Lecture Notes in Physics

The Editorial Policy for Edited Volumes

The series *Lecture Notes in Physics* (LNP), founded in 1969, reports new developments in physics research and teaching - quickly, informally but with a high degree of quality. Manuscripts to be considered for publication are topical volumes consisting of a limited number of contributions, carefully edited and closely related to each other. Each contribution should contain at least partly original and previously unpublished material, be written in a clear, pedagogical style and aimed at a broader readership, especially graduate students and nonspecialist researchers wishing to familiarize themselves with the topic concerned. For this reason, traditional proceedings cannot be considered for this series though volumes to appear in this series are often based on material presented at conferences, workshops and schools.

Acceptance

A project can only be accepted tentatively for publication, by both the editorial board and the publisher, following thorough examination of the material submitted. The book proposal sent to the publisher should consist at least of a preliminary table of contents outlining the structure of the book together with abstracts of all contributions to be included. Final acceptance is issued by the series editor in charge, in consultation with the publisher, only after receiving the complete manuscript. Final acceptance, possibly requiring minor corrections, usually follows the tentative acceptance unless the final manuscript differs significantly from expectations (project outline). In particular, the series editors are entitled to reject individual contributions if they do not meet the high quality standards of this series. The final manuscript must be ready to print, and should include both an informative introduction and a sufficiently detailed subject index.

Contractual Aspects

Publication in LNP is free of charge. There is no formal contract, no royalties are paid, and no bulk orders are required, although special discounts are offered in this case. The volume editors receive jointly 30 free copies for their personal use and are entitled, as are the contributing authors, to purchase Springer books at a reduced rate. The publisher secures the copyright for each volume. As a rule, no reprints of individual contributions can be supplied.

Manuscript Submission

The manuscript in its final and approved version must be submitted in ready to print form. The corresponding electronic source files are also required for the production process, in particular the online version. Technical assistance in compiling the final manuscript can be provided by the publisher‘s production editor(s), especially with regard to the publisher’s own LaTeX macro package which has been specially designed for this series.

LNP Homepage (springerlink.com)

On the LNP homepage you will find:
−The LNP online archive. It contains the full texts (PDF) of all volumes published since 2000. Abstracts, table of contents and prefaces are accessible free of charge to everyone. Information about the availability of printed volumes can be obtained.
−The subscription information. The online archive is free of charge to all subscribers of the printed volumes.
−The editorial contacts, with respect to both scientific and technical matters.
−The author’s / editor’s instructions.

W. Dieter Heiss (Ed.)

Quantum Dots: a Doorway to Nanoscale Physics

Springer

Editor

W. Dieter Heiss
University of Stellenbosch
Department of Physics
MATIELAND 7602
South Africa

W. Dieter Heiss (Ed.), *Quantum Dots: a Doorway to Nanoscale Physics*,
Lect. Notes Phys. **667** (Springer, Berlin Heidelberg 2005), DOI 10.1007/b103740

Library of Congress Control Number: 2005921338

ISSN 0075-8450
ISBN 3-540-24236-8 Springer Berlin Heidelberg New York

Typesetting: Camera-ready by the authors/editor
Data conversion by TechBooks
Cover design: *design & production*, Heidelberg

Printed on acid-free paper
54/3141/jl - 5 4 3 2 1 0

Preface

Nanoscale physics, nowadays one of the most topical research subjects, has two major areas of focus. One is the important field of potential applications bearing the promise of a great variety of materials having specific properties that are desirable in daily life. Even more fascinating to the researcher in physics are the fundamental aspects where quantum mechanics is seen at work; most macroscopic phenomena of nanoscale physics can only be understood and described using quantum mechanics. The emphasis of the present volume is on this latter aspect.

It fits perfectly within the tradition of the South African Summer Schools in Theoretical Physics and the fifteenth Chris Engelbrecht School was devoted to this highly topical subject. This volume presents the contents of lectures from four speakers working at the forefront of nanoscale physics. The first contribution addresses some more general theoretical considerations on Fermi liquids in general and quantum dots in particular. The next topic is more experimental in nature and deals with spintronics in quantum dots. The alert reader will notice the close correspondence to the South African Summer School in 2001, published in LNP 587. The following two sections are theoretical treatments of low temperature transport phenomena and electron scattering on normal-superconducting interfaces (Andreev billiards). The enthusiasm and congenial atmosphere created by the speakers will be remembered well by all participants. The beautiful scenery of the Drakensberg surrounding the venue contributed to the pleasant spirit prevailing during the school.

A considerable contingent of participants came from African countries outside South Africa and were supported by a generous grant from the Ford Foundation; the organisers gratefully acknowledge this assistance.

The Organising Committee is indebted to the National Research Foundation for its financial support, without which such high level courses would be impossible. We also wish to express our thanks to the editors of Lecture Notes in Physics and Springer for their assistance in the preparation of this volume.

Stellenbosch *WD Heiss*
February 2005

List of Contributors

R. Shankar
Sloane Physics Lab, Yale University,
New Haven CT 06520
r.shankar@yale.edu

J.M. Elzerman
Kavli Institute of Nanoscience Delft,
PO Box 5046, 2600 GA Delft,
The Netherlands

ERATO Mesoscopic Correlation
Project, University of Tokyo,
Bunkyo-ku, Tokyo 113-0033, Japan
elzerman@qt.tn.tudelft.nl

R. Hanson
Kavli Institute of Nanoscience Delft,
PO Box 5046, 2600 GA Delft,
The Netherlands

L.H.W. van Beveren
Kavli Institute of Nanoscience Delft,
PO Box 5046, 2600 GA Delft,
The Netherlands

ERATO Mesoscopic Correlation
Project, University of Tokyo,
Bunkyo-ku, Tokyo 113-0033, Japan

S. Tarucha
ERATO Mesoscopic Correlation
Project, University of Tokyo,
Bunkyo-ku, Tokyo 113-0033,
Japan

NTT Basic Research Laboratories,
Atsugi-shi, Kanagawa 243-0129,
Japan

L.M.K. Vandersypen
Kavli Institute of Nanoscience Delft,
PO Box 5046, 2600 GA Delft,
The Netherlands

L.P. Kouwenhoven
Kavli Institute of Nanoscience Delft,
PO Box 5046, 2600 GA Delft,
The Netherlands

ERATO Mesoscopic Correlation
Project, University of Tokyo,
Bunkyo-ku, Tokyo 113-0033, Japan

M. Pustilnik
School of Physics, Georgia Institute
of Technology,
Atlanta, GA 30332, USA

L.I. Glazman
William I. Fine Theoretical
Physics Institute,
University of Minnesota,
Minneapolis, MN 55455, USA

C.W.J. Beenakker
Instituut-Lorentz,
Universiteit Leiden,
P.O. Box 9506, 2300 RA Leiden,
The Netherlands

Contents

A Guide for the Reader

Quantum dots, often denoted artificial atoms, are the exquisite tools by which quantum behavior can be probed on a scale appreciably larger than the atomic scale, that is, on the nanometer scale. In this way, the physics of the devices is closer to classical physics than that of atomic physics but they are still sufficiently small to clearly exhibit quantum phenomena. The present volume is devoted to some of these fascinating aspects.

In the first contribution general theoretical aspects of Fermi liquids are addressed, in particular, the renormalization group approach. The choice of appropriate variables as a result of averaging over "unimportant" variables is presented. This is then aptly applied to large quantum dots. The all important scales, ballistic dots and chaotic motion are discussed. Nonperturbative methods and critical phenomena feature in this thorough treatise. The traditional phenomenological Landau parameters are given a more satisfactory theoretical underpinning.

A completely different approach is encountered in the second contribution in that it is a thorough experimental expose of what can be done or expected in the study of small quantum dots. Here the emphasis lies on the electron spin to be used as a qubit. The experimental steps toward using a single electron spin – trapped in a semiconductor quantum dot – as a spin qubit are described. The introduction contains a resume of quantum computing with quantum dots. The following sections address experimental implementations, the use of different quantum dot architectures, measurements, noise, sensitivity and high-speed performance. The lectures are based on a collaborative effort of research groups in the Netherlands and in Japan.

The last two contributions are again theoretical in nature and address particular aspects relating to quantum dots. In the third lecture series, mechanisms of low-temperature electronic transport through a quantum dot – weakly coupled to two conducting leads – are reviewed. In this case transport is dominated by electron–electron interaction. At moderately low temperatures (comparing with the charging energy) the linear conductance is suppressed by

the Coulomb blockade. A further lowering of the temperature leads into the Kondo regime.

The fourth series of lectures deals with a very specific and cute aspect of nanophysics: a peculiar property of superconducting mirrors as discovered by Andreev about forty years ago. The Andreev reflection at a superconductor modifies the excitation spectrum of a quantum dot. The difference between a chaotic and integrable billiard (quantum dot) is discussed and relevant classical versus quantum time scales are given. The results are a challenge to experimental physicists as they are not confirmed as yet.

The Renormalization Group Approach – From Fermi Liquids to Quantum Dots

R. Shankar

Sloane Physics Lab, Yale University, New Haven CT 06520
r.shankar@yale.edu

1 The RG: What, Why and How

Imagine that you have some problem in the form of a partition function

$$Z(a,b) = \int dx \int dy e^{-a(x^2+y^2)} e^{-b(x+y)^4} \tag{1}$$

where a, b are the parameters.

First consider $b = 0$, *the gaussian model*. Suppose that you are just interested in x, say in its fluctuations. Then you have the option of integrating out y and working with the new partition function

$$Z(a) = N \int dx e^{-ax^2} \tag{2}$$

where N comes from doing the y-integration. We will ignore such an x-independent pre-factor here and elsewhere since it will cancel in any averaging process.

Consider now the nongaussian case with $b \neq 0$. Here we have

$$\begin{aligned} Z(a', b' \ldots) &= \int dx \left[\int dy e^{-a(x^2+y^2)} e^{-b(x+y)^4} \right] \\ &\equiv \int dx e^{-a' \ x^2} e^{-b' x^4 - c' x^6 + \ldots} \end{aligned} \tag{3}$$

where a', b' etc., define the parameters of the effective field theory for x. These parameters will reproduce exactly the same averages for x as the original ones. *This evolution of parameters with the elimination of uninteresting degrees of freedom, is what we mean these days by renormalization, and as such has nothing to do with infinities; you just saw it happen in a problem with just two variables.*

R. Shankar: *The Renormalization Group Approach – From Fermi Liquids to Quantum Dots*, Lect. Notes Phys. **667**, 3–24 (2005)
www.springerlink.com

The parameters b, c etc., are called *couplings* and the monomials they multiply are called *interactions*. The x^2 term is called the *kinetic* or *free-field* term.

Notice that to get the effective theory we need to do a nongaussian integral. This can only be done perturbatively. At the simplest *tree Level*, we simply drop y and find $b' = b$. At higher orders, we bring down the nonquadratic exponential and integrate in y term by term and generate effective interactions for x. This procedure can be represented by Feynman graphs in which variables in the loop are limited to the ones being eliminated.

Why do we do this? Because certain tendencies of x are not so apparent when y is around, but surface to the top, as we zero in on x. For example, we are going to consider a problem in which x stands for low-energy variables and y for high energy variables. Upon integrating out high energy variables a numerically small coupling can grow in size (or initially impressive one diminish into oblivion), as we zoom in on the low energy sector.

This notion can be made more precise as follows. Consider the gaussian model in which we have just $a \neq 0$. We have seen that this value does not change as y is eliminated since x and y do not talk to each other. This is called a *fixed point of the RG*. Now turn on new couplings or "interactions" (corresponding to higher powers of x, y etc.) with coefficients b, c and so on. Let a', b' etc., be the new couplings after y is eliminated. The mere fact that $b' > b$ does not mean b is more important for the physics of x. This is because a' could also be bigger than a. So we rescale x so that the kinetic part, x^2, has the same coefficient as before. If the quartic term still has a bigger coefficient, (still called b'), we say it is a *relevant* interaction. If $b' < b$ we say it is irrelevant. This is because in reality y stands for many variables, and as they are eliminated one by one, the coefficient of the quartic term will run to zero. If a coupling neither grows not shrinks it is called *marginal*.

There is another excellent reason for using the RG, and that is to understand the phenomenon of universality in critical phenomena. I must regretfully pass up the opportunity to explain this and refer you to Professor Michael Fisher's excellent lecture notes in this very same school many years ago [1].

We will now see how this method is applied to interacting fermions in $d = 2$. Later we will apply these methods to quantum dots.

2 The Problem of Interacting Fermions

Consider a system of nonrelativistic spinless fermions in two space dimensions. The one particle hamiltonian is

$$H = \frac{K^2}{2m} - \mu \tag{4}$$

where the chemical potential μ is introduced to make sure we have a finite density of particles in the ground state: all levels up the Fermi surface, a circle defined by

$$K_F^2/2m = \mu \tag{5}$$

are now occupied and occupying these levels lowers the ground-state energy.

Notice that this system has gapless excitations above the ground state. You can take an electron just below the Fermi surface and move it just above, and this costs as little energy as you please. Such a system will carry a dc current in response to a dc voltage. An important question one asks is if this will be true when interactions are turned on. For example the system could develop a gap and become an insulator. What really happens for the $d = 2$ electron gas?

We are going to answer this using the RG. Let us first learn how to do RG for noninteracting fermions. To understand the low energy physics, we take a band of of width Λ on either side of the Fermi surface. This is the first great difference between this problem and the usual ones in relativistic field theory and statistical mechanics. Whereas in the latter examples low energy means small momentum, here it means small deviations from the Fermi surface. Whereas in these older problems we zero in on the origin in momentum space, here we zero in on a surface. The low energy region is shown in Fig. 1.

To apply our methods we need to cast the problem in the form of a path integral. Following any number of sources, say [2] we obtain the following expression for the partition function of free fermions:

$$Z_0 = \int d\psi d\overline{\psi} e^{S_0} \tag{6}$$

where

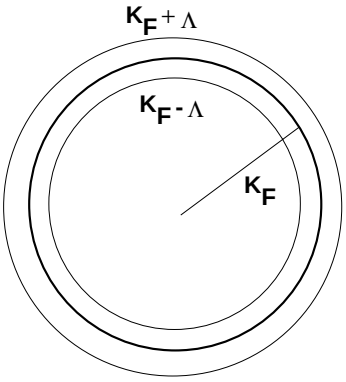

Fig. 1. The low energy region for nonrelativistic fermions lies within the annulus concentric with the Fermi circle

$$S_0 = \int d^2K \int_{-\infty}^{\infty} d\omega \overline{\psi}(\omega, \mathbf{K}) \left(i\omega - \frac{(K^2 - K_F^2)}{2m} \right) \psi(\omega, \mathbf{K}) \tag{7}$$

where ψ and $\overline{\psi}$ are called Grassmann variables. They are really weird objects one gets to love after some familiarity. Most readers can assume they are ordinary integration variables. The dedicated reader can learn more from [2].

We now adapt this general expression to the annulus to obtain

$$Z_0 = \int d\psi d\overline{\psi} e^{S_0} \tag{8}$$

where

$$S_0 = \int_0^{2\pi} d\theta \int_{-\infty}^{\infty} d\omega \int_{-\Lambda}^{\Lambda} dk \overline{\psi}(i\omega - v\ k)\psi\ . \tag{9}$$

To get here we have had to approximate as follows:

$$\frac{K^2 - K_F^2}{2m} \simeq \frac{K_F}{m} \cdot k = v_F\ k \tag{10}$$

where $k - K - K_F$ and v_F is the fermi velocity, hereafter set equal to unity. Thus Λ can be viewed as a momentum or energy cut-off measured from the Fermi circle. We have also replaced KdK by $K_F dk$ and absorbed K_F in ψ and $\overline{\psi}$. It will seen that neglecting k in relation to K_F is irrelevant in the technical sense.

Let us now perform mode elimination and reduce the cut-off by a factor s. Since this is a gaussian integral, mode elimination just leads to a multiplicative constant we are not interested in. So the result is just the same action as above, but with $|k| \leq \Lambda/s$. Let us now do make the following additional transformations:

$$\begin{aligned} (\omega', k') &= s(\omega, k) \\ (\psi'(\omega', k'), \overline{\psi}'(\omega', k')) &= s^{-3/2} \left(\psi \left(\frac{\omega'}{s}, \frac{k'}{s} \right), \overline{\psi} \left(\frac{\omega'}{s}, \frac{k'}{s} \right) \right) \ . \end{aligned} \tag{11}$$

When we do this, the action and the phase space all return to their old values. So what? Recall that our plan is to evaluate the role of quartic interactions in low energy physics as we do mode elimination. Now what really matters is not the absolute size of the quartic term, but its size relative to the quadratic term. Keeping the quadratic term identical before and after the RG action makes the comparison easy: if the quartic coupling grows, it is relevant; if it decreases, it is irrelevant, and if it stays the same it is marginal. The system is clearly gapless if the quartic coupling is irrelevant. Even a marginal coupling implies no gap since any gap will grow under the various rescalings of the RG.

Let us now turn on a generic four-Fermi interaction in path-integral form:

$$S_4 = \int \overline{\psi}(4)\overline{\psi}(3)\psi(2)\psi(1)u(4,3,2,1) \tag{12}$$

where $\int$ is a shorthand:

$$\int \equiv \prod_{i=1}^{3} \int d\theta_i \int_{-\Lambda}^{\Lambda} dk_i \int_{-\infty}^{\infty} d\omega_i \tag{13}$$

At the tree level, we simply keep the modes within the new cut-off, rescale fields, frequencies and momenta, and read off the new coupling. We find

$$u'(k', \omega', \theta) = u(k'/s, \omega'/s, \theta) \tag{14}$$

This is the evolution of the coupling function. To deal with coupling constants with which we are more familiar, we expand the functions in a Taylor series (schematic)

$$u = u_o + k u_1 + k^2 u_2 \ldots \tag{15}$$

where k stands for all the k's and ω's. An expansion of this kind is possible since couplings in the Lagrangian are nonsingular in a problem with short range interactions. If we now make such an expansion and compare coefficients in (14), we find that u_0 is marginal and the rest are irrelevant, as is any coupling of more than four fields. Now this is exactly what happens in ϕ_4^4, scalar field theory in four dimensions with a quartic interaction. The difference here is that we still have dependence on the angles on the Fermi surface:

$$u_0 = u(\theta_1, \theta_2, \theta_3, \theta_4)$$

Therefore in this theory we are going to get coupling functions and not a few coupling constants.

Let us analyze this function. Momentum conservation should allow us to eliminate one angle. Actually it allows us more because of the fact that these momenta do not come form the entire plane, but a very thin annulus near K_F. Look at the left half of Fig. 2. Assuming that the cutoff has been reduced to the thickness of the circle in the figure, it is clear that if two points 1 and 2 are chosen from it to represent the incoming lines in a four point coupling,

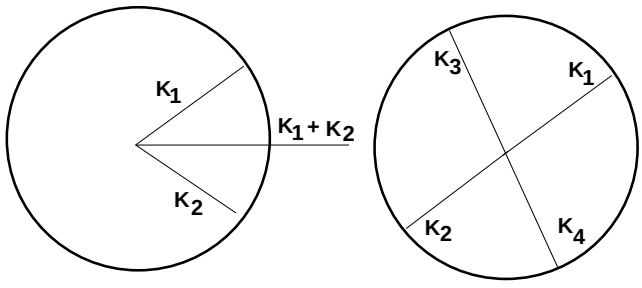

Fig. 2. Kinematical reasons why momenta are either conserved pairwise or restricted to the BCS channel

the outgoing ones are forced to be equal to them (not in their sum, but individually) up to a permutation, which is irrelevant for spinless fermions. Thus we have in the end just one function of two angles, and by rotational invariance, their difference:

$$u(\theta_1, \theta_2, \theta_1, \theta_2) = F(\theta_1 - \theta_2) \equiv F(\theta) \ . \tag{16}$$

About forty years ago Landau came to the very same conclusion [3] that a Fermi system at low energies would be described by one function defined on the Fermi surface. He did this without the benefit of the RG and for that reason, some of the leaps were hard to understand. Later detailed diagrammatic calculations justified this picture [4]. The RG provides yet another way to understand it. It also tells us other things, as we will now see.

The first thing is that the final angles are not slaved to the initial ones if the former are exactly opposite, as in the right half of Fig. 2. In this case, the final ones can be anything, as long as they are opposite to each other. This leads to one more set of marginal couplings in the BCS channel, called

$$u(\theta_1, -\theta_1, \theta_3, -\theta_3) = V(\theta_3 - \theta_1) \equiv V(\theta) \ . \tag{17}$$

The next point is that since F and V are marginal at tree level, we have to go to one loop to see if they are still so. So we draw the usual diagrams shown in Fig. 3. We eliminate an infinitesimal momentum slice of thickness $d\Lambda$ at $k = \pm\Lambda$.

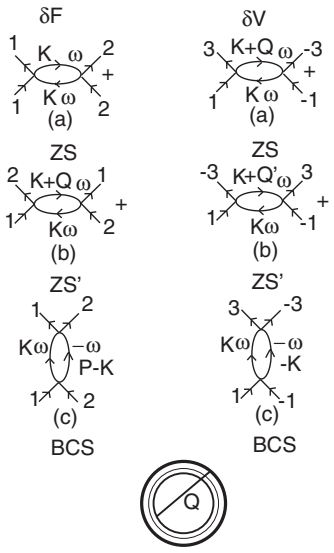

Fig. 3. One loop diagrams for the flow of F and V. The last at the *bottom* shows that a large momentum Q can be absorbed only at two particular initial angles (only one of which is shown) if the final state is to lie in the shell being eliminated

These diagrams are like the ones in any quartic field theory, but each one behaves differently from the others and its its traditional counterparts. Consider the first one (called ZS) for F. The external momenta have zero frequencies and lie of the Fermi surface since ω and k are irrelevant. The momentum transfer is exactly zero. So the integrand has the following schematic form:

$$\delta F \simeq \int d\theta \int dk d\omega \left(\frac{1}{(i\omega - \varepsilon(K))} \; \frac{1}{(i\omega - \varepsilon(K))} \right) \tag{18}$$

The loop momentum K lies in one of the two shells being eliminated. Since there is no energy difference between the two propagators, the poles in ω lie in the same half-plane and we get zero, upon closing the contour in the other half-plane. In other words, this diagram can contribute if it is a particle-hole diagram, but given zero momentum transfer we cannot convert a hole at $-\Lambda$ to a particle at $+\Lambda$. In the ZS' diagram, we have a large momentum transfer, called Q in the inset at the bottom. This is of order K_F and much bigger than the radial cut-off, a phenomenon unheard of in say ϕ^4 theory, where all momenta and transfers are bounded by Λ. This in turn means that the loop momentum is not only restricted in the direction to a sliver $d\Lambda$, but also in the angular direction in order to be able to absorb this huge momentum Q and land up in the other shell being eliminated (see bottom of (Fig. 3). So we have $du \simeq dt^2$, where $dt = d\Lambda/\Lambda$. The same goes for the BCS diagram. Thus F does not flow at one loop.

Let us now turn to the renormalization of V. The first two diagrams are useless for the same reasons as before, but the last one is special. Since the total incoming momentum is zero, the loop momenta are equal and opposite and no matter what direction K has, $-K$ is guaranteed to lie in the same shell being eliminated. However the loop frequencies are now equal and opposite so that the poles in the two propagators now lie in opposite half-planes. We now get a flow (dropping constants)

$$\frac{dv(\theta_1 - \theta_3)}{dt} = -\int d\theta v(\theta_1 - \theta) \; v(\theta - \theta_3) \tag{19}$$

Here is an example of a flow equation for a coupling function. However by expanding in terms of angular momentum eigenfunctions we get an infinite number of flow equations

$$\frac{dv_m}{dt} = -v_m^2 \; . \tag{20}$$

one for each coefficient. These equations tell us that if the potential in angular momentum channel m is repulsive, it will get renormalized down to zero (a result derived many years ago by Anderson and Morel) while if it is attractive, it will run off, causing the BCS instability. This is the reason the V's are not a part of Landau theory, which assumes we have no phase transitions. This is also a nice illustration of what was stated earlier: one could begin with a large positive coupling, say v_3 and a tiny negative coupling v_5. After much renormalization, v_3 would shrink to a tiny value and v_5 would dominate.

3 Large-N Approach to Fermi Liquids

Not only did Landau say we could describe Fermi liquids with an F function, he also managed to compute the response functions at small ω and q in terms of the F function even when it was large, say 10, in dimensionless units. Again the RG gives us one way to understand this. To this end we need to recall the the key ideas of "large-N" theories.

These theories involve interactions between N species of objects. The largeness of N renders fluctuations (thermal or quantum) small, and enables one to make approximations which are not perturbative in the coupling constant, but are controlled by the additional small parameter $1/N$.

As a specific example let us consider the Gross-Neveu model [5] which is one of the simplest fermionic large-N theories. This theory has N identical massless relativistic fermions interacting through a short-range interaction. The Lagrangian density is

$$\mathcal{L} = \sum_{i=1}^{N} \bar{\psi}_i \not{\partial} \psi_i - \frac{\lambda}{N} \left(\sum_{i=1}^{N} \bar{\psi}_i \psi_i \right)^2 \tag{21}$$

Note that the kinetic term conserves the internal index, as does the interaction term: any index that goes in comes out. You do not have to know much about the GN model to to follow this discussion, which is all about the internal indices.

Figure 4 shows the first few diagrams in the expression for the scattering amplitude of particle of isospin index i and j in the Gross-Neveu theory. The "bare" vertex comes with a factor λ/N. The one-loop diagrams all share a factor λ^2/N^2 from the two vertices. The first one-loop diagram has a free internal summation over the index k that runs over N values, with the contribution being identical for each value of k. Thus, this one-loop diagram acquires a compensating factor of N which makes its contribution of order λ^2/N, the *same order in* $1/N$ *as the bare vertex*. However, the other one-loop diagrams have no such free internal summation and their contribution

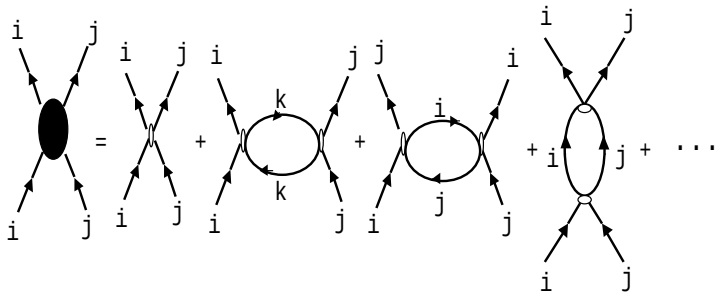

Fig. 4. Some diagrams from a large-N theory

is indeed of order $1/N^2$. Therefore, to leading order in $1/N$, one should keep only diagrams which have a free internal summation for every vertex, that is, iterates of the leading one-loop diagram, which are called bubble graphs. *For later use remember that in the diagrams that survive* (*do not survive*)*, the indices i and j of the incoming particles do not* (*do*) *enter the loops.* Let us assume that the momentum integral up to the cutoff Λ for one bubble gives a factor $-\Pi(\Lambda, q_{ext})$, where q_{ext} is the external momentum or frequency transfer at which the scattering amplitude is evaluated. To leading order in large-N the full expression for the scattering amplitude is

$$\Gamma(q_{ext}) = \frac{1}{N} \frac{\lambda}{1 + \lambda \Pi(\Lambda, q_{ext})} \tag{22}$$

Once one has the full expression for the scattering amplitude (to leading order in $1/N$) one can ask for the RG flow of the "bare" vertex as the cutoff is reduced by demanding that the physical scattering amplitude Γ remain insensitive to the cutoff. One then finds (with $t = \ln(\Lambda_0/\Lambda)$)

$$\frac{d\Gamma(q_{ext})}{dt} = 0 \Rightarrow \frac{d\lambda}{dt} = -\lambda^2 \frac{d\Pi(\Lambda, q_{ext})}{dt} \tag{23}$$

which is exactly the flow one would extract at one loop. Thus the one-loop RG flow is the exact answer to leading order in a large-N theory. All higher-order corrections must therefore be subleading in $1/N$.

3.1 Large-N Applied to Fermi Liquids

Consider now the $\bar{\psi}\psi - \bar{\psi}\psi$ correlation function (with vanishing values of external frequency and momentum transfer). Landau showed that it takes the form

$$\chi = \frac{\chi_0}{1 + F_0} \ , \tag{24}$$

where F_0 is the angular average of $F(\theta)$ and χ_0 is the answer when $F = 0$. Note that the answer is not perturbative in F.

Landau got this result by working with the ground-state energy as a functional of Fermi surface deformations. The RG provides us with not just the ground-state energy, but an effective hamiltonian (operator) for all of low-energy physics. This operator problem can be solved using large N-techniques.

The value of N here is of order K_F/Λ, and here is how it enters the formalism. Imagine dividing the annulus in (Fig. 1) into N patches of size (Λ) in the radial and angular directions. The momentum of each fermion $\mathbf{k}_i$ is a sum of a "large" part ($\mathcal{O}(k_F)$) centered on a patch labelled by a patch index $i = 1, \ldots N$ and a "small" momentum ($\mathcal{O}(\Lambda)$ within the patch [2].

Consider a four-fermion Green's function, as in (Fig. 4). The incoming momenta are labelled by the patch index (such as i) and the small momentum is not shown but implicit. We have seen that as $\Lambda \to 0$, the two outgoing

momenta are equal to the two incoming momenta up to a permutation. At small but finite Λ this means the patch labels are same before and after. Thus the patch index plays the role of a conserved isospin index as in the Gross-Neveu model.

The electron-electron interaction terms, written in this notation, (with **k** integrals replaced by a sum over patch index and integration over small momenta) also come with a pre-factor of $1/N$ ($\simeq \Lambda/K_F$).

