.

The phenomenon of fractional quantum Hall effect [7] was discovered in 1982 when a two-dimensional electron gas was subjected to a very high magnetic field at a very low temperature. The two-dimensional electron gas then exhibits additional plateaus in the Hall resistance at h/ne^2 with fractional values of n. It is now well-established that Coulomb correlation among the electrons becomes quit important for the explanation of the fractional quantum Hall effect and its origin has been traced back to the existence of correlated collective quasi-particle excitations [8]. An extensive review on this topic can be found in [5].

Persistent Currents

Electrons in mesoscopic normal metal rings can support non-decaying circulating current in thermodynamic equilibrium even at non-zero temperature and it is the well-known phenomenon of persistent current in mesoscopic normal metal rings. This exotic phenomenon is due to phase coherence of the electronic wave functions over the mesoscopic rings. This current depends on the magnetic flux ϕ enclosed by the ring and it exhibits the elementary flux quantum $\phi_0 = hc/e$ periodicity. The possibility of persistent current was predicted in the very early days of quantum mechanics by Hund [9], but their experimental evidences came much later only after the fabrication of mesoscopic systems. A detailed discussion on the phenomenon of persistent current in mesoscopic normal metal rings will be presented in Sect. 2.

Conductance Fluctuations

In the mesoscopic regime conductance of a disordered wire exhibits pronounced fluctuations as a function of the external parameters like magnetic field, Fermi energy, etc. These fluctuations that are observed [10] only at very

low temperatures are perfectly reproducible. Conductance fluctuations are due to quantum interference of the electronic wave functions corresponding to various pathways that the electrons can take while traversing the sample. The most important feature of conductance fluctuation is that its typical amplitude is universal in the diffusive regime [11] which actually gives the fingerprint of the underlying quantum interference effects. The fluctuations are always of the order of conductance quantum e^2/h and depend only on the basic symmetries of the system [12].

The conductance of a ballistic quantum point contact was found [13, 14] to be quantized in units of $2e^2/h$ and in a recent experiment [15] it has been demonstrated that conductance quantization can be observed in an extremely simple setup. A quantum point contact is a very narrow link between two conducting materials which can be formed by imposing a narrow constriction between them. With the decrease of the width (W) of constriction conductance goes down in quantized steps. This is due to the fact that although the width of constriction changes continuously, the number of sub-bands or transverse modes (M) changes in discrete steps. This discreteness is not evident if constriction is several thousands of wavelength wide as then a very small fractional change in W changes M by many integers.

2 Persistent Current in Mesoscopic Normal Metal Rings

The phenomenon of persistent current in mesoscopic normal metal rings provides an excellent testing ground for many ideas of basic physics. In thermodynamic equilibrium, a mesoscopic metallic ring threaded by magnetic flux ϕ supports a current in the ring which does not decay dissipatively even at nonzero temperature. It is the so-called persistent current which is a purely quantum mechanical effect and it gives an obvious demonstration of the Aharonov-Bohm effect [16]. In 1983 Büttiker et al. [17] predicted that persistent current can exist in mesoscopic normal metal rings threaded by magnetic flux even in the presence of impurity. In a pioneering experiment Levy et al. [18] first gave the experimental evidence of persistent current in mesoscopic normal metal rings and later its existence was further confirmed by other experiments [19–22]. It has been observed that the experimental results do not agree well with the theoretical predictions. The measured amplitudes of the currents are orders of magnitude larger than the theoretical estimates, both ϕ_0 and $\phi_0/2$ periodicities were observed, and, the low-field magnetic susceptibility were found to be diamagnetic as well as paramagnetic. Metals are intrinsically disordered and electron-electron interaction becomes important at the mesoscopic length scale as electrons are not screened much. In order to understand the behavior of persistent current in mesoscopic normal metal rings, one has to focus attention on the interplay of quantum phase coherence, electron-electron interaction and disorder. This is a highly complex problem

which has been addressed quite extensively over the last twenty years both experimentally [18–22] as well as theoretically [23–34].

2.1 Single-Channel Non-Interacting Rings

The system under consideration is a quantum ring of N atomic sites, as depicted schematically in Fig. 1. In the non-interacting picture, this system

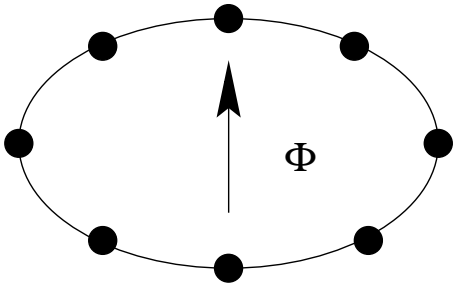

Fig. 1. Schematic view of a 1D mesoscopic ring threaded by magnetic flux ϕ

is usually modeled by a single-band tight-binding Hamiltonian

$$H = \sum_i \epsilon_i c_i^\dagger c_i + v \sum_{<ij>} \left[\mathrm{e}^{\mathrm{i}\theta} c_i^\dagger c_j + \mathrm{h.c.} \right] , \tag{2}$$

where $c_i^\dagger$ (c_i) is the creation (annihilation) operator of an electron in the ith Wannier state, ϵ_i's are the site energies, and v represents the nearest-neighbor hopping integral. Here $\theta = 2\pi\phi/N$ is the phase factor due to magnetic flux ϕ (measured in units of $\phi_0 = ch/e$, the elementary flux quantum). Unless mentioned explicitly, henceforth we will use the units $c = e = h = 1$.

In a ring geometry, periodic boundary condition leads to quantized energy levels $E_n(\phi)$ and the discreteness of the energy levels plays a very important role for the existence of persistent current. At sufficiently low temperature and/or for small system size, level spacings might become larger than the thermal energy $k_\mathrm{B}T$ and under this situation electrons will not suffer any inelastic collision. This in turn gives persistent current as the energy eigenstates carry currents which cannot decay dissipatively [23]. At zero temperature persistent current carried by the nth eigenstate is given by

$$I_n(\phi) = -\frac{\partial E_n(\phi)}{\partial \phi} . \tag{3}$$

For a system characterized by fixed number of electrons N_e, the total current is obtained by taking sum of the individual contributions of the lowest N_e energy levels. On the other hand when a system is specified with fixed

chemical potential μ, the total current is obtained by adding the individual contributions from all the energy levels with energies less than or equal to μ.

In perfect rings the current-flux characteristics exhibit saw-tooth like shapes with sharp discontinuities at $\phi = \pm p\phi_0/2$ and $\pm p\phi_0$ (p being an in-

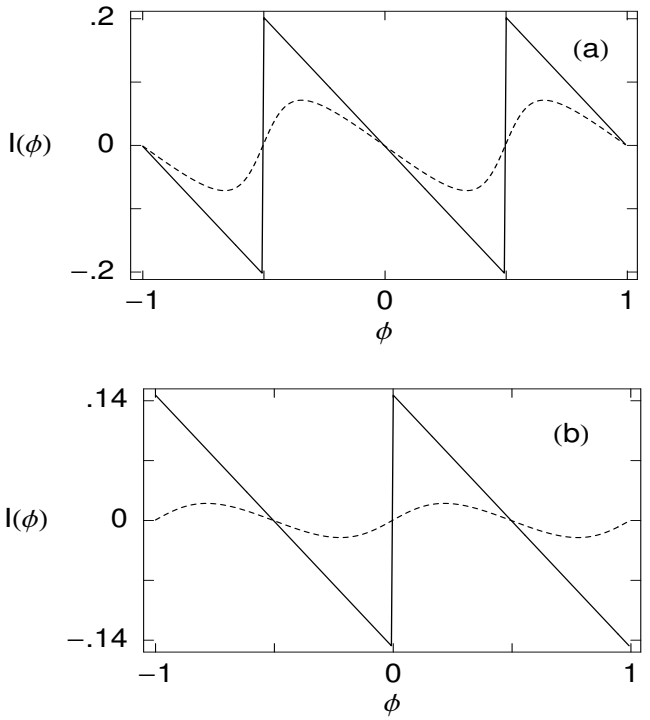

Fig. 2. $I - \phi$ curves of a 1D mesoscopic ring of size $N = 50$ with (a) $N_e = 15$ and (b) $N_e = 10$. The solid and dotted curves respectively corresponds to the perfect and disordered ($W = 1$) rings

teger), respectively for the cases with odd and even N_e (see the solid curves of Fig. 2). These discontinuities appear due to the degeneracy of the energy levels at these values of ϕ. On the other hand, in the presence of impurity all these degeneracies get lifted and persistent currents become continuous functions of ϕ (see the dotted curves of Fig. 2). It may be noted that the amplitudes of the currents are greatly reduced in the disordered rings due to the tendency of localization of the electrons. It may be noted that both the perfect and dirty rings exhibit the flux quantum ϕ_0 periodicity.

Though the phenomenon of persistent current in mesoscopic normal metal rings is well-understood within the one-electron picture [23–28], it fails to explain the experimental results [18–22]. One major discrepancy between the-

ory and experiment is that the amplitudes of the measured persistent currents are orders of magnitude larger than the theoretical predictions. Another controversial issue is that experimentally persistent currents exhibit both half-integer and integer flux quantum periodicity. The explanation of these results in terms of the ensemble averaged persistent currents is also quite intriguing and calculations show that the disorder averaged current crucially depends on the choice of the ensemble [23, 24]. It is believed that the interplay between electron-electron interaction and disorder has a major role in the enhancement of persistent currents and in the following sections this issue has been addressed in detail.

2.2 Single-Channel Interacting Rings

In order to reveal the role of disorder and electron-electron interaction on persistent current, in this subsection we calculate persistent currents in mesoscopic normal metal rings with aperiodic site potentials. These systems closely resemble the disorder systems and the advantage is that here we do not require any configuration averaging. A mesoscopic ring with incommensurate site potentials can be described by the tight-binding Hubbard Hamiltonian

$$H = \sum_{\sigma} \sum_{i=1}^{N} \epsilon \cos(i\lambda\pi) c_{i,\sigma}^{\dagger} c_{i,\sigma} + v \sum_{\sigma} \sum_{i=1}^{N} (c_{i,\sigma}^{\dagger} c_{i+1,\sigma} \mathrm{e}^{\mathrm{i}\theta} + \mathrm{h.c.}) + U \sum_{i=1}^{N} n_{i\uparrow} n_{i\downarrow} \ , \tag{4}$$

where λ is an irrational number and in the present calculation it has been taken as the golden mean $(1+\sqrt{5})/2$. Here U denotes the strength of Hubbard interaction and $n_{i\sigma}$ represents the number operator.

At zero temperature persistent current can be calculated from the expression

$$I(\phi) = -\frac{\partial E_0(\phi)}{\partial \phi} \ , \tag{5}$$

where $E_0(\phi)$ is the many-body ground state energy. We determine $I(\phi)$ from the above equation via exact numerical diagonalization of the Hamiltonian (4). To understand the precise role of electron-electron correlation and disorder on persistent current, here we present results for the simplest possible system containing two opposite spin $(\uparrow, \downarrow)$ electrons with incommensurate site potentials. In Fig. 3 we plot the current-flux characteristics for a 30-site ring where the solid, dotted and dashed curves correspond to the interaction strengths $U = 0$, 1 and 3, respectively. In the absence of electron-electron interaction disorder suppresses persistent current enormously and it becomes apparent from Fig. 3 as the I–ϕ curve for the non-interacting case (solid line) almost coincides with the ϕ-axis. This is due to the fact that in the presence of aperiodic site potentials electronic eigenstates are almost critical [35, 36] with

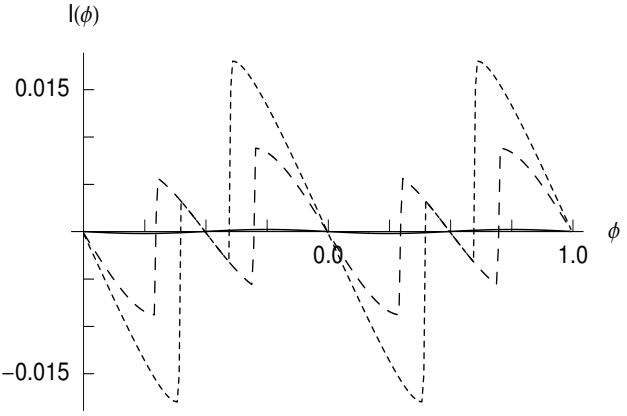

Fig. 3. $I - \phi$ curves of a 1D mesoscopic ring ($N = 30$) with incommensurate site potentials containing two opposite spin ($\uparrow, \downarrow$) electrons. The solid, dotted and dashed curves are respectively for $U = 0, 1$ and 3

power-law localization and persistent currents get reduced. But this situation changes quite dramatically as we switch on electron-electron interaction. Figure 3 clearly shows that electron-electron interaction enhances persistent current considerably for low values of U. This is because the repulsive Coulomb interaction does not favor confinement of the electrons due to localization in the presence of disorder. Thus mobility of the electrons increases as we introduce Hubbard correlation and persistent currents get enhanced. However, this enhancement of persistent current due to correlation is not enough to explain the experimental results. Also this enhancement ceases to occur after certain value of U due to the ring geometry and persistent current eventually decreases as we increase U further.

A new feature appears in the $I - \phi$ characteristics in the presence of electron-electron interaction. The $I - \phi$ characteristics exhibit kink-like structures and most interestingly we see that the currents inside these kinks are independent of the strength of Hubbard correlation U. This result can be analyzed as follows. For two opposite spin electrons, total spin S can have the values 0 and 1. The Hamiltonian of this system for any ϕ can be block diagonalized by proper choice of the basis states, and this can be achieved by taking all the basis states in one sub-space with $S = 0$, while those in the other sub-space with $S = 1$. It is easy to see that in the sub-space spanned by the basis set with $S = 1$, the block Hamiltonian becomes free from U, and, the corresponding energy eigenvalues and eigenstates become U-independent. In absence of interaction, these U-independent energy levels are always above the ground state energy of the system for any value of ϕ. But for non-zero values of U, one of these U-independent energy levels becomes the ground state energy of the system in certain domains of ϕ. In these regions we have

kinks in the I–ϕ curves and obviously persistent currents inside these kinks are independent of the strength of Hubbard correlation U. In a recent experiment Keyser et al. [37] reported similar oscillations in the conductance of small rings with less than ten electrons. The role of electron correlation on persistent current both in perfect and dirty rings with higher number of electrons has been discussed in [38, 39].

2.3 Enhancement of Persistent Current in Mesoscopic Rings

Free electron theory predicts that, at $T = 0$, an ordered one-dimensional normal metal mesoscopic ring threaded by magnetic flux ϕ supports persistent current with maximum amplitude $I_0 = ev_{\mathrm{F}}/L$, where v_{F} is the Fermi velocity and L is the circumference of the ring. Real samples are always disordered which tends to decrease the amplitudes of the currents, and calculations show that the disorder-averaged current $< I >$ crucially depends on the choice of the ensemble [23, 24]. The magnitude of current $< I^2 >^{1/2}$ is however

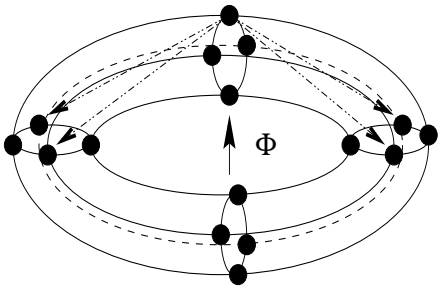

Fig. 4. Schematic view of a mesoscopic cylindrical ring threaded by magnetic flux ϕ. Filled circles represent the lattice sites

insensitive to the averaging issues and is of the order of $I_0 l/L$, l being the elastic mean free path of the electrons. This expression remains valid even if one takes into account the finite width of the ring by adding contributions from the transverse channels, since disorder leads to a compensation between the channels [23, 24].

Measurements on an ensemble of 10^7 Cu rings [18] reported a diamagnetic persistent current of average amplitude $3 \times 10^{-3}\, ev_{\mathrm{F}}/L$ with a half flux quantum periodicity. Such $\phi_0/2$ oscillations with diamagnetic response were also found in other persistent current experiments consisting of ensemble of isolated rings [22, 40]. The strange period-halving is due to the fact that the first harmonic averages out to zero while the second harmonic survives [24]. The measured average currents are comparable to the typical currents $< I^2 >^{1/2}$, but are one or two orders of magnitude larger than the ensemble averaged currents as expected from the free electron theory [23, 24, 26, 27].

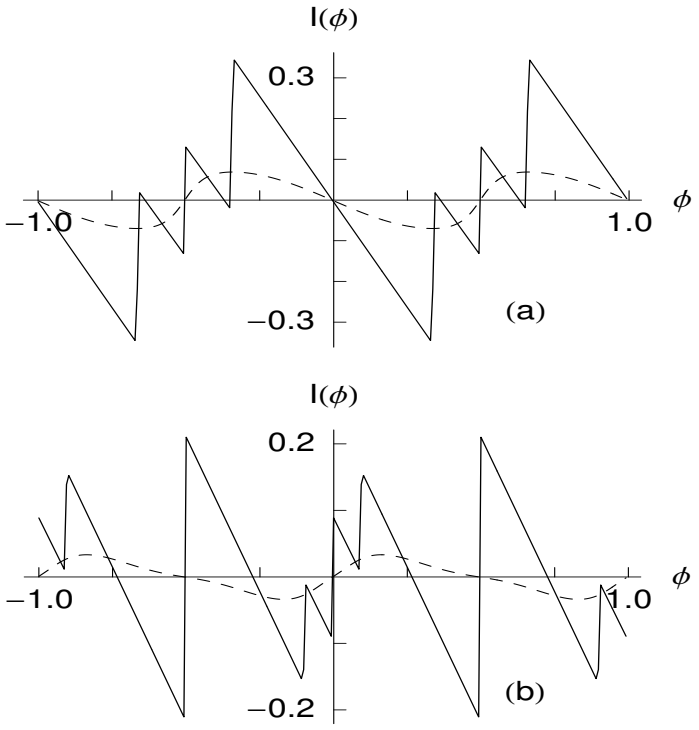

Fig. 5. I–ϕ curves of multi-channel mesoscopic cylindrical rings ($N = 50$, $M = 4$) with only NNH integrals for **(a)** $N_e = 45$ and **(b)** $N_e = 40$. The solid and dashed curves respectively corresponds to the perfect ($W = 0$) and dirty ($W = 1$) systems

On the other hand measurements on single isolated mesoscopic rings detected ϕ_0-periodic persistent currents with amplitudes of the order of $I_0 \sim ev_F/L$ (close to the perfect ring result). Theory and experiment [19] seem to agree only when disorder is weak. However, the amplitudes of the currents in the diffusive single-isolated-disordered gold rings [20] were two orders of magnitude larger than the theoretical estimates. This discrepancy initiated intense theoretical activity and it is still an open problem.

In the previous section it has been shown that the simple nearest-neighbor tight-binding Hamiltonian cannot explain the observed enhancement of persistent current even in the presence of electron-electron interaction [38,39,41,42]. Some recent works [43,44] show that a much more realistic model with higher order hopping integrals in the usual nearest-neighbor tight-binding Hamiltonian gives the desired enhancement of persistent current in the mesoscopic normal metal rings. Physically, the higher order hopping integrals try to delocalize the electrons even in one dimension preserving its phase coherence and hence prevents the reduction of persistent current due to disorder. The

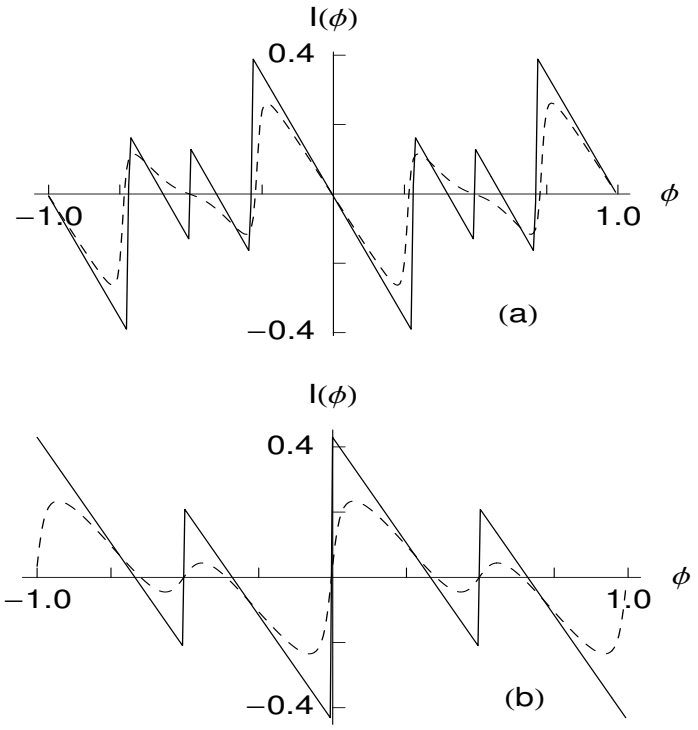

Fig. 6. I–ϕ curves of multi-channel mesoscopic cylindrical rings ($N = 50$, $M = 4$) with both NNH and SNH ($\alpha = 1$) integrals for **(a)** $N_e = 45$ and **(b)** $N_e = 40$. The solid and dashed curves respectively corresponds to the perfect ($W = 0$) and dirty ($W = 1$) systems

fluctuations of the persistent currents also get suppressed in the presence of higher order hopping integrals. As a result average amplitude of the persistent currents becomes comparable to I_0 and this is exactly what has been observed experimentally.