It can then be verified that in all Feynman diagrams of this cut-off theory the patch index plays the role of the conserved isospin index exactly as in a theory with N fermionic species. For example in (Fig. 4) in the first diagram, the external indices i and j do not enter the diagram (small momentum transfer only) and so the loop momentum is nearly same in both lines and integrated fully over the annulus, i.e., the patch index k runs over all N values. In the second diagram, the external label i enters the loop and there is a large momentum transfer ($\mathcal{O}(K_F)$). In order for both momenta in the loop to be within the annulus, and to differ by this large q, the angle of the loop momentum is limited to a range $\mathcal{O}(\Lambda/K_F)$. (This just means that if one momentum is from patch i the other has to be from patch j.) Similarly, in the last loop diagram, the angle of the loop momenta is restricted to one patch. In other words, the requirement that all loop momenta in this cut-off theory lie in the annulus singles out only diagrams that survive in the large N limit.

The sum of bubble diagrams, singled out by the usual large-N considerations, reproduces Landau's Fermi liquid theory. For example in the case of χ, one obtains a geometric series that sums to give $\chi = \frac{\chi_0}{1+F_0}$.

Since in the large N limit, the one-loop β-function for the fermion-fermion coupling is exact, it follows that the marginal nature of the Landau parameters F and the flow of V, (20), are both exact as $\Lambda \to 0$.

A long paper of mine [2] explains all this, as well as how it is to be generalized to anisotropic Fermi surfaces and Fermi surfaces with additional special features and consequently additional instabilities. Polchinski [6] independently analyzed the isotropic Fermi liquid (though not in the same detail, since it was a just paradigm or toy model for an effective field theory for him).

4 Quantum Dots

We will now apply some of these ideas, very successful in the bulk, to two-dimensional quantum dots [7, 9] tiny spatial regions of size $L \simeq 100 - 200$ nm, to which electrons are restricted using gates. The dot can be connected weakly or strongly to leads. Since many experts on dots are contributing to this volume, I will be sparing in details and references.

Let us get acquainted with some energy scales, starting with Δ, the mean single particle level spacing. The Thouless energy is defined as $E_T = \hbar/\tau$, where τ is the time it takes to traverse the dot. If the dot is strongly coupled to leads, this is the uncertainty in the energy of an electron as it traverses

the dot. Consequently the g (sharply defined) states of an isolated dot within E_T will contribute to conductance and lead to a (dimensionless) conductance $g = \frac{E_T}{\Delta}$.

The dots in question have two features important to us. First, motion within the dot is ballistic: L_{el}, the elastic scattering length is the same as L, the dot size, so that $E_T = \frac{\hbar v_F}{L}$, where v_F is the Fermi velocity. Next, the boundary of the dot is sufficiently irregular as to cause chaotic motion at the classical level. At the quantum level single-particle energy levels and wavefunctions (in any basis) within E_T of the Fermi energy will resemble those of a random hamiltonian matrix and be described Random Matrix Theory (RMT) [8]. We will only invoke a few results from RMT and they will be explained in due course.

At a generic value of gate voltage V_g the ground state has a definite number of particles N and energy $\mathcal{E}_N$. If $\mathcal{E}_{N+1} - \mathcal{E}_N = \alpha e V_g$ (α is a geometry-dependent factor) the energies of the N and $N+1$-particle states are degenerate, and a tunnelling peak occurs at zero bias. Successive peaks are separated by the second difference of $\mathcal{E}_N$, called Δ_2, the distribution of which is measured. Also measured are statistics of peak-height distributions [10, 11, 12], which depend on wavefunction statistics of RMT.

To describe the data one needs to write down a suitable hamiltonian

$$H_U = \sum_{\alpha} \varepsilon_\alpha c_\alpha^\dagger c_\alpha + \frac{1}{2} \sum_{\alpha\beta\gamma\delta} V_{\alpha\beta\gamma\delta} c_\alpha^\dagger c_\beta^\dagger c_\gamma c_\delta \tag{25}$$

(where the subscripts label the exact single particle states including spin) and try to extract its implications. Earlier theoretical investigations were confined to the noninteracting limit: $V \equiv 0$ and missed the fact that due to the small capacitance of the dot, adding an electron required some significant charging energy on top of the energy of order ε_α it takes to promote an electron by one level. Thus efforts have been made to include interactions [9, 13, 14, 15, 16].

The simplest model includes a constant charging energy $U_0 N^2/2$ [7, 13]. Conventionally U_0 is subtracted away in plotting Δ_2. This model predicts a bimodal distribution for Δ_2: Adding an electron above a doubly-filled (spin-degenerate) level costs $U_0 + \varepsilon$, with ε being the energy to the next single-particle level. Adding it to a singly occupied level costs U_0. While the second contribution gives a delta-function peak at 0 after U_0 has been subtracted, the first contribution is the distribution of nearest neighbor level separation ε, of the order of Δ. But simulations [16] and experiments [17, 18] produce distributions for Δ_2 which do not show any bimodality, and are much broader.

The next significant advance was the discovery of the Universal Hamiltonian [14, 15]. Here one keeps only couplings of the form $V_{\alpha\beta\beta\alpha}$ on the grounds that only they have a non-zero ensemble average (over disorder realizations). This seems reasonable in the limit of large g since couplings with zero average are typically of size $1/g$ according to RMT. The Universal Hamiltonian is thus

$$H_U = \sum_{\alpha,s} \varepsilon_\alpha c^\dagger_{\alpha,s} c_{\alpha,s} + \frac{U_0}{2} N^2 - \frac{J}{2}\mathbf{S}^2 + \lambda \left(\sum_\alpha c^\dagger_{\alpha,\uparrow} c^\dagger_{\alpha,\downarrow} \right) \left(\sum_\beta c_{\beta,\downarrow} c_{\beta,\uparrow} \right) \quad (26)$$

where s is single-particle spin and $\mathbf{S}$ is the total spin. The Cooper coupling λ does not play a major role, but the inclusion of the exchange coupling J brings the theoretical predictions [9, 14, 15] into better accord with experiments, especially if one-body "scrambling" [19, 20, 21, 22] and finite temperature effects are taken into account. However, some discrepancies still remain in relation to numerical [16] and experimental results [18] at $r_s \geq 2$.

We now see that the following dot-related questions naturally arise. Given that adding more refined interactions (culminating in the universal hamiltonian) led to better descriptions of the dot, should one not seek a more systematic way to to decide what interactions should be included from the outset? Does our past experience with clean systems and bulk systems tell us how to proceed? Once we have written down a comprehensive hamiltonian, is there a way to go beyond perturbation theory to unearth nonperturbative physics in the dot, including possible phases and transitions between them? What will be the experimental signatures of these novel phases and the transitions between them if indeed they do exist? These questions will now be addressed.

4.1 Interactions and Disorder: Exact Results on the Dot

The first crucial step towards this goal was taken by Murthy and Mathur [23]. Their ideas was as follows.

- **Step 1:** *Use the clean system RG described earlier [2] (eliminating momentum states on either side of the Fermi surface) to eliminate all states far from the Fermi surface till one comes down to the Thouless band, that is, within E_T of E_F.*
  We have seen that this process inevitably leads to Landau's Fermi liquid interaction (spin has been suppressed):

$$V = \sum_{m=0}^{\infty} \frac{u_m \Delta}{2} \sum_{k,k'} \cos[m(\theta - \theta')] c^\dagger_{\mathbf{k}} c_{\mathbf{k}} c^\dagger_{\mathbf{k}'} c_{\mathbf{k}'} \quad (27)$$

  where θ, θ' are the angles of $\mathbf{k}, \mathbf{k}'$ on the Fermi circle, and u_m is defined by

$$F(\theta) = \sum_m u_m e^{im\theta} \ . \quad (28)$$

  A few words before we proceed. First, some experts will point out that the interaction one gets from the RG allows for small momentum transfer, i.e., there should be an additional sum over a small values q in (27) allowing $\mathbf{k} \to \mathbf{k} + \mathbf{q}$ and $\mathbf{k}' \to \mathbf{k}' - \mathbf{q}$. It can be shown that in the large g limit

this sum has just one term, at $q = 0$. Unlike in a clean system, there is no singular behavior associated with $q \to 0$ and this assumption is a good one. Others have asked how one can introduce the Landau interaction that respects momentum conservation in a dot that does not conserve momentum or anything else except energy. To them I say this. Just think of a pair of molecules colliding in a room. As long as the collisions take place in a time scale smaller than the time between collisions with the walls, the interaction will be momentum conserving. That this is true for a collision in the dot for particles moving at v_F, subject an interaction of range equal to the Thomas Fermi screening length (the typical range) is readily demonstrated. Like it or not, momentum is a special variable even in a chaotic but ballistic dot since it is tied to translation invariance, and that that is operative for realistic collisions within the dot.

- **Step 2:** *Switch to the exact basis states of the chaotic dot, writing the kinetic and interaction terms in this basis. Run the RG by eliminating exact energy eigenstates within* E_T.

While this looks like a reasonable plan, it is not clear how it is going to be executed since knowledge of the exact eigenfunctions is needed to even write down the Landau interaction in the disordered basis:

$$V_{\alpha\beta\gamma\delta} = \tfrac{\Delta}{4} \sum_{\mathbf{k}\mathbf{k}'} u(\theta - \theta') \left[\phi^*_\alpha(\mathbf{k})\phi^*_\beta(\mathbf{k}') - \phi^*_\alpha(\mathbf{k}')\phi^*_\beta(\mathbf{k})\right] \times [\phi_\gamma(\mathbf{k}')\phi_\delta(\mathbf{k}) - \phi_\gamma(\mathbf{k})\phi_\delta(\mathbf{k}')] \tag{29}$$

where $\mathbf{k}$ and $\mathbf{k}'$ take g possible values. These are chosen as follows. Consider the momentum states of energy within E_T of E_F. In a dot momentum is defined with an uncertainty $\Delta k \simeq 1/L$ in either direction. Thus one must form packets in k space obeying this condition. It can be easily shown that g of them will fit into this band. One way to pick such packets is to simply take plane waves of precise $\mathbf{k}$ and chop them off at the edges of the dot and normalize the remains. The g values of $\mathbf{k}$ can be chosen with an angular spacing $2\pi/g$. It can be readily verified that such states are very nearly orthogonal. The wavefunction $\phi_\delta(\mathbf{k})$ is the projection of exact dot eigenstate δ on the state $\mathbf{k}$ as defined above.

We will see that one can go a long way without detailed knowledge of the wavefunctions $\phi_\delta(\mathbf{k})$.

First, one can take the view of the Universal Hamiltonian (UH) adherents and consider the ensemble average (enclosed in $\langle\ \rangle$) of the interactions. RMT tells us that to leading order in $1/g$,

$$\left\langle \phi^*_\alpha(\mathbf{k}_1)\phi^*_\beta(\mathbf{k}_2)\phi_\gamma(\mathbf{k}_3)\phi_\delta(\mathbf{k}_4)\right\rangle = \frac{\delta_{\alpha\delta}\delta_{\beta\gamma}\delta_{\mathbf{k}_1\mathbf{k}_4}\delta_{\mathbf{k}_2\mathbf{k}_3}}{g^2} + \frac{\delta_{\alpha\gamma}\delta_{\beta\delta}\delta_{\mathbf{k}_1\mathbf{k}_3}\delta_{\mathbf{k}_2\mathbf{k}_4}}{g^2} + \frac{\delta_{\alpha\beta}\delta_{\gamma\delta}\delta_{\mathbf{k}_1,-\mathbf{k}_2}\delta_{\mathbf{k}_3,-\mathbf{k}_4}}{g^2} \tag{30}$$

It is seen that only matrix elements in (29) for which the indices $\alpha\beta\gamma\delta$ are pairwise equal survive disorder-averaging, and also that the average has no

dependence on the energy of $\alpha\beta\gamma\delta$. In the spinless case, the first two terms on the right hand side make equal contributions and produce the constant charging energy in the Universal Hamiltonian of (26), while in the spinful case they produce the charging and exchange terms. The final term of (30) produces the Cooper interaction of (26).

Thus the UH contains the rotationally invariant part of the Landau interaction. The others, i.e., those that do not survive ensemble averaging, are dropped because they are of order $1/g$. But we have seen before in the BCS instability of the Fermi liquid that a term that is nominally small to begin with can grow under the RG. That this is what happens in this case was shown by the RG calculation of Murthy and Mathur. There was however one catch. The neglected couplings could overturn the UH description for couplings that exceeded a critical value. However the critical value is of order unity and so one could not trust either the location or even the very existence of this critical point based on their perturbative one-loop calculation. Their work also gave no clue as to what lay on the other side of the critical point.

Subsequently Murthy and I [24] showed that the methods of the large N theories (with g playing the role of N) were applicable here and could be used to show nonperturbatively in the interaction strength that the phase transition indeed exists. This approach also allowed us to study in detail the phase on the other side of the transition, as well as what is called the quantum critical region, to be described later.

Let us now return to Murthy and Mathur and ask how the RG flow is derived. After integrating some of the g states within E_T, we end up with $g' = ge^{-t}$ states. Suppose we compute a scattering amplitude $\Gamma_{\alpha\beta\gamma\delta}$ for the process in which two fermions originally in states $\alpha\beta$ are scattered into states $\gamma\delta$. This scattering can proceed directly through the vertex $V_{\alpha\beta\gamma\delta}(t)$, or via intermediate virtual states higher order in the interactions, which can be classified by a set of Feynman diagrams, as shown in Fig. 5. All the states in the diagrams belong to the g' states kept. *We demand that the entire amplitude be independent of g', meaning that the physical amplitudes should be the same in the effective theory as in the original theory.* This will lead to a set of flow

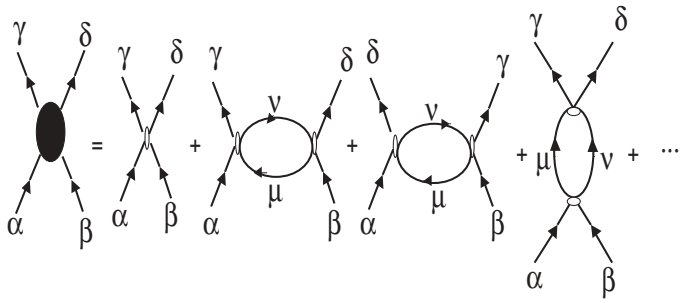

Fig. 5. Feynman diagrams for the full four-point amplitude $\Gamma_{\alpha\beta\gamma\delta}$

equations for the $V_{\alpha\beta\gamma\delta}$. In principle this flow equation will involve all powers of V but we will keep only quadratic terms (the one-loop approximation). Then the diagrams are limited to the ones shown in Fig. 5, leading to the following contributions to the scattering amplitude $\Gamma_{\alpha\beta\gamma\delta}$

$$\begin{aligned}\Gamma_{\alpha\beta\gamma\delta} = V_{\alpha\beta\gamma\delta} \\ + {\sum_{\mu,\nu}}' \frac{N_F(\nu) - N_F(\mu)}{\varepsilon_\mu - \varepsilon_\nu} \Big(V_{\alpha\nu\mu\delta} V_{\beta\mu\nu\gamma} - V_{\alpha\nu\mu\gamma} V_{\beta\mu\nu\delta} \Big) \\ - {\sum_{\mu\nu}}' \frac{1 - N_F(\mu) - N_F(\nu)}{\varepsilon_\mu + \varepsilon_\nu} V_{\alpha\beta\mu\nu} V_{\nu\mu\gamma\delta}\end{aligned} \tag{31}$$

where the prime on the sum reminds us that only the g' remaining states are to be kept and where $N_F(\alpha)$ is the Fermi occupation of the state α. We will confine ourselves to zero temperature where this number can only be zero or one. The matrix element $V_{\alpha\beta\gamma\delta}$ now explicitly depends on the RG flow parameter t.

Now we demand that upon integrating the two states at $\pm g'\Delta/2$ we recover the same $\Gamma_{\alpha\beta\gamma\delta}$. Clearly, since $g' = ge^{-t}$,

$$\frac{d}{dt} = -g' \frac{\delta}{\delta g'} \tag{32}$$

The effect of this differentiation on the loop diagrams is to fix one of the internal lines of the loop to be at the cutoff $\pm g'\Delta/2$, while the other one ranges over all smaller values of energy. In the particle-hole diagram, since μ or ν can be at $+g'\Delta/2$ or $-g'\Delta/2$, and the resulting summations are the same in all four cases, we take a single contribution and multiply by a factor of 4. The same reasoning applies to the Cooper diagram. Let us define the energy cutoff $\Lambda = g'\Delta/2$ to make the notation simpler. Since we are integrating out two states we have $\delta g' = 2$

$$\begin{aligned}0 = \frac{dV_{\alpha\beta\gamma\delta}}{dt} \\ - \frac{g'}{2} 4 {\sum_{\mu=\Lambda,\nu}}' \frac{N_F(\nu) - N_F(\mu)}{\varepsilon_\mu - \varepsilon_\nu} \Big(V_{\alpha\nu\mu\delta} V_{\beta\mu\nu\gamma} - V_{\alpha\nu\mu\gamma} V_{\beta\mu\nu\delta} \Big) \\ + \frac{g'}{2} 4 {\sum_{\mu=\Lambda,\nu}}' \frac{1 - N_F(\mu) - N_F(\nu)}{\varepsilon_\mu + \varepsilon_\nu} V_{\alpha\beta\mu\nu} V_{\nu\mu\gamma\delta}\end{aligned} \tag{33}$$

where $\mu = \Lambda$ means $\varepsilon_\mu = \Lambda$ and so on. The changed sign in front of the 1-loop diagrams reflects the sign of (32)

So far we have not made any assumptions about the form of $V_{\alpha\beta\gamma\delta}$, and the formulation applies to any finite system. In a generic system such as an atom, the matrix elements depend very strongly on the state being integrated over, and the flow must be followed numerically for each different set $\alpha\beta\gamma\delta$ kept in the low-energy subspace.

In our problem things have become so bad that are good once again: the wavefunctions $\phi(\mathbf{k})$ that enter the matrix elements above have so scrambled up by disorder that they can be handled by RMT. In particular it is possible to argue that the sum over the g' terms may be replaced by it ensemble average. In other words the flow equation is self-averaging. While the most convincing way to show this is to compute its variance, and see that it is of order $1/\sqrt{g}$ times its average, this fact can be motivated in the following way: There is a sum over $g' \gg 1$ values of ν with a slowly varying energy denominator, which makes the sum over ν similar to a spectral average, which in RMT is the same as an average over the disorder ensemble. A more sophisticated argument is presented in [25]. We can use the result

$$\left\langle \phi^*_\mu(\mathbf{p}_1)\phi^*_\nu(\mathbf{p}_2)\phi_\nu(\mathbf{p}_3)\phi_\mu(\mathbf{p}_4)\right\rangle = \frac{\delta_{14}\delta_{23}}{g^2} - \frac{\delta_{13}\delta_{24}}{g^3} - \frac{\delta_{1,-2}\delta_{3,-4}}{g^3} \tag{34}$$

to deal with the product of four wavefunctions inside the loop and deal with the energy sum as follows:

$$\sum_{\varepsilon_\nu=-\Lambda}^{0} \frac{1}{\Lambda+|\varepsilon_\nu|} \approx \int_0^\Lambda \frac{d\varepsilon}{\Delta}\frac{1}{\Lambda+\varepsilon} = \frac{\ln 2}{\Delta} \tag{35}$$

We are exploiting the fact that wavefunction correlations are energy independent in the large -g limit.

After we make this simplifications we find that there are many kinds of terms of which one kind dominates in the large -g limit.

Let us go back to the properly antisymmetrized matrix element defined in terms of the Fermi liquid interaction function (29). Since there is a product of two V's in each loop diagram, and each V contains 4 terms, it is clear that each loop contribution has 16 terms. Let us first consider a term of the leading type in the particle-hole diagram, which survives in the large-g limit. Putting in the full wavefunction dependencies (and ignoring factors other than g, g') we have the following contribution from this type of term

$$\begin{aligned}\frac{dV_{\alpha\beta\gamma\delta}}{dt}\text{Leading} &= g'\Delta^2 \sum_{\nu=-\Lambda}^{0} \frac{1}{\Lambda+|\varepsilon_\nu|}\sum_{\mathbf{kk'}}\sum_{\mathbf{pp'}} u\Big(\theta_{\mathbf{k}}-\theta_{\mathbf{p}}\Big)u\Big(\theta_{\mathbf{p'}}-\theta_{\mathbf{k'}}\Big) \\ &\times \phi^*_\alpha(\mathbf{k})\phi^*_\beta(\mathbf{k'})\phi_\gamma(\mathbf{k'})\phi_\delta(\mathbf{k})\phi^*_\mu(\mathbf{p})\phi^*_\nu(\mathbf{p'})\phi_\nu(\mathbf{p})\phi_\mu(\mathbf{p'})\end{aligned} \tag{36}$$

Substituting the appropriate momentum labels for the particle-hole diagram in (34), we see that the wavefunction average relevant to the sum over intermediate states is

$$\frac{\delta_{\mathbf{pp'}}}{g^2} - \frac{1+\delta_{\mathbf{p},-\mathbf{p'}}}{g^3} \tag{37}$$

Using the self-averaging, the first term of (37) forces $\mathbf{p} = \mathbf{p'}$ in (36). For large g, using

$$\sum_{\mathbf{p}} = g \int \frac{d\theta_{\mathbf{p}}}{2\pi} \tag{38}$$

we obtain a convolution of the two Fermi liquid functions

$$\sum_{\mathbf{p}} u(\theta_{\mathbf{k}} - \theta_{\mathbf{p}}) u(\theta_{\mathbf{p}} - \theta_{\mathbf{k}'}) = g \left(u_0^2 + \frac{1}{2} \sum_{m=1}^{\infty} u_m^2 \cos m(\theta - \theta') \right) \tag{39}$$

where we have reverted to the notation $\theta = \theta_{\mathbf{k}}, \theta' = \theta_{\mathbf{k}'}$. In the second term of (37), the $\delta_{\mathbf{p},-\mathbf{p}'}$ turns out to be subleading, while the other allows independent sums over $\mathbf{p}, \mathbf{p}'$. This means that only u_0 contributes to this term, (other avrerage to zero upon summing over all angles) which produces

$$-\sum_{\mathbf{p}\mathbf{p}'} u(\theta_{\mathbf{k}} - \theta_{\mathbf{p}}) u(\theta_{\mathbf{p}'} - \theta_{\mathbf{k}'}) = g^2 u_0^2 \tag{40}$$

Feeding this into full expression for this contribution to the particle-hole diagram, we find it to be

$$\frac{dV_{\alpha\beta\gamma\delta}}{dt}\Big|_{\text{Leading}} = \frac{g'}{g} \Delta \ln 2 \sum_{\mathbf{k}\mathbf{k}'} \left(\sum_{m=1}^{\infty} u_m^2 \cos m(\theta - \theta') \right) \phi_\alpha^*(\mathbf{k}) \phi_\beta^*(\mathbf{k}') \phi_\gamma(\mathbf{k}') \phi_\delta(\mathbf{k}) \tag{41}$$

Notice that the result is still of the Fermi liquid form. In other words the couplings $V_{\alpha\beta\gamma\delta}$ which were written in terms of Landau parameters u_m, flow into renormalized coupling once again expressible in terms of renormalized Landau parameters. By comparing the two sides, we see each u_m flows independently of the others as per

$$\frac{du_m}{dt} = -e^{-t} (\ln 2) u_m^2 \qquad m \neq 0 \tag{42}$$

The above equation can be written in a more physically transparent form by using a rescaled variable (for $m \neq 0$ only)

$$\tilde{u}_m = e^{-t} u_m \tag{43}$$

in terms of which the flow equation becomes

$$\frac{d\tilde{u}_m}{dt} = -\tilde{u}_m - (\ln 2) \tilde{u}_m^2 \equiv \beta(\tilde{u}_m) \tag{44}$$

where the last is a definition of the β-function.

The reason u_o does not flow is that the corresponding interaction commutes with the one-body "kinetic" part, and therefore does not suffer quantum fluctuations.

This is the answer at large g. We have dropped subleading contributions of the following type:

$$\left.\frac{dV_{\alpha\beta\gamma\delta}}{dt}\right|_{\rm sub} = -g'\Delta^2 \sum_{\nu=-\Lambda}^{0} \frac{1}{\Lambda+|\varepsilon_\nu|}\Bigg[\sum_{\mathbf{k}\mathbf{k}'}\sum_{\mathbf{p}\mathbf{p}'} u(\theta_{\mathbf{k}}-\theta_{\mathbf{p}})u(\theta_{\mathbf{p}'}-\theta_{\mathbf{k}'}) \times \phi^*_\alpha(\mathbf{p})\phi^*_\beta(\mathbf{k}')\phi_\gamma(\mathbf{k}')\phi_\delta(\mathbf{k})\phi^*_\mu(\mathbf{k})\phi^*_\nu(\mathbf{p}')\phi_\nu(\mathbf{p})\phi_\mu(\mathbf{p}')\Bigg] \quad (45)$$

Note that the momentum labels of ϕ^*_α and ϕ^*_μ have been exchanged compared to (36) and there is a minus sign, both coming from the antisymmetrization of (29). Once again we ensemble average the internal μ,ν sum, the wavefunction part of which gives

$$\left\langle \phi^*_\mu(\mathbf{k})\phi^*_\nu(\mathbf{p}')\phi_\nu(\mathbf{p})\phi_\mu(\mathbf{p}')\right\rangle = \frac{\delta_{\mathbf{k}\mathbf{p}'}\delta_{\mathbf{p}\mathbf{p}'}}{g^2} - \frac{\delta_{\mathbf{k}\mathbf{p}}+\delta_{\mathbf{k},-\mathbf{p}'}\delta_{\mathbf{p},-\mathbf{p}'}}{g^3} \quad (46)$$

It is clear that there is an extra momentum restriction in each term compared to (37), which means that one can no longer sum freely over $\mathbf{p}$ to get the factor of g in (39), or the factor of g^2 in (40). Therefore this contribution will be down by $1/g$ compared to that of (36).

Turning now to the Cooper diagrams, the internal lines are once again forced to have the same momentum labels as the external lines by the Fermi liquid vertex, therefore they do not make any leading contributions.

The general rule is that whenever a momentum label corresponding to an internal line in the diagram (here μ and ν) is forced to become equal to a momentum label corresponding to an external disorder label (here α,β,γ, or δ), the diagram is down by $1/g$, exactly as in the $1/N$ expansion. The fact that $1/g$ plays the role of $1/N$ was first realized by Murthy and Shankar [24]. Not only did this mean that the one loop flow of Murthy and Mathur was exact, it meant the disorder-interaction problem of the chaotic dot could be solved exactly in the large g limit. It is the only known case where the problem of disorder and interactions [26, 27, 28] can be handled exactly.

From (44) it can be seen that positive initial values of $\tilde{u}_m$ (which are equal to initial values of u_m inherited from the bulk) are driven to the fixed point at $\tilde{u}_m = 0$, as are negative initial values as long as $u_m(t=0) \geq u^*_m = -1/\ln 2$. Thus, the Fermi liquid parameters are *irrelevant* for this range of starting values. Recall that setting all $u_m = 0$ for $m \neq 0$ results in the Universal Hamiltonian. Thus, the range $u_m \geq u^*_m$ is the basin of attraction of the Universal Hamiltonian. On the other hand, for $u_m(t=0) \leq u^*_m$ results in a runaway flow towards large negative values of u_m, signalling a transition to a phase not perturbatively connected to the Universal Hamiltonian.

Since we have a large N theory here (with $N = g$), the one-loop flow and the new fixed point at strong coupling are parts of the final theory.[1] What is

[1] However the exact location of the critical point cannot be predicted, as pointed out to us by Professor Peet Brower. The reason is that the Landau couplings u_m are defined at a scale E_L much higher than E_T (but much smaller than E_F) and their flow till we come down to E_T, where our analysis begins, is not within

the nature of the state for $u_m(t=0) \leq u_m^*$? The formalism and techniques needed to describe that are beyond what was developed in these lectures, which has focused on the RG. Suffice it to say that it is possible to write the partition function in terms of a new collective field $\boldsymbol{\sigma}$ (which depends on all the particles) and that the action $S(\boldsymbol{\sigma})$ has a factor g in front of it, allowing us to evaluate the integral by saddle point(in the limit $g \to \infty$), to confidently predict the strong coupling phase and many of its properties. Our expectations based on the large g analysis have been amply confirmed by detailed numerical studies [25]. For now I will briefly describe the new phenomenon in qualitative terms for readers not accustomed to these ideas and give some references for those who are.

In the strong coupling region $\boldsymbol{\sigma}$ acquires an expectation value in the ground state. The dynamics of the fermions is affected by this variable in many ways: quasi-particle widths become broad very quickly above the Fermi energy, the second difference Δ_2 has occasionally very large values and can even be negative[2], and the system behaves like one with broken time-reversal symmetry if m is odd.

Long ago Pomeranchuk [29] found that if a Landau function of a pure system exceeded a certain value, the fermi surface underwent a shape transformation from a circle to an non-rotationally invariant form. Recently this transition has received a lot of attention [30, 31] The transition in question is a disordered version of the same. Details are given in [24, 25].

Details aside, there is another very interesting point: even if the coupling does not take us over to the strong-coupling phase, we can see vestiges of the critical point u_m^* and associated critical phenomena. This is a general feature of many *quantum critical points* [32], i.e., points like u_m^* where as a variable in a hamiltonian is varied, the system enters a new phase (in contrast to transitions wherein temperature T is the control parameter).

Figure 6 shows what happens in a generic situation. On the x-axis a variable (u_m in our case) along which the quantum phase transition occurs. Along y is measured a new variable, usually temperature T. Let us consider that case first. If we move from right to left at some value of T, we will first encounter physics of the weak-coupling phase determined by the weak-coupling fixed point at the origin. Then we cross into the *critical fan* (delineated by the V-shaped dotted lines), where the physics controlled by the quantum critical point. In other words we can tell there is a critical point on the x-axis

the regime we can control. In other words we can locate u^* in terms of what couplings we begin with, but these are the Landau parameters renormalized in a nonuniversal way as we come down from E_L to E_T.