So far we have confined our discussions to one-dimensional systems only which do not really correspond to the experimental situations. Enhancement of persistent current has been observed even in the single-isolated diffusive (disordered) metallic rings. But diffusion is not possible strictly in one dimension and hence it becomes necessary to consider the finite width of the samples. The simplest way of doing this is to consider a cylindrical mesoscopic ring threaded by magnetic flux ϕ. A schematic view of this system is presented in Fig. 4. Assuming that the lattice spacings both in the longitudinal and transverse directions are identical (i.e., the surface of the cylinder forms a square lattice), we can describe the system by the Hamiltonian

$$H = \sum_{x} \epsilon_x c_x^{\dagger} c_x + \sum_{<xx'>} \left[v_{xx'} \mathrm{e}^{\mathrm{i}\theta_{xx'}} c_x^{\dagger} c_{x'} + \mathrm{h.c.} \right] , \tag{6}$$

where ϵ_x is the site energy of the lattice point x of coordinate, say, (i, j). Here $v_{xx'}$ is the hopping integral between the lattice points x and x', and, $\theta_{xx'}$ is the phase acquired by the electron due to this hopping in the presence of magnetic flux ϕ. Let us investigate the role of just the second-neighbor hopping integrals on persistent current and neglect the effects of all the other higher order hopping integrals. Let v denotes the nearest-neighbor hopping integral and the second-neighbor hopping integral (across the diagonal of the square) is taken to have the exponential form $v \exp(-\alpha)$, where α is a decay constant.

In the absence of impurity, setting $\epsilon_x = 0$ for all x, the energy eigenvalue of the nth eigenstate can be expressed as

$$E_n(\phi) = 2v \cos\left[\frac{2\pi}{N}(n+\phi)\right] + 4v\mathrm{e}^{-\alpha} \cos\left[\frac{2\pi}{N}(n+\phi)\right] \cos\left[\frac{2\pi m}{M}\right] + 2v \cos\left[\frac{2\pi m}{M}\right] , \tag{7}$$

and the persistent current carried by this eigenstate is

$$I_n(\phi) = \left(\frac{4\pi v}{N}\right) \sin\left[\frac{2\pi}{N}(n+\phi)\right] + \left(\frac{8\pi v}{N}\right) \mathrm{e}^{-\alpha} \sin\left[\frac{2\pi}{N}(n+\phi)\right] \times \cos\left[\frac{2\pi m}{M}\right] , \tag{8}$$

where n and m are two integers in the intervals $-\lfloor N/2 \rfloor \leq n < \lfloor N/2 \rfloor$ and $-\lfloor M/2 \rfloor \leq m < \lfloor M/2 \rfloor$, respectively. The symbol $\lfloor \ldots \rfloor$ denotes the integral part of its argument. Here M and N are the number of sites along the longitudinal and transverse directions of the cylinder, respectively.

Let us first describe the behavior of persistent currents in the multi-channel cylindrical rings using the nearest-neighbor tight-binding Hamiltonian. Figure 5 shows the current-flux characteristics of some mesoscopic cylindrical rings with $N = 50$ and $M = 4$. Figures 5(a) and 5(b) respectively corresponds to the systems with the number of electrons $N_e = 45$ and 40, where the solid curves represent currents in the absence of any impurity and the dotted curves are for the disordered systems. In the disordered cases ϵ_x's are taken as random variables with uniform box distribution of width W. These figures show that persistent currents for the perfect systems (solid curves) have many discontinuities within each ϕ_0 period. These discontinuities are due to the existence of degenerate energy levels at certain values of ϕ and these degeneracies get lifted as we introduce impurity into the system. Accordingly persistent currents of the disordered systems become continuous functions of ϕ as shown by the dashed curves in Figs. 5(a) and 5(b). It is observed that, even in the multi-channel cylindrical rings, the nearest-neighbor tight-binding

model gives orders of magnitude reduction of the persistent currents in the disordered rings compared to its value in the ballistic cases.

The behavior of persistent current in the multi-channel disordered mesoscopic rings changes drastically as we switch on the second-neighbor hopping integrals. In Fig. 6 we plot the current-flux characteristics of some disordered multi-channel cylindrical rings in the presence of second-neighbor hopping integral with system size $M = 50$ and $N = 4$. Figures 6(a) and 6(b) respectively corresponds to the cases with $N_e = 45$ and 40, where the solid and dashed curves have the same meanings as those of Fig. 5. It is apparent from Figs. 6(a) and 6(b) that the amplitudes of persistent currents in the dirty systems (dashed curves) are comparable to those of the perfect systems (solid curves). Physically the higher order hopping integrals try to delocalize the electrons and the current amplitudes get enhanced, even by an order of magnitude, in comparison to its estimates in the dirty samples by the usual nearest-neighbor tight-binding Hamiltonian. The present study reveals that the higher order hopping integrals play a very significant role in the enhancement of persistent current in diffusive mesoscopic rings.

2.4 Low-Field Magnetic Susceptibility of Mesoscopic Rings

The diamagnetic or paramagnetic sign of the low-field persistent currents also becomes a controversial issue due to the discrepancy between theory and experiment. Cheung et al. [23] predicted that the sign of the low-field persistent current is quite random with respect to the total number of electrons N_e or the specific realization of the disordered configurations of the ring. The possibility of both the diamagnetic and paramagnetic responses were also predicted theoretically by Yu and Fowler [45] in mesoscopic Hubbard rings. They observed that rings with odd N_e exhibit paramagnetic response while those with even N_e have diamagnetic response in the limit $\phi \to 0$. In an experiment on 10^7 isolated mesoscopic Cu rings Levy et al. [18] reported diamagnetic response for the low-field currents, while with Ag rings Chandrasekhar et al. [20] obtained paramagnetic response. In another experiment Jariwala et al. [21] also detected diamagnetic persistent currents in an array of 30 diffusive mesoscopic gold rings. The diamagnetic sign of persistent current in the vicinity of zero magnetic field was also observed in an experiment [22] on 10^5 disconnected Ag ring. Thus the sign of the low-field current is *a priori* not consistent with the theoretical predictions. In this section we will study the behavior of low-field magnetic susceptibility of single- and multi-channel mesoscopic rings through some exact calculations.

The magnetic susceptibility of a mesoscopic ring is given by the general expression

$$\chi(\phi) = \frac{N^3}{16\pi^2} \left(\frac{\partial I(\phi)}{\partial \phi} \right) . \tag{9}$$

The sign of $\chi(\phi)$ determines the paramagnetic or diamagnetic nature of the persistent currents. Here we will present results for the systems either with fixed number of electrons (N_e) or with fixed chemical potential (μ).

Let us first study the behavior of low-field magnetic susceptibility of impurity free single-channel mesoscopic rings with fixed N_e. Figure 7(a) gives the variation of $\chi(\phi \to 0)$ with respect to N_e for a perfect ring. It is observed

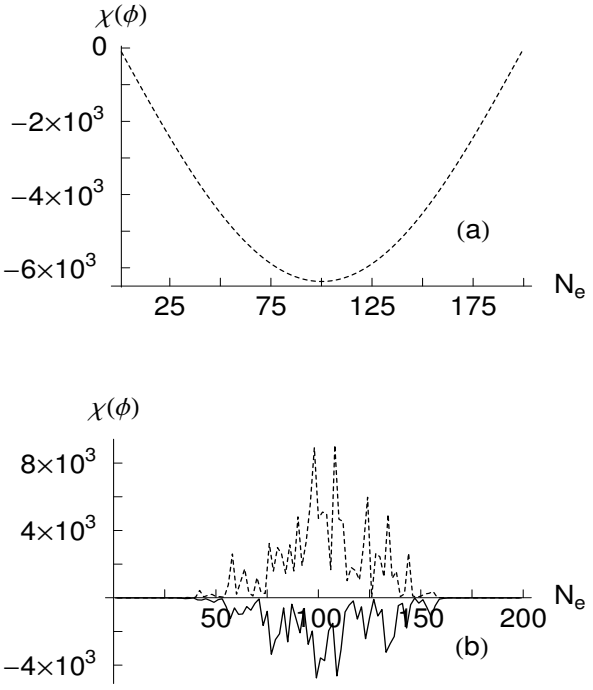

Fig. 7. Low-field magnetic susceptibility as a function of N_e for **(a)** perfect ($W = 0$) and **(b)** disordered ($W = 1$) rings of size $N = 200$. The solid and the dotted curves in Fig. 7(b) respectively corresponds to the cases with odd and even N_e

that the currents are always diamagnetic for both the cases with even and odd N_e. The effects of impurity on the low-field currents are quite interesting and we see that the sign of the currents can be predicted without any ambiguity even in the presence of disorder. Figure 7(b) is a plot of $\chi(\phi \to 0)$ as a function of N_e for the disordered rings. The solid and dotted curves in Fig. 7(b) respectively corresponds to the cases with odd and even N_e. Interestingly, we observe that the rings with odd N_e exhibit only diamagnetic sign while those with even N_e always have the paramagnetic sign. As disorder lifts the degeneracy of the energy levels of a perfect ring, the sharp discontinuities of the

I–ϕ characteristics (see the solid curves of Fig. 2) disappear and the slopes of the I–ϕ curves for even and odd N_e always have opposite signs near zero magnetic field (see the dashed curves of Fig. 2). It explains why the sign of the low-field currents of one-dimensional disordered rings with fixed number of electrons are quite independent of the specific realization of the disordered configurations and depends only on the oddness or evenness of N_e.

The above results hold true only at zero temperature. We now describe the effect of temperature on the sign of the low-field currents of single-channel mesoscopic rings. At finite temperature, the probability that an electron will

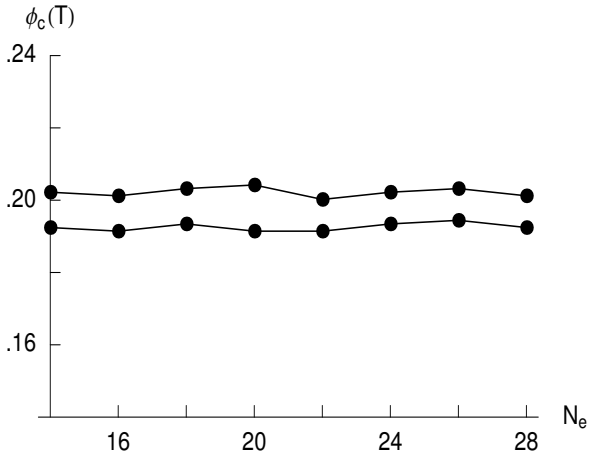

Fig. 8. $\phi_c(T)$ versus N_e curves of perfect rings with $N = 45$

occupy higher energy levels increases and the amplitude of persistent current decreases due to mutual cancellations of the opposite currents carried by the individual energy levels. Thermal excitations can scatter electrons inelastically and may randomize the phase of the electronic wave functions. This will try to destroy phase coherence of the electrons and may eliminate the quantum effects. So it becomes necessary to work at sufficiently low temperatures (defined by $T < T^{\star}$, where T^{*} is a characteristic temperature determined by the average level spacing Δ) such that phase coherence length of the electrons is larger than the circumference of the ring.

At finite temperature, we observe an interesting behavior in the low-field magnetic susceptibility of mesoscopic rings. Let us confine ourselves only to the systems with even Ne as we have observed that the systems with odd N_e also exhibit similar behavior. We see that the magnetic response of both the perfect and disordered rings with even N_e are always paramagnetic in the vicinity of zero magnetic field. The most interesting observation is that at finite temperature magnetic response of a given system becomes diamagnetic

beyond some critical magnetic flux $\phi_c(T)$. It may be noted that persistent currents do not show this kind of behavior at $T = 0$.

In Fig. 8 we plot $\phi_c(T)$ with respect to N_e for a perfect single-channel

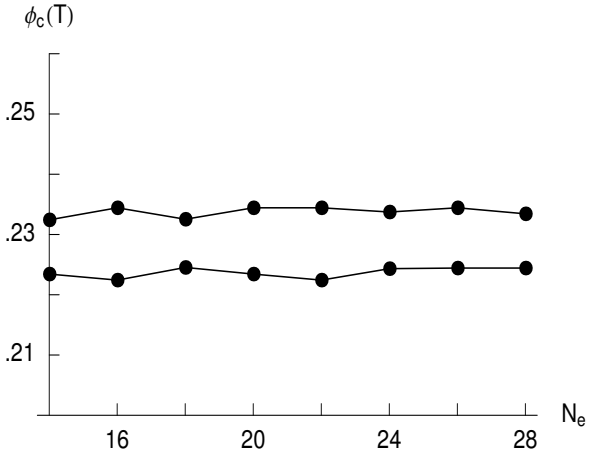

Fig. 9. $\phi_c(T)$ versus N_e curves of dirty ($W = 1$) rings with $N = 40$

ring of size $N = 45$. The curve with higher values of $\phi_c(T)$ corresponds to the temperature $T/T^\star = 1.0$, while the other curve corresponds to $T/T^\star = 0.5$. Figure 9 represents the behavior of $\phi_c(T)$ of the dirty samples ($W = 1$) for the same two temperatures $T/T^\star = 1.0$ (upper curve) and $T/T^\star = 0.5$ (lower curve). From Fig. 8 and Fig. 9 it is clear that the critical value of ϕ, where transition from paramagnetic to diamagnetic phase takes place, increases with the increase of temperature. Thus at finite temperature both the perfect and the dirty mesoscopic rings exhibit transition from paramagnetic to diamagnetic phase in the neighborhood of some critical values of magnetic flux.

The situation is quite different even at zero temperature when we describe the system by constant chemical potential, instead of fixed N_e. It may be noted that only for some particular values of μ, the system will have some fixed number of electrons, and, the sign of the low-field currents can be predicted according to the above prescriptions for these values of μ. While for other choices of μ, the total number of electrons varies even for a slight change of ϕ in the neighborhood of zero flux. Hence it is not possible to predict the sign of the low-field currents precisely even in the absence of any impurity in the system. Thus the sign of the low-field currents strongly depends on the choice of μ, the strength of disorder and the specific realization of the disordered configurations.

The study of the low-field magnetic responses of multi-channel mesoscopic rings reveals that it is not possible to predict the sign of the low-field currents precisely even for the impurity free cases with fixed number of electrons. So we

conclude that in the diffusive multi-channel mesoscopic rings, the sign of the low-field currents is a highly unpredictable quantity as it can be easily affected by the total number of electrons N_e, chemical potential μ, magnetic flux ϕ, the strength of disorder, the specific realization of the disordered configuration, etc. This is exactly what has been observed experimentally regarding the sign of the low-field currents.

3 Electron Transport in Molecular Bridge Systems

Remarkable progress in the fabrication of nanoscale quantum devices has allowed one to study electron transport through molecules in a very controllable way and these systems have attracted enormous attention in recent years since the molecules constitute the most promising building blocks for the future generation electronic devices. Electron conduction through molecules was first studied theoretically in 1970 [46]. Since then electronic transport properties of molecular bridge systems were studied in numerous experiments [47–56] placing the sample between two metallic electrodes with few nanometer separation. It becomes important to know how electron transport is affected by the structure of the molecule itself or by the nature of its coupling with the electrodes. Electronic transport through such two-terminal devices can also be controlled by some bias voltage. The current across the junction happens to be a strongly nonlinear function of the applied bias voltage understanding of which is a highly challenging problem.

In this section we will develop a simple analytical method for studying electronic transport properties of such molecular bridge systems within the framework of the tight-binding model. In this context it is worth mentioning that there are some excellent works on electron transport through molecular devices using *ab initio* methods [57–62] or in terms of some physical models [63–66].

3.1 Basic Theoretical Formulation

In this subsection we briefly describe how electronic transmission probability (T), conductance (g) and current (I) through a finite-size conducting system sandwiched between two semi-infinite metallic electrodes can be calculated using the Green's function technique. Let us first consider a one-dimensional conductor with N sites which is connected to the semi-infinite electrodes, the source and the drain, as shown schematically in Fig. 10. The conducting system in between the two electrodes can be a single molecule, an array of few molecules, an array of few quantum dots, etc. At low voltage and low temperature its conductance is given by the Landauer conductance formula

$$g = \frac{2e^2}{h} T \, . \tag{10}$$

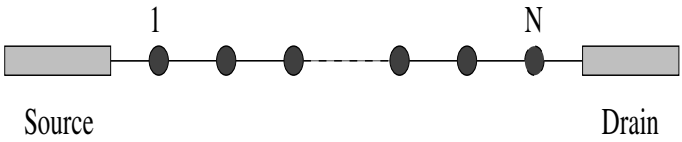

Fig. 10. Schematic view of a 1D conductor with N sites attached to the electrodes

The transmission probability T can be obtained from the relation [67]

$$T = \mathrm{Tr}\left[\Gamma_S G_c^r \Gamma_D G_c^a\right] \ , \tag{11}$$

where G_c^r and G_c^a are respectively the retarded and advanced Green's functions for the conductor including the effects of the electrodes. The matrix $\Gamma_{S(D)}$ depends on the coupling between conductor and source (drain), and can be expressed in terms of the self-energies of source (drain). The Green's function for the complete system (i.e., conductor with the electrodes) is defined as

$$G = (\epsilon - H)^{-1} \ , \tag{12}$$

where $\epsilon = E + \mathrm{i}\eta$, E being the energy of the source electrons and η is an infinitesimal real number. Evaluation of G requires the inversion of an infinite matrix as the system consists of the finite conductor and the semi-infinite electrodes. However, it can be partitioned into sub-matrices corresponding to the individual sub-systems and eliminating the degrees of freedom of the electrodes, the effective Green's function for the conductor can be expressed as

$$G_c = (\epsilon - H_c - \Sigma_S - \Sigma_D)^{-1} \ . \tag{13}$$

Here H_c is the bare Hamiltonian for the conductor which in the tight-binding model reads as

$$H_c = \sum_i \epsilon_i c_i^\dagger c_i + \sum_{<ij>} t \left(c_i^\dagger c_j + \mathrm{h.c.} \right) \ , \tag{14}$$

where ϵ_i's are the site energies and t denotes the nearest-neighbor hopping integral. In (13) the symbol $\Sigma_{S(D)}$ corresponds to the self-energy operator of source (drain). Explicitly we can write them as $\Sigma_S = h_{Sc}^\dagger g_S h_{Sc}$ and $\Sigma_D = h_{Dc} g_D h_{Dc}^\dagger$. Here $g_{S(D)}$ represents the Green's function for the source (drain) and $h_{Sc}(h_{Dc})$ is the coupling matrix between source (drain) and conductor. In terms of the retarded and advanced self-energies $\Sigma_{S(D)}^r$ and $\Sigma_{S(D)}^a$, the matrices Γ_S and Γ_D can be calculated from the expression

$$\Gamma_{S(D)} = \mathrm{i}\left[\Sigma_{S(D)}^r - \Sigma_{S(D)}^a\right] \ . \tag{15}$$

Tian et. al. [67] have shown that the self-energies can be written into the form

$$\Sigma_{S(D)}^r = \Lambda_{S(D)} - \mathrm{i}\Delta_{S(D)} \ . \tag{16}$$

The real part $\Lambda_{S(D)}$ of self-energy gives a shift to the energy levels of the conductor whereas the imaginary part $\Delta_{S(D)}$ produces some broadening to the energy levels. As these broadenings are much larger than the thermal broadening, we can restrict ourselves to zero temperature. The real and imaginary parts of the self-energies can be calculated in terms of the hopping integral $\tau_{S(D)}$ between the boundary sites of the conductor and the electrode $S(D)$, the energy E of the transmitting electron and the hopping integral v between the nearest-neighbor sites of the electrodes. All informations about the electrodes along with their couplings with the conductor are now contained within these two self-energies and this is just an extension of the Newns-Anderson chemisorption theory [63, 64].

Since the electrodes are connected to the conductor only at the boundary sites "1" and "N" as depicted in Fig. 10, we can express T as

$$T(E,V) = 4\Delta_{11}^{S}(E,V)\Delta_{NN}^{D}(E,V)|G_{1N}(E,V)|^2 \tag{17}$$

where $\Delta_{11} =< 1|\Delta|1 >$, $\Delta_{NN} =< N|\Delta|N >$ and $G_{1N} =< 1|G_c|N >$.

The current that passes through the conductor can be considered as a single-electron scattering process between the two reservoirs. Then current-voltage relation can be obtained from the following expression [68]

$$I(V) = \frac{e}{\pi\hbar}\int_{E_{\mathrm{F}}-eV/2}^{E_{\mathrm{F}}+eV/2} T(E,V)\mathrm{d}E\,, \tag{18}$$

where E_{F} is the equilibrium Fermi energy. For the sake of simplicity, here we assume that the entire voltage drop is across the conductor-electrode interfaces and this assumption does not greatly affect the qualitative aspects of the I–V characteristics. Using (17) we can express $I(V)$ as

$$I(V) = \frac{4e}{\pi\hbar}\int_{E_{\mathrm{F}}-eV/2}^{E_{\mathrm{F}}+eV/2} \Delta_{11}^{S}(E,V)\Delta_{NN}^{D}(E,V)|G_{1N}(E,V)|^2\mathrm{d}E\,. \tag{19}$$

Equations (10), (17) and (19) are respectively the basic working formulae for the calculation of transmission probability T, conductance g and current I through any finite size conductor sandwiched between two metallic electrodes.