[2] How can the cost of adding one particle be negative (after removing the charging energy)? The answer is that adding a new particle sometimes lowers the energy of the collective variable which has a life of its own. However, if we turn a blind eye to it and attribute all the energy to the single particle excitations, Δ_2 can be negative.

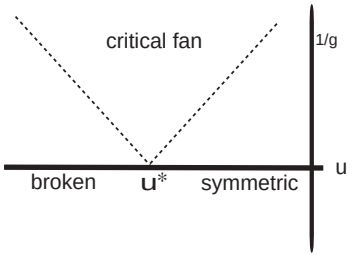

Fig. 6. The generic phase diagram for a second-order quantum phase transition. The *horizontal* axis represents the coupling constant which can be tuned to take one across the transition. The *vertical* axis is usually the temperature in bulk quantum systems, but is $1/g$ here, since in our system one of the roles played by g is that of the inverse temperature

without actually traversing it. As we move further to the left, we reach the strongly-coupled symmetry-broken phase, with a non-zero order parameter.

It can be shown that in our problem, $1/g$ plays the role of T. One way to see is this that in any large N theory N stands in front of the action when written in terms of the collective variable. That is true in this case well for g. (Here g also enters at a subdominant level inside the action, which makes it hard to predict the exact shape of the critical fan. The bottom line is that we can see the critical point at finite $1/g$. In addition one can also raise temperature or bias voltage to see the critical fan.

Subsequent work has shown, in more familiar examples that Landau interactions, that the general picture depicted here is true in the large g limit: upon adding sufficiently strong interactions the Universal Hamiltonian gives way to other descriptions with broken symmetry [33, 34].[3]

Acknowledgement

It is a pleasure to thank the organizers of this school especially Professors Dieter Heiss, Nithaya Chetty and Hendrik Geyer for their stupendous hospitality that made all of decide to revisit South Africa as soon as possible.

[3] The only result that is not exact in the large g limit is the critical value u_m^* since the input value of u_m at E_T is related in a non-universal way to the Landau parameter. In other words, the Landau coupling u_m settles down to its fixed point value at an energy scale that is much larger than E_T. We do not know how it flows as we reach energies inside E_T wherein our RMT assumptions are valid.

References

1. M. E. Fisher *Critical Phenomena*, F. W. J. Hahne, Editor, Lecture Notes Number 186, Springer-Verlag, Berlin, (1983). These notes are from the Engelbrecht Summer School of 1983! 4
2. R. Shankar, *Physica* **A177**, 530 (1991); R.Shankar, Rev. Mod. Phys. **66**, 129 (1994). 5, 6, 11, 12, 14
3. L. D. Landau, Sov. Phys. JETP **3**, 920 (1956); Sov. Phys. JETP **5**, 101 (1957). 8
4. A. A. Abrikosov, L. P. Gorkov, and I. E. Dzyaloshinski, *Methods of Quantum Field Theory in Statistical Physics*, Dover Publications, New York, 1963. 8
5. D. J. Gross and A. Neveu, Phys. Rev.**D10**, 3235 (1974). 10
6. P. Polchinski In: *TASI Elementary Particle Physics*, ed by J. Polchinski and J.Harvey (World Scientific, 1992.) 12
7. For recent reviews, see, T. Guhr, A. Müller-Groeling, and H. A. Weidenmüller, Phys. Rep. **299**, 189 (1998); Y. Alhassid, Rev. Mod. Phys. **72**, 895 (2000); A. D. Mirlin, Phys. Rep. **326**, 259 (2000). 12, 13
8. M. L. Mehta, *Random Matrices*, Academic Press, San Diego, 1991. K. B. Efetov, Adv. Phys. **32**, 53 (1983); B. L. Al'tshuler ad B. I. Shklovskii, Sov. Phys. JETP **64**, 127 (1986). 13
9. I. L. Aleiner, P. W. Brouwer, and L. I. Glazman, Phys. Rep. **358**, 309 (2002), and references therein; Y. Oreg, P. W. Brouwer, X. Waintal, and B. I. Halperin, cond-mat/0109541, "Spin, spin-orbit, and electron-electron interactions in mesoscopic systems". 12, 13, 14
10. R. A. Jalabert, A. D. Stone, and Y. Alhassid, Phys. Rev. Lett.**68**, 3468 (1992). 13
11. A. M. Chang, H. U. Baranger, L. N. Pfeiffer, K. W. West, and T. Y. Chang, Phys. Rev. Lett.**76**, 1695 (1996); J. A. Folk, S. R. Patel, S. F. Godjin, A. G. Huibers, S. M. Cronenwett, and C. M. Marcus, Phys. Rev. Lett.**76**, 1699 (1996). 13
12. S. R. Patel, D. R. Stewart, C. M. Marcus, M. Gokcedag, Y. Alhassid, A. D. Stone, C. I. Dururoz, and J. S. Harris, Phys. Rev. Lett.**81**, 5900 (1998). 13
13. D. V. Averin and K. K. Likharev, in *Mesoscopic Phenomena in Solids*, edited by B. L. Altshuler, P. A. Lee, and R. Webb (Elsevier, Amsterdam, 1991); C. W. J. Beenakker, Phys. Rev. **B44**, 1646 (1991). 13
14. A. V. Andreev and A. Kamenev, Phys. Rev. Lett.**81**, 3199 (1998), P. W. Brouwer, Y. Oreg, and B. I. Halperin, Phys. Rev. **B60**, R13977 (1999), H. U. Baranger, D. Ullmo, and L. I. Glazman, Phys. Rev. **B61**, R2425 (2000). 13, 14
15. I. L. Kurland, I. L. Aleiner, and B. L. Al'tshuler, Phys. Rev. **B 62**, 14886 (2000). 13, 14
16. O. Prus, A. Auerbach, Y. Aloni, U. Sivan, and R. Berkovits, Phys. Rev. **B54**, R14289 (1996), R. Berkovits, Phys. Rev. Lett. **81**, 2128 (1998), A. Cohen, K. Richter, and R. Berkovits, Phys. Rev. **B 60**, 2536 (1999), P. N. Walker, G. Montambaux, and Y. Gefen, *ibid*, 2541 (1999), S. Levit and D. Orgad, Phys. Rev. **B 60**, 5549 (1999), D. Ullmo and H. U. Baranger, Phys. Rev. **B 64**, 245324 (2001), V. Belinicher, E. Ginossar, and S. Levit, cond-mat/0109005, Y. Alhassid and S. Malhotra, cond-mat/0202453. 13, 14
17. U. Sivan *et al*, Phys. Rev. Lett. **77**, 1123 (1996); S. R. Patel *et al*, Phys. Rev. Lett. **80**, 4522 (1998). 13
18. F. Simmel *et al*, Phys. Rev. **B 59**, 10441 (1999); D. Abusch-Magder *et al*, Physica E **6**, 382 (2000). 13, 14
19. Y. M. Blanter, A. D. Mirlin, and B. A. Muzykantskii, Phys. Rev. Lett.**78**, 2449 (1997), R. O. Vallejos, C. H. Lewenkopf, and E. R. Mucciolo, Phys. Rev. Lett.**81**, 677 (1998). 14

20. S. Adam, P. W. Brouwer, J. P. Sethna, and X. Waintal, cond-mat/0203002. 14
21. J. B. French and S. S. M. Wong, Phys. Lett. **33B**, 447 (1970), " O. Bohigas and J. Flores, Phys. Lett. **34B**, 261 (1971), Y. Alhassid, Ph. Jacquod, and A. Wobst, Phys. Rev. **B 61**, 13357 (2000), Physica E **9**, 393 (2001), Y. Alhassid and A. Wobst, Phys. Rev. **B 65**, 041304 (2002). 14
22. A. Tscherich and K. B. Efetov, Phys. Rev.**E62**, 2042 (2000). 14
23. G. Murthy and H. Mathur, Phys. Rev. Lett. **89**, 126804 (2002). 14
24. G. Murthy and R. Shankar, Phys. Rev. Lett.**90**, 066801 (2003). 16, 20, 21
25. G Murthy, R. Shankar, Damir Herman, and Harsh Mathur, cond-mat 0306529, Phys. Rev. **B69**, 075321, (2004). 18, 21
26. A. M. Finkel'shtein, *Sov. Phys. JETP* **57**, 97 (1983); C. Castellani, C. Di Castro, P. A. Lee, and M. Ma, Phys. Rev. **B30**, 527 (1984). 20
27. For a review of the theory, see, D. Belitz and T. R. Kirkpatrick, Rev. Mod. Phys. **66**, 261 (1994). 20
28. D. Belitz and T. R. Kirkpatrick, Phys. Rev. **B53**, 14364 (1996); C. de C. Chamon and E. Mucciolo, Phys. Rev. Lett.**85**, 5607 (2000); C. Nayak and X. Yang, cond-mat/0302503. 20
29. I. I. Pomeranchuk, Sov. Phys. JETP **8**, 361 (1958). 21
30. C. M. Varma, Phys. Rev. Lett. **83**, 3538 (1999). 21
31. V. Oganesyan, S. A. Kivelson, and E. Fradkin, Phys. Rev. **B 64**, 195109 (2001). 21
32. S. Chakravarty, B. I. Halperin, and D. R. Nelson, Phys. Rev. Lett. **60**, 1057 (1988); Phys. Rev. **B 39**, 2344 (1989); For a detailed treatment of the generality of the phenomenon, see, S. Sachdev, *Quantum Phase Transitions*, Cambridge University Press, Cambridge 1999. 21
33. Shaffique Adam, Piet W. Brouwer, and Prashant Sharma Phys. Rev. **B** 68, 241311 (2003). 22
34. G. Murthy, Random Matrix Crossovers and Quantum Critical Crossovers for Interacting Electrons in Quantum Dotscond-mat-0406029. 22

Semiconductor Few-Electron Quantum Dots as Spin Qubits

J.M. Elzerman[1,2], R. Hanson[1], L.H.W. van Beveren[1,2], S. Tarucha[2,3], L.M.K. Vandersypen[1], and L.P. Kouwenhoven[1,2]

[1] Kavli Institute of Nanoscience Delft, PO Box 5046, 2600 GA Delft, The Netherlands
elzerman@qt.tn.tudelft.nl
[2] ERATO Mesoscopic Correlation Project, University of Tokyo, Bunkyo-ku, Tokyo 113-0033, Japan
[3] NTT Basic Research Laboratories, Atsugi-shi, Kanagawa 243-0129, Japan

The spin of an electron placed in a magnetic field provides a natural two-level system suitable as a qubit in a quantum computer [1]. In this work, we describe the experimental steps we have taken towards using a single electron spin, trapped in a semiconductor quantum dot, as such a spin qubit [2].

The outline is as follows. Section 1 serves as an introduction into quantum computing and quantum dots. Section 2 describes the development of the "hardware" for the spin qubit: a device consisting of two coupled quantum dots that can be filled with one electron (spin) each, and flanked by two quantum point contacts (QPCs). The system can be probed in two different ways, either by performing conventional measurements of transport through one dot or two dots in series, or by using a QPC to measure changes in the (average) charge on each of the two dots. This versatility has proven to be very useful, and the type of device shown in this section was used for all subsequent experiments.

In Sect. 3, it is shown that we can determine all relevant parameters of a quantum dot even when it is coupled very weakly to only one reservoir. In this regime, inaccessible to conventional transport experiments, we use a QPC charge detector to determine the tunnel rate between the dot and the reservoir. By measuring changes in the effective tunnel rate, we can determine the excited states of the dot.

In Sect. 4, the QPC as a charge detector is pushed to a faster regime ($\sim$100 kHz), to detect single electron tunnel events in real time. We also determine the dominant contributions to the noise, and estimate the ultimate speed and sensitivity that could be achieved with this very simple method of charge detection.

In Sect. 5, we develop a technique to perform single-shot measurement of the spin orientation of an individual electron in a quantum dot. This is done by

J.M. Elzerman et al.: *Semiconductor Few-Electron Quantum Dots as Spin Qubits*, Lect. Notes Phys. **667**, 25–95 (2005)
www.springerlink.com

combining fast QPC charge detection with "spin-to-charge conversion". This fully electrical technique to read out a spin qubit is then used to determine the relaxation time of the single spin, giving a value of 0.85 ms at a magnetic field of 8 Tesla.

Finally, Sect. 6 puts the results in perspective, arriving at a realistic path towards the experimental demonstration of single- and two-qubit gates and the creation of entanglement of spins in quantum dot systems.

1 Introduction

This section gives a brief introduction into quantum computing, continuing with a description of semiconductor quantum dots that covers their fabrication as well as their electronic behavior. We also describe our experimental setup for performing low-temperature transport experiments to probe such quantum dots.

1.1 Quantum Computing

More than three quarters of a century after its birth, quantum mechanics remains in many ways a peculiar theory [3]. It describes many physical effects and properties with great accuracy, but uses unfamiliar concepts like superposition, entanglement and projection, that seem to have no relation with the everyday world around us. The interpretation of these concepts can still cause controversy.

The inherent strangeness of quantum mechanics already emerges in the simplest case: a quantum two-level system. Unlike a classical two-level system, which is always either in state 0 or in state 1, a quantum two-level system can just as well be in a *superposition* of states $|0\rangle$ and $|1\rangle$. It is, in some sense, in both states at the same time.

Even more exotic states can occur when two such quantum two-level systems interact: the two systems can become *entangled*. Even if we know the complete state of the system as a whole, for example $(|01\rangle - |10\rangle)/\sqrt{2}$, which tells us all there is to know about it, we cannot know the state of the two subsystems individually. In fact, the subsystems do not even have a definite state! Due to this strong connection between the two systems, a measurement made on one influences the state of the other, even though it may be arbitrarily far away. Such spooky non-local correlations enable effects like "quantum teleportation" [4, 5].

Finally, the concept of measurement in quantum mechanics is rather special. The evolution of an isolated quantum system is deterministic, as it is governed by a first order differential equation – the Schrödinger equation. However, coupling the quantum system to a measurement apparatus forces it into one of the possible measurement eigenstates in an apparently non-deterministic way: the particular measurement outcome is random, only the

probability for each outcome can be determined [3]. The question of what exactly constitutes a measurement is still not fully resolved [6].

These intriguing quantum effects pose fundamental questions about the nature of the world we live in. The goal of science is to explore these questions. At the same time, this also serves a more opportunistic purpose, since it might allow us to actually *use* the unique features of quantum mechanics to do something that is impossible from the classical point of view.

And there are still many things that we cannot do classically. A good example is prime-factoring of large integers: it is easy to take two prime numbers and compute their product. However, it is difficult to take a large integer and find its prime factors. The time it takes any classical computer to solve this problem grows exponentially with the number of digits. By making the integer large enough, it becomes essentially impossible for any classical computer to find the answer within a reasonable time – such as the lifetime of the universe. This fact is used in most forms of cryptography nowadays [7].

In 1982, Richard Feynman speculated [8] that efficient algorithms to solve such hard computational problems might be found by making use of the unique features of quantum systems, such as entanglement. He envisioned a set of quantum two-level systems that are quantum mechanically coupled to each other, allowing the system as a whole to be brought into a superposition of different states. By controlling the Hamiltonian of the system and therefore its time-evolution, a computation might be performed in fewer steps than is possible classically. Essentially, such a quantum computer could take many computational steps at once; this is known as "quantum parallelism".

A simplified view of the difference between a classical and a quantum computer is shown in Fig. 1. A one-bit classical computer is a machine that

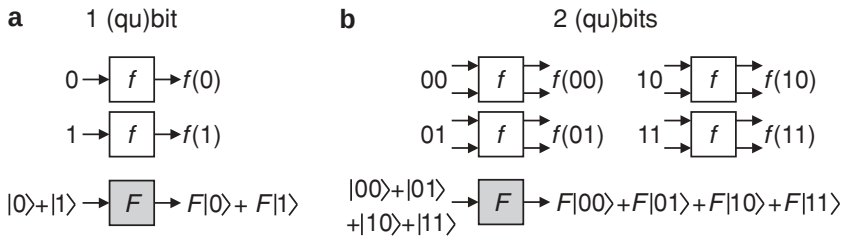

Fig. 1. Difference between a classical and a quantum computer. (**a**) To determine the function f for the two possible input states 0 and 1, a one-bit classical computer needs to evaluate the function twice, once for every input state. In contrast, a one-qubit quantum computer can have a superposition of $|0\rangle$ and $|1\rangle$ as an input, to end up in a superposition of the two output values, $F|0\rangle$ and $F|1\rangle$. It has taken only half the number of steps as its classical counterpart. (**b**) Similarly, a two-qubit quantum computer needs only a quarter of the number of steps that are required classically. The computing power of a quantum computer scales exponentially with the number of qubits, for a classical computer the scaling is only linear

takes one input value, 0 or 1, and computes the corresponding output value, $f(0)$ or $f(1)$. A quantum computer with one quantum bit (or "qubit") could take as an input value a superposition of $|0\rangle$ and $|1\rangle$, and due to the linearity of quantum mechanics the output would be a superposition of $F|0\rangle$ and $F|1\rangle$. So, in a sense it has performed two calculations in a single step. For a two-qubit system, the gain becomes even more significant: now the input can be a superposition of four states, so the quantum computer can perform four calculations in one step. In fact, it can be proved [9] that the computing power of a quantum computer scales exponentially with the number of qubits, whereas this scaling is only linear for a classical computer. Therefore, a large enough quantum computer can outperform any classical computer.

It might appear that a fundamental problem has been overlooked: according to quantum mechanics, a superposition of possible measurement outcomes can only exist before it is measured, and the measurement gives only one actual outcome. The exponential computing power thus appears inaccessible. However, by using carefully tailored quantum algorithms, an exponential speed-up can be achieved for some problems such as factoring integers [10] or simulating a quantum system [11]. For other tasks, such as searching a database, a quadratic speed-up is possible [12]. Using such quantum algorithms, a quantum computer can indeed be faster than a classical one.

Another fundamental problem is the interaction of the quantum system with the (uncontrolled) environment, which inevitably disturbs the desired quantum evolution. This process, known as "decoherence", results in errors in the computation. Additional errors are introduced by imperfections in the quantum operations that are applied. All these errors propagate, and after some time the state of the computer will be significantly different from what it should be. It would seem that this prohibits any long computations, making it impossible for a quantum computer to use its exponential power for a non-trivial task. Fortunately, it has been shown that methods to detect and correct any errors exist [13, 14], keeping the computation on track. Of course, such methods only help if the error rate is small enough, since otherwise the correction operations create more errors than they remove. This sets a so-called "accuracy threshold" [15, 16], which is currently believed to be around 10^{-4}. If the error per quantum operation is smaller than this threshold, any errors can be corrected and an arbitrarily long computation is possible.

Due to the development of quantum algorithms and error correction, quantum computation is feasible from a theoretical point of view. The challenge is building an actual quantum computer with a sufficiently large number of coupled qubits. Probably, more than a hundred qubits will be required for useful computations, but a system of about thirty qubits might already be able to perform valuable simulations of quantum systems.

1.2 Implementations

A number of features are required for building an actual quantum computer [17]:

1. A scalable physical system with well-characterized qubits
2. A "universal" set of quantum gates to implement any algorithm
3. The ability to initialize the qubits to a known pure state
4. A qubit-specific measurement capability
5. Decoherence times much longer than the gate operation time

Many systems can be found which satisfy some of these criteria, but it is very hard to find a system that satisfies all of them. Essentially, we have to reconcile the conflicting demands of good access to the quantum system (in order to perform fast and reliable operations or measurements) with sufficient isolation from the environment (for long coherence times). Current state-of-the-art is a seven-bit quantum computer that has factored the number 15 into its prime factors 3 and 5, in fewer steps than is possible classically [18]. This was done using an ensemble of molecules in liquid solution, with seven nuclear spins in each molecule acting as the seven qubits. These could be controlled and read out using nuclear magnetic resonance (NMR) techniques. Although this experiment constitutes an important proof-of-principle for quantum computing, practical limitations do not allow the NMR approach to be scaled up to more than about ten qubits.

Therefore, many other implementations are currently being studied [19]. For instance, trapped ions have been used to demonstrate a universal set of one- and two-qubit operations, an elementary quantum algorithm, as well as entanglement of up to three qubits and quantum teleportation [19]. Typically, microscopic systems such as atoms or ions have excellent coherence properties, but are not easily accessible or scalable – on the other hand, larger systems such as solid-state devices, which can be accessed and scaled more easily, usually lack long decoherence times. A solid-state device with a long decoherence time would represent the best of both worlds. Such a system could be provided by the spin of an electron trapped in a quantum dot: a *spin qubit*.

1.3 The Spin Qubit

Our programme to build a solid-state qubit follows the proposal by Loss and DiVincenzo [2]. This describes a quantum two-level system defined by the spin orientation of a single electron trapped in a semiconductor quantum dot. The electron spin can point "up" or "down" with respect to an external magnetic field. These eigenstates, $|\uparrow\rangle$ and $|\downarrow\rangle$, correspond to the two basis states of the qubit.

The quantum dot that holds the electron spin is defined by applying negative voltages to metal surface electrodes ("gates") on top of a semiconductor (GaAs/AlGaAs) heterostructure (see Fig. 2). Such gated quantum dots are

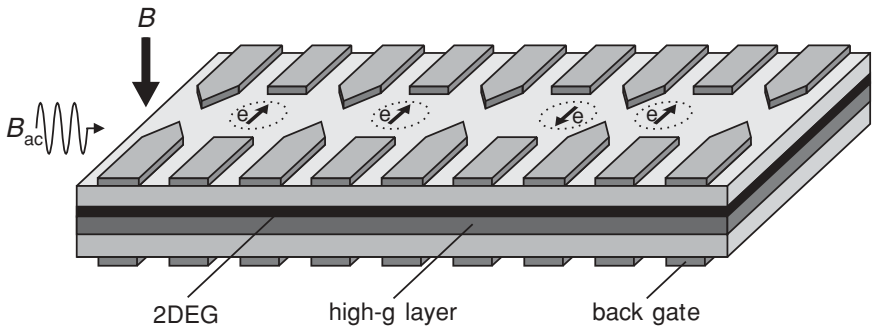

Fig. 2. Schematic picture of the spin qubit as proposed by Loss and DiVincenzo [2]. The array of metal electrodes on top of a semiconductor heterostructure, containing a two-dimensional electron gas (2DEG) below the surface, defines a number of quantum dots (*dotted circles*), each holding a single electron spin (*arrow*). A magnetic field, B, induces a Zeeman splitting between the spin-up and spin-down states of each electron spin. The spin state is controlled either via an oscillating magnetic field, B_{ac} (on resonance with the Zeeman splitting), or via an oscillating electric field created with the back gates, which can pull the electron wavefunction into a layer with a large g-factor. Coupling between two spins is controlled by changing the voltage on the electrodes between the two dots (Adapted from [2])

very controllable and versatile systems, which can be manipulated and probed electrically. Increasing the number of dots is straightforward, by simply adding more electrodes. Tuning all these gate voltages allows control of the number of electrons trapped on each dot, as well as the tunnel coupling between the dots. With the external magnetic field, B, we can tune the Zeeman splitting, $\Delta E_Z = g\mu_B B$, where $g \approx -0.44$ is the g-factor of GaAs, and $\mu_B = 9.27{\times}10^{-24}$ J/T is the Bohr magneton. In this way, we can control the energy levels of the qubit.

To perform single-qubit operations, different techniques are available. We can apply a microwave magnetic field on resonance with the Zeeman splitting, i.e. with a frequency $f = \Delta E_Z/h$, where h is Planck's constant. The oscillating magnetic component perpendicular to the static magnetic field B results in a spin nutation. By applying the oscillating field for a fixed duration, a superposition of $|\uparrow\rangle$ and $|\downarrow\rangle$ can be created. This magnetic technique is known as electron spin resonance (ESR).

A completely electrical alternative might be the emerging technique of g-tensor modulation [20]. In this scheme, an oscillating electric field is created by modulating the voltage applied to a (back) gate. The electric field does not couple to the spin directly, but it can push or pull the electron wavefunction somewhat into another semiconductor layer with a different g-factor. This procedure modulates the effective g-tensor felt by the electron spin. If the modulation frequency is resonant with the Zeeman splitting, the required spin nutation results and superpositions of spin states can again be created.

Two-qubit operations can be carried out purely electrically, by varying the gate voltages that control the potential barrier between two dots. It has been shown [2] that the system of two electron spins on neighboring dots, coupled via a tunnel barrier, can be mapped onto the Heisenberg exchange Hamiltonian $H = J\mathbf{S_1}\cdot\mathbf{S_2}$. This Hamiltonian describes an indirect interaction between the two spins, $\mathbf{S_1}$ and $\mathbf{S_2}$, mediated by the exchange interaction, J, which depends on the wavefunction overlap of the electrons. By lowering the tunnel barrier for some time and then raising it again, the effective spin-spin interaction is temporarily turned on. In this way, the two electron spins can be swapped or even entangled. Together with arbitrary single-spin rotations, the exchange interaction can be used to construct a universal set of quantum gates [2].

A last crucial ingredient is a method to read out the state of the spin qubit. This implies measuring the spin orientation of a single electron – a daunting task, since the electron spin magnetic moment is exceedingly small. Therefore, an indirect spin measurement is proposed [2]. First the spin orientation of the electron is correlated with its position, via "spin-to-charge conversion". Then an electrometer is used to measure the position of the charge, thereby revealing its spin. In this way, the problem of measuring the spin orientation has been replaced by the much easier measurement of charge.

The essential advantage of using the electron's spin degree of freedom to encode a qubit, lies in the fact that the spin is disturbed only weakly by the environment. The main source of spin decoherence and relaxation is predicted to be the phonon bath, which is coupled to the spin via the (weak) spin-orbit interaction [21, 22, 23]. In addition, fluctuations in the nuclear-spin configuration couple to the electron spin via the (even weaker) hyperfine coupling [21, 24]. In contrast, the electron's charge degree of freedom is much easier to manipulate and read out, but it is coupled via the strong Coulomb interaction to charge fluctuations, which are the source of the ubiquitous $1/f$ noise in the "dirty" semiconductor environment. This leads to typical charge decoherence times of a few nanoseconds [25, 26]. The spin decoherence and relaxation times are predicted to be about four orders of magnitude longer [22].

Finally, it should be stressed that our efforts to create a spin qubit are not purely application-driven. Aside from the search for a spin quantum computer, many interesting questions await exploration. If we have the ability to (coherently) control and read out a single electron spin in a quantum dot, this spin could be used as a local probe of the semiconductor environment. This could shed light for instance on many details of the spin-orbit interaction or the hyperfine coupling.

1.4 Quantum Dots

In this paragraph, the properties of semiconductor quantum dots are described in more detail [27]. In essence, a quantum dot is simply a small box that can be filled with electrons. The box is coupled via tunnel barriers to a source

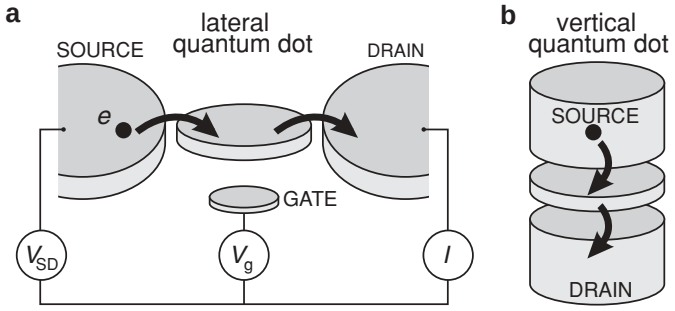

Fig. 3. Schematic picture of a quantum dot in a lateral (**a**) and a vertical (**b**) geometry. The quantum dot (represented by a disk) is connected to source and drain contacts via tunnel barriers, allowing the current through the device, I, to be measured in response to a bias voltage, V_{SD} and a gate voltage, V_g

and drain reservoir, with which particles can be exchanged (see Fig. 3). By attaching current and voltage probes to these reservoirs, we can measure the electronic properties of the dot. The box is also coupled capacitively to one or more "gate" electrodes, which can be used to tune the electrostatic potential of the dot with respect to the reservoirs. When the size of the box is comparable to the wavelength of the electrons that occupy it, the system exhibits a discrete energy spectrum, resembling that of an atom. As a result, quantum dots behave in many ways as *artificial atoms*.

Because a quantum dot is such a general kind of system, there exist quantum dots of many different sizes and materials: for instance single molecules trapped between electrodes, metallic or superconducting nanoparticles, self-assembled quantum dots, semiconductor lateral or vertical dots, and even semiconducting nanowires or carbon nanotubes between closely spaced electrodes. In this work, we focus on lateral (gated) semiconductor quantum dots. These lateral devices allow all relevant parameters to be controlled in the fabrication process, or tuned in situ.

Fabrication of gated quantum dots starts with a semiconductor heterostructure, a sandwich of different layers of semiconducting material (see Fig. 4a). These layers, in our case GaAs and AlGaAs, are grown on top of each other using molecular beam epitaxy (MBE), resulting in very clean crystals. By doping the n-AlGaAs layer with Si, free electrons are introduced. These accumulate at the interface between GaAs and AlGaAs, typically 100 nm below the surface, forming a two-dimensional electron gas (2DEG) – a thin (10 nm) sheet of electrons that can only move along the interface. The 2DEG can have a high mobility and relatively low electron density (typically 10^5–10^6 cm^2/Vs and $\sim 3 \times 10^{15}$ m^{-2}, respectively). The low electron density results in a large Fermi wavelength ($\sim$40 nm) and a large screening length, which allows us to locally deplete the 2DEG with an electric field. This electric field is created

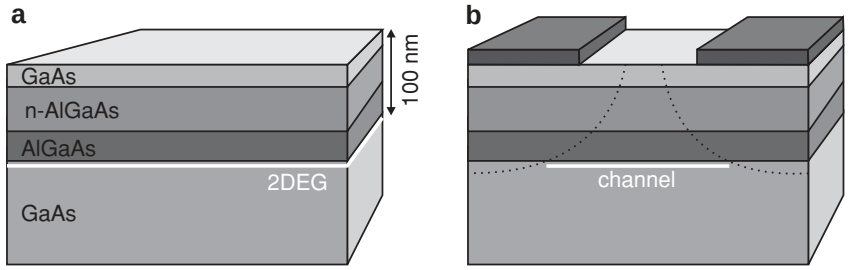

Fig. 4. Confining electrons in a semiconductor. (**a**) Semiconductor heterostructure containing a 2DEG (indicated in *white*) approximately 100 nm below the surface, at the interface between GaAs and AlGaAs. The electrons in the 2DEG result from Si donors in the n-AlGaAs layer. (The thickness of the different layers is not to scale.) (**b**) By applying negative voltages to the metal electrodes on the surface of the heterostructure, the underlying 2DEG can be locally depleted. In this way, electrons can be confined to one or even zero dimensions

by applying (negative) voltages to metal gate electrodes on top of the heterostructure (Fig. 4b).

To fabricate these electrodes, we first spin a layer of organic resists (typically poly-methyl-methacrylate, PMMA) on the heterostructure surface (Fig. 5a). Then the gate pattern is defined by writing with a focused electron beam in the electron-sensitive resist. This locally breaks up the polymer chains, so that the exposed parts can be removed by a developer. (Note that there is some undercut of the bottom resist layer, caused by electrons backscattering from the heterostructure during exposure to the electron beam.) In the next step, metal is evaporated, which only makes contact to the heterostructure at the places where the resist has been exposed and removed. In our devices, the metal gates consist of a thin (5 nm) "sticking" layer of titanium, with a 30 nm layer of gold on top. In the final so-called "lift-off" step, the remaining resist is removed with acetone. Now metal electrodes are left at the places that were exposed to the electron beam.

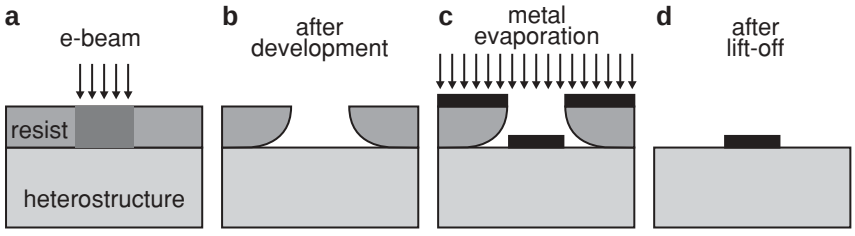

Fig. 5. Fabrication of metal electrodes on the surface of the heterostructure. (**a**) Writing a pattern in the resist layer with an electron beam. (**b**) After developing, the resist has been locally removed. (**c**) Evaporating metal. (**d**) After lift-off, a metal electrode remains

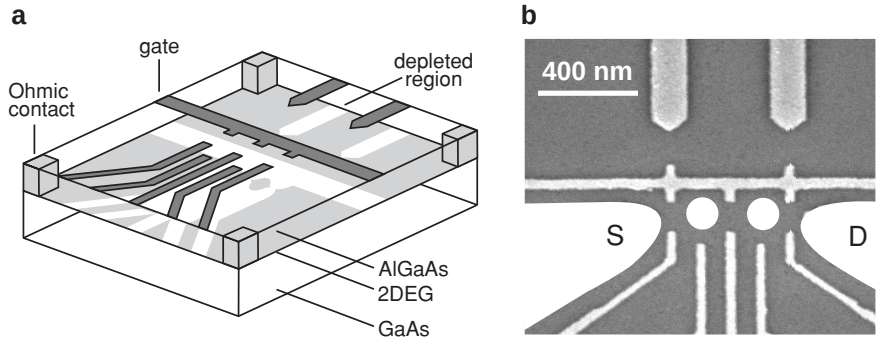

Fig. 6. Lateral quantum dot device defined by metal surface electrodes. (**a**) Schematic view of a device. Negative voltages applied to metal gate electrodes (*dark gray*) lead to depleted regions (*white*) in the 2DEG (*light gray*). Ohmic contacts (*light gray* columns) enable bonding wires (not shown) to make electrical contact to the 2DEG reservoirs. (**b**) Scanning electron microscope image of an actual device, showing the gate electrodes (*light gray*) on top of the surface (*dark gray*). The two *white dots* indicate two quantum dots, connected via tunable tunnel barriers to a source (S) and drain (D) reservoir, indicated in *white*. The two *upper* gates can be used to create two quantum point contacts, in order to detect changes in the number of electrons on the dot

The electron beam can accurately write very small patterns with a resolution of about 20 nm, allowing us to make very complicated gate structures (Fig. 6). By applying negative voltages to the gates, the 2DEG is locally depleted, creating one or more small islands that are isolated from the large 2DEG reservoirs. These islands are the quantum dots. In order to probe them, we need to make electrical contact to the reservoirs. For this, we use rapid thermal annealing to diffuse AuGeNi from the surface to the 2DEG below. This forms ohmic contacts that connect the 2DEG source and drain reservoirs electrically to metal bonding pads on the surface. Metal wires bonded to these pads run toward the current or voltage probes, enabling us to perform transport measurements.

1.5 Transport Though Quantum Dots

We use two different ways to probe the behavior of electrons on a quantum dot. In this work, we mostly rely on a nearby quantum point contact (QPC) to detect changes in the number of electrons on the dot. In addition, we can perform conventional transport experiments. These experiments are conveniently understood using the constant interaction (CI) model [27]. This model makes two important assumptions. First, the Coulomb interactions among electrons in the dot are captured by a single constant capacitance, C. This is the total capacitance to the outside world, i.e. $C = C_S + C_D + C_g$, where C_S is the capacitance to the source, C_D that to the drain, and C_g to the gate. Second,

the discrete energy spectrum is independent of the number of electrons on the dot. Under these assumptions the total energy of a N-electron dot with the source-drain voltage, V_{SD}, applied to the source (and the drain grounded), is given by

$$U(N) = \frac{[-|e|(N - N_0) + C_S V_{SD} + C_g V_g]^2}{2C} + \sum_{n=1}^{N} E_n(B) \tag{1}$$

where $-|e|$ is the electron charge and N_0 the number of electrons in the dot at zero gate voltage, which compensates the positive background charge originating from the donors in the heterostructure. The terms $C_S V_{SD}$ and $C_g V_g$ can change continuously and represent the charge on the dot that is induced by the bias voltage (through the capacitance C_S) and by the gate voltage V_g (through the capacitance C_g), respectively. The last term of (1) is a sum over the occupied single-particle energy levels $E_n(B)$, which are separated by an energy $\Delta E_n = E_n - E_{n-1}$. These energy levels depend on the characteristics of the confinement potential. Note that, within the CI model, only these single-particle states depend on magnetic field, B.

To describe transport experiments, it is often more convenient to use the electrochemical potential. This is defined as the energy required to add an electron to the quantum dot:

$$\mu(N) \equiv U(N) - U(N-1) = \left(N - N_0 - \frac{1}{2}\right) E_C - \frac{E_C}{|e|}(C_S V_{SD} + C_g V_g) + E_N$$

where $E_C = e^2/C$ is the charging energy. The electrochemical potential for different electron numbers N is shown in Fig. 7a. The discrete levels are spaced by the so-called addition energy:

$$E_{add}(N) = \mu(N+1) - \mu(N) = E_C + \Delta E \; . \tag{2}$$

The addition energy consists of a purely electrostatic part, the charging energy E_C, plus the energy spacing between two discrete quantum levels, ΔE. Note that ΔE can be zero, when two consecutive electrons are added to the same spin-degenerate level.

Of course, for transport to occur, energy conservation needs to be satisfied. This is the case when an electrochemical potential level falls within the "bias window" between the electrochemical potential (Fermi energy) of the source (μ_S) and the drain (μ_D), i.e. $\mu_S \geq \mu \geq \mu_D$ with $-|e|V_{SD} = \mu_S - \mu_D$. Only then can an electron tunnel from the source onto the dot, and then tunnel off to the drain without losing or gaining energy. The important point to realize is that since the dot is very small, it has a very small capacitance and therefore a large charging energy – for typical dots $E_C \approx$ a few meV. If the electrochemical potential levels are as shown in Fig. 7a, this energy is not available (at low temperatures and small bias voltage). So, the number of

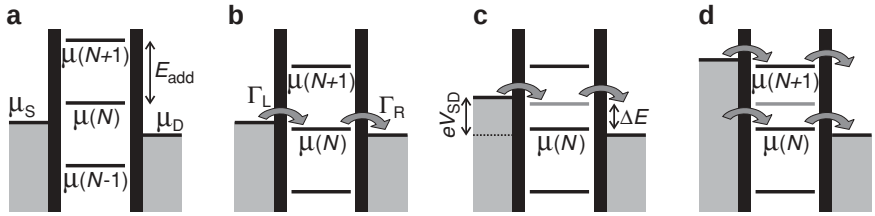

Fig. 7. Schematic diagrams of the electrochemical potential of the quantum dot for different electron numbers. (**a**) No level falls within the bias window between μ_S and μ_D, so the electron number is fixed at $N-1$ due to Coulomb blockade. (**b**) The $\mu(N)$ level is aligned, so the number of electrons can alternate between N and $N-1$, resulting in a single-electron tunneling current. The magnitude of the current depends on the tunnel rate between the dot and the reservoir on the *left*, Γ_L, and on the *right*, Γ_R. (**c**) Both the ground-state transition between $N-1$ and N electrons (*black line*), as well as the transition to an N-electron excited state (*gray line*) fall within the bias window and can thus be used for transport (though not at the same time, due to Coulomb blockade). This results in a current that is different from the situation in (b). (**d**) The bias window is so large that the number of electrons can alternate between $N-1$, N and $N+1$, i.e. two electrons can tunnel onto the dot at the same time

electrons on the dot remains fixed and no current flows through the dot. This is known as Coulomb blockade.

Fortunately, there are many ways to lift the Coulomb blockade. First, we can change the voltage applied to the gate electrode. This changes the electrostatic potential of the dot with respect to that of the reservoirs, shifting the whole "ladder" of electrochemical potential levels up or down. When a level falls within the bias window, the current through the device is switched on. In Fig. 7b $\mu(N)$ is aligned, so the electron number alternates between $N-1$ and N. This means that the Nth electron can tunnel onto the dot from the source, but only after it tunnels off to the drain can another electron come onto the dot again from the source. This cycle is known as single-electron tunnelling.

By sweeping the gate voltage and measuring the current, we obtain a trace as shown in Fig. 8a. At the positions of the peaks, an electrochemical potential level is aligned with the source and drain and a single-electron tunnelling current flows. In the valleys between the peaks, the number of electrons on the dot is fixed due to Coulomb blockade. By tuning the gate voltage from one valley to the next one, the number of electrons on the dot can be precisely controlled. The distance between the peaks corresponds to $E_C + \Delta E$, and can therefore give information about the energy spectrum of the dot.

A second way to lift Coulomb blockade is by changing the source-drain voltage, V_{SD} (see Fig. 7c). (In general, we keep the drain potential fixed, and change only the source potential.) This increases the bias window and also "drags" the electrochemical potential of the dot along, due to the capacitive coupling to the source. Again, a current can flow only when an electrochemical

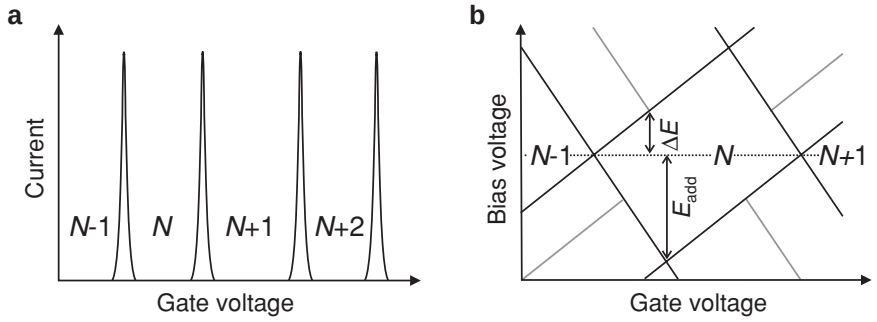

Fig. 8. Transport through a quantum dot. (**a**) Coulomb peaks in current versus gate voltage in the linear-response regime. (**b**) Coulomb diamonds in differential conductance, dI/dV_{SD}, versus V_{SD} and V_g, up to large bias. The edges of the diamond-shaped regions (*black*) correspond to the onset of current. *Diagonal* lines emanating from the diamonds (*gray*) indicate the onset of transport through excited states

potential level falls within the bias window. By increasing V_{SD} until both the ground state as well as an excited state transition fall within the bias window, an electron can choose to tunnel not only through the ground state, but also through an excited state of the N-electron dot. This is visible as a change in the total current. In this way, we can perform excited-state spectroscopy.

Usually, we measure the current or differential conductance while sweeping the bias voltage, for a series of different values of the gate voltage. Such a measurement is shown schematically in Fig. 8b. Inside the diamond-shaped region, the number of electrons is fixed due to Coulomb blockade, and no current flows. Outside the diamonds, Coulomb blockade is lifted and single-electron tunnelling can take place (or for larger bias voltages even double-electron tunnelling is possible, see Fig. 7d). Excited states are revealed as changes in the current, i.e. as peaks or dips in the differential conductance. From such a "Coulomb diamond" the excited-state splitting as well as the charging energy can be read off directly.

The simple model described above explains successfully how quantisation of charge and energy leads to effects like Coulomb blockade and Coulomb oscillations. Nevertheless, it is too simplified in many respects. For instance, the model considers only first-order tunnelling processes, in which an electron tunnels first from one reservoir onto the dot, and then from the dot to the other reservoir. But when the tunnel rate between the dot and the leads, Γ, is increased, higher-order tunnelling via virtual intermediate states becomes important. Such processes are known as "cotunnelling". Furthermore, the simple model does not take into account the spin of the electrons, thereby excluding for instance exchange effects. Also the Kondo effect, an interaction between the spin on the dot and the spins of the electrons in the reservoir, cannot be accounted for.

1.6 Spin Configurations in Few-Electron Quantum Dots

The fact that electrons carry spin determines the electronic states of the quantum dot. In the simplest case – a dot containing just a single electron – spin leads to a splitting of all orbitals into Zeeman doublets, with the ground state corresponding to the electron spin pointing up ($\uparrow$), and the excited state to the spin pointing down ($\downarrow$). The difference between the corresponding energy levels $E_\uparrow$ and $E_\downarrow$ is given by the Zeeman energy, $\Delta E_Z = g\mu_B B$, which is approximately 25 μeV/T in GaAs.

For two electrons in a quantum dot, the situation is more complicated. For a Hamiltonian without explicit spin-dependent terms, the two-electron state is the product of the orbital and spin state. Since electrons are fermions, the total two-electron state has to be anti-symmetric under exchange of the two particles. Therefore, if the orbital part is symmetric, the spin state must be anti-symmetric, and if the spin part is anti-symmetric, the orbital state must be symmetric. The anti-symmetric two-spin state is the so-called spin singlet (S):

$$S = \frac{|\uparrow\downarrow\rangle - |\downarrow\uparrow\rangle}{\sqrt{2}} \tag{3}$$

which has total spin $S = 0$. The symmetric two-spin states are the so-called spin triplets (T_+, T_0 and T_-):

$$T_+ = |\uparrow\uparrow\rangle T_0 = \frac{|\uparrow\downarrow\rangle + |\downarrow\uparrow\rangle}{\sqrt{2}} T_- = |\downarrow\downarrow\rangle \tag{4}$$

which have total spin $S = 1$ and a quantum number m_s (corresponding to the spin z-component) of 1, 0, and −1, respectively. In a finite magnetic field, the three triplet states are split by the Zeeman splitting, ΔE_Z.

Even at zero magnetic field, the energy of the two-electron system depends on its spin configuration, through the requirement of anti-symmetry of the total state. If we consider just the two lowest orbitals, ε_0 and ε_1, then there are six possibilities to fill these with two electrons (Fig. 9). At zero magnetic field [28], the two-electron ground state is always the spin singlet (Fig. 9a), and the lowest excited states are always the three spin triplets (Fig. 9b–d). The energy gain of T_0 with respect to the excited spin singlet S_1 (Fig. 9e) is known as the exchange energy, J. It essentially results from the fact that electrons in the triplet states tend to avoid each other, reducing their mutual Coulomb energy. As the Coulomb interaction is very strong, the exchange energy can be quite large (a few 100 μeV) [29].

The energy difference between T_0 and the lowest singlet S, the "singlet-triplet energy" E_{ST}, is thus considerably smaller than $\varepsilon_1 - \varepsilon_0$. In fact, besides the gain in exchange energy for the triplet states, there is also a gain in the direct Coulomb energy, related to the different occupation of the orbitals [29]. For a magnetic field above a few Tesla (perpendicular to the 2DEG plane), E_{ST} can even become negative, leading to a singlet-triplet transition of the two-electron ground state [30].

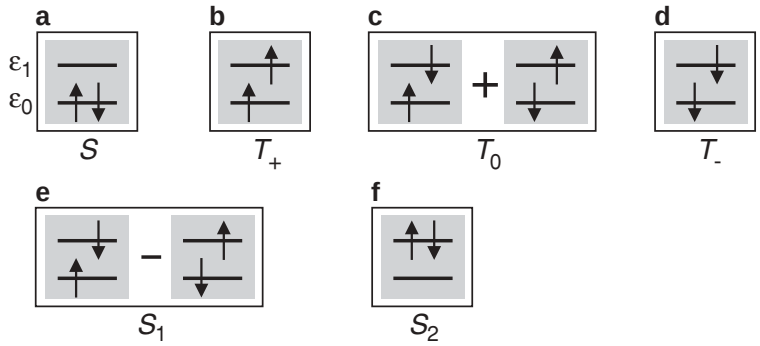

Fig. 9. Schematic energy diagrams depicting the spin states of two electrons occupying two spin degenerate single-particle levels (ε_0 and ε_1). (**a**) Spin singlet, which is the ground state at zero magnetic field. (**b**)–(**d**) Lowest three spin triplet states, T_+, T_0 and T_-, which have total spin $S = 1$ and quantum number $m_s = +1$, 0 and -1, respectively. In finite magnetic field, the triplet states are split by the Zeeman energy. (**e**) Excited spin singlet state, S_1, which has an energy J compared to triplet state T_0. (**f**) Highest excited spin singlet state, S_2

In the presence of a magnetic field, the energies of the lowest singlet and triplet states (Fig. 9a–d) can be expressed as:

$$\begin{aligned} E_S &= E_\uparrow + E_\downarrow + E_C = 2E_\uparrow + \Delta E_Z + E_C \\ E_{T_+} &= 2E_\uparrow + E_{ST} + E_C \\ E_{T_0} &= E_\uparrow + E_\downarrow + E_{ST} + E_C = 2E_\uparrow + E_{ST} + \Delta E_Z + E_C \\ E_{T_-} &= 2E_\downarrow + E_{ST} + E_C = 2E_\uparrow + E_{ST} + 2\Delta E_Z + E_C . \end{aligned}$$

Figure 10a shows the possible transitions between the one-electron spin-split orbital ground state and the two-electron states. We have omitted the transitions $\uparrow \leftrightarrow T_-$ and $\downarrow \leftrightarrow T_+$ since these require a change in the spin z-component of more than 1/2 and are thus spin-blocked [31]. From the energy diagram we can deduce the electrochemical potential ladder, which is shown in Fig. 10b. Note that $\mu_{\uparrow \leftrightarrow T_+} = \mu_{\downarrow \leftrightarrow T_0}$ and $\mu_{\uparrow \leftrightarrow T_0} = \mu_{\downarrow \leftrightarrow T_-}$. Consequently, the *three* triplet states lead to only *two* resonances in first order transport through the dot.

For more than two electrons, the spin states can be much more complicated. However, in some cases and for certain magnetic field regimes they might be well approximated by a one-electron Zeeman doublet (when N is odd) or by two-electron singlet or triplet states (when N is even). But there are still differences – for instance, if $N > 2$ the ground state at zero field can be a spin triplet, due to Hund's rule [32].

The eigenstates of a two-electron double dot are also spin singlets and triplets. We can again use the diagrams in Fig. 9, but now the single-particle eigenstates ε_0 and ε_1 represent the symmetric and anti-symmetric combination of the lowest orbital on each of the two dots, respectively. Due to

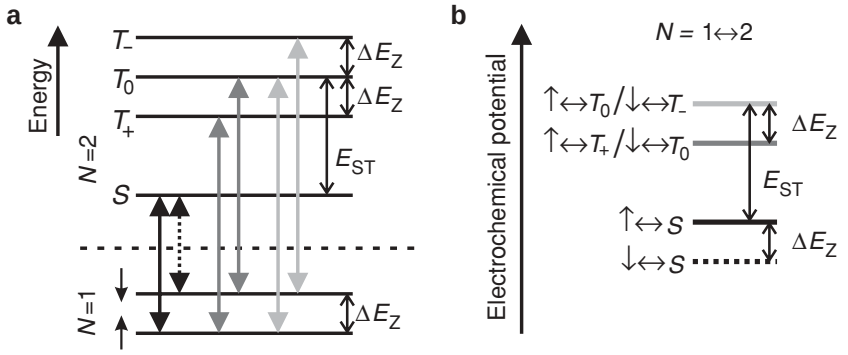

Fig. 10. One- and two-electron states and transitions at finite magnetic field. (**a**) Energy diagram for a fixed gate voltage. By changing the gate voltage, the one-electron states (*below the dashed line*) shift up or down relative to the two-electron states (*above the dashed line*). The six transitions that are allowed (i.e. not spin-blocked) are indicated by vertical *arrows*. (**b**) Electrochemical potentials for the transitions between one- and two-electron states. The six transitions in (**a**) correspond to only four different electrochemical potentials. By changing the gate voltage, the whole ladder of levels is shifted up or down

tunnelling between the dots, with tunnelling matrix element t, ε_0 (the "bonding state") and ε_1 (the "anti-bonding state") are split by an energy $2t$. By filling the two states with two electrons, we again get a spin singlet ground state and a triplet first excited state (at zero field). However, the singlet ground state is not purely S (Fig. 9a), but also contains a small admixture of the excited singlet S_2 (Fig. 9f). The admixture of S_2 depends on the competition between inter-dot tunnelling and the Coulomb repulsion, and serves to lower the Coulomb energy by reducing the double occupancy of the dots [33].

If we focus only on the singlet ground state and the triplet first excited states, then we can describe the two spins $\mathbf{S_1}$ and $\mathbf{S_2}$ by the Heisenberg Hamiltonian, $H = J\mathbf{S_1} \cdot \mathbf{S_2}$. Due to this mapping procedure, J is now defined as the energy difference between the triplet state T_0 and the singlet ground state, which depends on the details of the double dot orbital states. From a Hund-Mulliken calculation [34], J is approximately given by $4t^2/U + V$, where U is the on-site charging energy and V includes the effect of the long-range Coulomb interaction. By changing the overlap of the wavefunctions of the two electrons, we can change t and therefore J. Thus, control of the inter-dot tunnel barrier would allow us to perform operations such as swapping or entangling two spins.

1.7 Measurement Setup

Dilution Refrigerator

To resolve small energies such as the Zeeman splitting, the sample has to be cooled down to temperatures well below a Kelvin. We use an Oxford Kelvinox 300 dilution refrigerator, which has a base temperature of about 10 mK, and a cooling power in excess of 300 μW (at 100 mK). The sample holder is connected to a cold finger and placed in a copper can (36 mm inner diameter) in the bore of a superconducting magnet that can apply a magnetic field up to 16 T.

Measurement Electronics

A typical measurement involves applying a source-drain voltage over (a part of) the device, and measuring the resulting current as a function of the voltages applied to the gates. The electrical circuits for the voltage-biased current measurement and for applying the gate voltages are shown in Fig. 11 and Fig. 12, respectively. The most important parts of the measurement electronics – i.e. the current-to-voltage (IV) convertor, isolation amplifier, voltage source and digital-to-analog convertors (DACs) – were all built by Raymond Schouten at Delft University. The underlying principle of the setup is to isolate the sample electrically from the measurement electronics. This is achieved via optical isolation at both sides of the measurement chain, i.e. in the voltage source, the isolation amplifier, as well as the DACs. In all these units, the electrical signal passes through analog optocouplers, which first convert it to an optical signal using an LED, and then convert the optical signal back using a photodiode. In this way, there is no galvanic connection between the two sides. In addition, all circuitry at the sample side is analog (even the DACs have no clock circuits or microprocessors), battery-powered, and uses a single clean ground (connected to the metal parts of the fridge) which is separated from the ground used by the "dirty" electronics. All these features help to eliminate ground loops and reduce interference on the measurement signal.

Measurements are controlled by a computer running LabView. It sends commands via a fiber link to two DAC-boxes, each containing 8 digital-to-analog convertors, and powered by a specially shielded transformer. Most of the DACs are used to generate the voltages applied to the gate electrodes (typically between 0 and −5 V). One of the DACs controls the source-drain voltage for the device. The output voltage of this DAC (typically between +5 and −5 V) is sent to a voltage source, which attenuates the signal by a factor 10, 10^2, 10^3 or 10^4 and provides optical isolation. The attenuated voltage is then applied to one of the ohmic contacts connected to the source reservoir of the device.

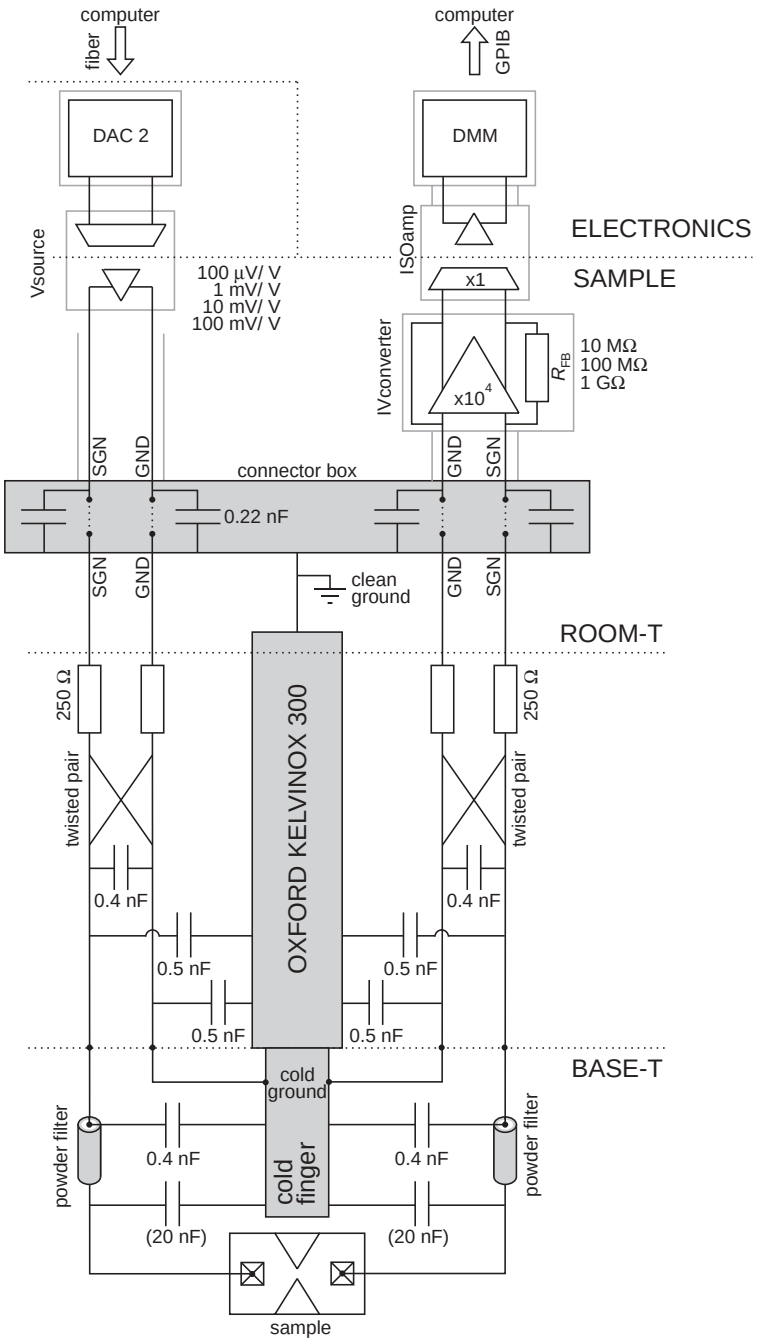

Fig. 11. Electrical circuit for performing a voltage-biased current measurement. Elements shown in *gray* are connected to ground. *Gray* lines indicate the shielding of the measurement electronics and wires

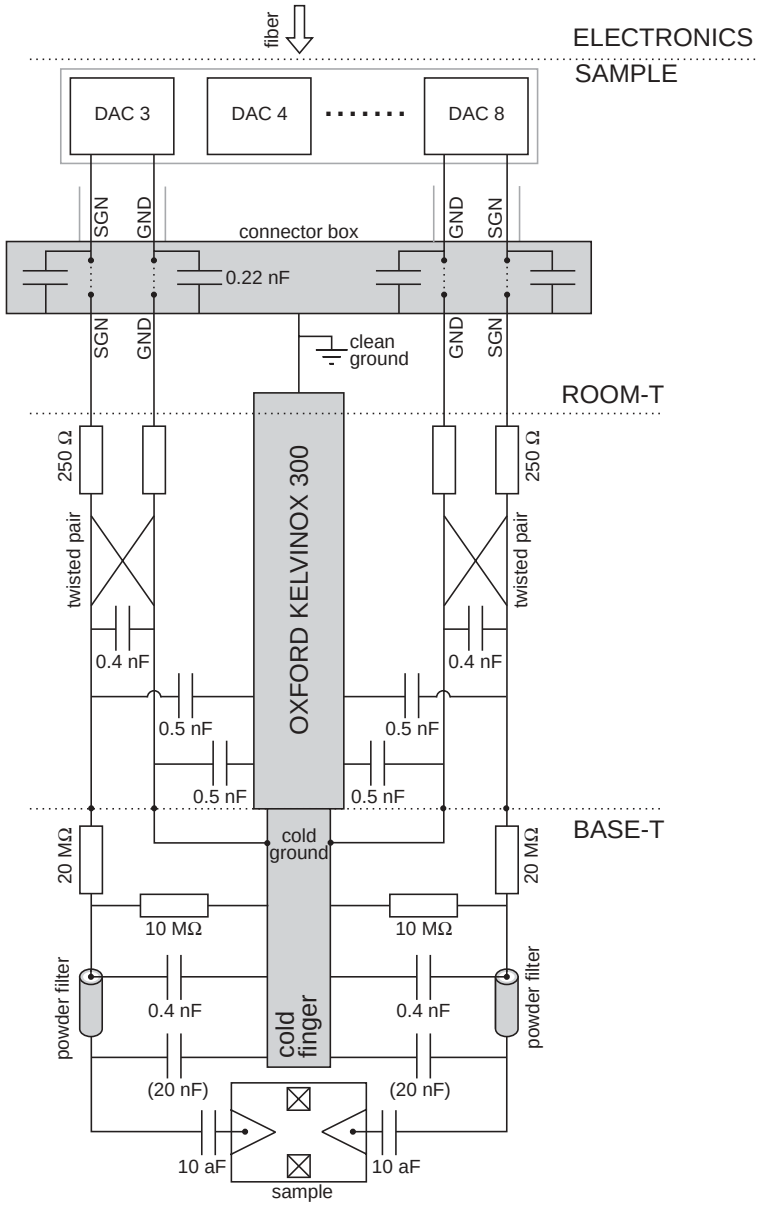

Fig. 12. Electrical circuit for applying voltages to the gate electrodes. Elements shown in *gray* are connected to ground. *Gray* lines indicate the shielding of the measurement electronics and wires

The resulting current coming from the drain reservoir is fed to a low-noise IV convertor. In this work we use two types, depending on the desired bandwidth. The first one is designed for low-frequency measurements. It has a bandwidth of about 1 kHz, and a noise floor of $\sim$5 fA/Hz$^{1/2}$. The feedback resistance can be set to 10 MΩ, 100 MΩ or 1 GΩ, with an input resistance that is a factor 10^3 or 10^4 smaller (for the "low noise" or "low input resistance" setting, respectively). The faster IV convertor has a bandwidth of about 150 kHz, and a current noise of $\sim$1 pA/Hz$^{1/2}$ at 100 kHz. The feedback resistance is 10 MΩ, corresponding to an input resistance of 1.3 kΩ. More characteristics are given in Sect. 4.