3.2 Electron Transport Through Organic Molecules

In a recent experiment Reed et al. [55] studied the behavior of conductance and the current-voltage characteristics of benzene-1, 4-dithiolate (BDT) molecule in a two terminal geometry and obtained many interesting features that are highly reproducible. We will see that the qualitative behavior of the electronic transport properties of BDT molecules can be explained using the nearest-neighbor tight-binding model.

Transport through a molecule strongly depends on the geometry of its contacts with the electrodes, and also on the strengths of coupling between

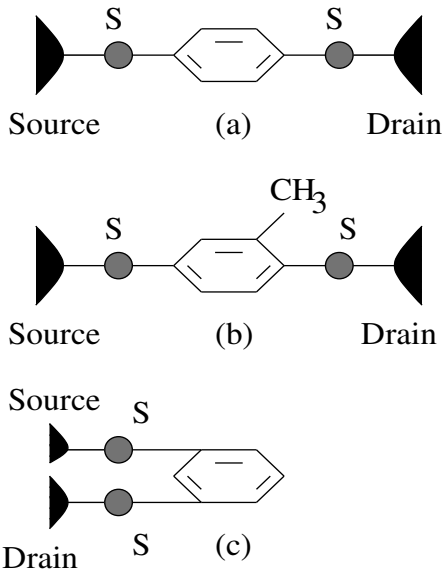

Fig. 11. Schematic view of a single benzene molecule attached to the electrodes by the thiol (S-H) groups. **(a)** The benzene molecule is attached symmetrically to the electrodes, **(b)** the symmetry is broken by adding the chemical substituent CH_3 in upper arm of the molecular ring and **(c)** the benzene molecule is attached asymmetrically to the electrodes

molecule and the electrodes. The behavior of electron transport through such molecular devices can be distinguished into two distinct regimes. One is the so-called weak coupling regime $\tau_{S(D)} << t$ while the other one is the strong coupling regime $\tau_{S(D)} \sim t$, where $\tau_{S(D)}$ is the hopping integral between the boundary sites of molecule and source (drain). In the subsequent discussions, the parameters in these two regimes are chosen as $\tau_S = \tau_D = 0.5$, $t = 3$ (in the weak coupling regime) and $\tau_S = \tau_D = 2.5$, $t = 3$ (in the strong coupling regime). The hopping integrals within the electrodes are taken as $v = 4$. Here all the parameters are measured in arbitrary units.

Our scheme of study in this section is as follows. In part A, we describe the conductance-energy and current-voltage characteristics of a single benzene molecule with contacts at non-equivalent geometrical points as well as with chemical substituents, while in part B we study electron transport through an array of benzene molecules. Lastly, in part C we investigate the effect of transverse magnetic field on electron transport through a single benzene molecule.

A) A Single Benzene Molecule

A single benzene molecule attached to the electrodes by thiol (S-H) groups are shown schematically in Fig. 11 for various inequivalent electrode-molecule-electrode geometries. Usually in the experiments gold (Au) electrodes are attached to the molecule by thiol (S-H) groups via chemisorption technique in which hydrogen atoms get removed and the sulfur (S) atoms reside. In Fig. 11(a) thiol groups are attached symmetrically at 1 and 4 positions of the

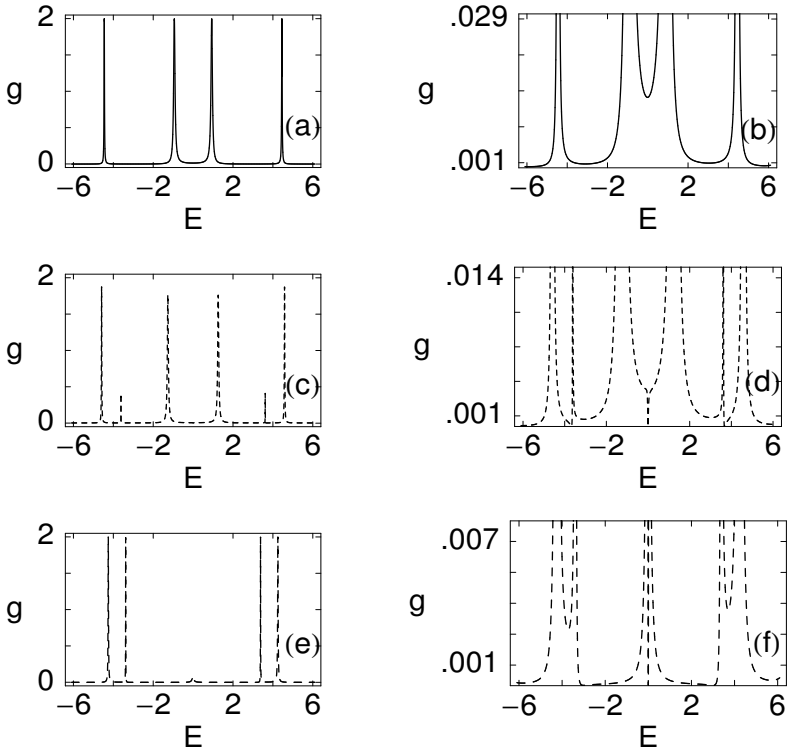

Fig. 12. *g-E* characteristics of the molecular bridges of Fig. 11 in the weak coupling regime. Figures **(a)**, **(c)** and **(e)** respectively corresponds to Figs. 11(a), 11(b) and 11(c). The second column shows the lower portions of the curves given in the first column to exhibit the anti-resonances of the conductance spectra

benzene molecule and it is known as benzene-1, 4-dithiolate (BDT) molecule. This symmetry can be broken either by adding some chemical substituent (say, CH_3) at any one arm of the molecular ring as shown in Fig. 11(b) or by sandwiching the molecule asymmetrically between the electrodes as illustrated in Fig. 11(c). In this way one can change the interference conditions

for the electronic wave functions which in turn greatly influence the behavior of electron transport through the molecular bridge systems.

In Fig. 12 we plot the conductance g of benzene molecule as a function of E in the weak coupling regime. Figures 12(a), 12(c) and 12(e) correspond to the

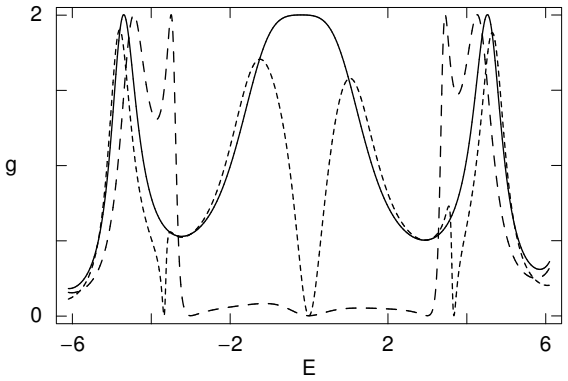

Fig. 13. g–E characteristics of the molecular bridges of Fig. 11 in the strong coupling regime. The solid, dotted and dashed curves respectively corresponds to Figs. 11(a), 11(b) and 11(c)

systems of Figs. 11(a), 11(b) and 11(c), respectively. In the second column of Fig. 12, we have magnified the lower portions of the curves given in Figs. 12(a), 12(c) and 12(e) to exhibit the various resonance and anti-resonance (zero-conductance) peaks more clearly. In the weak coupling regime conductance spectra exhibit sharp peaks. It has been observed that the resonance peaks appear exactly at the energy eigenvalues of the benzene molecule and hence the conduction spectrum may be considered as a fingerprint of the energy spectrum of the molecule itself.

It becomes apparent from the curves of Fig. 12 that quantum interference of the electronic wave functions plays a very significant role on electron conduction through such molecular devices. At some resonances conductance reaches its maximum possible value 2 with transmission probability being unity. However, there are other resonance peaks where conductance no-longer has the maximum possible value. This behavior can be easily understood from quantum interference of the electronic wave functions. As electrons are carried from source to drain through the molecule, there might be constructive or destructive interference due to superposition of the electronic wave functions corresponding to the two arms of the molecular ring. Accordingly electron transmission through the molecular device strongly depends on the electrode-molecule-electrode geometry. The anti-resonances (or zeros) of the conductance spectra are due to exact cancellation of the transmittances along the two paths.

In the strong coupling regime resonance peaks of the conduction spectra acquire substantial widths as shown in Fig. 13. The solid, dotted and dashed curves are respectively for the molecular bridges given in Figs. 11(a), 11(b) and 11(c). The strong coupling between molecule and the leads broadens the molecular energy levels and the resonance peaks become wider.

The most remarkable feature of the conductance spectra of these molecular bridge systems is that there exist conductance (transmittance) zeros both

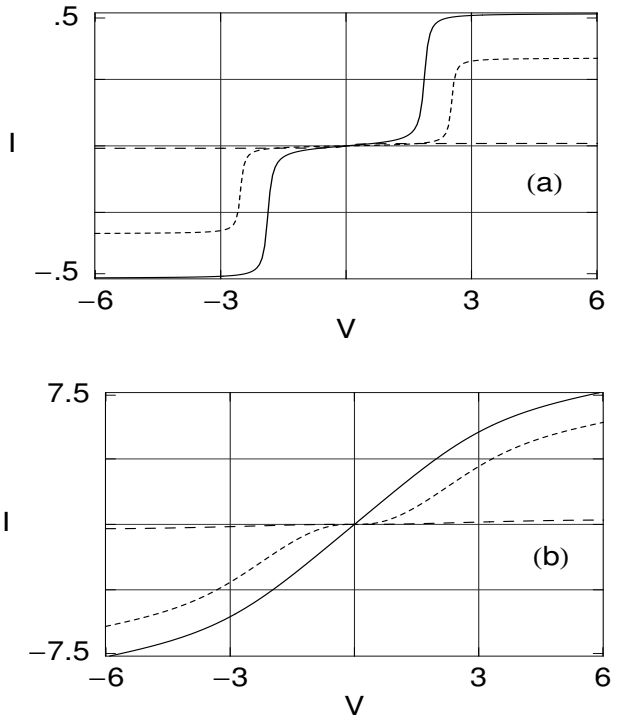

Fig. 14. $I - V$ curves of the molecular bridges of Fig. 11 in the **(a)** weak and **(b)** strong coupling regimes. The solid, dotted and dashed curves respectively corresponds to Figs. 11(a), 11(b) and 11(c)

in the weak and strong coupling regimes. Such anti-resonances are specific to the interferometric nature of the scattering states and do not occur in the usual one-dimensional scattering problems involving potential barriers. It may be noted that the positions of the anti-resonances on the energy scale are quite independent of the electrode-molecule-electrode coupling strengths (see Fig. 12 and Fig. 13). Since the width of the anti-resonance peaks are very small, they do not have any significant contribution in the current-voltage

characteristics. However, any change of interference condition has strong effect on the magnitude of currents that flow through the molecular bridge systems.

The scenario of electron transport through molecular junctions can be understood more from the current-voltage characteristics. Current that flows through a molecular junction can be computed from (19) and in Fig. 14 we display the current-voltage characteristics of the molecular bridge systems given in Fig. 11. Figures 14(a) and 14(b) respectively corresponds to the weak and strong coupling regimes. The solid, dotted and dashed curves are for the molecular bridges shown in Figs. 11(a), 11(b) and 11(c) respectively. It is observed that in the weak coupling regime currents exhibit staircase-like structures with sharp steps. This is due to the fact that as voltage increases, the electrochemical potentials of the electrodes gradually get shifted and a current channel opens up once they cross one of the molecular energy levels. Accordingly, a jump appears in the I–V characteristic. The shape and height of the steps depend on the widths of the resonance peaks. In the weak coupling regime we get steep jumps in the $I - V$ characteristics due to the presence of sharp resonance peaks in the conductance spectra. With the increase of electrode-molecule-electrode coupling strengths, the magnitude of current increases but its step-like behavior gradually disappears. Another feature of the $I - V$ characteristics is that the amplitudes of the currents are reduced as we introduce asymmetry in the system (see the dotted and dashed curves). It is solely due to quantum interference between the electronic wave functions corresponding to the two arms of the molecular ring. It is clear from the present study that electron transport through a molecular device strongly depends on the geometry of the junction and also on the coupling strengths between molecule and electrodes. Most importantly the model calculations provide a physical insight to the mechanism of electron conduction through the molecular bridge systems and the predictions are in good agreement with the qualitative features of the $g - E$ and $I - V$ characteristics that were observed experimentally [55].

B) An Array of Benzene Molecules

Recent experimental techniques allow one to connect an array of molecules between the electrodes [69], and it opens up the possibility to measure electron conduction through such array of molecules bounded chemically to the electrodes. In this subsection we will study electron transport through some arrays of benzene molecules within the tight-binding framework. Some typical arrays of benzene molecules are shown schematically in Fig. 15 which are attached with the electrodes by thiol (S-H) groups. Figure 15(a) is an array of benzene molecules without any chemical substituent, while Figs. 15(b), 15(c) and 15(d) are three inequivalent arrays with the chemical substituent CH_3.

The conductance spectra of these systems exhibit resonance and anti-resonance peaks similar to those of the single molecule bridges and hence we do not display them here explicitly. In these cases conductance spectra

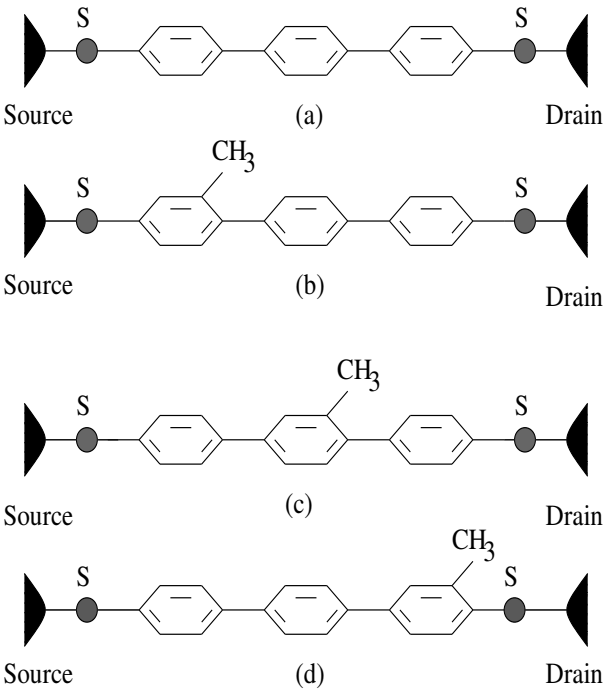

Fig. 15. Schematic view of array of benzene molecules attached to the electrodes by the thiol (S-H) groups, where **(a)** without any chemical substituent, **(b)** with chemical substituent (CH_3) at 1st molecule, **(c)** with chemical substituent (CH_3) at 2nd molecule and **(d)** with chemical substituent (CH_3) at 3rd molecule

have dense resonance peaks compared to those of Fig. 12 and Fig. 13 as the energy spectra of these systems are much more dense. Anti-resonance peaks appear in the conductance spectra of the bridges given in Figs. 15(b), 15(c) and 15(d) due to the breaking of molecular symmetry by the chemical substituent CH_3, In the weak coupling regime resonance peaks become very sharp and they broaden up in the strong coupling regime.

The current-voltage characteristics for these arrays of benzene molecules are shown in Fig. 16, where, Fig. 16(a) corresponds to the weak coupling cases while Fig. 16(b) corresponds to the strong coupling cases. The solid, dotted, small dashed and dashed curves are respectively for the molecular bridges given in Figs. 15(a), 15(b), 15(c) and 15(d). Currents have sharp staircase-like structures (see Fig. 16(a)) in the weak coupling regime and it is associated with the fact that the broadening of the molecular energy levels are not too much in this regime. With the increase of coupling strength current increases and the I–V curves becomes much more smooth as shown in Fig. 16(b). In the presence of chemical substituent CH_3 amplitude of the currents get reduced as a result of destructive quantum interference.

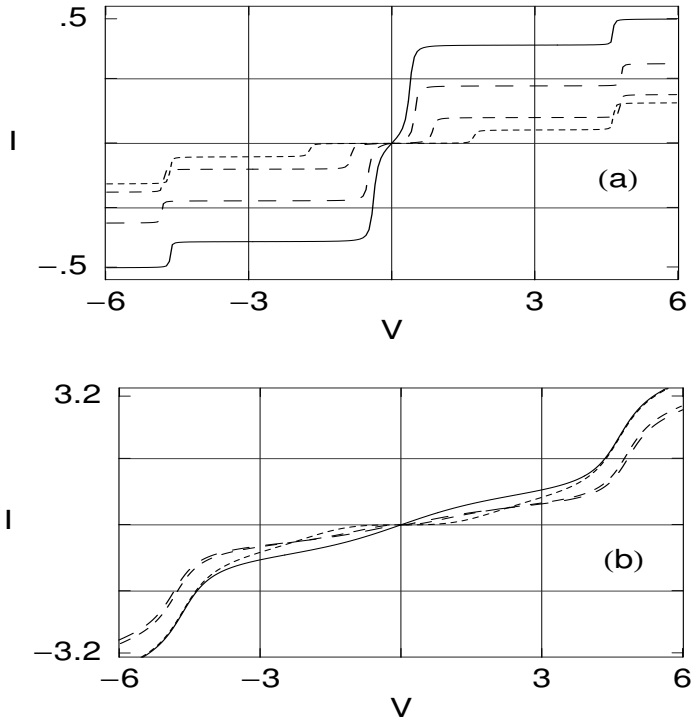

Fig. 16. $I-V$ curves for the systems given in Fig. 15 in the **(a)** weak and **(b)** strong coupling regimes. The solid, dotted, small dashed and dashed curves are respectively for the bridges shown in Figs. 15(a), 15(b), 15(c) and 15(d)

C) A Single Benzene Molecule Threaded by Magnetic Flux

In this subsection we will investigate the transport properties of electrons through a single benzene molecule in the presence of magnetic flux. The systems that we consider are the molecular bridges shown in Fig. 11. We assume that the magnetic flux ϕ threads the molecular ring in such a way that it does not penetrate the ring circumference and hence we can neglect the Zeeman term. Due to this magnetic flux ϕ, the electronic wave functions corresponding to the two arms of the molecular ring acquire different phase factors. Using gauge transformation, the Hamiltonian of the system can be expressed as

$$H_c = \sum_i \epsilon_i c_i^\dagger c_i + \sum_{<i,j>} t \left(\mathrm{e}^{\mathrm{i}\theta} c_i^\dagger c_j + \mathrm{h.c.} \right) , \tag{20}$$

where $\theta = 2\pi\phi/N$ is a phase factor due to the magnetic flux ϕ (measured in units of $\phi_0 = hc/e$). In Fig. 17, we plot g as a function of ϕ where the first, second and third rows respectively corresponds to the molecular bridges

shown in Figs. 11(a), 11(b) and 11(c). The left panel of Fig. 17 corresponds to the injection energy $E = 0.75$ of the electrons, while that in the right panel is $E = 1.0$. The solid and dotted curves respectively corresponds to the strong and weak coupling regimes. We observe that conductance exhibits the flux

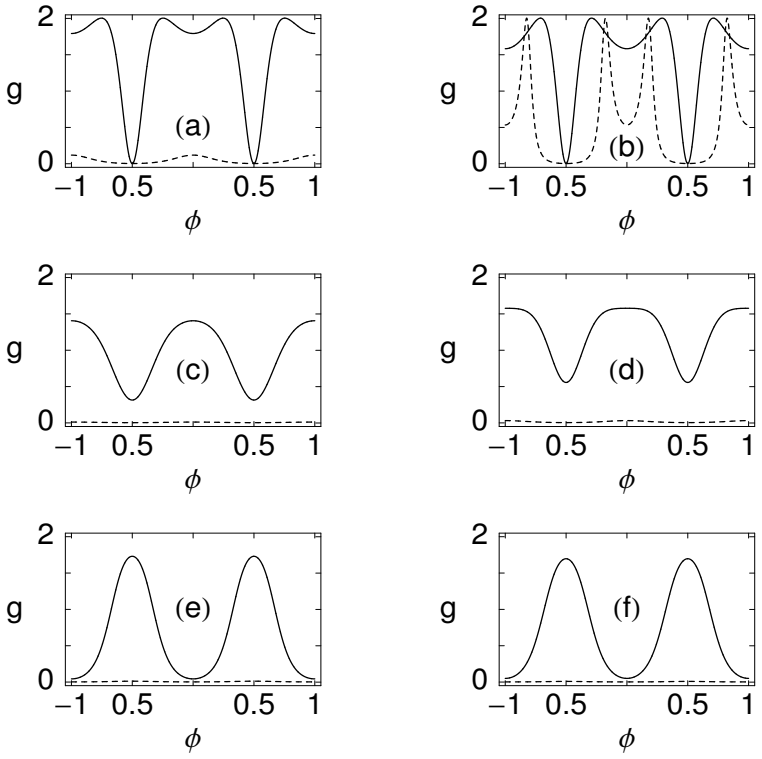

Fig. 17. $g-\phi$ curves for the systems given in Fig. 11 in the presence of magnetic flux. The first, second and third rows are respectively for the molecular bridges shown in Figs. 11(a), 11(b) and 11(c). The first and second columns respectively corresponds to $E = 0.75$ and $E = 1.0$. The solid and dotted curves respectively corresponds to the cases with strong and weak couplings between molecule and the electrodes

quantum ϕ_0 periodicity with extrema at half-integer values flux quantum and these features are quite independent of the strengths of coupling between the molecule and the electrodes. Conductance is very low in the weak coupling regime (dotted curves) whereas it is quite high in the strong coupling regime (solid curves). Though magnetic flux affects conductance spectra considerably, there is no appreciable change in the current-voltage characteristics. It shows that current across such molecular bridge junctions cannot be controlled effectively by applying magnetic field.