The signal from the IV convertor is then sent to an isolation amplifier, to provide optical isolation and possibly gain. Again we can choose a low-frequency version (up to $\sim$1 kHz) or a high-frequency one (up to $\sim$300 kHz). The voltage from the isolation amplifier is finally measured by a digital multimeter (Keithley 2700) and sent to the computer via GPIB interface. Alternatively, we can use a lock-in amplifier (Stanford EG&G 5210) if the signal to be measured is periodic, or an ADwin Gold module for very fast measurements (up to 2.2×10^6 14-bit samples per second).

Measurement Wires

To make contact to the sample, 2×12 twisted pairs of wires run from two connector boxes at room temperature all the way down to the "cold finger" at base temperature. The diameter and material of these wires is chosen to minimize the heat load on the mixing chamber. From room temperature to 1 Kelvin, 2×9 pairs consist of manganine wires (100 µm diameter), and 2×3 pairs of copper wires (90 µm diameter). The copper wires can be used if a large current has to be applied. From 1 Kelvin to the mixing chamber, superconducting "Niomax" wires (50 µm diameter) are used. From the mixing chamber to the bottom of the cold finger, where thermal conductivity is no longer a constraint, we have standard copper wires. At base temperature, one wire of each twisted pair is connected to "cold ground" (i.e. the cold finger), which is electrically connected to clean ground via the metal parts of the fridge.

All wires are thermally anchored to the fridge, by carefully wrapping them around copper posts, at several temperature stages (4 K, 1 K, $\sim$100 mK and $\sim$10 mK). At room temperature, the resistance of the wires is about 250 Ω or 150 Ω for the manganine or copper wires, respectively. At low temperature it is about 50 Ω. The wires have various parasitic capacitances to their twisted partner and to ground, as indicated in Fig. 11 and Fig. 12.

Filtering

The wires connect the device to the measurement electronics at room temperature, so they have to be carefully filtered to avoid that the electrons in

the sample heat up due to spurious noise and interference. Several filtering stages are required for different frequency ranges (see Fig. 11 and Fig. 12). In the connector box at room temperature, all wires are connected to ground via 0.22 nF "feedthrough capacitors". At base temperature, all signal wires run through "copper powder filters" [35]. These are copper tubes filled with copper powder, in which 4 signal wires with a length of about 2 meters each are wound. The powder absorbs the high-frequency noise very effectively, leading to an attenuation of more than −60 dB from a few 100 MHz up to more than 50 GHz [36].

To remove the remaining low-frequency noise, we solder a 20 nF capacitor between each signal wire and the cold finger ground. In combination with the ∼100 Ω resistance of the wires, this forms a low-pass RC filter with a cut-off frequency of about 100 kHz (even 10 kHz for the wire connected to the IV convertor, due to its input resistance of about 1.3 kΩ). These filters are used for the wires connecting to ohmic contacts (although they were taken out to perform some of the high-bandwidth measurements described in this work). For the wires connecting to gate electrodes, a 1:3 voltage divider is present (consisting of a 20 MΩ resistance in the signal line and a 10 MΩ resistance to ground). In this way, the gate voltages are filtered by a low-pass RC filter with a cut-off frequency of about 1 Hz. By combining all these filters, the electrons in the sample can be cooled to an effective temperature below 50 mK (if no extra heat loads such as coaxial cables are present).

High-Frequency Signals

High-frequency signals can be applied to gate electrodes via two coaxial cables. They consist of three parts, connected via standard 2.4 mm Hewlett Packard connectors (specified up to 50 GHz). From room temperature to 1 Kelvin, a 0.085 inch semi-rigid Be-Cu (inner and outer conductor) coaxial cable is used. From 1 Kelvin to the mixing chamber, we use 0.085 inch semi-rigid superconducting Nb. From the mixing chamber to the sample holder, flexible tin plated Cu coaxial cables are present. The coaxes are thermally anchored at 4 K, 1 K, ∼800 mK, ∼100 mK and base temperature, by clamping each cable firmly between two copper parts. To thermalize also the inner conductor of the coax, we use Hewlett Packard 8490D attenuators (typically −20 dB) at 1 K. These attenuators cannot be used at the mixing chamber, as they tend to become superconducting below about 100 mK. We have also tried using Inmet 50EH attenuators at the mixing chamber, but these showed the same problem.

To generate the high-frequency signals, we use a microwave source (Hewlett Packard 83650A) that goes up to 50 GHz (or 75 GHz, in combination with a "frequency doubler"); a pulse generator (Hewlett Packard 8133A), which generates simple 10 ns to 1 µs pulses with a rise time of 60 ps; and an arbitrary

waveform generator (Sony Tektronix AWS520), which can generate more complicated pulses with a rise time of 1.5 ns. With the cables described above, the fastest pulse flank we can transmit to the sample is about 200 ps. Microwave signals are transmitted with about 10 dB loss at 50 GHz.

Special care needs to be given to the connection from the coaxial cable to the chip, in order to minimize reflections. The sample holder we use, has an SMA connector that can be connected to the 2.4 mm coaxial cable. At the other end, the pin of the SMA connector sticks through a small hole in the chip carrier. This allows it to be soldered to a metal pad on the chip carrier, from which we can then bond to the chip. This sample holder is used to apply pulses or microwave signals to a gate electrode.

1.8 Sample Stability

A severe experimental difficulty that is not related to the measurement setup, but to the sample itself, is the problem of "charge switching". It shows up in measurements as fluctuations in the position of a Coulomb peak, or as sudden jumps in the QPC-current that are not related to charging or discharging of a nearby quantum dot. Generally, these switches are attributed to (deep) traps in the donor layer that capture or release an electron close to the quantum dot [37]. This well-known but poorly understood phenomenon is a manifestation of 1/f noise in semiconductors, which causes the electrostatic potential landscape in the 2DEG to fluctuate.

The strength of the fluctuations can differ enormously. In some samples, switching occurs on a time scale of seconds, making only the most trivial measurements possible, whereas in other samples, no switches are visible on a time scale of hours. It is not clear what exactly determines the stability. It certainly depends on the heterostructure, as some wafers are clearly better than others. A number of growth parameters could be important, such as the Al concentration in the AlGaAs, the doping density and method (modulation doping or delta doping), the thickness of the spacer layer between the n-AlGaAs and GaAs, the depth of the 2DEG below the surface, a possible surface layer, and many more. We have recently started a collaboration with the group of Professor Wegscheider in Regensburg to grow and characterize heterostructures in which some of these parameters are systematically varied. In this way we hope to find out what makes certain heterostructures stable.

Even for the same heterostructure, some samples are more quiet than others. The reasons for this are not clear. There are reports that stability is improved if the sample is cooled down slowly, while applying a positive voltage (about +280 mV) on all gates that are going to be used in the experiment. This procedure effectively "freezes in" a negative charge around the gates, such that less negative gate voltages are sufficient to define the quantum dot at low temperatures. Most samples described in this work have been cooled down from room temperature to 4 K slowly (in one to two days) with all gates

grounded. We find that in general samples get more quiet during the first week of applying the gate voltages. Finally, sample stability also involves an element of luck: Fig. 13 shows two Coulomb diamonds that were measured immediately after each other under identical conditions. Measurement Fig. 13a is reasonably quiet, but in Fig. 13b the effects of an individual two-level fluctuator are visible. This particular fluctuator remained active for a week, until the sample was warmed up.

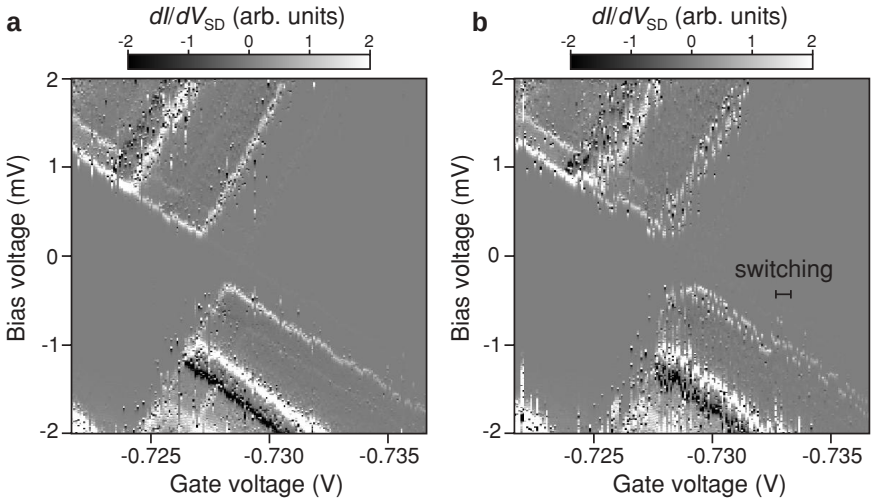

Fig. 13. Charge switching in a large-bias measurement in the few-electron regime, for $B = 12$ T. (**a**) Differential conductance, dI/dV_{SD} (*in grayscale*), as a function of bias voltage and gate voltage. This measurement is considered reasonably stable. (**b**) Identical measurement, taken immediately after (**a**). A single two-level fluctuator has become active, causing the effective gate voltage to fluctuate between two values at any position in the figure, and leading to an apparent splitting of all the lines. This is considered a measurement of poor stability

Switching has made all experiments we performed more difficult, and has made some experiments that we wanted to perform impossible. Better control over heterostructure stability is therefore essential for the increasingly difficult steps towards creating a quantum dot spin qubit.

2 Few-Electron Quantum Dot Circuit with Integrated Charge Read-Out

In this section, we report on the realization of few-electron double quantum dots defined in a two-dimensional electron gas by means of surface gates on top of a GaAs/AlGaAs heterostructure. The double quantum dots are flanked

by two quantum point contacts, serving as charge detectors. These enable determination of the precise number of conduction electrons on each dot. This number can be reduced to zero while still allowing transport measurements through the double dot. Even in the few-electron case, the tunnel coupling between the two dots can be controlled over a wide range, from the weak-coupling to the strong-coupling regime. In addition, we use microwave radiation to pump an electron from one dot to the other by absorption of a single photon. The experiments demonstrate that this quantum dot circuit can serve as a good starting point for a scalable spin-qubit system.

2.1 Few-Electron Quantum Dots

The experimental development of a quantum computer is presently at the stage of realizing few-qubit circuits. In the solid state, particular success has been achieved with superconducting devices, in which two macroscopic quantum states are used as a qubit two-level system (see [38] and references therein). The opposite alternative would be the use of two-level systems defined by microscopic variables, for instance the spin (or charge) state of single electrons confined in semiconductor quantum dots [27]. For the control of one-electron quantum states by electrical voltages, the first requirement is to realize an appropriate quantum dot circuit containing just a single conduction electron.

Single-electron quantum dots have been created in self-assembled structures [39] and also in small vertical pillars defined by etching [40]. (Recently, also semiconductor nanowires and carbon nanotubes have been used for this purpose.) The disadvantage of these types of quantum dots is that they are hard to integrate into circuits with a controllable coupling between the elements, although integration of vertical quantum dot structures is currently being pursued [41, 42]. Alternatively, we can use a system of lateral quantum dots defined in a two-dimensional electron gas (2DEG) by surface gates on top of a semiconductor heterostructure [27]. Here, integration of multiple dots is straightforward, by simply increasing the number of gate electrodes. In addition, the tunnel coupling between the dots can be tuned in situ, since it is controlled by the gate voltages. The challenge is to reduce the number of electrons to one per quantum dot. This has long been impossible, since reducing the electron number tends to be accompanied by a decrease in the tunnel coupling, resulting in a current too small to be measured [43].

In this section, we demonstrate double quantum dot devices containing a voltage-controllable number of electrons, down to a single electron. We have integrated these devices with charge detectors that can read out the charge state of the double quantum dot with a sensitivity better than a single electron charge. The importance of the present circuit is that it can serve as a fully tunable two-qubit system, following the proposal by Loss and DiVincenzo [2], which describes an optimal combination of the single-electron charge degree of freedom (for convenient manipulation using electrical voltages) and the

spin degree of freedom (which promises a long coherence time, essential for encoding quantum information).

2.2 Samples

We have fabricated and measured several few-electron double quantum dots, of three different designs (Fig. 14). The first two types have only been used once as few-electron *single* dots. In both cases, one of the gate electrodes was not functioning, which prevented us from testing if these devices also function as few-electron *double* dots. The third type of device (Fig. 14c) did function as a double dot, and was used for all subsequent few-electron experiments.

To verify that the first device (Fig. 14a) can operate as a few-electron single quantum dot, we performed a large-bias measurement of the differential conductance through the dot. Going towards more negative gate voltage, a series of "Coulomb diamonds" is revealed (Fig. 15a), in which the number of electrons on the dot, N, is constant. This is followed by a region in which the "diamond" does not close, even up to a source-drain voltage of 10 mV, i.e. several times larger than the typical charging energy for a small dot ($\sim$2 meV). Therefore, in this region $N = 0$.

The tunnel coupling between the dot and the source and drain reservoirs could be changed by simply readjusting the gate voltages. For strong coupling, a zero-bias peak – hallmark of the Kondo effect – became visible throughout the one-electron diamond (Fig. 15b). From the width of the zero-bias peak (Fig. 15c) we found a Kondo temperature of about 0.4 K. The appearance of a one-electron Kondo effect (unpublished) implies that this quantum dot design allows the tunnel coupling to be tuned over a wide range, even in the few-electron regime. In addition, it is striking evidence that we can confine a single *spin* in a lateral quantum dot.

In the second quantum dot design (Fig. 14b), the narrow "plunger" gates approach the dot more from the sides, rather than from below. In this way, they are further away from the central tunnel barrier, reducing the effect they have on the tunnel rate. Also, the gate coming from the top of the picture was made thinner, in order to make the tunnel barriers more easily controllable [43]. Thirdly, the characteristic gates ending in circles (see Fig. 14a) were left out. This device was quite easily tunable.

In the rest of this section, we use the third design (Fig. 14c). Two nominally identical devices are studied, both as shown in Fig. 16a. They consist of a double quantum dot flanked by two quantum point contacts (QPCs), defined in a 2DEG that is present below the surface of a GaAs/AlGaAs heterostructure. The layout of the double quantum dot is an extension of previously reported single dot devices [43]. The double dot is defined by applying negative voltages to the 6 central gates. Gate T in combination with the left (right) gate, L (R), controls the tunnel barrier from the left (right) dot to drain 1 (source 2). Gate T in combination with the middle gate, M, controls the tunnel barrier between the two dots. The narrow "plunger" gates, P_L and

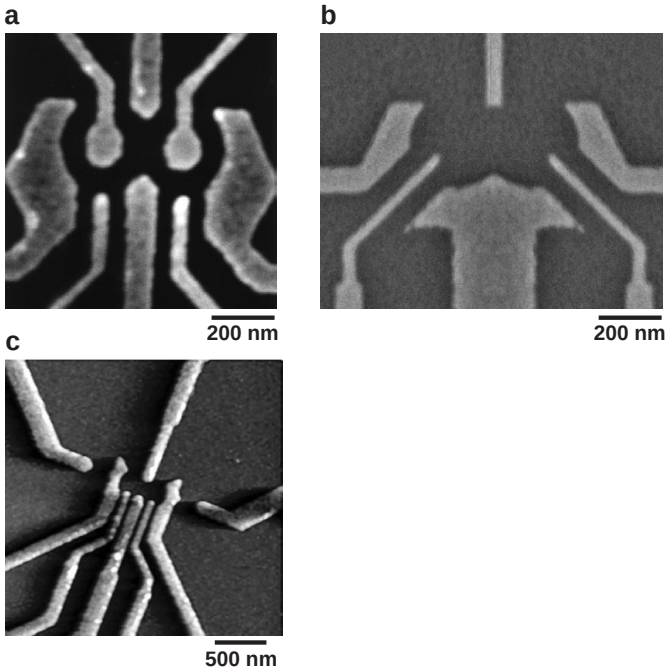

Fig. 14. Few-electron quantum dot devices. (**a**) Scanning electron microscope image of the first sample, showing the metal gate electrodes (*light*) on top of a GaAs/AlGaAs heterostructure (*dark*) that contains a 2DEG 90 nm below the surface (with electron density $2.9 \times 10^{11}\,\mathrm{cm}^{-2}$). This device was used only as a few-electron *single* dot. Due to the similarity of the image to characters from the Japanese "Gundam" animation, this has become known as the Gundam design. The two gates coming from the top and ending in small circles (the "eyes") were meant to make the dot confinement potential steeper, by applying a positive voltage to them (up to $\sim$0.5 V). The gates were not very effective, and were left out in later designs. (The device was fabricated by Wilfred van der Wiel at NTT Basic Research Laboratories.) (**b**) Scanning electron microscope image of the second device, made on a similar heterostructure. It was used only as a few-electron *single* dot, and was more easily tunable than the first one. (The device was fabricated by Wilfred van der Wiel and Ronald Hanson at NTT Basic Research Laboratories.) (**c**) Atomic force microscope image of the third device, made on a similar heterostructure. This design, with two extra side gates to form two quantum point contacts, was operated many times as a *single* dot, and twice as a few-electron *double* dot. It was used for all subsequent measurements. A zoom-in of the gate structure is shown in Fig. 16a. (The device was fabricated by Ronald Hanson and Laurens Willems van Beveren at NTT Basic Research Laboratories)

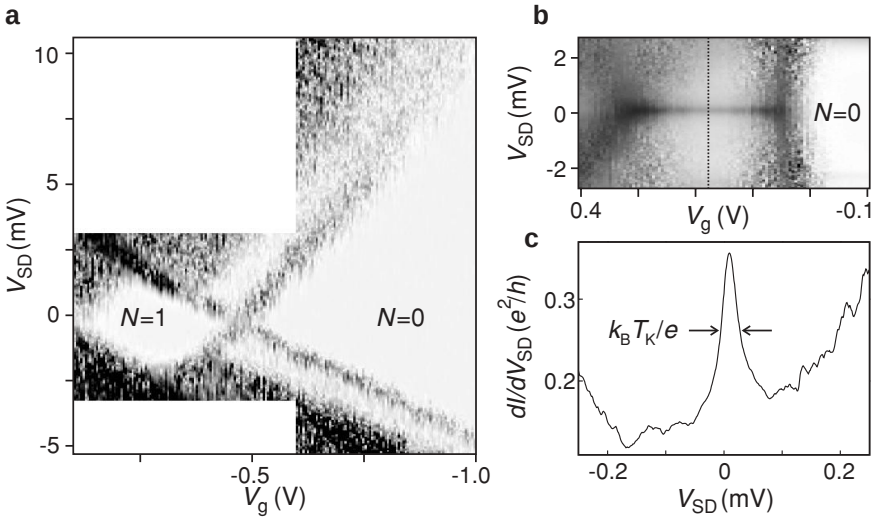

Fig. 15. Kondo effect in a one-electron lateral quantum dot of the type shown in Fig. 14a. (**a**) Differential conductance (*in grayscale*) versus source-drain voltage, V_{SD}, and plunger gate voltage, V_g. In the white diamond and the white region to the right (indicated by $N = 1$ and $N = 0$, respectively), no current flows due to Coulomb blockade. The $N = 0$ region opens up to more than 10 mV, indicating that the dot is really empty here. (**b**) Close-up of the $N = 1$ diamond for stronger coupling to the reservoirs. A sharp Kondo resonance is visible at zero source-drain voltage. Although charge switching is very severe in this sample, the position of the Kondo resonance is very stable, as it is pinned to the Fermi energy of the reservoirs. (**c**) Kondo zero-bias peak in differential conductance, taken at the position indicated by the *dotted* line in (**b**)

P_R, are used to change the electrostatic potential of the left and right dot, respectively. The left plunger gate is connected to a coaxial cable, so that we can apply high-frequency signals. In the present experiments we do not apply dc voltages to P_L. In order to control the number of electrons on the double dot, we use gate L for the left dot and P_R or R for the right dot. All measurements are performed with the sample cooled to a base temperature of about 10 mK inside a dilution refrigerator.

We first study sample 1. The individual dots are characterized using standard Coulomb blockade experiments [27], i.e. by measuring I_{DOT}. We find that the energy cost for adding a second electron to a one-electron dot is 3.7 meV. The one-electron excitation energy (i.e. the difference between the ground state and the first orbital excited state) is 1.8 meV at zero magnetic field. For a two-electron dot the energy difference between the spin singlet ground state and the spin triplet excited state is 1.0 meV at zero magnetic field. Increasing the field (perpendicular to the 2DEG) leads to a transition from a singlet to a triplet ground state at about 1.7 Tesla.

2.3 Quantum Point Contact as Charge Detector

As an alternative to measuring the current through the quantum dot, we can also measure the charge on the dot using one of the QPCs [44, 45]. To demonstrate this functionality, we first define only the left dot (by grounding gates R and P_R), and use the left QPC as a charge detector. The QPC is formed by applying negative voltages to $Q - L$ and L. This creates a narrow constriction in the 2DEG, with a conductance, G, that is quantized when sweeping the gate voltage V_{Q-L}. The last plateau (at $G = 2e^2/h$) and the transition to complete pinch-off (i.e. $G = 0$) are shown in Fig. 16b. We tune the QPC to the steepest point ($G \approx e^2/h$), where the QPC-conductance has a maximum sensitivity to changes in the electrostatic environment, including changes in the charge of the nearby quantum dot.

To change the number of electrons in the left dot, we make gate voltage V_M more negative (see Fig. 16c). This reduces the QPC current, due to the capacitive coupling from gate M to the QPC constriction. In addition, the changing gate voltage periodically pushes an electron out of the dot. The associated sudden change in charge lifts the electrostatic potential at the QPC constriction, resulting in a step-like feature in I_{QPC} (see the expansion in Fig. 16c, where the linear background is subtracted). This step indicates a change in the electron number. So, even without passing current through the dot, I_{QPC} provides information about the charge on the dot.

To enhance the charge sensitivity we apply a small modulation (0.3 mV at 17.7 Hz) to V_M and use lock-in detection to measure dI_{QPC}/dV_M [45]. The steps in I_{QPC} now appear as dips in dI_{QPC}/dV_M. Figure 16d shows the resulting dips, as well as the corresponding Coulomb peaks measured in the current through the dot. The coincidence of the Coulomb peaks and dips demonstrates that the QPC indeed functions as a charge detector. From the height of the step in Fig. 16c (∼50 pA, typically 1–2% of the total current) compared to the noise (∼5 pA for a measurement time of 100 ms), we estimate the sensitivity of the charge detector to be about $0.1e$, with e being the single electron charge. The unique advantage of QPC charge detection is that it provides a signal even when the tunnel barriers of the dot are so opaque that I_{DOT} is too small to be measured [44, 45]. This allows us to study quantum dots even when they are virtually isolated from the reservoirs.

2.4 Double Dot Charge Stability Diagram

The QPC can also detect changes in the charge configuration of the *double* dot. To demonstrate this, we use the QPC on the right to measure dI_{QPC}/dV_L versus V_L and V_{PR} (Fig. 17a), where V_L controls (mainly) the number of electrons on the left dot, and V_{PR} (mainly) that on the right. Dark lines in the figure signify a dip in dI_{QPC}/dV_L, corresponding to a change in the total number of electrons on the double dot. Together these lines form the so-called "honeycomb diagram" [46, 47]. The almost-horizontal lines correspond to a

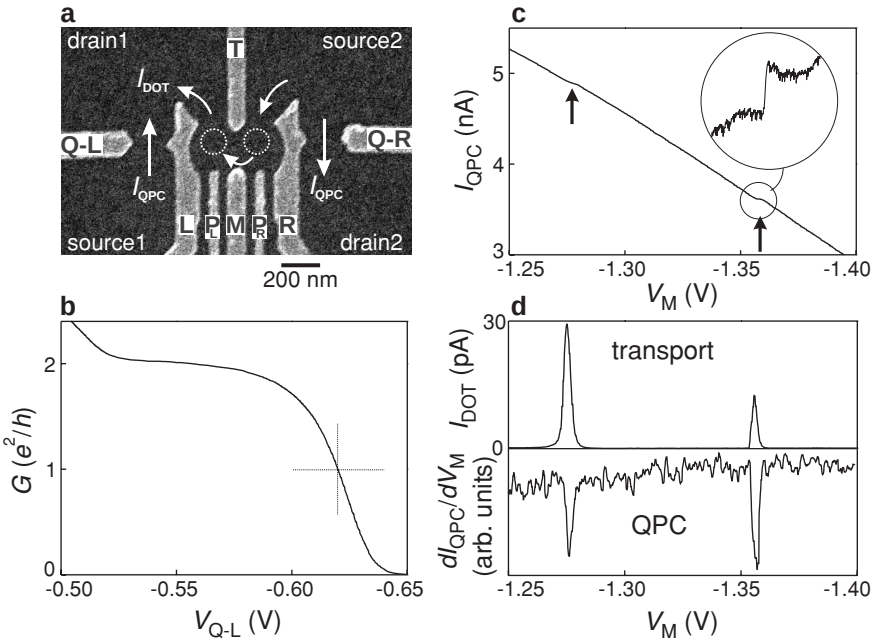

Fig. 16. Operating the QPC as a charge detector of a single dot. (**a**) Scanning electron microscope image of the device, showing metallic surface gates (*light gray*) on top of a GaAs/AlGaAs heterostructure (*dark gray*). The device contains a 2DEG 90 nm below the surface, with an electron density of $2.9 \times 10^{11}\,\mathrm{cm}^{-2}$. *White dotted* circles indicate the two quantum dots, *white arrows* show the possible current paths. A bias voltage, V_{DOT}, can be applied between source 2 and drain 1, leading to current through the dot(s), I_{DOT}. A bias voltage, V_{SD1} (V_{SD2}), between source 1 (source 2) and drain 1 (drain 2), yields a current, I_{QPC} through the *left* (*right*) QPC. (**b**) Conductance, G, of the left QPC versus gate voltage, V_{Q-L}, showing the last quantized plateau (at $G = 2e^2/h$) and the transition to complete pinch-off ($G = 0$). The QPC is set to the point of highest charge sensitivity, at $G \approx e^2/h$ (indicated by the *dashed cross*). (**c**) Current through the left QPC, I_{QPC}, versus left-dot gate voltage, V_M, with $V_{SD1} = 250\,\mu\mathrm{V}$ and $V_{SD2} = V_{DOT} = 0$. Steps indicated by arrows correspond to changes in the number of electrons on the left dot. Encircled inset: the last step (∼50 pA high), with the linear background subtracted. (**d**) *Upper* panel: Coulomb peaks measured in transport current through the left dot, with $V_{DOT} = 100\,\mu\mathrm{V}$ and $V_{SD1} = V_{SD2} = 0$. *Lower* panel: changes in the number of electrons on the left dot measured with the left QPC, with $V_{SD1} = 250\,\mu\mathrm{V}$ and $V_{SD2} = V_{DOT} = 0$)

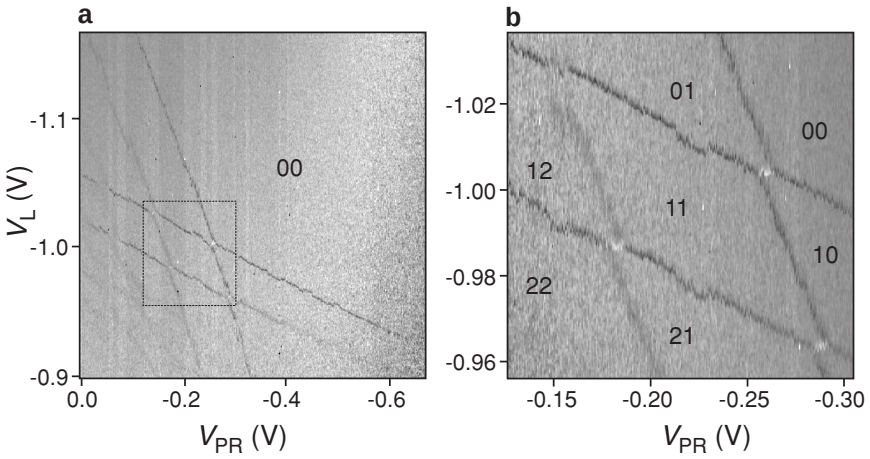

Fig. 17. Using the QPC to measure the charge configuration of a double quantum dot in the few-electron regime. (**a**) dI_{QPC}/dV_L (*in grayscale*) versus V_L and V_{PR}, with $V_{SD2} = 100\,\mu\text{V}$ and $V_{SD1} = V_{DOT} = 0$. A small modulation (0.3 mV at 17.77 Hz) is applied to V_L, and the resulting modulation in I_{QPC} is measured with a lock-in amplifier to give dI_{QPC}/dV_L directly. The label "00" indicates the region where the double dot is completely empty. In the *bottom left* corner the *dark* lines are poorly visible. Here the tunnel rates to the reservoirs are quite large, leading to smearing of the steps in the QPC current, and therefore to smaller dips in dI_{QPC}/dV_L. (**b**) Zoom-in of Fig. 17a, showing the "honeycomb" diagram for the first few electrons in the double dot. The *black* labels indicate the charge configuration, with "21" meaning 2 electrons in the left dot and 1 on the right

change in the number of electrons on the left dot, whereas almost-vertical lines indicate a change in the electron number on the right. In the upper left region the "horizontal" lines are not present, even though the QPC can still detect changes in the charge, as demonstrated by the presence of the "vertical" lines. We conclude that in this region the *left* dot contains zero electrons. Similarly, a disappearance of the "vertical" lines occurs in the lower right region, showing that here the *right* dot is empty. In the upper right region, the absence of lines shows that here the *double* dot is completely empty.