Let us now summarize all the results that we have obtained in this subsection on electron transport through BDT molecular bridge systems. Molecular

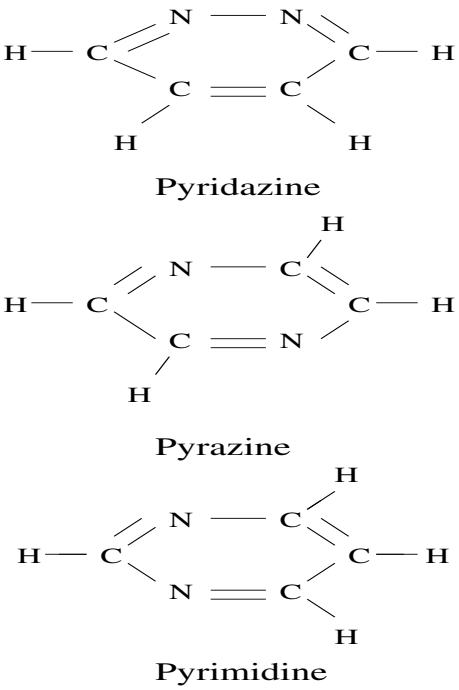

Fig. 18. Schematic view of the isomer molecules pyridazine, pyrazine and pyrimidine

geometry and electrode-molecule-electrode couplings have very important role on electron transport through the molecular bridge systems. In the symmetric cases conductance spectra have only resonance peaks and they appear exactly at the molecular energy levels. When molecular symmetry is broken by chemical substituent or by sandwiching the molecule asymmetrically between the electrodes, anti-resonance peaks appear in the conductance spectra due to quantum interference and the positions of these anti-resonance dips are quite independent of the coupling strengths between the molecule and the electrodes. In the weak coupling regime I–V characteristics have staircase-like structures due to the presence of sharp conductance resonances, while, in the strong coupling cases $I - V$ characteristics become smooth due to the broadening of the molecular energy levels. For same coupling strength between molecule and the electrodes, the amplitude of current is higher in a symmetrically connected molecule than that of a asymmetrically connected one. One can get even an order of magnitude large current just by increasing the coupling strengths between molecule and the electrodes. It has also

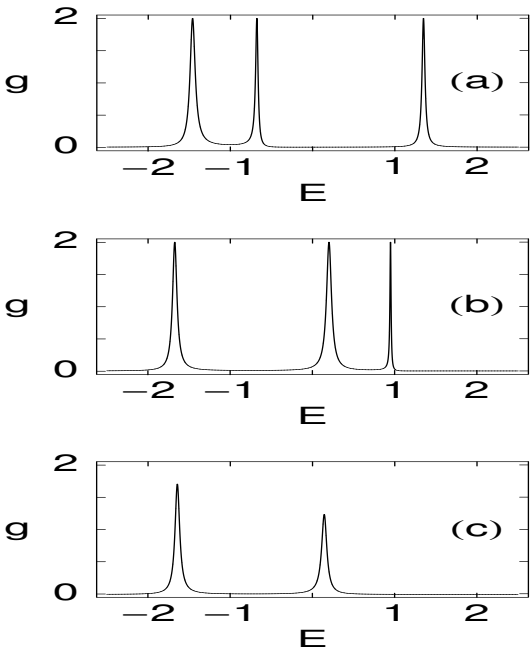

Fig. 19. $g - E$ curves for the isomer molecules **(a)** pyridazine, **(b)** pyrazine and **(c)** pyrimidine in the weak coupling regime

been observed that one cannot control current in the molecular bridge systems quite efficiently just by applying magnetic field.

3.3 Electron Transport Through Isomer Molecules

Electron transport through isomer molecules strongly depends on the arrangement of the atoms in these molecules even though they have the same chemical formula. This is due to the fact that the structural differences are associated with certain changes in the electronic structure of these molecules. Here we study the electron transport properties of pyridazine, pyrazine and pyrimidine molecules which are three different isomer molecules. Pyridazine, pyrazine and pyrimidine molecules are shown schematically in Fig. 18 indicating the position of the nitrogen atom in each of these molecules. Each molecule is connected to the electrodes via the thiol (S-H) groups (not shown explicitly in the figure). We assume that the electrodes are connected to the extreme left and to the extreme right hydrogen atoms of each molecule via the sulfur atoms of the thiol groups, where the hydrogen atoms of the thiol groups are removed by the chemisorption procedure.

We study electron transport properties of the isomer molecules using the nearest-neighbor tight-binding model. In this particular study, the Hamilto-

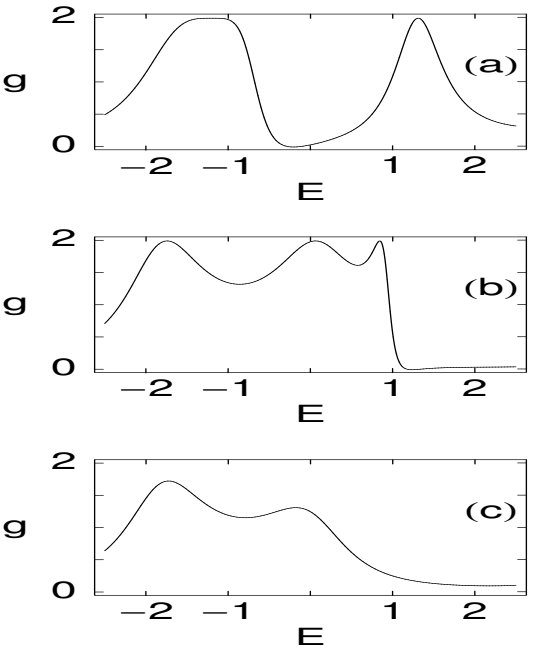

Fig. 20. $g - E$ curves for the isomer molecules **(a)** pyridazine, **(b)** pyrazine and **(c)** pyrimidine in the strong coupling regime

nian parameters for the weak coupling cases are taken as $\tau_S = \tau_D = 0.5$, $t = 3$, while those for the strong coupling cases are set to $\tau_S = \tau_D = 2$, $t = 3$. The hopping integrals inside the electrodes are taken as $v = 4$.

Figure 19 shows the variation of conductance g as a function of E for all the three isomer molecules in the weak coupling regime. Figures 19(a), 19(b) and 19(c) respectively corresponds to the pyridazine, pyrazine and pyrimidine molecules. Conductance spectra exhibit sharp resonance peaks and we observe that at the resonances transmission probability of pyridazine and pyrazine molecules becomes unity, while that for the pyrimidine molecule is less than unity. The most significant result is that both for the pyrazine and pyrimidine molecules, the systems are conducting even at low energy values (close to zero), and therefore these two molecular bridges allow electron conduction as soon as we apply a small voltage. On the other hand for the pyridazine molecule electron conduction starts only after some threshold energy and hence it becomes conducting only beyond some finite threshold voltage. It indicates that just by changing the arrangements of atoms in the molecule, one can go from a metallic phase to an insulating phase.

In Fig. 20, we plot the $g - E$ characteristics for the isomer molecules in the strong coupling regime, where Figs. 20(a), 20(b) and 20(c) are respectively

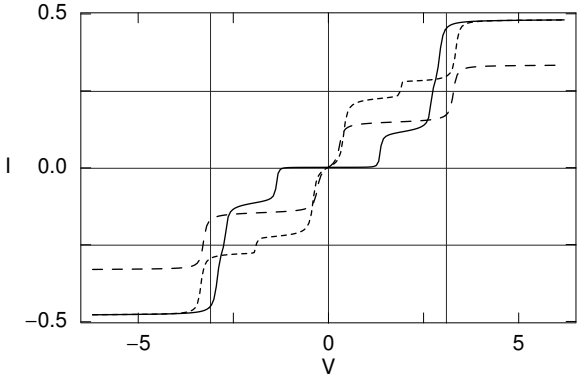

Fig. 21. $I - V$ curves of pyridazine (solid), pyrazine (dotted) and pyrimidine (dashed) molecules in the weak coupling regime

for the pyridazine, pyrazine and pyrimidine molecules. Though the resonance peaks become wider in the strong coupling regime, the most striking result is that the pyridazine molecule becomes conducting only beyond some finite threshold voltage whereas the other two isomer molecules start conducting as soon as we apply some non-zero voltage.

The above observations also become apparent from the $I - V$ characteristics of the isomer molecules. In Fig. 21 and Fig. 22 we plot the $I - V$ characteristics of these molecules in the weak and strong coupling regimes, respectively. The solid, dotted and dashed curves respectively corresponds

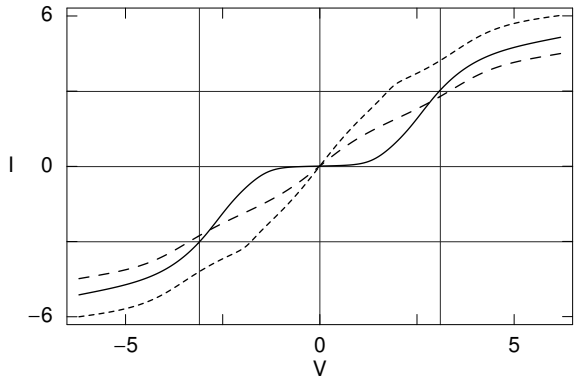

Fig. 22. $I - V$ curves of pyridazine (solid), pyrazine (dotted) and pyrimidine (dashed) molecules in the strong coupling regime

to the pyridazine, pyrazine and pyrimidine molecules. These figures clearly show that the bridges with the pyrazine and pyrimidine molecules have finite

currents (dotted and dashed curves) as soon as the applied voltage becomes non-zero while the molecular bridge with pyridazine molecule allows current only beyond some threshold voltage (solid curves). These properties of the isomer molecules might be useful in the fabrication of molecular switches.

3.4 Electron Transport Through a Quantum Wire Attached with a Mesoscopic Ring

Electron transport through a molecular bridge depends on several aspects, like, the geometry of the molecule, coupling strengths between molecule and the electrodes, isomeric variety, etc. and it can be tuned quite effectively just by adjusting the parameters of the system or its geometry. Now we describe a system where electron transport can be controlled by some external fields instead of changing the parameters of the system. It is a quantum wire attached with a mesoscopic ring as depicted in Fig. 23, where the ring is placed in an

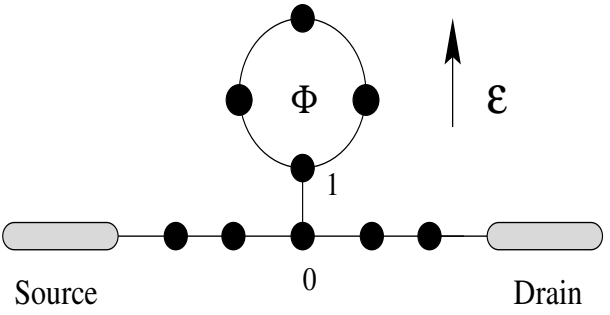

Fig. 23. A schematic view of a quantum wire attached with a quantum ring, the wire being connected to the electrodes. The ring is placed in an in-plane electric field $\mathcal{E}$ and encloses a magnetic flux ϕ

in-plane electric field $\mathcal{E}$ and it also encloses a magnetic flux ϕ. We describe the system by the tight-binding Hamiltonian

$$H_c = H_W + H_R + H_{WR} \,, \tag{21}$$

where

$$H_W = \sum_i \epsilon_i^w c_i^\dagger c_i + \sum_{<ij>} t_w \left(c_i^\dagger c_j + \text{h.c.} \right) \,,$$

$$H_R = \sum_m \epsilon_i^r c_m^\dagger c_m + \sum_{<mn>} t_r \left(c_m^\dagger c_n \mathrm{e}^{\mathrm{i}\theta} + \text{h.c.} \right) \,,$$

$$H_{WR} = t_0 \left(c_1^\dagger c_0 + c_0^\dagger c_1 \right) \,. \tag{22}$$

Here $\epsilon_i^w(\epsilon_i^r)$ is the site energy of the atoms belonging to ring(wire), $t_w(t_r)$ is the hopping integral between two nearest-neighbor sites of the ring(wire)

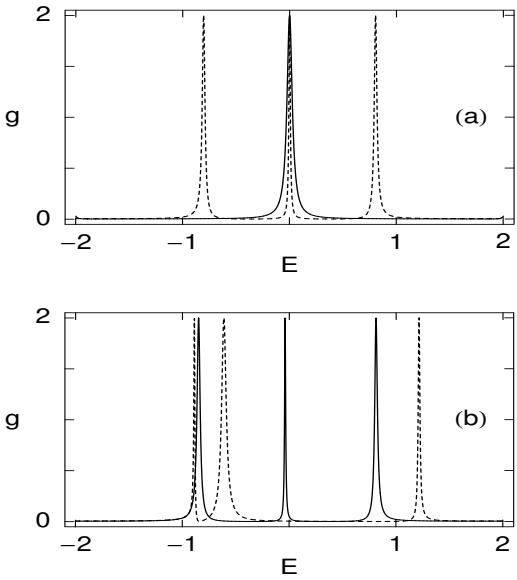

Fig. 24. $g - E$ curves for the system depicted in Fig. 23 in the weak coupling regime with $N = 10$ for the cases: **(a)** In the absence of any electric filed with $\phi = 0$ (solid) and 0.4 (dotted) and **(b)** in the presence of magnetic flux $\phi = 0.4$ with $\mathcal{E} = 2$ (solid) and 4 (dotted)

and the phase factor $\theta = 2\pi\phi/N$ is due to the magnetic flux ϕ. The ring is connected with the wire via the hopping integral t_0 between the sites "1" and "0" as shown in Fig. 23.

In the presence of an in-plane electric field $\mathcal{E}$ perpendicular to the wire, the site potentials of the atoms that belong to ring can be taken as

$$\epsilon_i^r = (e\mathcal{E}aN/2\pi)\cos\left[2\pi(i-1)/N\right] \tag{23}$$

where a is lattice the spacing and N is the number of sites in the ring. For simplicity we set $\epsilon_i^w = 0$ and take $t_w = t_r = t_0 = t$ (say). We investigate the electronic transport properties of this system for the following two cases: $\tau_S = \tau_D = 0.5$, $t = 3$ (in the weak coupling regime) and $\tau_S = \tau_D = 2$, $t = 3$ (in the strong coupling regime).

In Fig. 24, we plot conductance g as a function of energy E for this system in the weak coupling regime. Figure 24(a) shows the nature of the conductance spectra in the absence of any electric filed, where the solid and the dotted curves are respectively for $\phi = 0$ and 0.4. In Fig. 24(b) we present conductance spectra for non-zero electric fields keeping magnetic flux fixed at some typical value $\phi = 0.4$, where the solid and the dotted curves respectively corresponds to $\mathcal{E} = 2$ and 4. We observe that conductance vanishes everywhere excepting the resonances where it approaches the maximum possible value 2. For zero

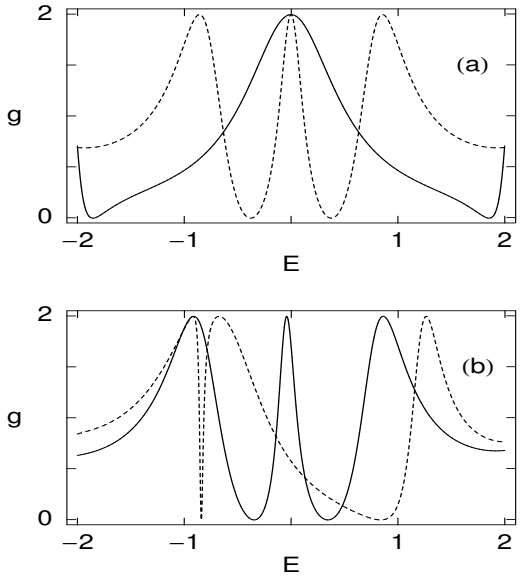

Fig. 25. $g - E$ curves for the system depicted in Fig. 23 in the strong coupling regime with $N = 10$ for the cases: **(a)** In the absence of any electric filed with $\phi = 0$ (solid) and 0.4 (dotted) and **(b)** in the presence of magnetic flux $\phi = 0.4$ with $\mathcal{E} = 2$ (solid) and 4 (dotted)

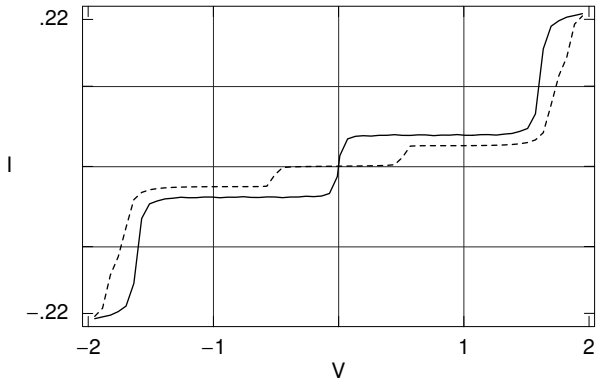

Fig. 26. $I - V$ curves of the system depicted in Fig. 23 in the weak coupling regime with $N = 10$ and $\phi = 0.4$. The solid and dotted curves respectively corresponds to $\mathcal{E} = 0$ and $\mathcal{E} = 3$

electric field conductance spectrum shows only one resonance peak at $E = 0$ when ϕ is zero (see the solid curve of Fig. 24(a)), while more resonance peaks appear in the spectrum when ϕ is non-zero (see the dotted curve of Fig. 24(a)). These additional peaks appear due to removal of degeneracy of the energy

levels in the presence of magnetic flux. When we apply electric field, the resonance peaks get shifted and conductance spectra become asymmetric with respect to E (see Fig. 24(b)).

In the strong coupling regime, the resonance peaks become wider [67] as shown in Fig. 25 where the solid and dotted curves correspond to the same cases as those in Fig. 24. We see that the behavior of the $g-E$ characteristics in this regime are very similar to those of the weak coupling regime. It may be noted from Fig. 24 and Fig. 25 that just by adjusting the in-plane electric field $\mathcal{E}$ or magnetic flux ϕ, the device can be made either conducting or non-conducting at some particular energy.

In Fig. 26 and Fig. 27 we plot the $I-V$ characteristics of this system in the weak and strong coupling regimes respectively with $\mathcal{E}=0$ (solid curve) and $\mathcal{E}=3$ (dotted curve) keeping magnetic flux fixed at $\phi=0.4$. We observe that in both the cases electric field suppresses current and most importantly

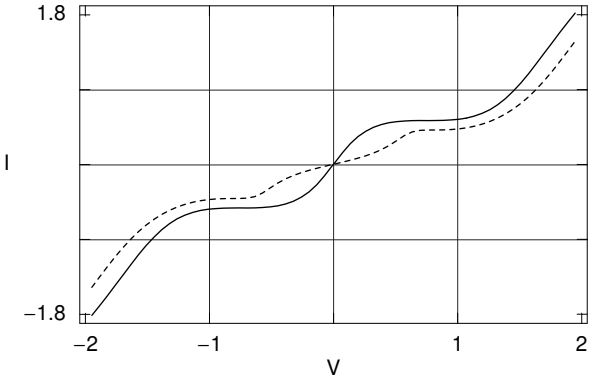

Fig. 27. $I-V$ curves of the system depicted in Fig. 23 in the strong coupling regime with $N=10$ and $\phi=0.4$. The solid and dotted curves respectively corresponds to $\mathcal{E}=0$ and $\mathcal{E}=3$

in the presence of electric field the system becomes conducting only beyond some threshold voltage. Thus the junction can be made either conducting or insulating just by applying an external electric field and this prediction on the basis our model calculations might be useful towards the fabrication of nanoscale switches.

4 Conclusions

In this article we have studied first the phenomenon of persistent current in mesoscopic normal metal rings in some details and show that the observed enhancement of persistent current in the diffusive mesoscopic rings can be explained by the simple tight-binding model just by including second-neighbor

hopping integrals in the Hamiltonian. We have also investigated the role of disorder and electron-electron interaction on persistent current through some exact numerical calculations. We observe that the dia- or para-magnetic sign of persistent current cannot be predicted precisely in the disordered samples. The second topic that has been discussed in this article is on electron transport through nanoscale devices. The qualitative behavior of the conductance-energy or current-voltage characteristics can be understood well using the tight-binding model. Our study based on the model calculations provides a physical insight to the behavior of electron conduction through such molecular devices. One importance result is the appearance of conductance (transmittance) zeros in the conductance spectra. These anti-resonance states do not occur in the conventional one-dimensional scattering problems involving potential barriers and are solely due to quantum interference of the electronic wave functions in these nanoscale systems. As a result we see that the positions of the anti-resonance peaks in the conductance spectra are quite independent of the coupling strengths between the device and the electrodes. We have studied electron transport through several molecular bridge systems (e.g., benzene molecules, isomer molecules, quantum wire, etc.) and show that quantum interference depends on the electrode-molecule-electrode geometry, position of the chemical substituents in the molecule, configuration of atoms of the isomer molecules, etc. We observer that quantum interference of the electronic wave functions also depends on external electric and magnetic fields and thus the device can be made either conducting or non-conducting just by adjusting these external fields.