We are now able to identify the exact charge configuration of the double dot in every honeycomb cell, by simply counting the number of "horizontal" and "vertical" lines that separate it from the 00 region. In Fig. 17b the first few honeycomb cells are labelled according to their charge configuration, with e.g. the label "21" meaning 2 electrons in the left dot and 1 on the right. Besides the dark lines, also short bright lines are visible, signifying a peak in dI_{QPC}/dV_L. These bright lines correspond to an electron being transferred from one dot to the other, with the total electron number remaining the same. (The fact that some charge transitions result in a dip in dI_{QPC}/dV_L and others in a peak, derives from the fact that we use the QPC on the *right* and apply

the modulation to the gate on the *left*. When an electron is pushed out of the double dot by making V_L more negative, the QPC opens up and dI_{QPC}/dV_L displays a dip. When V_L pushes an electron from the left to the right dot, the QPC is closed slightly, resulting in a peak.) The visibility of all lines in the honeycomb pattern demonstrates that the QPC is sufficiently sensitive to detect even *inter-dot* transitions.

2.5 Tunable Tunnel Barriers in the Few-Electron Regime

In measurements of transport through lateral double quantum dots, the few-electron regime has never been reached [47]. The problem is that the gates that are used to deplete the dots also strongly influence the tunnel barriers. Reducing the electron number would therefore always lead to the Coulomb peaks becoming unmeasurably small, but not necessarily due to an empty double dot. The QPC detectors now permit us to compare charge and transport measurements.

Figure 18a shows the current through the double dot in the same region as shown in Fig. 17b. In the bottom left region the gates are not very negative, hence the tunnel barriers are quite open. Here the resonant current at the charge transition points is quite high ($\sim$100 pA, dark gray), and lines due to cotunnelling are also visible [47]. Towards the top right corner the gate voltages become more negative, thereby closing off the barriers and reducing the current peaks (lighter gray). The last "triple points" [47] that are visible ($<$1 pA) are shown in the dashed square. Using the dotted lines, extracted from the measured charge transition lines in Fig. 17b, we label the various regions in the figure according to the charge configuration of the double dot. Apart from a small shift, the dotted lines correspond nicely to the regions where a transport current is visible. This allows us to be confident that the triple points in the dashed square are really the last ones before the double quantum dot is empty. We are thus able to measure transport through a one-electron double quantum dot.

Even in the few-electron regime, the double dot remains fully tunable. By changing the voltage applied to gate T, we can make the tunnel barriers more transparent, leading to a larger current through the device. We use this procedure to increase the current at the last set of triple points. For the gate voltages used in Fig. 18b, the resonant current is very small ($<$1 pA), and the triple points are only faintly visible. By making V_T less negative, the resonant current peaks grow to about 5 pA (Fig. 18c). The two triple points are clearly resolved and the cotunnelling current is not visible. By changing V_T even more, the current at the last triple points can be increased to $\sim$70 pA (Fig. 18d). For these settings, the triple points have turned into lines, due to the increased cotunnelling current. This sequence demonstrates that we can tune the few-electron double dot from being nearly isolated from the reservoirs, to being very transparent.

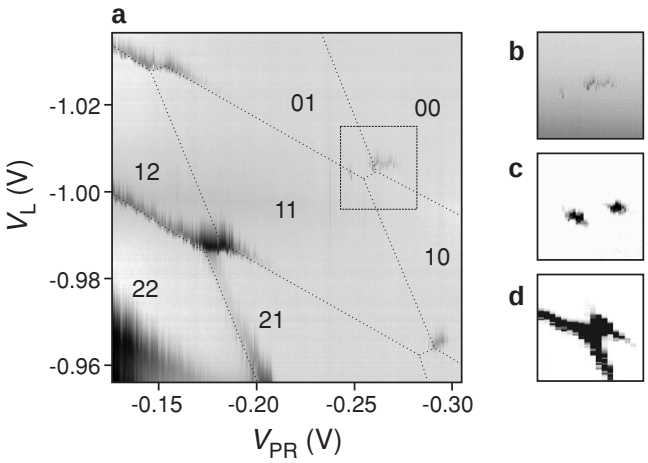

Fig. 18. Current through the double quantum dot in the few-electron regime. (**a**) I_{DOT} (in logarithmic *grayscale*) versus V_L and V_{PR} in the same region as shown in Fig. 17b, with $V_{DOT} = 100\,\mu V$ and $V_{SD1} = V_{SD2} = 0$. *Dotted* lines are extracted from Fig. 17b. *Dark gray* indicates a current flowing, with the darkest regions (in the *bottom left* corner) corresponding to ∼100 pA. In the *light gray* regions current is zero due to Coulomb blockade. Inside the *dashed* square, the last triple points are faintly visible (∼1 pA). (A smoothly varying background current due to a small leakage current from a gate to the 2DEG has been subtracted from all traces.) (**b**) Close-up of the region inside the *dashed* square in (**a**), showing the last two triple points before the double dot is completely empty. The current at these triple points is very small (<1 pA) since the tunnel barriers are very opaque. (**c**) Same two triple points for different values of the voltage applied to the gates defining the tunnel barriers. For these settings, the two individual triple points are well resolved, with a height of about 5 pA. The cotunnelling current is not visible. (**d**) Same two triple points, but now with the gate voltages such that the tunnel barriers are very transparent. The current at the triple points is about 70 pA, and the cotunnelling current is clearly visible

We can also control the inter-dot coupling, by changing the voltage applied to gate M. This is demonstrated with a QPC charge measurement (performed on sample 2). We apply a square wave modulation of 3 mV at 235 Hz to the rightmost plunger gate, P_R, and measure dI_{QPC}/dV_{PR} using a lock-in amplifier. Figure 19a shows the familiar honeycomb diagram in the few-electron regime. All lines indicating charge transitions are very straight, implying that for the gate settings used, the tunnnel-coupling between the two dots is negligible compared to the capacitive coupling. This is the so-called weak-coupling regime. (We note that the regular shape of the honeycomb pattern demonstrates that the double dot as a whole is still quite well-coupled to the leads, so that the total number of electrons can always find its lowest-energy value, unlike in [48].) By making V_M less negative, the tunnel barrier between the

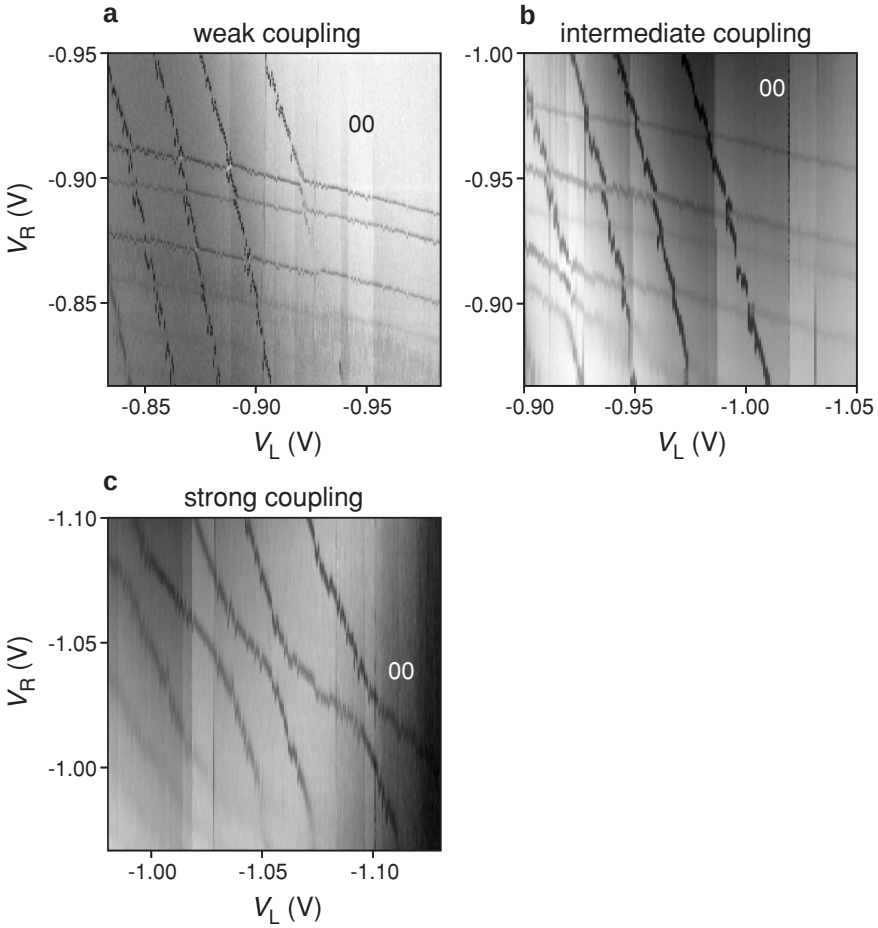

Fig. 19. Controlling the inter-dot coupling (in sample 2) with V_M. These charge stability diagrams of the double quantum dot are measured using the QPC on the left. A small modulation (3 mV at 235 Hz) is applied to gate PR, and dI_{QPC}/dV_{PR} is measured with a lock-in amplifier and plotted in *grayscale* versus V_L and V_R. A magnetic field of 6 Tesla is applied in the plane of the 2DEG. (**a**) Weak-coupling regime. V_M is such that all *dark* lines indicating charge transitions are straight. The tunnel-coupling between the two dots is therefore negligible compared to the capacitive coupling. (**b**) Intermediate-coupling regime. V_M is 0.07 V less negative than in (**a**), such that lines in the *bottom left* corner are slightly curved. This signifies that here the inter-dot tunnel-coupling is comparable to the capacitive coupling. (**c**) Strong-coupling regime. V_M is 0.1 V less negative than in (**b**), such that all lines are very curved. This implies that the tunnel-coupling is dominating over the capacitive coupling and the double dot behaves as a single dot

two dots is made more transparent, and the intermediate-coupling regime is reached (Fig. 19b). Most lines are still straight, except in the bottom left corner, where they are slightly curved. This signifies that here the inter-dot tunnel-coupling is comparable to the capacitive coupling. If we make V_M even less negative, we reach the strong-coupling regime (Fig. 19c). In this case, all lines are very curved, implying that the tunnel-coupling is dominating over the capacitive coupling. In this regime the double dot behaves like a single dot.

2.6 Photon-Assisted Tunnelling

The use of gated quantum dots for quantum state manipulation in time requires the ability to modify the potential at high frequencies. We investigate the high-frequency behavior in the region around the last triple points (Fig. 20a), with a 50 GHz microwave-signal applied to gate P_L. At the dotted line the 01 and 10 charge states are degenerate in energy, so one electron can tunnel back and forth between the two dots. Away from this line there is an energy difference and only one charge state is stable. However, if the energy difference matches the photon energy, the transition to the other dot is possible by absorption of a single photon. Such photon-assisted tunnelling events give rise to the two lines indicated by the arrows. At the lower (higher) line electrons are pumped from the the left (right) dot to the other one, giving rise to a negative (positive) photon-assisted current. We find that the distance (in terms of gate voltage) between the two photon-assisted tunnelling lines, ΔV_L, scales linearly with frequency (Fig. 20b), as expected in the weak-coupling regime [47]. From the absence of bending of the line in Fig. 20b down to a frequency of 6 GHz, it follows that the inter-dot tunnel coupling is smaller than about 12 μeV.

The realization of a controllable few-electron quantum dot circuit represents a significant step towards controlling the coherent properties of single electron spins in quantum dots [2, 49]. Integration with the QPCs permits charge read-out of closed quantum dots. We note that charge read-out only affects the spin state indirectly, via the spin-orbit interaction. The back-action on the spin should therefore be small (until spin-to-charge conversion is initiated), and can be further suppressed by switching on the charge detector only during the read-out stage. Experiments described in the following sections focus on increasing the speed of the charge measurement, such that single-shot read-out of a single electron spin can be accomplished [49, 50].

3 Excited-State Spectroscopy on a Nearly Closed Quantum Dot via Charge Detection

In this section, we demonstrate a method for measuring the discrete energy spectrum of a quantum dot connected very weakly to a single lead. A train of

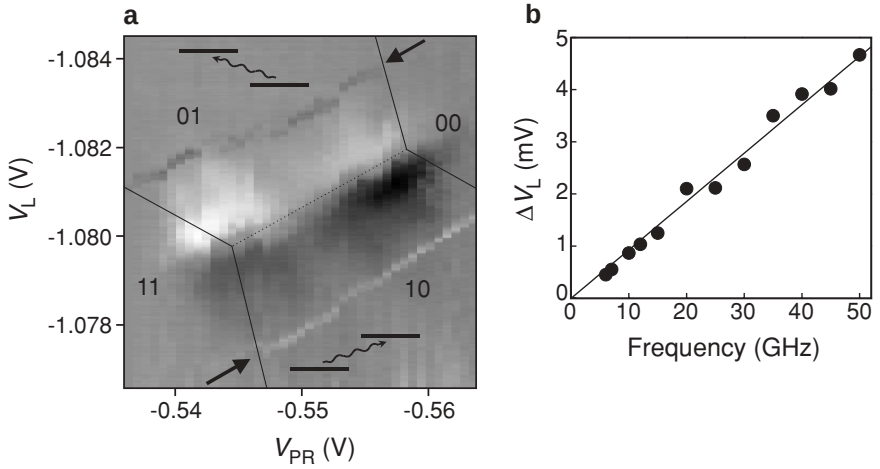

Fig. 20. Photon-assisted tunnelling in a one-electron double quantum dot. (**a**) Current through the double dot at the last set of triple points, with zero bias voltage ($V_{DOT} = V_{SD1} = V_{SD2} = 0$). A microwave signal of 50 GHz is applied to P_L. The microwaves pump a current, I_{DOT}, by absorption of single photons [47]. This photon-assisted current shows up as two lines, indicated by the two arrows. The *white* line (*bottom*) corresponds to electrons being pumped from the left to the right reservoir, the *dark* line (*top*) corresponds to pumping in the reverse direction. In the middle, around the *dotted* line separating the 01 from the 10 configuration, a finite current is induced by an unwanted voltage drop over the double dot, due to asymmetric coupling of the ac-signal to the two leads. (**b**) Separation between the two photon-assisted tunnelling lines versus microwave frequency. The dependence is linear down to the lowest frequency of about 6 GHz, from which it follows that the inter-dot tunnel coupling (half the energy difference between bonding and anti-bonding state) is smaller than $\sim$12 μeV

voltage pulses applied to a metal gate induces tunnelling of electrons between the quantum dot and a reservoir. The effective tunnel rate depends on the number and nature of the energy levels in the dot made accessible by the pulse. Measurement of the charge dynamics thus reveals the energy spectrum of the dot, as demonstrated for a dot in the few-electron regime.

3.1 Introduction

Few-electron quantum dots are considered as qubits for quantum circuits, where the quantum bit is stored in the spin or orbital state of an electron in a single or double dot. The elements in such a device must have functionalities such as initialization, one- and two-qubit operations and read-out [2]. For all these functions it is necessary to have precise knowledge of the qubit energy levels. Standard spectroscopy experiments involve electron transport through the quantum dot while varying both a gate voltage and the source-drain

voltage [27]. This requires that the quantum dot be connected to two leads with a tunnel coupling large enough to obtain a measurable current [43].

Coupling to the leads unavoidably introduces decoherence of the qubit: even if the number of electrons on the dot is fixed due to Coulomb blockade, an electron can tunnel out of the dot and be replaced by another electron through a second-order tunnelling process, causing the quantum information to be irretrievably lost. Therefore, to optimally store qubits in quantum dots, higher-order tunnelling has to be suppressed, i.e. the coupling to the leads must be made as small as possible. Furthermore, real-time observation of electron tunnelling, important for single-shot read-out of spin qubits via spin-to-charge conversion, also requires a small coupling of the dot to the leads. In this regime, current through the dot would be very hard or even impossible to measure. Therefore an alternative spectroscopic technique is needed, which does not rely on electron transport through the quantum dot.

Here we present spectroscopy measurements using charge detection. Our method resembles experiments on superconducting Cooper-pair boxes and semiconductor disks which have only one tunnel junction so that no net current can flow. Information on the energy spectrum can then be obtained by measuring the energy for adding an electron or Cooper-pair to the box, using a single-electron transistor (SET) operated as a charge detector [51, 52, 53]. We are interested in the excitation spectrum for a given number of electrons on the box, rather than the addition spectra. We use a quantum point contact (QPC) as an electrometer [44] and excitation pulses with repetition rates comparable to the tunnel rates to the lead, to measure the discrete energy spectrum of a nearly isolated one- and two-electron quantum dot.

3.2 Tuning the Tunnel Barriers

The quantum dot and QPC are defined in the two-dimensional electron gas (2DEG) in a GaAs/$Al_{0.27}Ga_{0.73}As$ heterostructure by dc voltages on gates T, M, R and Q (Fig. 21a). The dot's plunger gate, P, is connected to a coaxial cable, to which we can apply voltage pulses (rise time 1.5 ns). The QPC charge detector is operated at a conductance of about e^2/h with source-drain voltage $V_{SD} = 0.2\,\mathrm{mV}$. All data are taken with a magnetic field $B_{//} = 10$ T applied in the plane of the 2DEG, at an effective electron temperature of about 300 mK.

We first describe the procedure for setting the gate voltages such that tunnelling in and out of the dot take place through one barrier only (i.e. the other is completely closed), and the remaining tunnel rate be well controlled. For gate voltages far away from a charge transition in the quantum dot, a pulse applied to gate P (Fig. 21b) modulates the QPC current via the cross-capacitance only (solid trace in Fig. 21c). Near a charge transition, the dot can become occupied with an extra electron during the high stage of the pulse (Fig. 21d). The extra electron on the dot reduces the current through the QPC. The QPC response to the pulse is thus smaller when tunnelling takes place

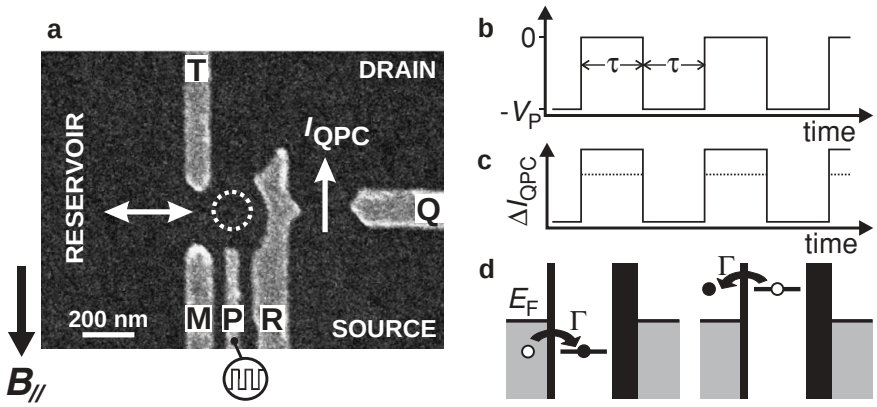

Fig. 21. QPC response to a pulse train applied to the plunger gate. (**a**) Scanning electron micrograph of a quantum dot and quantum point contact, showing only the gates used in the present experiment (the complete device is described in [55]) and Sect. 2. (**b**) Pulse train applied to gate P. (**c**) Schematic response in QPC current, ΔI_{QPC}, when the charge on the dot is unchanged by the pulse (*solid line*) or increased by one electron charge during the "high" stage of the pulse (*dashed*). (**d**) Schematic electrochemical potential diagrams during the high (*left*) and low (*right*) pulse stage, when the ground state is pulsed across the Fermi level in the reservoir, E_F

(dotted trace in Fig. 21c). We denote the amplitude of the difference between solid and dotted traces as the "electron response".

Now, even when tunnelling is allowed energetically, the electron response is only non-zero when an electron has sufficient time to actually tunnel into the dot during the pulse time, τ. By measuring the electron response as a function of τ, we can extract the tunnel rate, Γ, as demonstrated in Fig. 22a. We apply a pulse train to gate P with equal up and down times, so the repetition rate is $f = 1/(2\tau)$ (Fig. 21b). The QPC response is measured using lock-in detection at frequency f [45], and is plotted versus the dc voltage on gate M. For long pulses (lowest curves) the traces show a dip, which is due to the electron response when crossing the zero-to-one electron transition. Here, $f \ll \Gamma$ and tunnelling occurs quickly on the scale of the pulse duration. For shorter pulses the dip gradually disappears. We find analytically[1] that the dip height is proportional to $1 - \pi^2/(\Gamma^2\tau^2 + \pi^2)$, so the dip height should equal half its maximum value when $\Gamma\tau = \pi$. From the data (inset to Fig. 22a), we find that this happens for $\tau \approx 120\,\mu$s, giving $\Gamma \approx (40\,\mu\text{s})^{-1}$. Using this value

[1] This expression is obtained by multiplying the probability that the dot is empty, $P(t)$, with a sine-wave of frequency f (as is done in the lock-in amplifier), and averaging the resulting signal over one period. $P(t)$ is given by $\exp(-\Gamma t)(1 - \exp(-\Gamma\tau))/(1-\exp(-2\Gamma\tau))$ during the high stage of the pulse, and by $1-P(t-\tau)$ during the low stage.

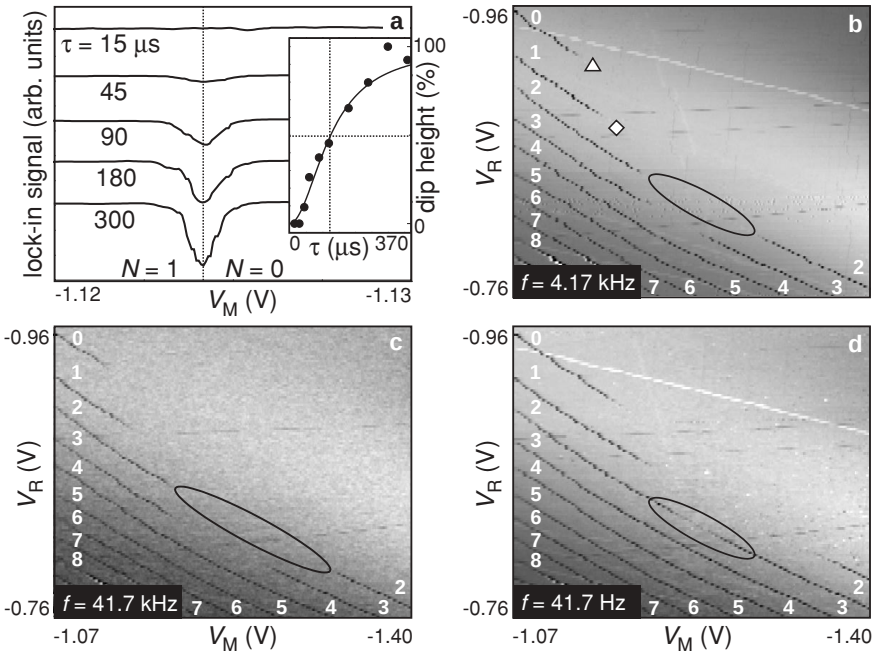

Fig. 22. Lock-in detection of electron tunnelling. (**a**) Lock-in signal at $f = 1/(2\tau)$ versus V_M for different pulse times, τ, with $V_P = 1\,\mathrm{mV}$. The dip due to the electron response disappears for shorter pulses. (Individual traces have been lined up horizontally to compensate for a fluctuating offset charge, and have been given a vertical offset for clarity.) (Inset) Height of the dip versus τ, as a percentage of the maximum height (obtained at long τ). Circles: experimental data. *Dashed* lines indicate the pulse time ($\tau \approx 120\,\mu\mathrm{s}$) for which the dip size is half its maximum value. *Solid* line: calculated dip height using $\Gamma = (40\,\mu\mathrm{s})^{-1}$. (**b**) Lock-in signal in grayscale versus V_M and V_R for $V_P = 1\,\mathrm{mV}$ and $f = 4.17\,\mathrm{kHz}$. *Dark* lines correspond to dips as in (**a**), indicating that the electron number changes by one. *White* labels indicate the absolute number of electrons on the dot. (**c**) Same plot as in (**b**), but with larger pulse repetition frequency ($f = 41.7\,\mathrm{kHz}$). (**d**) Same plot as in (**b**), but with smaller pulse repetition frequency ($f = 41.7\,\mathrm{Hz}$)

for Γ in the analytical expression given above, we obtain the solid line in the inset to Fig. 22a, which nicely matches the measured data points.

We explore several charge transitions in Fig. 22b, which shows the lock-in signal in grayscale for $\tau = 120\,\mu\mathrm{s}$, i.e. $f = 4.17\,\mathrm{kHz}$. The slanted dark lines correspond to dips as in Fig. 22a. From the absence of further charge transitions past the topmost dark line, we obtain the absolute electron number starting from zero. In the top left region of Fig. 22b, the right tunnel barrier (between gates R and T) is much more opaque than the left tunnel barrier (between M and T). Here, charge exchange occurs only with the left reservoir (indicated as "reservoir" in Fig. 21a). Conversely, in the lower right region

charge is exchanged only with the drain reservoir. In the middle region, indicated for the two-to-three electron transition by an ellipse, both barriers are too opaque and no charge can flow into or out of the dot during the 120 µs pulse; consequently the electron response becomes zero and thus the dark line disappears. For shorter pulses, i.e. larger pulse repetition frequency, the region where the dark line disappears becomes wider (ellipse in Fig. 22c). For longer pulses the dark line reappears (Fig. 22d). By varying the voltages on gates M and R, we can thus precisely set the tunnel rate to the left or right reservoir for each charge transition.

3.3 Excited-State Spectroscopy for $N = 1$

For spectroscopy measurements on a one-electron dot, we set the gate voltages near the zero-to-one electron transition at the point indicated as △ in Fig. 22b. At this point, the dot is operated as a charge box, with all tunnel events occurring through just a single barrier. The pulse repetition rate is set to 385 Hz, so that the dip height is half its maximum value. The electron response is then very sensitive to changes in the tunnel rate, which occur when an excited state becomes accessible for tunnelling.

Figure 23a shows the electron response for a pulse amplitude larger than was used for the data in Fig. 22. The dip now exhibits a shoulder on the right side (indicated by "b"), which we can understand as follows. Starting from the right ($N = 0$), the dip develops as soon as the ground state (GS) is pulsed across the Fermi level E_F and an electron can tunnel into the dot (Fig. 23b). As V_M is made less negative, we reach the point where both the GS and an excited state (ES) are pulsed across E_F (Fig. 23c). The effective rate for tunnelling on the box is now the sum of the rate for tunnelling in the GS and for tunnelling in the ES, and as a result the dip becomes deeper (the electron response increases). When V_M is made even less negative, the one-electron GS lies below E_F during both stages of the pulse, so there is always one electron on the dot. The electron response is now zero and the dip ends.

The derivative of a set of curves as in Fig. 23a is plotted in Fig. 23d. Three lines are observed. The right vertical, dark line corresponds to the right flank of the dip in Fig. 23a, the onset of tunnelling to the GS. The slanted bright line corresponds to the left flank of the dip in Fig. 23a (with opposite sign in the derivative) and reflects the pulse amplitude. The second, weaker, but clearly visible dark vertical line represents an ES. The distance between the two vertical lines is proportional to the energy difference between GS and ES.