References

1. S. Washburn, R. A. Webb: Adv. Phys. **35**, 375 (1986)
2. R. A. Webb, S. Washburn: Physics Today, **41**, 46 (1988)
3. K. von Klitzing, G. Dorda, M. Pepper: Phys. Rev. Lett. **45**, 494 (1980)
4. R. E. Prange, S. M. Girvin: *The Quantum Hall Effect* (Springer-Verlag, New York 1987)
5. T. Chakraborty, P. Pietiläinen: *The Quantum Hall Effects* (Solid State Sciences, Springer, Berlin 1995)
6. Yoseph Imry: *Introduction to Mesoscopic Physics* (Oxford University Press, New York 1997)
7. D. C. Tsui, H. L. Stormer, A. C. Gossard: Phys. Rev. Lett. **48**, 1559 (1982)
8. R. B. Laughlin: Phys. Rev. Lett. **50**, 1395 (1983)
9. F. Hund: Ann. Phys. (Leipzig) **32**, 102 (1938)
10. A. B. Fowler, A. Hartstein, R. A. Webb: Phys. Rev. Lett. **48**, 196 (1982)
11. P. A. Lee, A. Douglas Stone, H. Fukuyama: Phys. Rev. B **35**, 1039 (1987)
12. D. Mailly, M. Sanquer, J.-L. Pichard, P. Pari: Europhys. Lett. **8**, 471 (1989)
13. B. J. van Wees, H. van Houten, C. W. J. Beenakker, J. G. Williamson, L. P. Kouwenhoven, D. van der Marel, C. T. Foxon: Phys. Rev. Lett. **60**, 848 (1988)
14. D. A. Wharam, T. J. Thornton, R. Newbury, M. Pepper, H. Ahmed, J. E. F. Frost, D. G. Hasko, D. C. Peacock, D. A. Ritchie, G. A. C. Jones: J. Phys. C **21**, L209 (1988)

15. J. L. Costa-Krämer, N. Garcia, P. Garcia-Mochales, P. A. Serena: Surf. Sci. **342**, L1144 (1995)
16. Y. Aharonov, D. Bohm: Phys. Rev. **115**, 485 (1959)
17. M. Büttiker, Y. Imry, R. Landauer: Phys. Lett. A **96**, 365 (1983)
18. L. P. Levy, G. Dolan, J. Dunsmuir, H Bouchiat: Phys. Rev. Lett. **64**, 2074 (1990)
19. D. Mailly, C. Chapelier, A. Benoit: Phys. Rev. Lett. **70**, 2020 (1993)
20. V. Chandrasekhar, R. A. Webb, M. J. Brady, M. B. Ketchen, W. J. Gallagher, A. Kleinsasser: Phys. Rev. Lett. **67**, 3578 (1991)
21. E. M. Q. Jariwala, P. Mohant, M. B. Ketchen, R. A. Webb: Phys. Rev. Lett. **86**, 1594 (2001)
22. R. Deblock, R. Bel, B. Reulet, H. Bouchiat, D. Mailly: Phys. Rev. Lett. **89**, 206803 (2002)
23. H. F. Cheung, Y. Gefen, E. K. Riedel, W. H. Shih: Phys. Rev. B **37**, 6050 (1988)
24. G. Montambaux, H. Bouchiat, D. Sigeti, R. Friesner: Phys. Rev. B **42**, 7647 (1990)
25. R. Landauer, M. Büttiker: Phys. Rev. Lett. **54**, 2049 (1985)
26. F. von Oppen, E. K. Riedel: Phys. Rev. Lett. **66**, 84 (1991)
27. B. L. Altshuler, Y. Gefen, Y. Imry: Phys. Rev. Lett. **66**, 88 (1991)
28. A Schmid: Phys. Rev. Lett. **66**, 80 (1991)
29. H-F Cheung, E. K. Riedel, Y. Gefen: Phys. Rev. Lett. **62**, 587 (1989)
30. M. Büttiker: Phys. Rev. B **32**, 1846 (1985)
31. N. Byers, C. N. Yang: Phys. Rev. Lett. **7**, 46 (1961)
32. H. Bouchiat, G. Montambaux: J. Phys. (Paris) **50**, 2695 (1989)
33. M. Abraham, R. Berkovits: Phys. Rev. Lett. **70**, 1509 (1993)
34. A. Müller-Groeling, H. A. Weidenmuller: Phys. Rev. B **49**, 4752 (1994)
35. M. Kohmoto, B. Sutherland, C. Tang: Phys. Rev. B **35**, 1020 (1987)
36. A. Chakrabarti, S. N. Karmakar, R. K. Moitra: Phys. Lett. A **168**, 301 (1992)
37. U. F. Keyser, C. Fühner, S. Borck, R. J. Haug: Phys. Rev. Lett. **90**, 196601 (2003)
38. S. K. Maiti, J. Chowdhury, S. N. Karmakar: Phys. Lett. A **332**, 497 (2004)
39. S. K. Maiti, J. Chowdhury, S. N. Karmakar: Solid State Commun. **135**, 278 (2005)
40. B. Reulet, M. Ramin, H. Bouchiat, D. Mailly: Phys. Rev. Lett. **75**, 124 (1995)
41. G. Bouzerar, D. Poilblanc, G. Montambaux: Phys. Rev. B **49**, 8258 (1994)
42. T. Giamarchi, B. S. Shastry: Phys. Rev. B **51**, 10915 (1995)
43. S. K. Maiti, J. Chowdhury, S. N. Karmakar: Synthetic Metals **155**, 430 (2005)
44. S. K. Maiti, J. Chowdhury, S. N. Karmakar: J. Phys.: Condens Matter **18**, 5349 (2006)
45. N. Yu, M. Fowler: Phys. Rev. B **45**, 11795 (1992)
46. A. Aviram, M. Ratner: Chem. Phys. Lett. **29**, 277 (1974)
47. J. Reichert et al: Phys. Rev. Lett. **88**, 176804 (2002)
48. R. H. M. Smit et al: Nature **419**, 906 (2002)
49. J. Park et al: Nature **417**, 722 (2002)
50. W. Liang, M. P. Bockrath, M. Long, J. R. Park: Nature **417**, 725 (2002)
51. H. Park et al: Nature **407**, 57 (2000)
52. R. M. Metzger et al: J. Am. Chem. Soc. **119**, 10455 (1997)
53. C. M. Fischer, M. Burghard, S. Roth, K. V. Klitzing: Appl. Phys. Lett. **66**, 3331 (1995)

54. J. Chen, M. A. Reed, A. M. Rawlett, J. M. Tour: Science **286**, 1550 (1999)
55. M. A. Reed, C. Zhou, C. J. Muller, T. P. Burgin, J. M. Tour: Science **278**, 252 (1997)
56. T. Dadosh, Y. Gordin, R. Krahne, I. Khivrich, D. Mahalu, V. Frydman, J. Sperling, A. Yacoby, I. Bar-Joseph: Nature **436**, 677 (2005)
57. S. N. Yaliraki, A. E. Roitberg, C. Gonzalez, V. Mujica, M. A. Ratner: J. Chem. Phys. **111**, 6997 (1999)
58. M. Di Ventra, S. T. Pantelides, N. D. Lang: Phys. Rev. Lett. **84**, 979 (2000)
59. Y. Xue, S. Datta, M. A. Ratner: J. Chem. Phys. **115**, 4292 (2001)
60. J. Taylor, H. Gou, J. Wang: Phys. Rev. B **63**, 245407 (2001)
61. P. A. Derosa, J. M. Seminario: J. Phys. Chem. B **105**, 471 (2001)
62. P. S. Damle, A. W. Ghosh, S. Datta: Phys. Rev. B **64**, R201403 (2001)
63. V. Mujica, M. Kemp, M. A. Ratner: J. Chem. Phys. **101**, 6849 (1994)
64. V. Mujica, M. Kemp, A. E. Roitberg, M. A. Ratner: J. Chem. Phys. **104**, 7296 (1996)
65. M. P. Samanta, W. Tian, S. Datta, J. I. Henderson, C. P. Kubiak: Phys. Rev. B **53**, R7626 (1996)
66. M. Hjort, S. Stafström: Phys. Rev. B **62**, 5245 (2000)
67. W. Tian, S. Datta, S. Hong, R. Reifenberger, J. I. Henderson, C. I. Kubiak: J. Chem. Phys. **109**, 2874 (1998)
68. S. Datta: *Electronic Transport in Mesoscopic Systems* (Cambridge University Press, Cambridge 1997)
69. C. Kergueris et al: Phys. Rev. B **59**, 12505 (1999)

2D Disordered Electronic System in the Presence of Strong Magnetic Field

S. Sil[1], S. N. Karmakar[2], and Efrat Shimshoni[3]

[1] Department of Physics, Visva-Bharati University, Santiniketan 731 235, India.
sreekantha.sil@saha.ac.in
[2] Saha Institute of Nuclear Physics, 1/AF, Bidhannagar, Kolkata 700 064, India.
sachindranath.karmakar@saha.ac.in
[3] Department of Math-Physics, Oranim – University of Haifa, Tivon 36006, Israel.
efrats@research.haifa.ac.il

1 Introduction

The quantum Hall effect is one of the most remarkable phenomena in condensed matter physics discovered in the second half of the 20th century. In 1980 Klaus von Klitzing [1], who was investigating the magneto-galvanometric properties of the two-dimensional electron gas in high quality Silicon MOSFET in presence of high magnetic field, observed plateaus at integer multiples of the fundamental conductance quantum e^2/h in the Hall conductance and a vanishing longitudinal resistivity at a very low temperature ($\sim 1^o$ K). This phenomena is known as Integer Quantum Hall Effect (IQHE) (see Fig. 1).

It turns out, that the quantization of Hall resistivity was exact to one part in 10^8 [2]. This incredibly accurate quantization of Hall resistance in disordered two-dimensional semiconductors produced an important impact in physics (see reviews [2–4]), and necessitated a complete revision of the theory of electronic transport and Anderson localization [5–8]. Then just when everybody started to believe that two-dimensional electron systems were truly understood, this illusion was shattered in 1982 by the discovery of Fractional Quantum Hall Effect (FQHE) [9], where the Hall conductance is quantized in fractional multiples of e^2/h (see Fig. 2). Again this phenomena showed that existing transport theory of two-dimensional electronic system is inadequate to explain the surprises of nature and an improvement of the existing electronic transport theory is required.

In the early eighties a simple explanation was found for the integer quantum Hall effect in terms of non-interacting electrons in presence of disorder [3, 10, 11]. It was realized that the combination of a random potential, created by the impurities in a sample, and a strong magnetic field gives rise

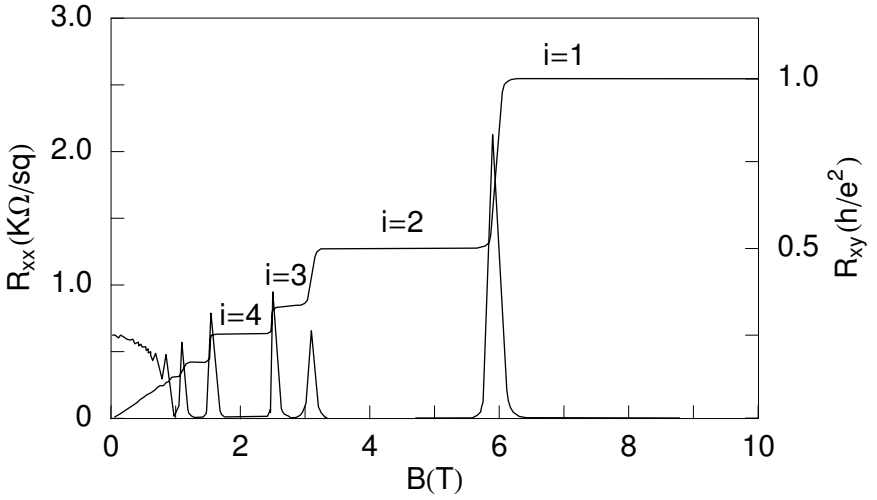

Fig. 1. The stepwise behavior of the transverse resistivity superimposed with the oscillatory behavior of the longitudinal resistance as a function of magnetic field. The plateaus of transverse resistivity coincide with the dissipationless behavior of the longitudinal resistance at each integer value of ν. Figure from [4]

to special coexistence of localized and extended states. Basically, the magnetic field controls the dimensionality of the wave function and makes the states lying near the equipotential lines of the potential in the strong magnetic field. Thus the states associated with closed loops of the equipotential lines are localized and states associated with the percolating equipotential lines are extended. When the Fermi level lies in an energy interval devoid of extended states, called a mobility gap, a change in the electron density or the magnetic field can only result in a different occupation of localized states, which does not affect the conductance. Thus a plateau is developed with respect to the change of electron density or magnetic field. Only when the Fermi-level reaches the extended states of the Landau sub-bands the conductance starts changing. The fact that the plateau of the quantum Hall effect lie exactly at integer multiple of e^2/h can be explained by relating the effect of impurity scattering on the localized electrons. A detailed investigation of the electron-impurity scattering in presence of a strong magnetic field shows that the extended states which are scattered by the impurity potential carry additional current and the sum of the excess current carried by the these delocalized electrons just exactly compensate the loss of the current caused by the localized states [2, 12, 13].

The fractional quantum Hall effect is the result of quite different underlying physics involving Coulomb interactions and correlation among the electrons. Many fancy ideas have been proposed to understand the role of electronic cor-

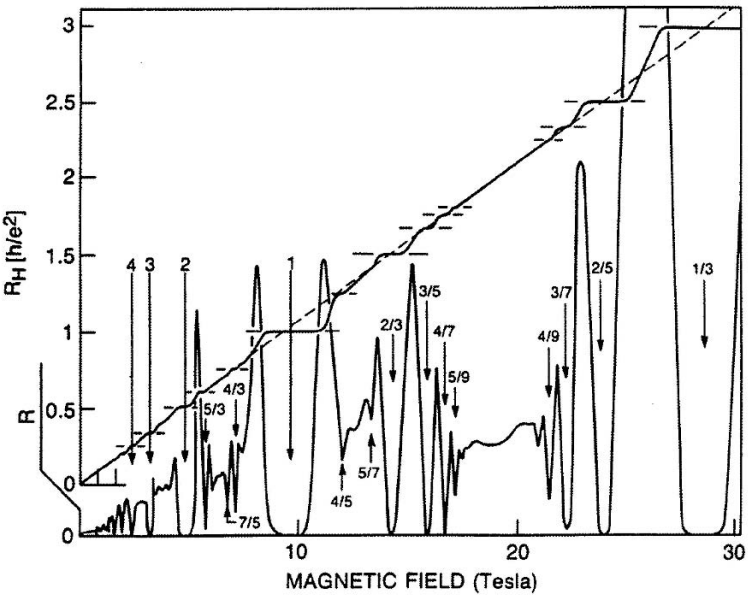

Fig. 2. The stepwise behavior of the transverse resistivity superimposed with the longitudinal resistance as a function of magnetic field, both for integer and fractional values of ν. Figure from [10]

relation in quantum Hall systems [14–19]. In 1983 Laughlin [14] introduced the famous Laughlin wave function to describe the fractional quantum Hall effect. It turns out that the ground state of the interacting electrons in strong magnetic field is an incompressible quantum fluid and the excitations of such fluid have the property of being fractional quantum numbers including fractional charge and fractional statistics that are neither Bose nor Fermi statistics. This correlated electron state can be interpreted in terms of electrons attached to magnetic flux quanta, which behave like composite particle. Since a large part of the magnetic field has gone into the formation of composite particle, the field that these particles feel is much smaller than the original one; in fact it exactly mimics the integer quantum Hall effects field strength. In this way the fractional quantum Hall effect is explained as the integer effect for composite particles.

The review presented here is an introduction to the basics of the integer and fractional quantum Hall effects. In the Sect. 2 and Sect. 3 we briefly present an explanation of the integer quantum Hall effect followed by the role played by edge states in a quantum Hall bar. In Sect. 4 and Sect. 5 we discuss the fractional quantum Hall effect followed by an introduction to the theory of edge states in fractional quantum Hall effect.

2 Integer Quantum Hall Effect

In this section we will discuss the behavior of non-relativistic electrons in presence of high magnetic field and disorder in the two-dimensional plane. In the presence of a strong magnetic field we can assume that the electron spins are polarized, so that their spin degrees of freedom can be discarded. When the short-range disorder potential is strong enough, the motion of the charge-carriers are mostly governed by the electron-disorder scattering and the coulomb repulsion between the electrons can be neglected. Therefore, the Hamiltonian for an electron in a strong homogeneous field and disorder can be expressed as

$$H = \frac{1}{2m}((p_x + eA_x)^2 + (p_y + eA_y)^2) + V(x, y) \ , \tag{1}$$

where, $p_x = -\mathrm{i}\hbar\frac{\partial}{\partial x} = -\mathrm{i}\hbar\partial_x$ and $p_y = -\mathrm{i}\hbar\frac{\partial}{\partial y} = -\mathrm{i}\hbar\partial_y$ are the x- and y-component of the kinetic momentum, and A_x and A_y are the components of the vector potential describing the magnetic field $\boldsymbol{B} = (0, 0, -B)$. The corresponding covariant momentum is

$$\begin{aligned} P_x &= -\mathrm{i}\hbar\partial_x + eA_x \ , \\ P_y &= -\mathrm{i}\hbar\partial_y + eA_y \ . \end{aligned} \tag{2}$$

We decompose the electron coordinate $\boldsymbol{r} = (x, y)$ to the guiding center (X, Y) and the relative coordinate $(\zeta, \eta) = (-\frac{1}{eB}P_y, \frac{1}{eB}P_x)$ as

$$\begin{aligned} x &= X + \zeta \ , \\ y &= Y + \eta \ . \end{aligned} \tag{3}$$

In the presence of a magnetic field B we observe the following commutation relations,

$$[X, Y] = -\mathrm{i}l_B^2 \ , \qquad [\zeta, \eta] = \mathrm{i}l_B^2 \ , \tag{4}$$

$$[X, \eta] = [X, \zeta] = [Y, \eta] = [Y, \zeta] = 0 \ , \tag{5}$$

with $l_B = \sqrt{\hbar/eB}$ is the fundamental length-scale of the Quantum Hall (QH) system. We now construct two pairs of Landau level ladder operators as

$$\begin{aligned} a &= \frac{l_B}{\sqrt{2}\hbar}(P_x + \mathrm{i}P_y), \ a^\dagger = \frac{l_B}{\sqrt{2}\hbar}(P_x - \mathrm{i}P_y), \\ b &= \frac{1}{\sqrt{2}l_B}(X - \mathrm{i}Y), \ b^\dagger = \frac{1}{\sqrt{2}l_B}(X + \mathrm{i}Y), \end{aligned} \tag{6}$$

obeying

$$[a, a^\dagger] = [b, b^\dagger] = 1 \ ,$$

$$[a, b] = [a^\dagger, b] = 0 \ .$$

The Hamiltonian can be rewritten as

$$H = \hbar\omega_c(a^\dagger a + \frac{1}{2}) + V(X + \zeta, Y + \eta) \ , \tag{7}$$

with the cyclotron frequency $\omega_c = eB/m$. In the absence of impurity potential ($V(x, y) = 0$) eigenvalues of the energy are similar to those of the harmonic oscillator with frequency $\omega = \omega_c$,

$$E_n = \hbar\omega_c(n + \frac{1}{2}) \ . \tag{8}$$

They are called the Landau levels. There exists a degeneracy in each Landau level since the Hamiltonian (7) does not depend on b and $b^\dagger$. The degeneracy is proportional to the size of the system and is given by AeB/h, where A is the area of the system and h is the Planck's constant. The motion of an electron within one Landau level is specified by the guiding center (X, Y). Since coordinates X and Y do not commute, it is necessary to choose one of them or appropriate combination of them to construct the eigenstates. This is related to the choice of gauge. We take the Landau gauge which is the best gauge to analyze the QH system in a rectangular geometry. In the Landau gauge the external magnetic field is expressed as

$$A_x = By, \ A_y = 0 \ . \tag{9}$$

Now it is straight forward to obtain the wave function of the system as

$$\psi_{n,k} = \frac{1}{\sqrt{L}} \exp[-\mathrm{i}kx] \exp[-(y - y_k)^2/2l_B^2] H_n(y/l_B) \ , \tag{10}$$

where L is the length of the system, $y_k = kl_B^2$ and $H_n(y/l_B)$ is the nth Hermite polynomial. However, a real QH system consists of some kind of disorder potential due to lattice defects or ionized donors. Moreover, electrons interact with each other to produce correlation effects on the motion of the electrons. In the case of integer quantum Hall effect, the effect of disorder on the motion of the electron dominates over the electron-electron interaction and we can ignore the role of electron-electron interaction in the system (for fractional quantum Hall effect the electronic correlations are important). In the presence of disorder the degeneracy of the Landau levels are lifted and the delta peaks in the density of states transform into structures with finite width. A well-known approach to calculate the shape of the disorder broadened Landau levels is the self-consistent Born approximation (SCBA), where only single scattering events are taken into account. Within the SCBA approach, the density of states of the integer quantum Hall system is found to be [20]

$$\rho(E) = \frac{1}{2\pi l_B^2} \sum_n \frac{2\pi}{\Gamma_n} \exp\left[-2(E - E_n)^2/\Gamma_n^2\right] \ , \tag{11}$$

where Γ_n is a Landau level dependent width. In addition to the broadening of the Landau levels a disorder potential changes the nature of the eigenstates significantly. In two dimensions (for zero magnetic field and non-interacting electrons) all states are localized even for arbitrary weak disorder. The essence of this weak localization effect is the current 'echo' associated with the quantum interference correction to Boltzmann transport. These quantum interference effects rely crucially on the existence of time-reversal symmetry. In the presence of a strong magnetic field, time-reversal symmetry is destroyed and the localization properties of the disordered 2D electron gas is expected to be destroyed. This is in contradiction to the experimental observation of quantum Hall plateau which is a manifestation of localized states at the tail of Landau sub-bands and extended states at the middle. To understand the nature of eigenstates in the Landau sub-band we consider the limiting case of very high magnetic field ($B \to \infty$) so that the magnetic length is small on the scale over which the potential varies.