We identify the ground and first excited state observed in this spectroscopy experiment as the spin-up and spin-down state of a single electron on the quantum dot. For $B_{//} = 10$ T, the Zeeman energy is about 0.21 meV [54], while the excitation energy of the first orbital excited state is of order 1 meV. The distance between the two vertical lines can, in principle, be converted to energy and directly provide the spin excitation energy. However, it is difficult

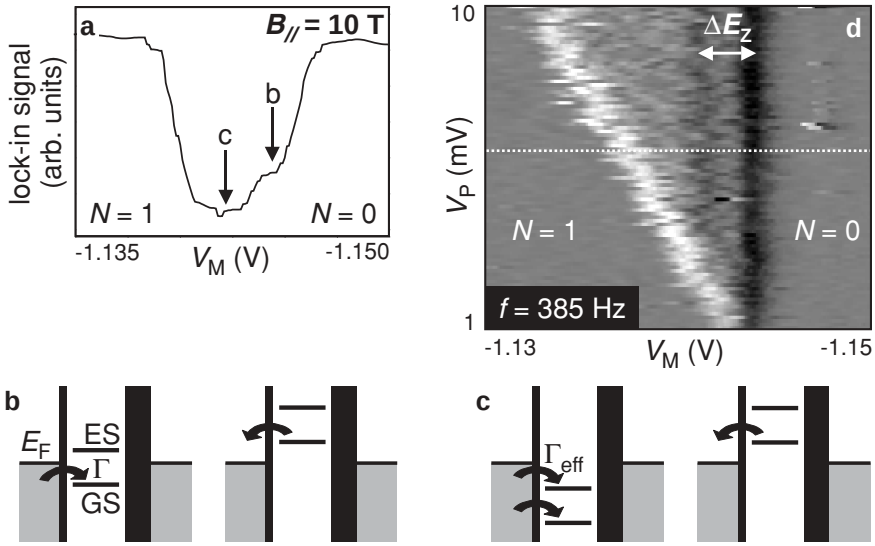

Fig. 23. Excited-state spectroscopy in a one-electron dot. (**a**) Lock-in signal at $f = 385\,\mathrm{Hz}$ versus V_M, with $V_P = 6\,\mathrm{mV}$. The dip is half the maximum value (obtained at low f and small V_P) from which we conclude that $\Gamma \approx 2.4\,\mathrm{kHz}$. (**b**) Schematic electrochemical potential diagrams for the case that only the GS is pulsed across E_F. (**c**) Idem when both the GS and an ES are pulsed across E_F. (**d**) Derivative of the lock-in signal with respect to V_M, plotted as a function of V_M and V_P (individual traces have been lined up to compensate for a fluctuating offset charge). The curve in (**a**) is taken at the *dotted* line. The Zeeman energy splitting between the one-electron GS (spin-up) and first ES (spin-down) is indicated by ΔE_Z

to determine independently the conversion factor between gate voltage and energy in this regime of a nearly closed quantum dot. Instead we take the measured Zeeman splitting from an earlier transport measurement [54] and deduce the conversion factor from gate voltage to energy, $\alpha = 105\,\mathrm{meV/V}$. This value will be used below, to convert the two-electron data to energy.

3.4 Excited-State Spectroscopy for $N = 2$

Figure 24a shows pulse spectroscopy data for the one-to-two electron transition, taken with the gate settings indicated by $\diamond$ in Fig. 22b. The rightmost vertical line corresponds to transitions between the one-electron GS (spin-up) and the two-electron GS (spin singlet) only. As V_P is increased above 5 mV, the two-electron ES (spin triplet) also becomes accessible, leading to an enhanced tunnel rate[2]. This gives rise to the left vertical line, and the distance between the two vertical lines corresponds to the singlet-triplet energy splitting ΔE_{ST}. Converted to energy, we obtain $\Delta E_{ST} = 0.49\,\mathrm{meV}$.

[2] The expected Zeeman splitting of the triplet state is not resolved here.

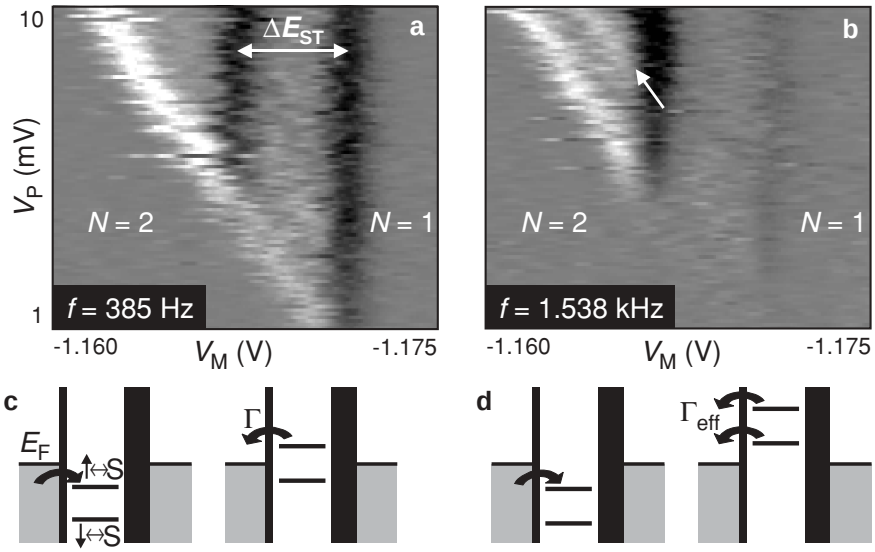

Fig. 24. Excited state spectroscopy in a two-electron dot. (**a**) Similar to Fig. 23d, but for the one-to-two electron transition. Again, $f = 385\,\text{Hz}$. We clearly observe the singlet-triplet splitting ΔE_{ST} (individual traces in (**a**) and (**b**) have been lined up). (**b**) Same experiment but with $f = 1.538\,\text{kHz}$, which increases the contrast for excited states. An extra *slanted* line appears (*arrow*), corresponding to the $N = 1$ ES, spin-down. (**c**) Schematic electrochemical potential diagram for the case that only the spin-down electron can leave from the two-electron GS (spin singlet). This occurs to the left of the *bright* line indicated by the *arrow* in (**b**). (**d**) Idem when either the spin-up or the spin-down electron can leave from the spin singlet. This occurs to the right of the *arrow* in (**b**), and leads to a larger effective tunnel rate

Excitations of the one-electron dot can be made visible at the one-to-two electron transition as well, by changing the pulse frequency to 1.538 kHz (Fig. 24b). This is too fast for electrons to tunnel if only the GS is accessible, so the rightmost line almost vanishes. However, a second slanted line becomes visible (indicated by the arrow in Fig. 24b), corresponding not to an increased tunnel rate into the dot (due to an $N = 2$ ES), but to an increased tunnel rate out of the dot (due to an $N = 1$ ES). Specifically, if the pulse amplitude is sufficiently large, either the spin-up or the spin-down electron can tunnel out of the two-electron dot. This is explained schematically in Fig. 24c and d.

Similar experiments at the transition between two and three electrons, and for tunnel rates to the reservoir ranging from 12 Hz to 12 kHz, yield similar excitation spectra.

The experiments described in this section demonstrate that an electrometer such as a QPC can reveal not only the charge state of a quantum dot, but also its tunnel coupling to the outside world and the energy level spectrum of

its internal states. We can thus access all the relevant properties of a quantum dot, even when it is almost completely isolated from the leads.

4 Real-Time Detection of Single Electron Tunnelling using a Quantum Point Contact

In this section, we observe individual tunnel events of a single electron between a quantum dot and a reservoir, using a nearby quantum point contact (QPC) as a charge meter. The QPC is capacitively coupled to the dot, and the QPC conductance changes by about 1% if the number of electrons on the dot changes by one. The QPC is voltage biased and the current is monitored with an IV-convertor at room temperature. At present, we can resolve tunnel events separated by only 8 µs, limited by noise from the IV-convertor. Shot noise in the QPC sets a 10 ns lower bound on the accessible timescales.

4.1 Charge Detectors

Fast and sensitive detection of charge has greatly propelled the study of single-electron phenomena. The most sensitive electrometer known today is the single-electron transistor (SET) [56], incorporated into a radio-frequency resonant circuit [57]. Such RF-SETs can be used for instance to detect charge fluctuations on a quantum dot, capacitively coupled to the SET island [58, 59]. Already, real-time electron tunnelling between a dot and a reservoir has been observed on a sub-µs timescale [58].

A much simpler electrometer is the quantum point contact (QPC). The conductance, G_Q, through the QPC channel is quantized, and at the transitions between quantized conductance plateaus, G_Q is very sensitive to the electrostatic environment, including the number of electrons, N, on a dot in the vicinity [44]. This property has been exploited to measure fluctuations in N in real-time, on a timescale from seconds [60] down to about 10 ms [61].

Here we demonstrate that a QPC can be used to detect single-electron charge fluctuations in a quantum dot in less than 10 µs, and analyze the fundamental and practical limitations on sensitivity and bandwidth.

4.2 Sample and Setup

The quantum dot and QPC are defined in the two-dimensional electron gas (2DEG) formed at a GaAs/$Al_{0.27}Ga_{0.73}As$ interface 90 nm below the surface, by applying negative voltages to metal surface gates (Fig. 25a). The device is attached to the mixing chamber of a dilution refrigerator with a base temperature of 20 mK, and the electron temperature is ~ 300 mK in this measurement. The dot is set near the $N = 0$ to $N = 1$ transition, with the gate voltages tuned such that the dot is isolated from the QPC drain, and has a small

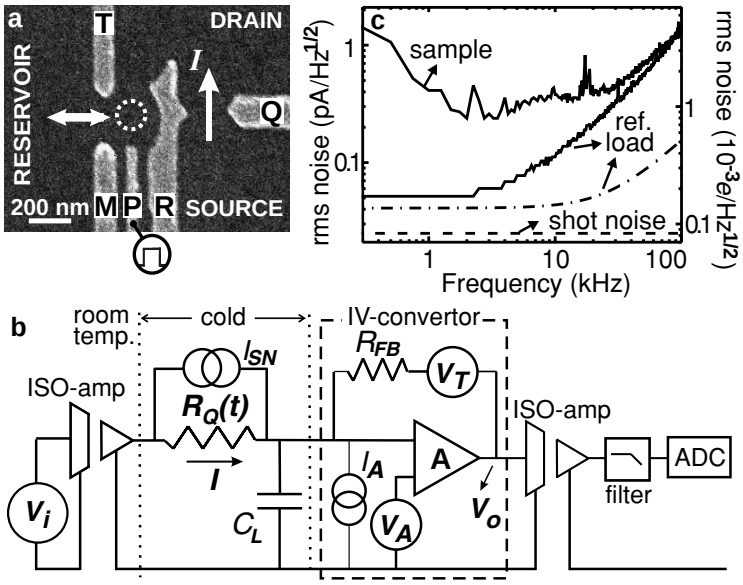

Fig. 25. Characterization of the experimental setup. (**a**) Scanning electron micrograph of a device as used in the experiment (gates which are grounded are hidden). Gates T, M and R define the quantum dot (*dotted circle*), and gates R and Q form the QPC. Gate P is connected to a pulse source via a coaxial cable. See [55] for a more detailed description. (**b**) Schematic of the experimental set-up, including the most relevant noise sources. The QPC is represented by a resistor, R_Q. (**c**) Noise spectra measured when the IV-convertor is connected to the sample (*top solid* trace), and, for reference, to an open-ended 1 m twisted pair of wires (*lower solid* trace). The latter represents a 300 pF load, if we include the 200 pF measured amplifier input capacitance. The diagram also shows the calculated noise level for the 300 pF reference load (*dotted-dashed*) and the shot noise limit (*dashed*). The *left* and *right* axes express the noise in terms of current through the QPC and electron charge on the dot respectively

tunnel rate, Γ, to the reservoir. Furthermore, the QPC conductance is set at $G_Q = 1/R_Q \approx (30\,\mathrm{k\Omega})^{-1}$, roughly halfway the transition between $G_Q = 2e^2/h$ and $G_Q = 0$, where it is most sensitive to the electrostatic environment[3].

A schematic of the electrical circuit is shown in Fig. 25b. The QPC source and drain are connected to room temperature electronics by signal wires, which run through Cu-powder filters at the mixing chamber to block high frequency noise (>100 MHz) coming from room temperature. Each signal wire is twisted with a ground wire from room temperature to the mixing chamber. A voltage, V_i, is applied to the source via a home-built opto-coupled isolation stage. The current through the QPC, I, is measured via an IV-convertor connected to the drain, and an opto-coupled isolation amplifier, both

[3] Despite a $B = 10$ T field in the plane of the 2DEG, no spin-split plateau is visible.

home-built as well. The IV-convertor is based on a dual low-noise JFET (Interfet 3602). Finally, the signal is AC-coupled to an 8th-order elliptic low-pass filter (SRS650), and the current fluctuations, ΔI, are digitized at 2.2×10^6 14-bit samples per second (ADwin Gold).

The measurement bandwidth is limited by the low-pass filter formed by the capacitance of the line and Cu-powder filters, $C_L \approx 1.5\,\mathrm{nF}$, and the input impedance of the IV-convertor, $R_i = R_{FB}/A$. Thermal noise considerations (below) impose $R_{FB} = 10\,\mathrm{M\Omega}$. We choose the amplifier gain $A = 10000$, such that $1/(2\pi R_i C_L) \approx 100\,\mathrm{kHz}$. The bandwidth of the amplifier inside the IV-convertor is 500 kHz, and the output ISO-amp bandwidth is 300 kHz. However, we shall see that the true limitation to measurement speed is not the bandwidth but the signal-to-noise ratio.

4.3 Sensitivity and Speed

The measured signal corresponding to a single electron charge on the dot amounts to $\Delta I \approx 0.3$ nA with the QPC biased at $V_i = 1\,\mathrm{mV}$, a 1% change in the overall current I ($I \approx 30\,\mathrm{nA}$, consistent with the series resistance of R_Q, $R_i = 1\,\mathrm{k\Omega}$ and the resistance of the Ohmic contacts of about $2\,\mathrm{k\Omega}$). Naturally, the signal strength is proportional to V_i, but we found that for $V_i \geq 1\,\mathrm{mV}$, the dot occupation was affected, possibly due to heating. We therefore proceed with the analysis using $I = 30\,\mathrm{nA}$ and $\Delta I = 0.3\,\mathrm{nA}$.

The most relevant noise sources [62] are indicated in the schematic of Fig. 25b. In Table 1, we give an expression and value for each noise contribution in terms of rms current at the IV-convertor input, so it can be compared directly to the signal, ΔI. We also give the corresponding value for the rms charge noise on the quantum dot. Shot noise, I_{SN}, is intrinsic to the QPC and therefore unavoidable. Both I_{SN} and ΔI are zero at QPC transmission $T = 0$ or $T = 1$, and maximal at $T = 1/2$; here we use $T \leq 1/2$. The effect of thermal noise, V_T, can be kept small compared to other noise sources by

Table 1. Contributions to the noise current at the IV-convertor input. By dividing the noise current by 300 pA (the signal corresponding to one electron charge leaving the dot), we obtain the rms charge noise on the dot

Noise Source	RMS Noise Current		RMS Charge Noise
	Expression	A/$\sqrt{\mathrm{Hz}}$	$e/\sqrt{\mathrm{Hz}}$
I_{SN}	$\sqrt{T(1-T)2eI}$	49×10^{-15}	1.6×10^{-4}
V_T	$\sqrt{4k_BT/R_{FB}}$	41×10^{-15}	1.4×10^{-4}
V_A	$V_A \frac{1+j2\pi f R_Q C_L}{R_Q}$		
V_A, low f	V_A/R_{FB}	32×10^{-15}	1.1×10^{-4}
V_A, high f	$V_A 2\pi f C_L$	$7.5 \times 10^{-18} f$	$2.5 \times 10^{-8} f$
I_A	I_A	–	–

choosing R_{FB} sufficiently large; here $R_{FB} = 10$ MΩ. The JFET input voltage noise is measured to be $V_A = 0.8\ \text{nV}/\sqrt{\text{Hz}}$. As a result of V_A, it is as if a noise current flows from the IV-convertor input leg to ground, through the QPC in parallel with the line capacitance. Due to the capacitance, C_L, the rms noise current resulting from V_A increases with frequency; it equals ΔI at 120 kHz. There is no specification available for the JFET input current noise, I_A, but usually I_A is small in JFETs.

We summarize the expected noise spectrum in Fig. 25c, and compare this with the measured noise spectrum in the same figure. For a 300 pF reference load, the noise level measured below a few kHz is $52\ \text{fA}/\sqrt{\text{Hz}}$, close to the noise current due to V_T, as expected; at high frequencies, the measured noise level is significantly higher than would be caused by V_A in combination with the 300 pF load, and appears to be dominated by I_A. With the sample connected, we observe substantial $1/f^2$ noise ($1/f$ in the noise amplitude), presumably from spurious charge fluctuations near the QPC, as well as interference at various frequencies. Near 100 kHz, the spectrum starts to roll off because of the 100 kHz low-pass filter formed by $C_L = 1.5$ nF and $R_i = 1$ kΩ (for the reference load, C_L is only 300 pF so the filter cut-off is at 500 kHz).

From the data, we see that the measured charge noise integrated from DC is comparable to e at 80 kHz, and 2.5 times smaller than e around 40 kHz. We set the cut-off frequency of the external low-pass filter at 40 kHz, so we should see clear steps in time traces of the QPC current, corresponding to single electrons tunnelling on or off the dot.

4.4 Real-Time Single Electron Tunnelling

We test this experimentally, in the regime where the electrochemical potential in the dot is nearly lined up with the electrochemical potential in the reservoir. The electron can then *spontaneously* tunnel back and forth between the dot and the reservoir, and the QPC current should exhibit a random telegraph signal (RTS). This is indeed what we observe experimentally (Fig. 26). In order to ascertain that the RTS really originates from electron tunnel events between the dot and the reservoir, we verify that (1) the dot potential relative to the Fermi level determines the fraction of the time an electron resides in the dot (Fig. 26a) and (2) the dot-reservoir tunnel barrier sets the RTS frequency (Fig. 26b). The shortest steps that clearly reach above the noise level are about 8 μs long. This is consistent with the 40 kHz filter frequency, which permits a rise time of 8 μs.

Next, we *induce* tunnel events by pulsing the dot potential, so N predictably changes from 0 to 1 and back to 0. The response of the QPC current to such a pulse contains two contributions (Fig. 27a). First, the shape of the pulse is reflected in ΔI, as the pulse gate couples capacitively to the QPC. Second, some time after the pulse is started, an electron tunnels into the dot and ΔI goes down by about 300 pA. Similarly, ΔI goes up by 300 pA when an electron leaves the dot, some time after the pulse ends. We observe that

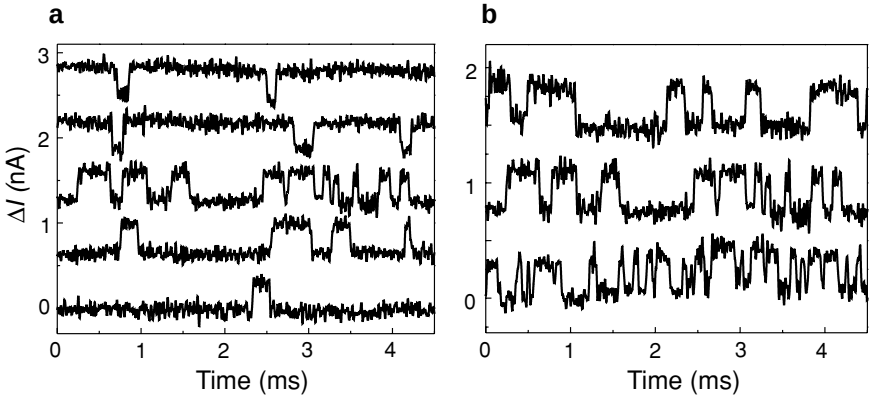

Fig. 26. Measured changes in the QPC current, ΔI, with the electrochemical potential in the dot and in the reservoir nearly equal. ΔI is "high" and "low" for 0 and 1 electrons on the dot respectively ($V_i = 1\,\mathrm{mV}$; the steps in ΔI are $\approx 300\,\mathrm{pA}$). Traces are offset for clarity. (**a**) The dot potential is lowered from *top* to *bottom*. (**b**) The tunnel barrier is lowered from *top* to *bottom*

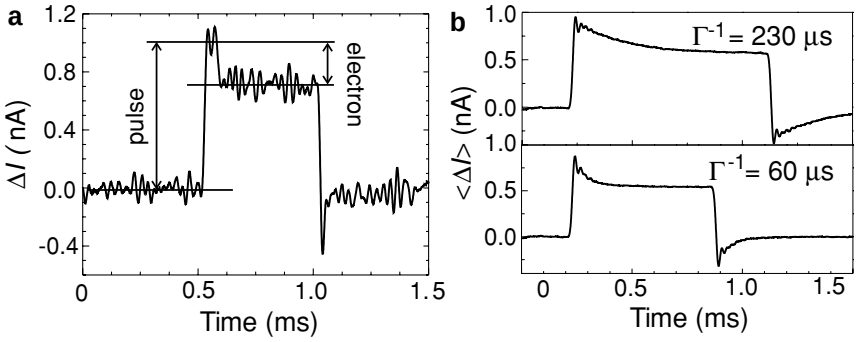

Fig. 27. QPC pulse response. (**a**) Measured changes in the QPC current, ΔI, when a pulse is applied to gate P, near the degeneracy point between 0 and 1 electrons on the dot ($V_i = 1\,\mathrm{mV}$). (**b**) Average of 286 traces as in (**a**). The *top* and *bottom* panel are taken with a different setting of gate M. The damped oscillation following the pulse edges is due to the 8th-order 40 kHz filter

the time before tunnelling takes place is randomly distributed, and obtain a histogram of this time simply by averaging over many single-shot traces (Fig. 27b). The measured distribution decays exponentially with the tunnel time, characteristic of a Poisson process. The average time before tunnelling corresponds to Γ^{-1}, and can be tuned by adjusting the tunnel barrier.

4.5 QPC Versus SET

Our measurements clearly demonstrate that a QPC can serve as a fast and sensitive charge detector. Compared to an SET, a QPC offers several practical advantages. First, a QPC requires fabrication and tuning of just a single additional gate when integrated with a quantum dot defined by metal gates, whereas an SET requires two tunnel barriers, and a gate to set the island potential. Second, QPCs are more robust and easy to use in the sense that spurious, low-frequency fluctuations of the electrostatic potential hardly change the QPC sensitivity to charges on the dot (the transition between quantized conductance plateaus has an almost constant slope over a wide range of electrostatic potential), but can easily spoil the SET sensitivity.

With an RF-SET, a sensitivity to charges on a quantum dot of $\approx 2 \times 10^{-4} e/\sqrt{\text{Hz}}$ has been reached [58], and theoretically even a ten times better sensitivity is possible [57]. Could a QPC reach similar sensitivities?

The noise level in the present measurement could be reduced by a factor of two or three using a JFET input-stage which better balances input voltage noise and input current noise. Further improvements can be obtained by lowering the capacitance of the filters in the line, or the line capacitance itself, by placing the IV-convertor close to the sample, inside the refrigerator.

Much more significant reductions in the instrumentation noise could be realized by embedding the QPC in a resonant electrical circuit and measuring the damping of the resonator. We estimate that with an "RF-QPC" and a low-temperature HEMT amplifier, a sensitivity of $2 \times 10^{-4} e/\sqrt{\text{Hz}}$ could be achieved with the present sample. The noise from the amplifier circuitry is then only 2.5 times larger than the shot noise level.

To what extent the signal can be increased is unclear, as we do not yet understand the mechanism through which the dot occupancy is disturbed for $V_i > 1\,\text{mV}$[4]. Certainly, the capacitive coupling of the dot to the QPC channel can easily be five times larger than it is now by optimizing the gate design [60]. Keeping $V_i = 1\,\text{mV}$, the sensitivity would then be $4 \times 10^{-5} e/\sqrt{\text{Hz}}$, and a single electron charge on the dot could be measured within a few ns.

Finally, we point out that a QPC can reach the quantum limit of detection [63, 64], where the measurement induced decoherence takes the minimum value permitted by quantum mechanics. Qualitatively, this is because (1) information on the charge state of the dot is transferred only to the QPC current and not to degrees of freedom which are not observed, and (2) an external perturbation in the QPC current does not couple back to the charge state of the dot.

[4] The statistics of the RTS were altered for $V_i > 1\,\text{mV}$, irrespective of (1) whether V_i was applied to the QPC source or drain, (2) the potential difference between the reservoir and the QPC source/drain, and (3) the QPC transmission T.

5 Single-Shot Read-Out of an Individual Electron Spin in a Quantum Dot

Spin is a fundamental property of all elementary particles. Classically it can be viewed as a tiny magnetic moment, but a measurement of an electron spin along the direction of an external magnetic field can have only two outcomes: parallel or anti-parallel to the field [65]. This discreteness reflects the quantum mechanical nature of spin. Ensembles of many spins have found diverse applications ranging from magnetic resonance imaging [66] to magneto-electronic devices [67], while individual spins are considered as carriers for quantum information. Read-out of single spin states has been achieved using optical techniques [68], and is within reach of magnetic resonance force microscopy [69]. However, electrical read-out of single spins [2, 49, 70, 71, 72, 73, 74, 75] has so far remained elusive. Here, we demonstrate electrical single-shot measurement of the state of an individual electron spin in a semiconductor quantum dot [40]. We use spin-to-charge conversion of a single electron confined in the dot, and detect the single-electron charge using a quantum point contact; the spin measurement visibility is ∼65%. Furthermore, we observe very long single-spin energy relaxation times (up to ∼0.85 ms at a magnetic field of 8 Tesla), which are encouraging for the use of electron spins as carriers of quantum information.

5.1 Measuring Electron Spin in Quantum Dots

In quantum dot devices, single electron charges are easily measured. Spin states in quantum dots, however, have only been studied by measuring the average signal from a large ensemble of electron spins [54, 68, 77, 78, 79, 80]. In contrast, the experiment presented here aims at a single-shot measurement of the spin orientation (parallel or antiparallel to the field, denoted as spin-↑ and spin-↓, respectively) of a particular electron; only one copy of the electron is available, so no averaging is possible. The spin measurement relies on spin-to-charge conversion [54, 79] followed by charge measurement in a single-shot mode [58, 59]. Figure 28a schematically shows a single electron spin confined in a quantum dot (circle). A magnetic field is applied to split the spin-↑ and spin-↓ states by the Zeeman energy. The dot potential is then tuned such that if the electron has spin-↓ it will leave, whereas it will stay on the dot if it has spin-↑. The spin state has now been correlated with the charge state, and measurement of the charge on the dot will reveal the original spin state.

5.2 Implementation

This concept is implemented using a structure [55] (Fig. 28b) consisting of a quantum dot in close proximity to a quantum point contact (QPC). The quantum dot is used as a box to trap a single electron, and the QPC is

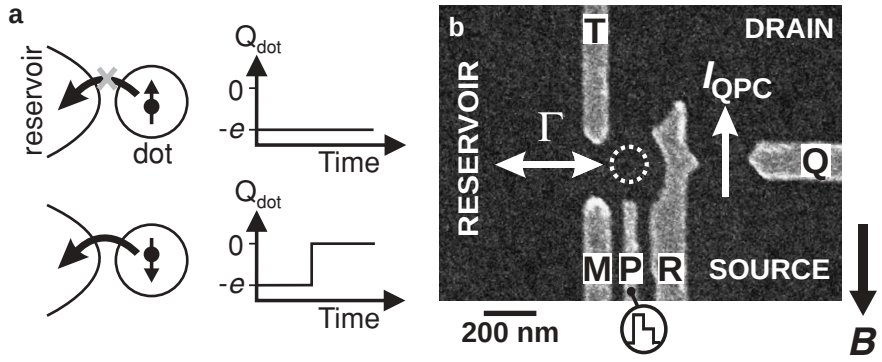

Fig. 28. Spin-to-charge conversion in a quantum dot coupled to a quantum point contact. (**a**) Principle of spin-to-charge conversion. The charge on the quantum dot, Q_{dot}, remains constant if the electron spin is ↑, whereas a spin-↓ electron can escape, thereby changing Q_{dot}. (**b**) Scanning electron micrograph of the metallic gates on the surface of a GaAs/$Al_{0.27}Ga_{0.73}As$ heterostructure containing a two-dimensional electron gas (2DEG) 90 nm below the surface. The electron density is 2.9×10^{15} m^{-2}. (Only the gates used in the present experiment are shown, the complete device is described in [55].) Electrical contact is made to the QPC source and drain and to the reservoir via Ohmic contacts. By measuring the current through the QPC channel, I_{QPC}, we can detect changes in Q_{dot} that result from electrons tunnelling between the dot and the reservoir (with a tunnel rate Γ). With a source-drain bias voltage of 1 mV, I_{QPC} is about 30 nA, and an individual electron tunnelling on or off the dot changes I_{QPC} by ∼0.3 nA. The QPC-current is sent to a room temperature current-to-voltage convertor, followed by a gain 1 isolation amplifier, an AC-coupled 40 kHz SRS650 low-pass filter, and is digitized at a rate of 2.2×10^6 samples/s. With this arrangement, the step in I_{QPC} resulting from an electron tunnelling is clearly larger than the rms noise level, provided it lasts at least 8 µs. A magnetic field, B, is applied in the plane of the 2DEG

operated as a charge detector in order to determine whether the dot contains an electron or not. The quantum dot is formed in the two-dimensional electron gas (2DEG) of a GaAs/AlGaAs heterostructure by applying negative voltages to the metal surface gates M, R, and T. This depletes the 2DEG below the gates and creates a potential minimum in the centre, that is, the dot (indicated by a dotted white circle). We tune the gate voltages such that the dot contains either zero or one electron (which we can control by the voltage applied to gate P). Furthermore, we make the tunnel barrier between gates R and T sufficiently opaque that the dot is completely isolated from the drain contact on the right. The barrier to the reservoir on the left is set [81] to a tunnel rate $\Gamma \approx (0.05\,\mathrm{ms})^{-1}$. When an electron tunnels on or off the dot, it changes the electrostatic potential in its vicinity, including the region of the nearby QPC (defined by R and Q). The QPC is set in the tunnelling regime, so that the current, I_{QPC}, is very sensitive to electrostatic changes [44]. Recording

changes in I_{QPC} thus permits us to measure on a timescale of about 8 µs whether an electron resides on the dot or not (L.M.K.V. et al., see Sect. 4). In this way the QPC is used as a charge detector with a resolution much better than a single electron charge and a measurement timescale almost ten times shorter than $1/\Gamma$.