Since the expectation value of ζ and η in the Hamiltonian (7) are of the order of $l_B \sim 1/\sqrt{B}$, we can neglect ζ and η in the argument of the impurity potential $V(X + \zeta, Y + \eta)$ in the limit $B \to \infty$. Thus in the very strong magnetic field the Hamiltonian (7) reduces to

$$H = \hbar\omega_c(a^\dagger a + \frac{1}{2}) + V(X, Y) \ . \tag{12}$$

Furthermore, in the limit $B \to \infty$ the commutator of X and Y vanishes like $1/B$ and thus quantities X and Y can be treated as classical variables. Therefore,

$$\frac{\mathrm{d}V}{\mathrm{d}t} = \frac{\partial V}{\partial X}\frac{\mathrm{d}X}{\mathrm{d}t} + \frac{\partial V}{dY}\frac{\mathrm{d}Y}{\mathrm{d}t} = 0 \tag{13}$$

because of the canonical equations of motion

$$\begin{aligned} \frac{\mathrm{d}X}{\mathrm{d}t} &= \frac{l_B^2}{\hbar}\frac{\partial V}{\partial Y} \ , \\ \frac{\mathrm{d}Y}{\mathrm{d}t} &= -\frac{l_B^2}{\hbar}\frac{\partial V}{\partial X} \ . \end{aligned} \tag{14}$$

Since $\mathrm{d}V/\mathrm{d}t = 0$, the electron moves along the equipotential orbits $V(X, Y) =$ constant. If the random potential $V(x, y)$ is symmetrically distributed around $V = 0$ and its variation $\Delta V = V_{\max} - V_{\min}$ is small compared to the Landau level splitting ($\Delta V \ll \hbar\omega_c$), all equipotential orbits with $V \neq 0$ are closed and open equipotential orbits exist only for $V = 0$, i.e., at energy $\epsilon_n = \hbar\omega_c(n + \frac{1}{2})$ (A sketch of the equipotential orbits in a randomly distributed disordered potential is shown in Fig. 3 [21]). Therefore, we would expect localized states

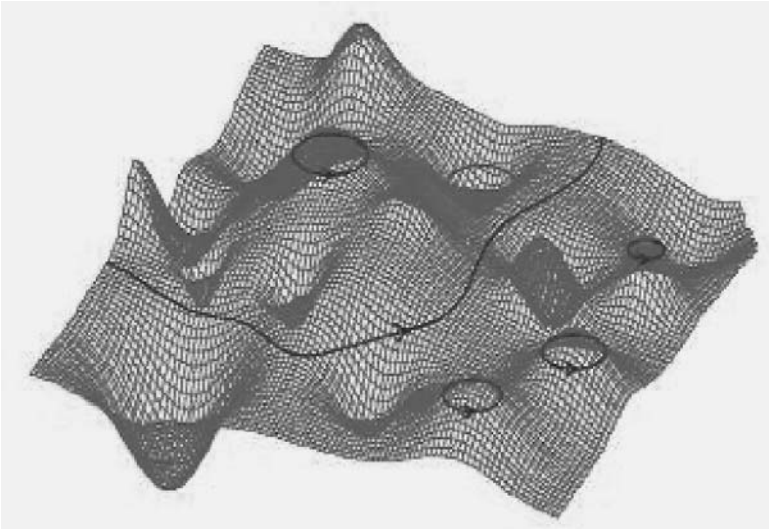

Fig. 3. Disorder potential with closed orbits (localized states) and open orbits (extended states). Figure from [21]

at the tail of the Landau sub-bands while extended states exist at the middle. The electronic density of states for the disordered Landau model is sketched in Fig. 4.

In the presence of an infinitely weak external electric field the closed equipotential orbits do not contribute to the net charge transfer; all the average current is carried by the open orbits. As long as temperature T is very low and the Fermi energy ϵ_{F} varies in a range of closed equipotential orbits (localized states) the Hall current remains constant (plateau). Around $\epsilon_{\mathrm{F}} = \epsilon_n$ the Hall conductivity increases with increasing number of occupied extended states and $1/R_H$ rises from one plateau value to the next (where $R_H \equiv R_{xy}$ is the Hall resistance). A detailed numerical studies on the nature of the eigenstates in the integer quantum Hall system is given by Ando [22]. Aers and Mcdonald [23] also confirms the existence of localized states, confined to a small region centered on an impurity at the tail of the Landau sub-bands and extended states, that spread through the whole system at the center of Landau sub-bands.

An electron in a two-dimensional plane in the presence of a perpendicular magnetic field makes a cyclotron motion around the guiding center. The guiding center induces a current in response to an external electric field $\boldsymbol{E} = (E_x, E_y)$. In the absence of any impurity potential, the Hamiltonian of the system is given by

$$H = \frac{1}{2m}(P_x^2 + P_y^2) + eE_x x + eE_y y \; . \tag{15}$$

The x- and y-component of the drift velocity (v_{D_x}, v_{D_y}) can be obtained as

$$v_{D_x} = \frac{\mathrm{d}X}{\mathrm{d}t} = \frac{1}{\mathrm{i}\hbar}[X, H] = \frac{eE_y}{\mathrm{i}\hbar}[X, Y] = -\frac{1}{B}E_y \ ,$$
$$v_{D_y} = \frac{\mathrm{d}Y}{\mathrm{d}t} = \frac{1}{\mathrm{i}\hbar}[Y, H] = \frac{eE_x}{\mathrm{i}\hbar}[Y, X] = \frac{1}{B}E_x \ . \tag{16}$$

When the system consists of a homogeneous distribution of electrons with density ρ_0, the current density is

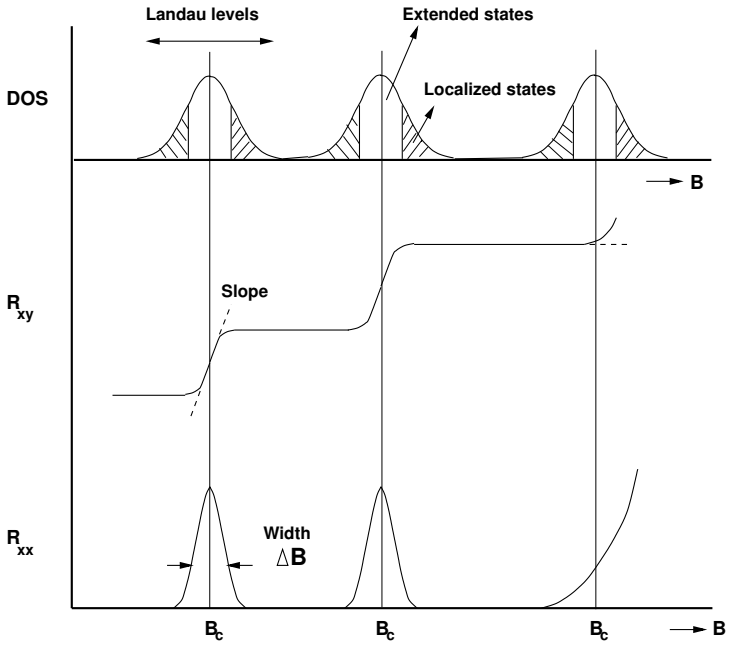

Fig. 4. Schematic density of states for the disordered Landau model. The shaded regions represent localized states

$$J_x = -e\rho_0\dot{X} = \frac{e\rho_0}{B}E_y = \frac{e^2\nu}{h}E_y \ ,$$
$$J_y = -e\rho_0\dot{Y} = -\frac{e\rho_0}{B}E_x = -\frac{e^2\nu}{h}E_x \ , \tag{17}$$

where $\nu = h\rho_0/eB$ is the filling factor. Note that the current flows in the direction perpendicular to the direction of the applied electric field. This is the Hall current. The Hall resistivity R_H is

$$R_H \equiv R_{xy} = \frac{h}{\nu e^2} \ . \tag{18}$$

When the filling fraction ν attains an integer value, i.e., when Landau sub-bands are completely filled, the Hall resistivity shows value h/ne^2 which is the Hall resistivity in the plateau region of the integer quantum system. However, in the plateau region of the integer quantum Hall system the filling fraction is not an integer. In a real quantum Hall system we have shown that there are localized states in the Landau sub-band which do not contribute to conductivity. So one would expect that the Hall conductivity in the plateau region will be less that the quantized value ne^2/h which is different from the experimental observation. From the experimental results it follows that the loss of the contribution of current due to localized states is exactly compensated by the electrons in the remaining delocalized (extended) states. A very important result in this context has been given by Prange [11]. This is an exact solution of a single δ function impurity potential inside an otherwise ideal system. Such a potential (attractive or repulsive) binds a single localized state from each Landau level. The localized state does not carry current. Each of the remaining extended state, however, carries a little bit more current, exactly compensating the loss of the companion. As discussed before, in the presence of disorder some states will be localized around impurities, while others, the states near the center of the Landau sub-bands, will be slightly modified. Energies in the scattering states ϵ_n will differ from those of the corresponding states of ϵ_n^0 of the pure system by a small amount $\Delta\epsilon_n^0$. Energies of localized states on the other hand will change appreciably. The important point is that energies of the extended states depend on the gauge potential, while those of localized states do not.

In the presence of scattering the wave function of the incoming wave $\psi_0(\boldsymbol{r}, E)$ goes through a phase shift $\delta(E)$ in the asymptotic region [24], i.e.,

$$\psi(\boldsymbol{r}, E) \to \mathrm{e}^{\mathrm{i}\delta(E)}\psi_0(\boldsymbol{r}, E) \; . \tag{19}$$

Let the wave packet, before it reaches the scattering center, have the form

$$W_0(\boldsymbol{r}, t) = \int \mathrm{d}E \; F(E)\psi_0(\boldsymbol{r}, E)\mathrm{e}^{-\mathrm{i}Et} \tag{20}$$

where $F(E)$ is an envelop function strongly peaked around E_0. The form for W after the scattering is

$$\begin{aligned} W(\boldsymbol{r}, t) &= \int \mathrm{d}E \; F(E)\psi_0(\boldsymbol{r}, E)\mathrm{e}^{\mathrm{i}\delta(E)-\mathrm{i}Et} \\ &= \int \mathrm{d}E \; F(E)\psi_0(\boldsymbol{r}, E) \exp\left[\mathrm{i}\left(\delta(E_0) + (E - E_0)\left.\frac{\mathrm{d}\delta}{\mathrm{d}E}\right|_{E_0} - Et\right)\right] \\ &= \exp\left[\mathrm{i}\left(\delta(E_0) - E_0\left.\frac{\mathrm{d}\delta}{\mathrm{d}E}\right|_{E_0}\right)\right] W_0(\boldsymbol{r}, t + t_A) \; , \end{aligned} \tag{21}$$

where $t_A = -\left.\frac{\mathrm{d}\delta}{\mathrm{d}E}\right|_{E_0}$ and W_0 is the form of the wave packet in the absence of scattering. The packet is then advanced from its unscattered position if

$\frac{\mathrm{d}\delta}{\mathrm{d}E}\big|_{E_0}$ is negative, which is typical when the scattering potential has bound states. This implies that a wave packet moving along a line at constant speed v will suffer an acceleration due to scattering. This acceleration makes the extended state carry an extra current to compensate the losses due to the localized states associated with the scattering center.

To see the exact current compensation we will calculate the contribution to the current due to the extended states when Fermi-energy lies in the region of localized states [24]. In the presence of a magnetic field along $z-$direction and electric filed $(0, E_y, 0)$, the current operator

$$j_x = \frac{e}{m} P_x = \frac{e}{m}(-\mathrm{i}\hbar\frac{\partial}{\partial x} + eA_x) \tag{22}$$

can be written as $\frac{\delta H}{\delta A_x}$. The Hellman-Feynman theorem states that if the Hamiltonian H depends on arbitrary parameter λ, then

$$\frac{\partial E_n}{\partial \lambda} = \left\langle \phi_n(\lambda) \left| \frac{\partial H}{\partial \lambda} \right| \phi_n(\lambda) \right\rangle \tag{23}$$

where $H(\lambda)|\phi_n(\lambda) >= E_n(\lambda)\phi_n(\lambda)$. Thus

$$j_{x,\alpha} = \frac{\partial E_\alpha}{\partial A_0} \, . \tag{24}$$

Here, A_0 is the additional (constant) vector potential so that $A'_x = -By + A_0$. The introduction of A_0 can be considered by a gauge transformation as well, as a change in the boundary conditions:

$$\psi(x = 0, y) = \psi(x = L_x, y) \rightarrow \psi(x = 0, y) = \mathrm{e}^{-\mathrm{i}\theta L_x}\psi(x = L_x, y) \, . \tag{25}$$

Owing to the magnetic field in the system the degeneracy due to time-reversal symmetry is lacking. This means that the elastic scattering cannot change the direction of propagation of wave packet, i.e., there is only forward scattering. In the forward scattering along the x-axis the outgoing plane wave differs from the incoming plane wave by a certain phase shift δ. The periodic boundary condition in this direction requires the phase of the scattered wave to satisfy the condition

$$k_{x,\mu} L_x + \delta(k_{x,\mu}) - \theta L_x = 2\pi\mu \, , \tag{26}$$

where μ is an integer. The variation of A_0 can be performed by a differential change of θ from $\theta \rightarrow (\theta + \frac{2\pi}{L_x})$. Now, the shifting θ by $\frac{2\pi}{L_x}$ is equivalent to changing μ by unity. Therefore the change of $k_{x,\mu}$ because of change of θ by $\frac{2\pi}{L_x}$ is given by

$$k_{x,\mu+1} - k_{x,\mu} = \frac{2\pi}{L_x}\left[1 - \frac{1}{2\pi}\left(\delta(k_{x,\mu+1}) - \delta(k_{x,\mu})\right)\right] \tag{27}$$

with

$$E_n(k_{x,\mu+1}) - E_n(k_{x,\mu}) = \hbar v_{D_x}(k_{x,\mu+1} - k_{x,\mu}) \\ = \frac{2\pi}{L_x}\hbar v_{D_x}\left[1 - \frac{1}{2\pi}\left(\delta(k_{x,\mu+1}) - \delta(k_{x,\mu})\right)\right] . \quad (28)$$

Therefore, the x-component of the current (Hall current) carried by the state $|n, k_x>$ is

$$-\frac{\hbar j_{x,k_x}}{e} = \frac{L_x}{2\pi}\left(E_n(k_{x,\mu+1}) - E_n(k_{x,\mu})\right) \\ = \hbar v_{D_x} - \frac{\hbar v_{D_x}}{2\pi}\left(\delta(k_{x,\mu+1}) - \delta(k_{x,\mu})\right) . \quad (29)$$

The total current in the system is the sum over all current carrying states within the band,

$$j_x = -ev_{D_x}(N - N_B) + e\frac{v_{D_x}}{2\pi}\sum_{n,k_x}^{\text{occu}}\left(\delta(k_{x,\mu+1}) - \delta(k_{x,\mu})\right) \quad (30)$$

where N is the total number density of states and N_B is the total number density of the bound states of the filled or partially filled bands. Now, according to the Levinson theorem [25], the total phase shift of the scattering states within a band is exactly given by the number of localized states in the band multiplied by (2π). Thus, the total current is found to be

$$j_x = -ev_{D_x}(N - N_B) - ev_{D_x}N_B = -eNv_{D_x} = \frac{e^2 n_0}{h}E_y , \quad (31)$$

where n_0 is the number of occupied Landau sub-bands, i.e., the same as that carried by a pure system with filled bands.

3 Edge States in the Integer Quantum Hall Effect

A realistic quantum Hall system has finite width which gives rise to edge effects. The surprising accuracy at quantized values of the Hall resistance necessitates an investigation on the role of the boundaries (edges) on the Hall resistance. To study the edge effects on the integer quantum Hall effect we consider a two-dimensional electron gas confined to a finite strip in y-direction, in the presence of a magnetic field along the z-direction (for reference, see [10, 26–28]). For simplicity, we assume translation invariance in the x-direction. The confinement of the electron can be represented by a confining potential $U(y)$ (shown in Fig. 5(a)). In the Landau gauge the eigenfunctions $\psi(x, y) = \exp(\mathrm{i}k_x x)\phi(y)$ satisfy the Schrödinger equation

$$\left[-\frac{\hbar^2}{2m}\frac{\partial^2}{\partial y^2} + \frac{1}{2}m\omega_c^2(y - y_{k_x})^2 + U(y)\right]\phi(y) = E\phi(y) , \quad (32)$$

where $y_{k_x} = k_x l_B^2$. It should be noted that in the absence of a confining potential $U(y)$, the eigenstates of the system with equal eigenvalues are centered around different locations y_{k_x} in the transverse direction and with a spatial extent of $\approx l_B$. Assuming that the potential $U(y)$ is nearly constant over the extent of each state, i.e., $\frac{\partial U(y)}{\partial y} \ll \hbar\omega_c/l_B$, we can replace $U(y)$ by $U(y_{k_x})$ and express the energy eigenvalues as

$$E(n, k_x) = \hbar\omega_c(n + \frac{1}{2}) + U(y_{k_x}) \ . \tag{33}$$

Figure 5(b) shows a sketch of the dispersion relation $E(n, k_x)$ versus y_{k_x}.

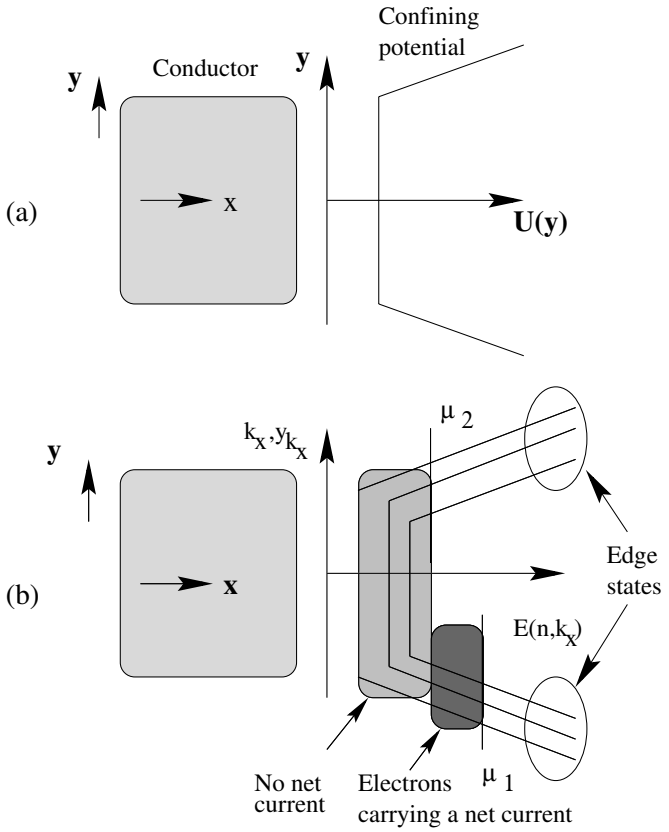

Fig. 5. The confining potential $U(y)$ of the finite boundary quantum Hall system is shown in figure **(a)**. Figure **(b)** shows the energy levels of the two-dimensional electronic system, confined by the potential $U(y)$, as a function of y_{k_x}. The two edges of the system is in equilibrium with the chemical potential μ_1 and μ_2 with $\mu_1 > \mu_2$

In the middle of the sample the states just look like the Landau levels of an unconfined 2D conductor. Near the two edges there are allowed states with a continuous distribution of energies. These are referred to as the edge states. The velocities of these edge states are given by

$$v_x(n,k_x) = \frac{1}{\hbar}\frac{\partial E(n,k_x)}{\partial k_x} = \frac{1}{\hbar}\frac{\partial U(y_{k_x})}{\partial k_x} = \frac{1}{\hbar}\frac{\partial U(y)}{\partial y}\frac{y_{k_x}}{\partial k_x} = \frac{1}{eB}\frac{\partial U(y)}{\partial y}\,. \tag{34}$$

Since $\frac{\partial U(y)}{\partial y}$ has different sign at the two edges of the sample, the edge states located at the two edges carry currents in opposite directions. We now consider a situation like in Fig. 6, where the chemical potential μ of the two electron reservoirs, namely the left ($\mu = \mu_{\rm L}$) and the right ($\mu = \mu_{\rm R}$), differ by a value $\delta\mu = \mu_{\rm L} - \mu_{\rm R}$.

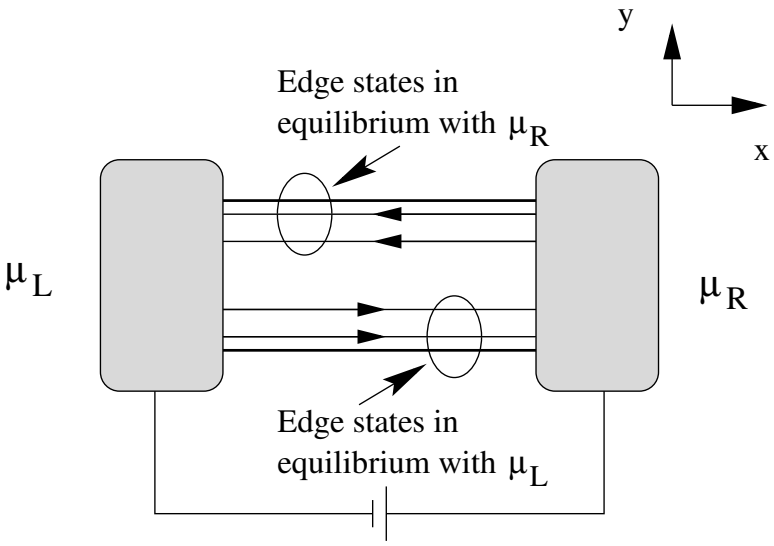

Fig. 6. A two-dimensional electronic system is connected to the electrodes with chemical potential $\mu_{\rm R}$ and $\mu_{\rm L}$. The current injected at the upper corner is in equilibrium with chemical potential $\mu_{\rm R}$ and the current injected at the lower corner is in equilibrium with chemical potential $\mu_{\rm L}$. The potential difference between these to electrodes is the Hall voltage

This chemical potential difference is the driving force for the current flow and is the voltage difference between the reservoirs times e. The situation is quite similar to that in an ordinary conductor carrying current. The positive k_x-states are occupied to a higher quasi-Fermi level than the negative k_x-states. The only difference in the QH edge states and the states of ordinary

conductor is that here the states carrying current in one direction are spatially separated from those carrying current in opposite direction. The resistance at low temperatures arises because of the momentum relaxation in the positive k_x states. However, to relax momentum in QH edge states an electron has to be scattered from the left of the sample to the right of the sample. This is practically impossible since the overlap between the wave functions with positive and negative momentum is exponentially small and there are no allowed states in the interior of the sample in this energy range ($\mu_{\rm L} > E > \mu_{\rm R}$). As a result of this complete suppression of backscattering, electrons originating in the left contact that enter the edge states carry current to the right while electrons in right contact that enter the edge states carry current to the left. Consequently, the edge states carrying current to the right are in complete equilibrium with the left contact and have Fermi energy equal to $\mu_{\rm L}$ and the edge states carrying current to the left originating from the right contact have Fermi energy equal to $\mu_{\rm R}$. Clearly the longitudinal voltage drop V_L as measured by two voltage probes located any where on the same side of the sample is zero, while the transverse (Hall) voltage V_H measured by two probes located any where on opposite edges of the sample in the y direction is equal to the applied voltage

$$eV_H = \mu_{\rm L} - \mu_{\rm R} \ . \tag{35}$$

The above mention phenomena arises only when the electrochemical potential lie between two bulk Landau levels. If the electrochemical potential lie on a bulk Landau level, there is a continuous distribution of allowed states from one edge to the other and electrons can scatter from the left of the sample to the right of the sample through the allowed energy states in the interior of the sample.

We now turn our attention to the current carried by the edge electrons in the system. The contribution to the current from a single-channel of the edge state is

$$\begin{aligned} I_i &= e\int_{\mu_{\rm R}}^{\mu_{\rm L}} \frac{{\rm d}n}{{\rm d}E} v_x(E){\rm d}E \\ &= e\int_{\mu_{\rm R}}^{\mu_{\rm L}} \frac{{\rm d}n}{{\rm d}k_x}\frac{{\rm d}k_x}{{\rm d}E}\frac{1}{\hbar}\frac{{\rm d}E}{{\rm d}k_x}{\rm d}E \\ &= \frac{e}{2\pi\hbar}\int_{\mu_{\rm R}}^{\mu_{\rm L}} {\rm d}E = \frac{e}{h}(\mu_{\rm L} - \mu_{\rm R}) \ , \end{aligned} \tag{36}$$

where $\frac{dn}{dE}$ is the density of states for the edge states, and $v_x(E)$ is the x-component of the velocity of the edge state with energy E. The total current due to edge states is given by

$$I = \sum_i I_i = \frac{Me^2}{h} V_H \ , \tag{37}$$

where M signifies the number of edge states at the Fermi energy (which is equal to the number of the filled Landau sub-bands in the bulk below the Fermi energy). M takes integer values that decreases as the magnetic field is increased. Hence the Hall resistance of the one-dimensional edge channel is

$$r_H = \frac{V_H}{I} = \frac{h}{Me^2} \,. \tag{38}$$

This quantized nature of Hall resistivity explains the striking accuracy (better than one part per million) even in presence of sample boundaries.

4 Fractional Quantum Hall Effect

Tsui et al. [9] created two-dimensional electron gases in GaAs-AlGaAs with electron mobilities of the order of $5 \times 10^5\,\mathrm{cm^2V^{-1}s^{-1}}$, which is 100 times larger than the mobility of the samples where the integer quantum Hall effect was observed. The importance of impurities was correspondingly less, but instead of eliminating plateaus in the Hall voltage, a host of new plateaus appeared. This is a very counter intuitive physical phenomenon and is known as fractional quantum Hall effect (FQHE). It implies that many electrons acting in concert, can create new quasi-particles having charge smaller than the charge of any individual electron and to be more precise they are exactly $\frac{1}{3}$ or $\frac{1}{5}$ or $\frac{1}{7}$ etc. of an electronic charge. The fractional quantum Hall state is something unprecedented - a new state of matter. Its phenomenology, however, is the same as that of the integral quantum Hall state in almost every detail: There is a plateau. The Hall conductance in the plateau is accurately a pure number times e^2/h. The parallel resistance and conductance are both zero in the plateau. The only qualitative difference between the IQHE and FQHE is the value of the quantum Hall conductance. This can only be explained on the basis of electron-electron interactions, in which the Coulomb repulsion favors a state with an energy gap at these rational fractional fillings. Laughlin [14] proposed a generalization of the free electron state in which the density of states within filled Landau level is eB/phc. In a metal, the Coulomb energy would be much larger than the magnetic energy at any realistic field strength, but in Gallium Arsenide heterostructures magnetic fields of a few Tesla are sufficient for magnetic energy to dominate. The significance of reaching the high-field limit is that the true ground state of the system of interacting electrons can accurately be written using only states in the lowest Landau level as basis. States in the higher Landau levels cost energy $\hbar\omega_c$, and Coulomb repulsion is too small an energy to force any electron to make the transition.

To construct the ground state of the interacting electrons, it is useful to represent the Hamiltonian for electrons in magnetic field in symmetric gauge (the vector potential $\boldsymbol{A}(\boldsymbol{r}) = (By/2, -Bx/2, 0)$) and in terms of complex variable $z = (x + \mathrm{i}y)/2l_B$ and $z^* = (x - \mathrm{i}y)/2l_B$ [29]. The harmonic oscillator operators $a, a^\dagger$, $b, b^\dagger$ and the angular momentum operator (L) now reads

$$L = xp_y - yp_x = \frac{eB}{2}(X^2 + Y^2) - \frac{1}{2eB}(P_x^2 + P_y^2) = (b^\dagger b - a^\dagger a)\hbar \ , \quad (39)$$

where

$$a = -\frac{\mathrm{i}}{\sqrt{2}}(z + \frac{\partial}{\partial z^*}) \ , \qquad a^\dagger = \frac{\mathrm{i}}{\sqrt{2}}(z^* - \frac{\partial}{\partial z}) \ ,$$
$$b = \frac{1}{\sqrt{2}}(z^* + \frac{\partial}{\partial z}) \ , \qquad b^\dagger = \frac{1}{\sqrt{2}}(z - \frac{\partial}{\partial z^*}) \ .$$

The degeneracy in each Landau level may be attributed to the arbitrariness in the angular momentum, where $b^\dagger$ increases (b decreases) the angular momentum by one unit. As $X^2 + Y^2 = 2(b^\dagger b + 1/2)l_B^2$ the position of guiding center determines the angular momentum of the particle. A state in the nth Landau level is given by

$$|n, l >= \sqrt{\frac{1}{n!(l+n)!}}(a^\dagger)^n (b^\dagger)^{l+n}|0> \ , \quad (40)$$

which has the energy $E = (n + \frac{1}{2})\hbar\omega_c$ and angular momentum $L = l\hbar$, where $|0>$ is the Fock vacuum, $a|0 >= b|0 >= 0$. The state $|n, l>$ is also an eigenstate of the operator $X^2 + Y^2$.

The ground state satisfies $H|\phi >= \frac{1}{2}\hbar\omega_c|\phi>$ with $a|\phi >= 0$, i.e.,

$$a\phi(x, y) = -\frac{\mathrm{i}}{\sqrt{2}}(z + \frac{\partial}{\partial z^*})\phi(x, y) = 0 \quad (41)$$

where $\phi(x, y) =< x, y|\phi >$. The general solution of the above equation is

$$\phi(x, y) = f(z)\exp(-zz^*) \ , \quad (42)$$

where, $f(z)$ is an arbitrary analytic function. The zero angular-momentum state in the lowest Landau level characterizes

$$b\phi_0(x, y) = \frac{1}{\sqrt{2}}(z^* + \frac{\partial}{\partial z})\phi_0(x, y) = 0 \quad (43)$$

which is solved as

$$\phi_0(x, y) = \frac{1}{\sqrt{2\pi l_B^2}}\mathrm{e}^{-|z|^2} = \frac{1}{\sqrt{2\pi l_B^2}}\exp(-\frac{r^2}{4l_B^2}) \ . \quad (44)$$

The angular momentum state $|l>$ in the lowest Landau level is described by the wave function,

$$\phi_l(x, y) = \sqrt{\frac{1}{l!}}(b^\dagger)^l\phi_0(x, y) = C_l z^l \exp(-|z|^2) \quad (45)$$

with C_l the normalization factor, $C_l = \sqrt{2^l/2\pi l_B^2 l!}$.

The probability of finding the electron at $r = 2l_B|z|$ is given by

$$|\phi_l(x,y)|^2 = |C_l|^2|z|^{2l}\exp(-2|z|^2) \propto r^{2l}\exp(-\frac{r^2}{2l_B^2}) \,, \tag{46}$$

which has a sharp peak at $r = \sqrt{2l}l_B$. The states $\phi_l(x,y)$ then represents a series of concentric rings of charge. The difference of area ΔS of two consecutive rings is equal for all l and is given by $\pi r_{l+1}^2 - \pi r_l^2 = 2\pi l_B^2$. The position of an electron cannot be localized within an area smaller than $|\Delta S|$ because the guiding centers of the electron do not commute and in the lowest Landau level $[X,Y] = -\mathrm{i}l_B^2$. When we put several electrons into the system in the lowest Landau level, they occupy the angular momentum states. Due to the Pauli exclusion principle one state accommodates one electron. Since it turns out that the origin of the FQHE is interaction between electrons, the correlation between the relative motion of the electrons should be taken into account. Therefore the generic N-body state can be expressed as

$$\Psi = f(z_1, z_2, \ldots z_N)\exp\left(-\sum_{r=1}^{N}|z_r|^2\right) \,, \tag{47}$$

where $f(z_1, z_2, \ldots z_N)$ is an analytic function of all its arguments. Laughlin proposed that a trial many-body wave function, called Jastrow function can describe the underline feature of the Coulomb correlation between the electrons. The ground state wave function of the interacting electronic system is

$$\Psi = \Pi_{r<s}(z_r - z_s)^m \exp\left(-\sum_{r=1}^{N}|z_r|^2\right) \,. \tag{48}$$

The restriction that the wave function is analytic implies that m must be an integer, and for the wave function to be antisymmetric m must be restricted to the odd integers. The most important feature of this wave function is that it locks electron density at exactly $1/(2\pi m l_B^2)$ in the limit the number of particles N becomes thermodynamically large. This fact can be observed from the comparison of the equivalent probability distribution function obtained from Laughlin conjecture and classical one-component plasma. Expressing

$$|\Psi|^2 = \mathrm{e}^{-\beta\Phi(z_1,\ldots z_N)} \tag{49}$$

and choosing $\beta = 1/m$ we get

$$\Phi = -2m^2\sum_{j<k}^{N}\ln|z_j - z_k| + 2m\sum_{j}^{N}|z_j|^2 \,. \tag{50}$$

This is the potential energy of particles of "charge" m repelling each other logarithmically - the natural Coulomb potential in two dimensions - and being

attracted to the uniform "charge" density $1/(2\pi l_B^2)$. In order to keep the local electrical neutrality the particles must have density $\rho = \rho_m = 1/(2\pi m l_B^2)$. This is equivalent to filling fraction $\nu = 1/m$. Numerical studies have shown that every Laughlin state lies at a deep energy minimum; states with filling other than ρ_m have much higher energies. This energy difference is caused by the electron-electron repulsion. When the Coulomb interaction is slowly turned off, the depth of the minim gradually decreases to zero and we re-obtain the non-interacting picture. In the case like FQHE, there exists a gap for density fluctuation and the system becomes incompressible quantum fluid. The Laughlin wave function is a particular state in which all electrons, having their own cyclotron motions, stay away from each other as much as possible. This wave function naturally builds in good correlations among the electrons because each particle sees an $m-$fold zero at the position of all other particles. The wave function vanishes extremely rapidly if any two particles approach each other and minimizes the expectation value of the Coulomb energy.

A typical term in the Laughlin wave function is

$$z_1^{(N-1)m} z_2^{(N-2)m} \cdots z_N^0 \ \exp(-\sum_{r=1}^{N} |z_r|^2) \ , \tag{51}$$

where N is the number of electrons. Because the state (45) has the angular momentum $l\hbar$, the total angular momentum of the state (51) is

$$L_{\text{total}} = m\hbar \sum_{j=0}^{N-1} j = \frac{(N-1)Nm}{2}\hbar \ . \tag{52}$$

As all the terms in the correlation function $f(z_1 ... z_N)$ are the polynomial of degree $(N-1)Nm/2$, the Laughlin wave function has total angular momentum given by (52). The maximum angular momentum $l_{\max}\hbar$ that a single electron can have is the maximum power of one of the variables in the wave function

$$l_{\max} = m(N-1) \ . \tag{53}$$

Therefore the area of the system S and electronic filling of the Laughlin state are:

$$S = \pi(X_{\max}^2 + Y_{\max}^2) = \pi(2l_{\max} + 1)l_B^2 = 2\pi m(N-1)l_B^2 \ , \tag{54}$$

$$\nu = \frac{N}{l_{\max}} = \frac{N}{m(N-1)} = \frac{1}{m} \ , \tag{55}$$

for $N \gg 1$. Thus there are $\frac{eBS}{2\pi\hbar N} = m$ magnetic flux quanta per electron.

Laughlin furthermore proposed that the excitation of the fractional quantum Hall state carry fractional charge. This can be justified intuitively. Let us imagine that the magnetic flux $\phi(t)$ as shown in Fig. 7 is changed adiabatically through a thin solenoid passing through the point (x_0, y_0) in the sample, so

that the wave function is an eigenstate of the changing Hamiltonian at any moment in time. When the magnetic flux through (x_0, y_0) is exactly one flux quantum $\phi_D = \pm h/e$, the electronic wave function will go through a phase change 2π and no physical measurement can detect the presence of the tube.

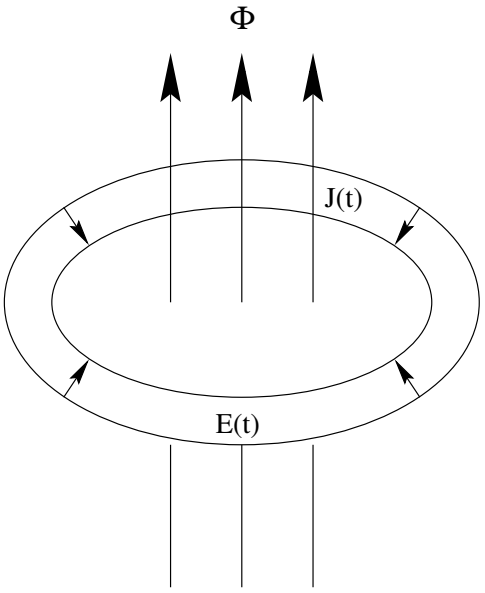

Fig. 7. Construction of a Laughlin quasi-particle by adiabatically threading flux Φ through a point in the sample. Faraday induction gives an azimuthal electric field $E(t)$ which in turn produces a radial current $J(t)$

Therefore the electron gas after the tube reaches flux ϕ_D must be in an eigenstate of the original Hamiltonian. This state cannot be the ground state because turning on the flux tube at (x_0, y_0) drives charge away from the center (x_0, y_0), and we are left with an excited state of the original Hamiltonian. Now the time dependent magnetic flux gives rise to an electric field

$$E(t) = -\frac{1}{2\pi r}\frac{\mathrm{d}\phi}{\mathrm{d}t} \tag{56}$$

which is detected azimuthally around the solenoid. In the quantum Hall state the electric filed generates a Hall current perpendicular to it,

$$J(t) = \frac{e\rho_0}{B}E(t)\ , \tag{57}$$

which holds far away from the flux tube. When the solenoid is removed the total change in the charge is

$$Q = 2\pi r \int \mathrm{d}t J(t) = -\frac{e\rho_0}{B}\int \mathrm{d}t \frac{\mathrm{d}\phi}{\mathrm{d}t} = -\frac{e\rho_0}{B}\phi_D\ \mathrm{sgn}(\phi) = -e\nu\ \mathrm{sgn}(\phi)\ . \tag{58}$$

This kind of excitation is called a quasi-particle or quasi-hole. We call it "quasi-particle (quasi-hole)" because the excitation is not a true particle, but a collective behavior of many particles. This quasi-particle (quasi-hole) has fractional charges $Q = -e\nu = -e/m$ at the fractional filling $\nu = 1/m$. The excitation in the form of generating a quasi-particle has the effect of increasing the angular momentum of each particle by 1 and in the case of quasi-holes decreasing the angular momentum of each particle by 1.

When the number density of the electron is changed by $\Delta\rho$ with the magnetic field B fixed, the filling factor is changed as

$$\nu = \frac{2\pi(\rho_0 + \Delta\rho)\hbar}{eB} = \frac{1}{m}(1 + \frac{\Delta\rho}{\rho_0}) \, . \tag{59}$$

Because the electronic charge of the quasi-particle is $\pm e/m$, m quasi-particles are created per electron on the Laughlin state at $\nu = 1/m$. The number density $\rho_{\rm qp}$ of quasi-particles created is

$$\rho_{\rm qp} = m|\Delta\rho| \, . \tag{60}$$

Quasi-particles are quasi-electrons (quasi-holes) when $\Delta\rho > 0$ ($\Delta\rho < 0$). Alternatively, a change in the perpendicular magnetic field B by ΔB with fixed number density gives rise to the change of the filling factor

$$\nu = \frac{2\pi\rho_0\hbar}{e(B + \Delta B)} = \frac{1}{m}\left(1 + \frac{\Delta B}{B}\right)^{-1} \, . \tag{61}$$

Since one quasi-particle carries one flux quanta $\pm\phi_D$, the number density of quasi-particles created is

$$\rho_{\rm qp} = \frac{|\Delta B|}{\phi_D} = \frac{e|\Delta B|}{2\pi\hbar} \, . \tag{62}$$

Quasi-particles are quasi-electrons (quasi-holes) when $\Delta B > 0$ ($\Delta B < 0$). If the density of the quasi-particle is low, there is effectively no Coulomb repulsion between the quasi-particles as they are far apart. Impurities will have the same effects on this non-interacting particles in the fractional quantum Hall effect as on electrons in the integer quantum Hall effect. The localization mechanisms will act in the same way and consequently, the current will be independent of the filling fraction in an interval around $\nu = 1/m$. This leads to a plateau formation in the Hall resistivity R_{xy} and vanishing of the diagonal resistivity R_{xx} around $\nu = 1/m$. For higher quasi-particle densities the sum of the quasi-particle states correspond to extended states which contribute to the current. This situation corresponds to the transition region between the plateaus.

5 Edge States in Fractional Quantum Hall Effect

We learned in our study of the IQHE that gapless edge excitations exist even when the bulk has a large excitation gap. Since the fractional QH state

is an incompressible liquid confined within a finite domain, the low energy excitations are the deformation of the boundary shape which preserve the area of the liquid as illustrated in Fig. 8. At the edge, the Landau levels are bent by the confining potential $U(y)$, cross the Fermi level, and form edge channels. Edge excitations are gapless modes since infinitesimal deformation of the boundary costs an infinitesimal energy.

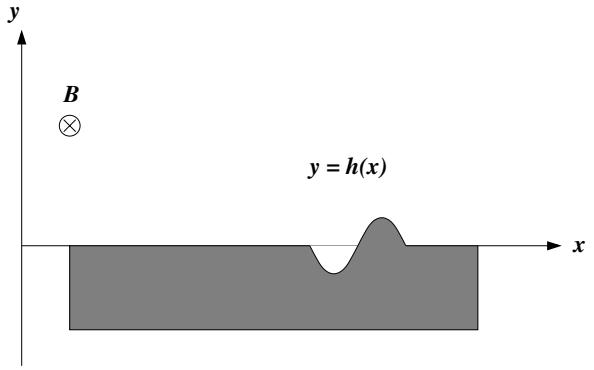

Fig. 8. The edge excitation of the QH droplet is a deformation of the boundary. Its shape is parametrized by a function $y = h(x)$. The electron density is constant within the droplet

The simplest way to understand the dynamics of the edge excitation is to use the hydrodynamical approach [30–33]. In the hydrodynamical approach we first study the classical theory of surface waves on the QH droplet. Then we quantize the classical theory to obtain the quantum description of the edge excitations. Let us consider a QH droplet with filling fraction ν confined by a smooth potential well. The electric field generates a Hall current $\boldsymbol{j} = \sigma_{xy}\hat{\boldsymbol{z}} \times \boldsymbol{E} = \rho_0 e(\hat{\boldsymbol{z}} \times \boldsymbol{E})/B$ which produces a drift velocity of the edge current $v = E/B$. Thus the edge wave (i.e., the deformation at the surface) propagates with velocity v. Let us consider one-dimensional density $\rho(x) = \rho_0 h(x)$ to describe the edge wave where $h(x)$ is the displacement of the edge at the position x. This deformation at the edge of the QH droplet can be expressed by the Hamiltonian

$$H = \frac{1}{2}\int \mathrm{d}x \rho(x) h(x) eE = \frac{\pi\hbar v}{\nu}\int \mathrm{d}x \rho^2(x)\ , \tag{63}$$

where ρ satisfies the hydrodynamic wave equation

$$\frac{\partial \rho(x)}{\partial t} - v\frac{\partial \rho(x)}{\partial x} = 0\ . \tag{64}$$

It should be noted that the edge excitations are chiral excitation, i.e., propagate in one direction only. In the momentum space the Hamiltonian (63) and the wave equation (64) reduces to

$$H = \frac{2\pi\hbar v}{\nu} \int_0^\infty \mathrm{d}k_x \rho(k_x)\rho(-k_x) \ , \tag{65}$$

$$\frac{\partial \rho(k_x)}{\partial t} = \mathrm{i}vk_x\rho(k_x) \ , \tag{66}$$

where $\rho(k_x) = \frac{1}{\sqrt{L}} \int \mathrm{d}x e^{\mathrm{i}k_x x}\rho(x)$ and L is the length of the edge. The system is now quantized according to the standard prescription. Let us identify $\rho(k_x)$ $(k_x > 0)$ as the coordinate. From equation (66) the conjugate momentum can be evaluated as

$$\pi(k_x) = \frac{\partial H}{\partial \dot{\rho}(k_x)} = -\mathrm{i}\frac{2\pi\hbar}{\nu k_x}\rho(-k_x) \ for \ k_x > 0 \ . \tag{67}$$

The theory can now be quantized by imposing the canonical quantization relation $[\rho(k_x), \pi(l_x)] = \mathrm{i}\hbar\delta(k_x - l_x)$ or

$$[\rho(k_x), \rho(l_x)] = -\frac{k_x\nu}{2\pi}\delta(k_x + l_x) \ . \tag{68}$$

In the coordinate space the above commutation relation reads

$$[\rho(x), \rho(x')] = \frac{\mathrm{i}\nu}{2\pi}\frac{\partial\delta(x - x')}{\partial x} \ . \tag{69}$$

The above commutation relation is known as $U(1)$ Kac-Moody algebra [34]. The Hamiltonian (63) together with the commutation relation defines a chiral version of the Luttinger model.

Let us now define the annihilation and creation operator for edge excitation of left going electron by

$$a(k_x) = \sqrt{\frac{2\pi}{\nu k_x}}\rho(k_x), \ a^\dagger(k_x) = \sqrt{\frac{2\pi}{\nu k_x}}\rho(-k_x) \tag{70}$$

respectively. $a^\dagger(k_x)$ and $a(k_x)$ are indeed creation and annihilation operators because they obey the commutation relation

$$[a(k_x), a^\dagger(l_x)] = \delta(k_x - l_x) \ . \tag{71}$$

In this notation ground state is defined by $a(k_x)|0>= 0$. The Hamiltonian H now reduces to

$$H = \int_0^\infty \mathrm{d}k_x \epsilon(k_x) a^\dagger(k_x)a(k_x) \ , \tag{72}$$

where the energy dispersion $\epsilon(k_x) = \hbar v k_x$. Here normal ordering is made so that the filled Fermi sea has zero energy. The Hilbert space is constructed as a Fock space by operating $a^\dagger(k_x)$ on the vacuum $|0>$.

We find that the gap-less excitations in QH states are described by free chiral bosons represented by the creation and annihilation operators $a^\dagger(k_x)$ and $a(k_x)$. This chiral boson in general can be expressed by the Lagrangian

$$\mathcal{L} = \frac{\pi\hbar}{\nu}(2\partial_t\phi(x,t)\partial_x\phi(x,t) - v\partial_x\phi(x,t)\partial_x\phi(x,t)) \ , \tag{73}$$

where $\phi(x,t)$ is the corresponding bosonic field. The canonical momentum $\pi(x,t) = \partial\mathcal{L}/\partial\dot{\phi}(x,t) = (2\pi\hbar/\nu)\partial_x\phi(x,t)$ obeys the commutation relation

$$[\phi(x,t), \pi(x',t)] = \mathrm{i}\hbar\delta(x-x') \tag{74}$$

or,

$$[\phi(x,t), \rho(x',t)] = \frac{\mathrm{i}\nu}{2\pi}\delta(x-x') \tag{75}$$

with $\rho(x,t) = \partial_x\phi(x,t)$. The Hamiltonian is

$$H = \int \mathrm{d}x\pi(x,t)\dot{\phi}(x,t) - \mathcal{L} = \frac{\pi\hbar v}{\nu}\int \mathrm{d}x\rho(x,t)^2 \tag{76}$$

and the corresponding Heisenberg equation of motion is

$$\mathrm{i}\hbar\partial_t\phi(x,t) = [\phi(x,t), H] = \mathrm{i}\hbar v\partial_x\phi(x,t) \tag{77}$$

or,

$$\partial_t\phi(x,t) = v\partial_x\phi(x,t) \ . \tag{78}$$

The above wave equation (78) is the hydrodynamical equation (64). From the commutation relation (75) we obtain the Kac-Moody algebra

$$[\rho(x,t), \rho(x',t)] = \frac{\mathrm{i}\nu}{4\pi}\partial_x\delta(x-x') \ , \tag{79}$$

Integrating the Kac-Moody algebra (79) we get

$$[\phi(x,t), \phi(x',t)] = -\mathrm{i}\frac{\nu}{4\pi}\ \mathrm{sgn}(x-x') \ . \tag{80}$$

Note that the standard quantization theory is not applicable to the Lagrangian because $\phi(x,t)$ and $\phi(x',t)$ do not commute. It implies that $\phi(x,t)$ is a non-local field operator. It is precisely the left chiral component of the massless Klein-Gordon theory, where the left and right components travel independently. The chiral boson can also be represented by the chiral Fermion system defined by the Lagrangian

$$\mathcal{L} = \mathrm{i}\hbar\psi^\dagger(x,t)(\partial_t - v\partial_x)\psi(x,t) \tag{81}$$

and the anti-commutation relation

$$[\psi(x,t), \psi^\dagger(x',t)]_+ = \delta(x-x'), \ [\psi(x,t), \psi(x',t)]_+ = 0 \ . \tag{82}$$

The corresponding Hamiltonian and the equation of motion are then given by

$$H = \mathrm{i}\hbar v \int \mathrm{d}x \psi^{\dagger}(x,t)\partial_x \psi(x,t) \ , \tag{83}$$

$$\mathrm{i}\hbar\partial_t \psi(x,t) = [\psi(x,t), H] = \mathrm{i}\hbar v \partial_x \psi(x,t) \ . \tag{84}$$

The equivalence between the fermion theory (83, 84) and the boson theory (73, 77) is provided by the bosonization formula [35–37]

$$\psi(x,t) = \lim_{\epsilon \to 0} \sqrt{\frac{\nu}{2\pi\epsilon}} \mathrm{e}^{\mathrm{i}\frac{2\pi\phi(x,t)}{\nu}} \tag{85}$$

and $\rho(x,t) = \psi^{\dagger}(x,t)\psi(x,t)$. Using the Kac-Moody algebra and Hausdorff formula

$$\mathrm{e}^{A}\mathrm{e}^{B} = \mathrm{e}^{[A,B]}\mathrm{e}^{B}\mathrm{e}^{A} \ , \tag{86}$$

with $A = \mathrm{i}\frac{2\pi}{\nu}\phi(x,t)$ and $B = \pm\mathrm{i}\frac{2\pi}{\nu}\phi(x',t)$, and

$$B\mathrm{e}^{A} = \mathrm{e}^{A}B - [A,B]\mathrm{e}^{A} \tag{87}$$

with $A = \pm\mathrm{i}\frac{2\pi}{\nu}\phi(x',t)$ and $B = \rho(x,t)$, we obtain the commutation relation

$$[\rho(x,t), \psi(x',t)] = -\delta(x-x')\psi(x,t) \ . \tag{88}$$

The physical meaning of the commutation relation (88) is that $\psi(x',t)$ annihilates one electron at the position x' by decreasing density $\rho(x,t)$ at $x = x'$. Thus the fermionic operator $\psi(x)$ is identified with the one-dimensional electronic field on the edge. For the fractional quantum Hall state $\nu = 1/m$, where m is an odd-integer, the electronic field $\psi(x,t)$ may be regarded as a composite of m basic fields.

$$\psi_{\mathrm{qe}}(x,t) = \lim_{\epsilon \to 0} \sqrt{\frac{1}{2\pi\epsilon m}} \mathrm{e}^{\mathrm{i}2\pi\phi(x,t)} \ . \tag{89}$$

Using the Kac-Moody algebra and Hausdorff formula (86, 87) we obtain

$$[\rho(x,t), \psi_{\mathrm{qe}}(x',t)] = -\frac{1}{m}\delta(x-x')\psi_{\mathrm{qe}}(x,t) \tag{90}$$

The physical meaning of (90) is that $\psi_{\mathrm{qe}}(x',t)$ annihilates a $1/m$ electron at the position x' by decreasing the density $\rho(x,t)$ at $x = x'$ with electric charge $-e/m$. We can identify it with the quasi-electron field on the edge. We also calculate the bosonic correlation function

$$< 0|\phi(x,t)\phi(x,0)|0 >= -\nu \ln(x - vt) + \text{const.} \ , \tag{91}$$

which yields the fermionic correlation function

$$< 0|T(\psi^{\dagger}(x,t)\psi(x,0))|0 > = \exp\left[\frac{1}{\nu^2} < 0|\phi(x,t)\phi(x,0)|0 >\right]$$
$$\propto \frac{1}{(x-vt)^m} \,. \tag{92}$$

The electron propagator are standard ones in the integer QH state at $\nu = 1/m = 1$, where edge state is described by non-interacting electrons and forms a Fermi-liquid. However, the electron propagator on the edge of the fractional QH state are strongly correlated and cannot be described by Fermi-liquid theory. This type of an electronic state is known as chiral Luttinger liquid. Note, however, that unlike the conventional Luttinger liquid, where the non-trivial power-law in the electron propagator depends on the strength of the electron-electron interactions - here it is dictated by the bulk filling factor characterizing the QH state.

References

1. K. von Klitzing, G. Dorda, M. Pepper: Phys. Rev. Lett. **45**, 494 (1980)
2. *The Quantum Hall Effect*, ed by R. E. Prange, S. M. Girvin (Springer-Verlag, New York 1990)
3. M. Janβen, O. Viehweger, U. Fastenrath, J. Hajdu: *Introduction to the Theory of Integer Quantum Hall Effect* (VCH Verlagsgesellschaft mbH, Weinheim, Federal Republic of Germany 1994)
4. D. C. Tsui: Rev. Mod. Phys. **71**, 891 (1999)
5. P. W. Anderson: Phys. Rev. **109**, 1492 (1958)
6. G. Bergmann: Phys. Rep. **107**, 1 (1984)
7. T. Ando, A. Fowler, F. Stern: Rev. Mod. Phys. **54**, 437 (1982)
8. A. M. M. Pruiskin: Phys. Rev. Lett. **61**, 1297 (1988)
9. D. C. Tsui, H. L. Störmer, A. C. Gossard: Phys. Rev. Lett. **48**, 1559 (1982)
10. S. M. Girvin: Cond-mat/9907002 (1999)
11. R. E. Prange: Phys. Rev. B **23**, 4802 (1981)
12. R. B. Laughlin: Phys. Rev. B **23**, 5632 (1981)
13. B. I. Halperin: Phys. Rev. B **25**, 2185 (1982)
14. R. B. Laughlin: Phys. Rev. Lett. **50**, 1395 (1983); Phys. Rev. Lett. **52**, 2304 (1984)
15. F. D. M. Halden, E. H. Rezayi: Phys. Rev. Lett. **54**, 237 (1985); Phys. Rev. B **31**, 2529 (1985)
16. B. I. Halperin: Phys. Rev. Lett. **52**, 1583 (1984)
17. J. K. Jain: Phys. Rev. Lett. **63**, 199 (1989)
18. S. C. Zhang, T. H. Hansson, S. Kivelson: Phys. Rev. Lett. **62**, 980 (1989)
19. T. Chakraborty, P. Pietiläinen: *The Fractional Quantum Hall Effect* (Springer-Verlag Berlin, New York 1988)
20. R. B. Gerhardts: Z. Phys. B **21**, 275, 285 (1975)
21. M. Weiss: Quantum Hall effect in strongly disordered system. Dissertation (2004)
22. T. Ando: J. Phys. Soc. **52**, 1740 (1983); **53**, 3101, 3126 (1984)
23. G. C. Aers, A. H. MacDonald: J. Phys. C **17**, 5491 (1984)

24. R. Joynt, R. E. Prange: Phys. Rev. B **29**, 3303 (1984)
25. W. Thirring: *Lehr der Mathmatischen Physics 3* (Springer Verlag, Berlin 1979)
26. S. Datta: *Electronic Transport in Mesoscopic Systems* (Cambridge University Press 1995)
27. C. L. Kane, Matthew P. A. Fisher: Edge-state transport. In: *Perspectives in Quantum Hall Effects*, ed by S. Das Sarma, A. Pinczuk (John Wiely & Sons, Inc. 1997) pp 109–159
28. Efrat Shimshoni: cond-mat/0406703 (2004)
29. Z. F. Ezawa: *Quantum Hall Effects: Field Theoretical Approach and Related Topics* (World Scientific 2000)
30. X. G. Wen: Phys. Rev. B **41**, 12838 (1990)
31. D. H. Lee, X. G. Wen: Phys. Rev. Lett. **60**, 1765 (1991)
32. X. G. Wen: Int. J. Mod. Phys. B **6**, 1711 (1992)
33. M. Milovanovic, Efrat Shimshoni: Phys. Rev. B **58**, 4694 (1998); Phys. Rev. B **59**, 10757 (1999)
34. P. Goddard, D. Olive: Int. J. Mod. Phys. A **1**, 303 (1986)
35. E. Fradkin: *Field Theory of Condensed Matter Systems* (Addison - Wesley Reading MA 1991)
36. A. Ludwig: In: *Low Dimensional Quantum Field Theories for Condensed Matter Physicists*, ed by S. Lundqvist, G. Morandi, Yu Lu (World Scientific 1995)
37. R. Shankar: In: *Current Topics in Condensed Matter Physics*, ed by J. Pati, Q. Shafi, Yu Lu (World Scientific 1993)

Index

Springer Series in

SOLID-STATE SCIENCES

Series Editors:

M. Cardona P. Fulde K. von Klitzing R. Merlin H.-J. Queisser H. Störmer

91 **Electronic Properties and Conjugated Polymers III**
Editors: H. Kuzmany, M. Mehring, and S. Roth

92 **Physics and Engineering Applications of Magnetism**
Editors: Y. Ishikawa and N. Miura

93 **Quasicrystals**
Editor: T. Fujiwara and T. Ogawa

94 **Electronic Conduction in Oxides**
2nd Edition By N. Tsuda, K. Nasu, A. Fujimori, and K. Siratori

95 **Electronic Materials**
A New Era in MaterialsScience
Editors: J.R. Chelikowski and A. Franciosi

96 **Electron Liquids**
2nd Edition By A. Isihara

97 **Localization and Confinement of Electrons in Semiconductors**
Editors: F. Kuchar, H. Heinrich, and G. Bauer

98 **Magnetism and the Electronic Structure of Crystals**
By V.A. Gubanov, A.I. Liechtenstein, and A.V. Postnikov

99 **Electronic Properties of High-T_c Superconductors and Related Compounds**
Editors: H. Kuzmany, M. Mehring, and J. Fink

100 **Electron Correlations in Molecules and Solids**
3rd Edition By P. Fulde

101 **High Magnetic Fields in Semiconductor Physics III**
Quantum Hall Effect, Transport and Optics By G. Landwehr

102 **Conjugated Conducting Polymers**
Editor: H. Kiess

103 **Molecular Dynamics Simulations**
Editor: F. Yonezawa

104 **Products of Random Matrices**
in Statistical Physics By A. Crisanti, G. Paladin, and A. Vulpiani

105 **Self-Trapped Excitons**
2nd Edition
By K.S. Song and R.T. Williams

106 **Physics of High-Temperature Superconductors**
Editors: S. Maekawa and M. Sato

107 **Electronic Properties of Polymers**
Orientation and Dimensionality of Conjugated Systems
Editors: H. Kuzmany, M. Mehring, and S. Roth

108 **Site Symmetry in Crystals**
Theory and Applications
2nd Edition
By R.A. Evarestov and V.P. Smirnov

109 **Transport Phenomena in Mesoscopic Systems**
Editors: H. Fukuyama and T. Ando

110 **Superlattices and Other Heterostructures**
Symmetry and Optical Phenomena
2nd Edition
By E.L. Ivchenko and G.E. Pikus

111 **Low-Dimensional Electronic Systems**
New Concepts
Editors: G. Bauer, F. Kuchar, and H. Heinrich

112 **Phonon Scattering in Condensed Matter VII**
Editors: M. Meissner and R.O. Pohl

113 **Electronic Properties of High-T_c Superconductors**
Editors: H. Kuzmany, M. Mehring, and J. Fink

Springer Series in

SOLID-STATE SCIENCES

Series Editors:

M. Cardona P. Fulde K. von Klitzing R. Merlin H.-J. Queisser H. Störmer

114 **Interatomic Potential and Structural Stability**
Editors: K. Terakura and H. Akai

115 **Ultrafast Spectroscopy of Semiconductors and Semiconductor Nanostructures**
By J. Shah

116 **Electron Spectrum of Gapless Semiconductors**
By J.M. Tsidilkovski

117 **Electronic Properties of Fullerenes**
Editors: H. Kuzmany, J. Fink, M. Mehring, and S. Roth

118 **Correlation Effects in Low-Dimensional Electron Systems**
Editors: A. Okiji and N. Kawakami

119 **Spectroscopy of Mott Insulators and Correlated Metals**
Editors: A. Fujimori and Y. Tokura

120 **Optical Properties of III–V Semiconductors**
The Influence of Multi-Valley Band Structures By H. Kalt

121 **Elementary Processes in Excitations and Reactions on Solid Surfaces**
Editors: A. Okiji, H. Kasai, and K. Makoshi

122 **Theory of Magnetism**
By K. Yosida

123 **Quantum Kinetics in Transport and Optics of Semiconductors**
By H. Haug and A.-P. Jauho

124 **Relaxations of Excited States and Photo-Induced Structural Phase Transitions**
Editor: K. Nasu

125 **Physics and Chemistry of Transition-Metal Oxides**
Editors: H. Fukuyama and N. Nagaosa

126 **Physical Properties of Quasicrystals**
Editor: Z.M. Stadnik

127 **Positron Annihilation in Semiconductors**
Defect Studies
By R. Krause-Rehberg and H.S. Leipner

128 **Magneto-Optics**
Editors: S. Sugano and N. Kojima

129 **Computational Materials Science**
From Ab Initio to Monte Carlo Methods. By K. Ohno, K. Esfarjani, and Y. Kawazoe

130 **Contact, Adhesion and Rupture of Elastic Solids**
By D. Maugis

131 **Field Theories for Low-Dimensional Condensed Matter Systems**
Spin Systems and Strongly Correlated Electrons
By G. Morandi, P. Sodano, A. Tagliacozzo, and V. Tognetti

132 **Vortices in Unconventional Superconductors and Superfluids**
Editors: R.P. Huebener, N. Schopohl, and G.E. Volovik

133 **The Quantum Hall Effect**
By D. Yoshioka

134 **Magnetism in the Solid State**
By P. Mohn

135 **Electrodynamics of Magnetoactive Media**
By I. Vagner, B.I. Lembrikov, and P. Wyder

136 **Nanoscale Phase Separation and Colossal Magnetoresistance**
The Physics of Manganites and Related Compounds
By E. Dagotto

137 **Quantum Transport in Submicron Devices**
A Theoretical Introduction
By W. Magnus and W. Schoenmaker