The device is placed inside a dilution refrigerator, and is subject to a magnetic field of 10 T (unless noted otherwise) in the plane of the 2DEG. The measured Zeeman splitting in the dot [54], $\Delta E_Z \approx 200\,\mu\text{eV}$, is larger than the thermal energy (25 µeV) but smaller than the orbital energy level spacing (1.1 meV) and the charging energy (2.5 meV).

5.3 Two-Level Pulse Technique

To test our single-spin measurement technique, we use an experimental procedure based on three stages: (1) empty the dot, (2) inject one electron with unknown spin, and (3) measure its spin state. The different stages are controlled by voltage pulses on gate P (Fig. 29a), which shift the dot's energy levels (Fig. 29c). Before the pulse the dot is empty, as both the spin-↑ and spin-↓ levels are above the Fermi energy of the reservoir, E_F. Then a voltage pulse pulls both levels below E_F. It is now energetically allowed for an electron to tunnel onto the dot, which will happen after a typical time $\sim\Gamma^{-1}$. The particular electron can have spin-↑ (shown in the lower diagram) or spin-↓ (upper diagram). (The tunnel rate for spin-↑ electrons is expected to be larger than that for spin-↓ electrons [82], i.e. $\Gamma_\uparrow > \Gamma_\downarrow$, but we do not assume this a priori.) During this stage of the pulse, lasting t_{wait}, the electron is trapped on the dot and Coulomb blockade prevents a second electron to be added. After t_{wait} the pulse is reduced, in order to position the energy levels in the read-out configuration. If the electron spin is ↑, its energy level is below E_F, so the electron remains on the dot. If the spin is ↓, its energy level is above E_F, so the electron tunnels to the reservoir after a typical time $\sim\Gamma_\downarrow^{-1}$. Now Coulomb blockade is lifted and an electron with spin-↑ can tunnel onto the dot. This occurs on a timescale $\sim\Gamma_\uparrow^{-1}$ (with $\Gamma = \Gamma_\uparrow + \Gamma_\downarrow$). After t_{read}, the pulse ends and the dot is emptied again.

The expected QPC-response, ΔI_{QPC}, to such a two-level pulse is the sum of two contributions (Fig. 29b). First, due to a capacitive coupling between pulse-gate and QPC, ΔI_{QPC} will change proportionally to the pulse amplitude. Thus, ΔI_{QPC} versus time resembles a two-level pulse. Second, ΔI_{QPC} tracks the charge on the dot, i.e. it goes up whenever an electron tunnels off the dot, and it goes down by the same amount when an electron tunnels on the dot. Therefore, if the dot contains a spin-↓ electron at the start of the read-out stage, ΔI_{QPC} should go up and then down again. We thus expect a characteristic step in ΔI_{QPC} during t_{read} for spin-↓ (dotted trace inside gray circle). In contrast, ΔI_{QPC} should be flat during t_{read} for a spin-↑ electron. Measuring whether a step is present or absent during the read-out stage constitutes our spin measurement.

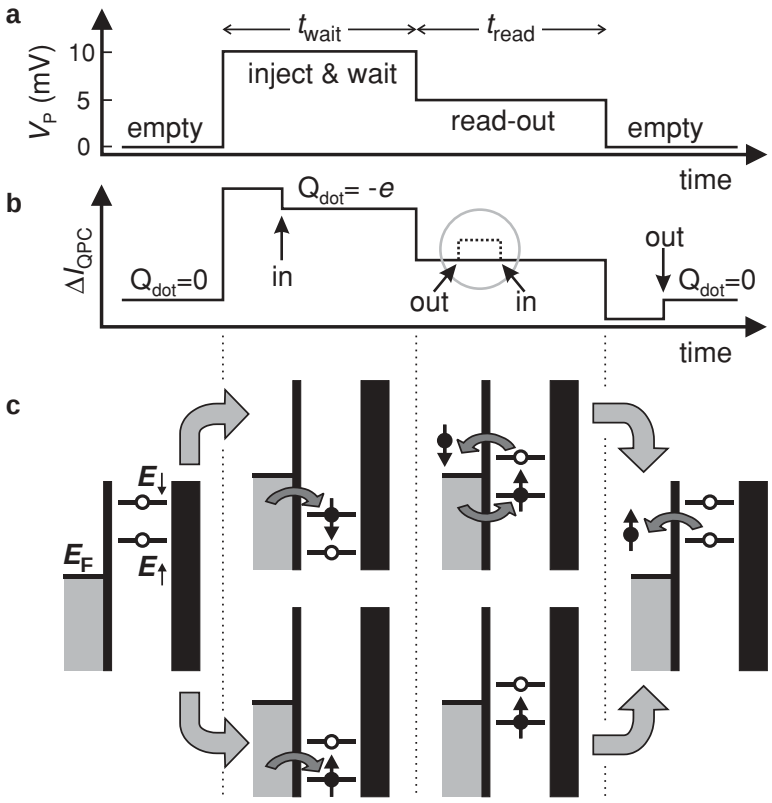

Fig. 29. Two-level pulse technique used to inject a single electron and measure its spin orientation. (**a**) Shape of the voltage pulse applied to gate P. The pulse level is 10 mV during t_{wait} and 5 mV during t_{read} (which is 0.5 ms for all measurements). (**b**) Schematic QPC pulse-response if the injected electron has spin-↑ (*solid line*) or spin-↓ (*dotted line*; the difference with the *solid* line is only seen during the read-out stage). *Arrows* indicate the moment an electron tunnels into or out of the quantum dot. (**c**) Schematic energy diagrams for spin-↑ ($E_\uparrow$) and spin-↓ ($E_\downarrow$) during the different stages of the pulse. *Black vertical* lines indicate the tunnel barriers. The tunnel rate between the dot and the QPC-drain on the right is set to zero. The rate between the dot and the reservoir on the left is tuned to a specific value, Γ. If the spin is ↑ at the start of the read-out stage, no change in the charge on the dot occurs during t_{read}. In contrast, if the spin is ↓, the electron can escape and be replaced by a spin-↑ electron. This charge transition is detected in the QPC-current (*dotted line* inside *gray* circle in (**b**))

5.4 Tuning the Quantum Dot into the Read-Out Configuration

To perform spin read-out, V_M has to be fine-tuned so that the position of the energy levels with respect to E_F is as shown in Fig. 29c. To find the correct settings, we apply a two-level voltage pulse and measure the QPC-response

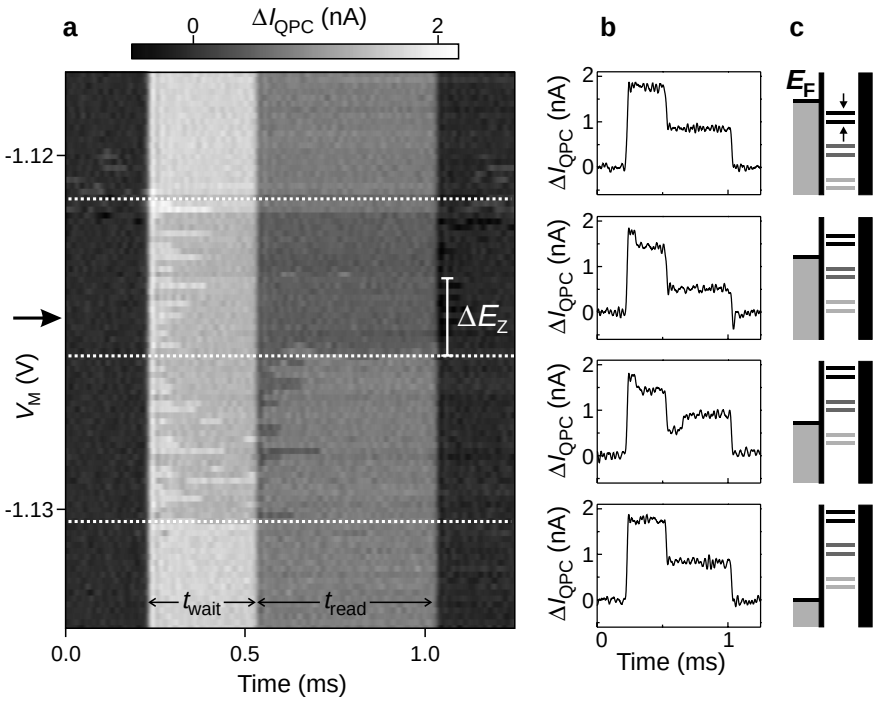

Fig. 30. Tuning the quantum dot into the spin read-out configuration. We apply a two-stage voltage pulse as in Fig. 29a ($t_{wait} = 0.3\,\text{ms}$, $t_{read} = 0.5\,\text{ms}$), and measure the QPC-response for increasingly negative values of V_M. (**a**) QPC-response (in colour-scale) versus V_M. Four different regions in V_M can be identified (separated by *white dotted lines*), with qualitatively different QPC-responses. (**b**) Typical QPC-response in each of the four regions. This behaviour can be understood from the energy levels during all stages of the pulse. (**c**) Schematic energy diagrams showing $E_\uparrow$ and $E_\downarrow$ with respect to E_F before and after the pulse (*upper pair*), during t_{wait} (*lower pair*) and during t_{read} (*middle pair*), for four values of V_M. For the actual spin read-out experiment, V_M is set to the optimum position (indicated by the *arrow* in a)

for increasingly negative values of V_M (Fig. 30a). Four different regions in V_M can be identified (separated by white dotted lines), with qualitatively different QPC-responses. The shape of the typical QPC-response in each of the four regions (Fig. 30b) allows us to infer the position of $E_\uparrow$ and $E_\downarrow$ with respect to E_F during all stages of the pulse (Fig. 30c).

In the top region, the QPC-response just mimics the applied two-level pulse, indicating that here the charge on the dot remains constant throughout the pulse. This implies that $E_\uparrow$ remains below E_F for all stages of the pulse, thus the dot remains occupied with one electron. In the second region from the top, tunnelling occurs, as seen from the extra steps in ΔI_{QPC}. The dot is empty before the pulse, then an electron is injected during t_{wait}, which escapes

after the pulse. This corresponds to an energy level diagram similar to before, but with $E_\uparrow$ and $E_\downarrow$ shifted up due to the more negative value of V_M in this region. In the third region from the top, an electron again tunnels on the dot during t_{wait}, but now it can escape already during t_{read}, irrespective of its spin. Finally, in the bottom region no electron-tunnelling is seen, implying that the dot remains empty throughout the pulse.

Since we know the shift in V_M corresponding to shifting the energy levels by ΔE_Z, we can set V_M to the optimum position for the spin read-out experiment (indicated by the arrow). For this setting, the energy levels are as shown in Fig. 29c, i.e. E_F is approximately in the middle between $E_\uparrow$ and $E_\downarrow$ during the read-out stage.

5.5 Single-Shot Read-Out of One Electron Spin

Figure 31a shows typical experimental traces of the pulse-response recorded after proper tuning of the DC gate voltages (see Fig. 30). We emphasize that each trace involves injecting one particular electron on the dot and subsequently measuring its spin state. Each trace is therefore a single-shot measurement. The traces we obtain fall into two different classes; most traces qualitatively resemble the one in the top panel of Fig. 31a, some resemble the one in the bottom panel. These two typical traces indeed correspond to the signals expected for a spin-↑ and a spin-↓ electron (Fig. 29b), a strong indication that the electron in the top panel of Fig. 31a was spin-↑ and in the bottom panel spin-↓. The distinct signature of the two types of responses in ΔI_{QPC} permits a simple criterion for identifying the spin[5]: if ΔI_{QPC} goes above the threshold value (red line in Fig. 31a and chosen as explained below), we declare the electron "spin-down"; otherwise we declare it "spin-up". Figure 31b shows the read-out section of twenty more "spin-down" traces, to illustrate the stochastic nature of the tunnel events.

The random injection of spin-↑ and spin-↓ electrons prevents us from checking the outcome of any individual measurement. Therefore, in order to further establish the correspondence between the actual spin state and the outcome

[5] The automated data analysis procedure first corrects for the offset of each trace. This offset, resulting from low-frequency interference signals or charge switches, is found by making a histogram of the QPC current during the read-out stage of a particular trace. The histogram typically displays a peak due to fluctuations around the average value corresponding to an occupied dot. The center of a gaussian fit to the histogram gives the offset. Then each trace is checked to make sure that an electron was injected during the injection stage, by evaluating if the signal goes below the injection threshold (dotted horizontal line in Fig. 33a). If not, the trace is disregarded. Finally, to determine if a trace corresponds to "spin-up" or "spin-down", we disregard all points that lie below the previous point (since these could correspond to points on the falling pulse flank at the end of the injection stage), and check if any of the remaining points are above the threshold.

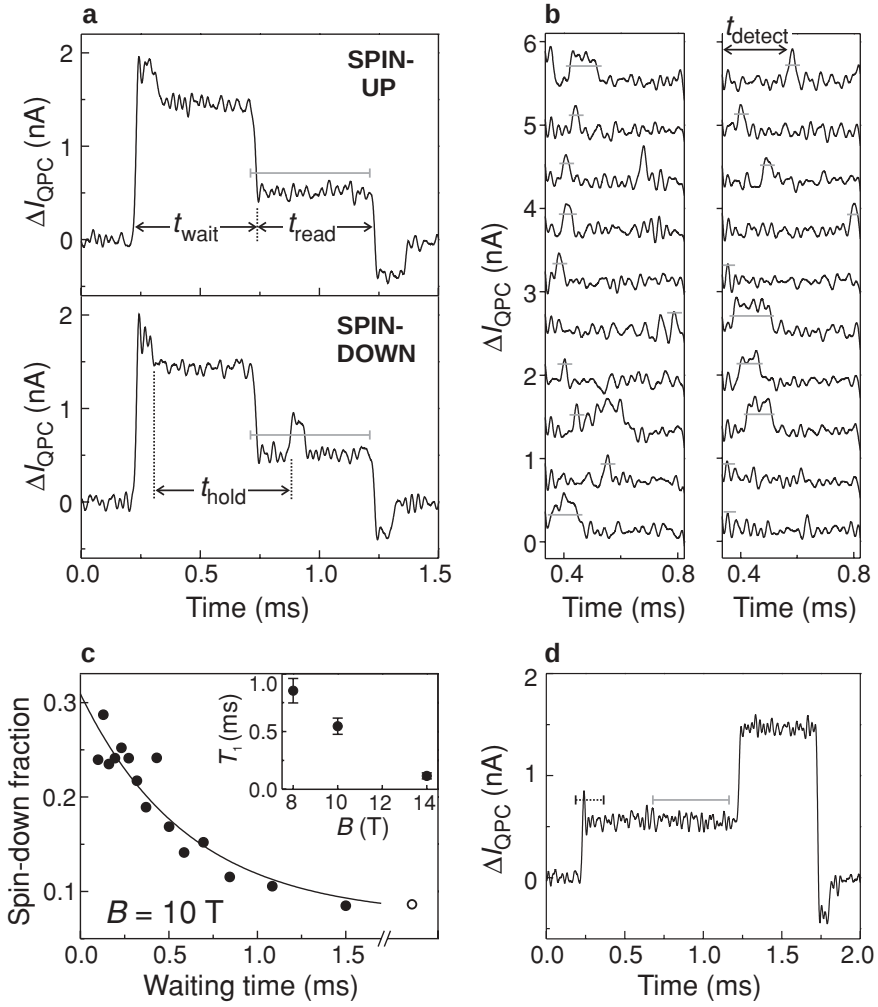

Fig. 31. Single-shot read-out of one electron spin. (**a**) Time-resolved QPC measurements. *Top* panel: an electron injected during t_{wait} is declared "spin-up" during t_{read}. *Bottom* panel: the electron is declared "spin-down". (**b**) Examples of "spin-down" traces (for $t_{wait} = 0.1$ ms). Only the read-out segment is shown, and traces are offset for clarity. The time when ΔI_{QPC} first crosses the threshold, t_{detect}, is recorded to make the histogram in Fig. 34a. (**c**) Fraction of "spin-down" traces versus t_{wait}, out of 625 traces for each waiting time. *Open* dot: spin-down fraction using modified pulse shape (**d**). *Solid* line: exponential fit to the data. Inset: T_1 versus B. (**d**) Typical QPC-signal for a "reversed" pulse, with the same amplitudes as in Fig. 29a, but a reversed order of the two stages. The leftmost threshold (*dotted line*) is used in Fig. 34b

of our spin measurement, we change the probability to have a spin-↓ at the beginning of the read-out stage, and compare this with the fraction of traces in which the electron is declared "spin-down". As t_{wait} is increased, the time between injection and read-out, t_{hold}, will vary accordingly ($t_{hold} \approx t_{wait}$). The probability for the spin to be ↓ at the start of t_{read} will thus decay exponentially to zero, since electrons in the excited spin state will relax to the ground state ($k_B T \ll \Delta E_Z$). For a set of 15 values of t_{wait} we take 625 traces for each t_{wait}, and count the fraction of traces in which the electron is declared "spin-down" (Fig. 31c). The fact that the expected exponential decay is clearly reflected in the data confirms the validity of the spin read-out procedure.

We extract a single-spin energy relaxation time, T_1, from fitting the datapoints in Fig. 31c (and two other similar measurements) to $\alpha + C\exp(-t_{wait}/T_1)$, and obtain an average value of $T_1 \approx (0.55 \pm 0.07)$ ms at 10 Tesla. This is an order of magnitude longer than the lower bound on T_1 established earlier [54], and clearly longer than the time needed for the spin measurement (of order $1/\Gamma_\downarrow \approx 0.11$ ms). A similar experiment at 8 Tesla gives $T_1 \approx (0.85 \pm 0.11)$ ms and at 14 Tesla we find $T_1 \approx (0.12 \pm 0.03)$ ms (Fig. 32). More experiments are needed in order to test the theoretical prediction that relaxation at high magnetic fields is dominated by spin-orbit interactions [22, 83, 84], with smaller contributions resulting from hyperfine interactions with the nuclear spins [83, 85] (cotunnelling is insignificant given the very small tunnel rates). We note that the obtained values for T_1 refer to our entire device under active operation: i.e. a single spin in a quantum dot subject to continuous charge detection by a QPC.

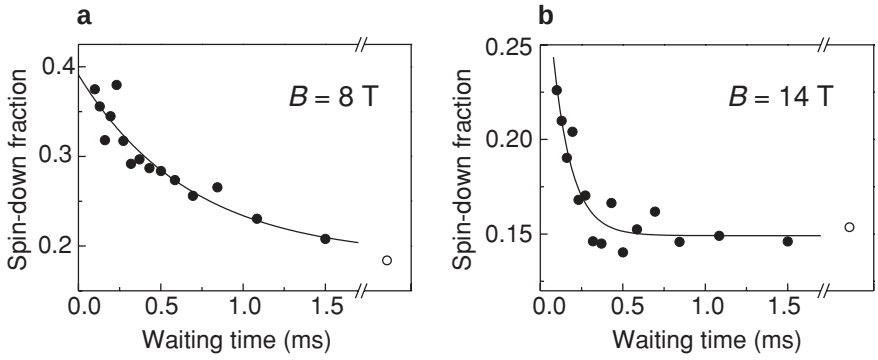

Fig. 32. Measurement of the spin-relaxation time as in Fig. 31c, but at different magnetic fields. Averaging the results of an exponential fit (as shown) over three similar measurements yields (**a**), $T_1 = (0.85 \pm 0.11)$ ms at 8 T and (**b**), $T_1 = (0.12 \pm 0.03)$ ms at 14 T

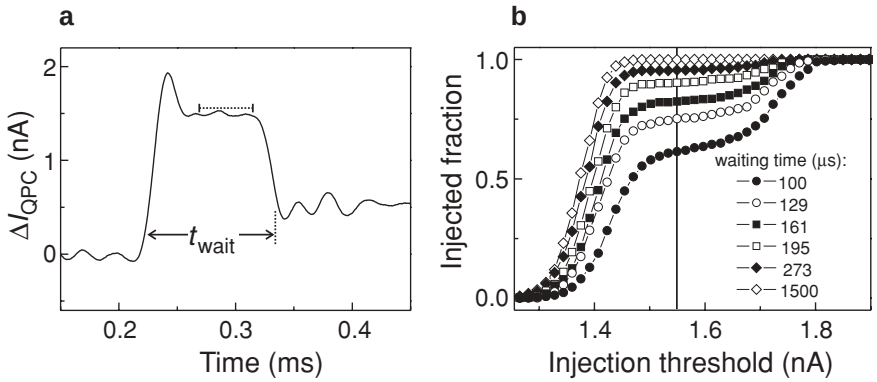

Fig. 33. Setting the injection threshold. (**a**) Example of QPC-signal for the shortest waiting time used (0.1 ms). The *dotted horizontal* line indicates the injection threshold. Injection is declared successful if the QPC-signal is below the injection threshold for a part or all of the last 45 μs before the end of the injection stage (t_{wait}). Traces in which injection was not successful, i.e. no electron was injected during t_{wait}, are disregarded. (**b**) Fraction of traces in which injection was successful, out of a total of 625 taken for each waiting time. The threshold chosen for analysing all data is indicated by the *vertical* line

5.6 Measurement Fidelity

For applications in quantum information processing it is important to know the accuracy, or fidelity, of the single-shot spin read-out. The measurement fidelity is characterised by two parameters, α and β (inset to Fig. 34a), which we now determine for the data taken at 10 T.

The parameter α corresponds to the probability that the QPC-current exceeds the threshold even though the electron was actually spin-↑, for instance due to thermally activated tunnelling or electrical noise (similar to "dark counts" in a photon detector). The combined probability for such processes is given by the saturation value of the exponential fit in Fig. 31c, α, which depends on the value of the threshold current. We analyse the data in Fig. 31c using different thresholds, and plot α in Fig. 34b.

The parameter β corresponds to the probability that the QPC-current stays below the threshold even though the electron was actually spin-↓ at the start of the read-out stage. Unlike α, β cannot be extracted directly from the exponential fit (note that the fit parameter $C = p(1-\alpha-\beta)$ contains two unknowns: $p = \Gamma_\downarrow/(\Gamma_\uparrow+\Gamma_\downarrow)$ and β). We therefore estimate β by analysing the two processes that contribute to it. First, a spin-↓ electron can relax to spin-↑ before spin-to-charge conversion takes place. This occurs with probability $\beta_1 = 1/(1+T_1\Gamma_\downarrow)$. From a histogram (Fig. 34a) of the actual detection time, t_{detect} (see Fig. 31b), we find $\Gamma_\downarrow^{-1} \approx 0.11$ ms, yielding $\beta_1 \approx 0.17$. Second, if the spin-↓ electron does tunnel off the dot but is replaced by a spin-↑ electron within about 8 μs, the resulting QPC-step is too small to be detected. The

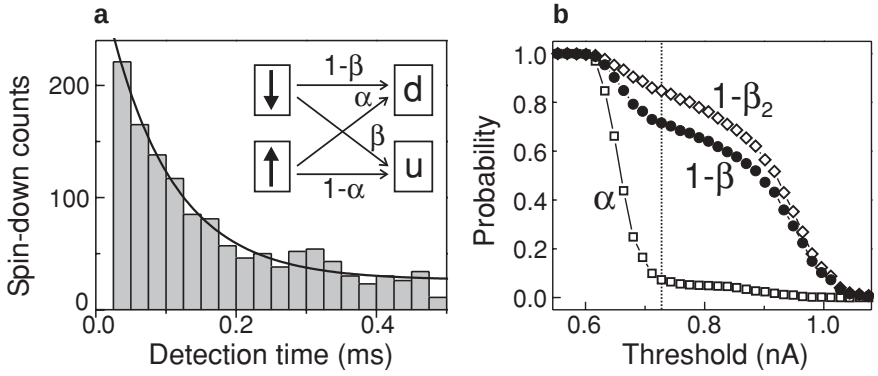

Fig. 34. Measurement fidelity. (**a**) Histogram showing the distribution of detection times, t_{detect}, in the read-out stage (see Fig. 31b for definition t_{detect}). The exponential decay is due to spin-↓ electrons tunnelling out of the dot (rate = $\Gamma_\downarrow$) and due to spin flips during the read-out stage (rate = $1/T_1$). *Solid* line: exponential fit with a decay time $(\Gamma_\downarrow + 1/T_1)^{-1}$ of 0.09 ms. Given that $T_1 = 0.55$ ms, this yields $\Gamma_\downarrow^{-1} \approx 0.11$ ms. Inset: fidelity parameters. A spin-↓ electron is declared "down" (d) or "up" (u) with probability $1-\beta$ or β, respectively. A spin-↑ electron is declared "up" or "down" with probability $1-\alpha$ or α, respectively. (**b**) *Open* squares represent α, obtained from the saturation value of exponential fits as in Fig. 31c for different values of the read-out threshold. A current of 0.54 nA (0.91 nA) corresponds to the average value of ΔI_{QPC} when the dot is occupied (empty) during t_{read}. *Open* diamonds: measured fraction of "reverse-pulse" traces in which ΔI_{QPC} crosses the injection threshold (*dotted black line* in Fig. 31d). This fraction approximates $1-\beta_2$, where β_2 is the probability of identifying a spin-↓ electron as "spin-up" due to the finite bandwidth of the measurement setup. *Filled* circles: total fidelity for the spin-↓ state, $1-\beta$, calculated using $\beta_1 = 0.17$. The *vertical dotted* line indicates the threshold for which the visibility $1-\alpha-\beta$ (separation between *filled* circles and *open* squares) is maximal. This threshold value of 0.73 nA is used in the analysis of Fig. 31

probability that a step is missed, β_2, depends on the value of the threshold. It can be determined by applying a modified ("reversed") pulse (Fig. 31d). For such a pulse, we know that in each trace an electron is injected in the dot, so there should always be a step at the start of the pulse. The fraction of traces in which this step is nevertheless missed, i.e. ΔI_{QPC} stays below the threshold (dotted black line in Fig. 31d), gives β_2. We plot $1-\beta_2$ in Fig. 34b (open diamonds). The resulting total fidelity for spin-↓ is given by $1-\beta \approx (1-\beta_1)(1-\beta_2)+(\alpha\beta_1)$. The last term accounts for the case when a spin-↓ electron is flipped to spin-↑, but there is nevertheless a step in ΔI_{QPC} due to the dark-count mechanism[6]. In Fig. 34b we also plot the extracted value of $1-\beta$ as a function of the threshold.

[6] Let us assume there is a spin-↓ electron on the dot at the start of the read-out stage. The probability that the ↓-electron tunnels out (i.e. that it does not relaxto

We now choose the optimal value of the threshold as the one for which the visibility $1-\alpha-\beta$ is maximal (dotted vertical line in Fig. 34b). For this setting, $\alpha \approx 0.07$, $\beta_1 \approx 0.17$, $\beta_2 \approx 0.15$, so the measurement fidelity for the spin-$\uparrow$ and the spin-$\downarrow$ state is $\sim$0.93 and $\sim$0.72 respectively. The measurement visibility in a single-shot measurement is thus at present 65%.

Significant improvements in the spin measurement visibility can be made by lowering the electron temperature (smaller α) and especially by making the charge measurement faster (smaller β). Already, the demonstration of single-shot spin read-out and the observation of T_1 of order 1 ms are encouraging results for the use of electron spins as quantum bits.

6 Semiconductor Few-Electron Quantum Dots as Spin Qubits

In the previous sections we have described experiments aimed at creating a quantum dot spin qubit according to the proposal by Loss and DiVincenzo [2] (see also paragraph 1.3). The key ingredients for these experiments – performed over the last two years – are a fully tunable few-electron double quantum dot and a quantum point contact (QPC) charge detector. We have operated the QPC in three different ways:

1. By measuring its DC conductance, changes in the average charge on the double dot are revealed, which can be used to identify the charge configuration of the system.
2. By measuring the conductance in real-time (with a bandwidth of $\sim$100 kHz), we can detect individual electrons tunnelling on or off the dot (in less than 10 µs).
3. By measuring the QPC response to a gate voltage pulse train (with the proper frequency) using a lock-in amplifier, we can determine the tunnel rate between the dot and a reservoir. In addition, by using a large pulse amplitude and measuring changes in the effective tunnel rate, we can identify excited states of the dot.

Using these techniques, we have demonstrated that our GaAs/AlGaAs quantum dot circuit is a promising candidate for a spin qubit. However, we do not have a fully functional qubit yet, as coherent manipulation of a single- or a two-spin system has so far remained elusive. In this section, we evaluate the experimental status of the spin qubit project in terms of the DiVincenzo

spin-$\uparrow$) is given by $1-\beta_1$. The probability that this tunnel event is detected (i.e. is not too fast) is given by $1-\beta_2$. Therefore, the probability that a spin-$\downarrow$ electron tunnels out and is detected, is $(1-\beta_1)(1-\beta_2)$. In addition, there is the possibility that the $\downarrow$-electron relaxes, with probability β_1, but a step in the QPC signal is nevertheless detected, with probability α, due to the "dark count" mechanism. Therefore, the total probability that a spin-$\downarrow$ electron is declared "spin-down" is given by $(1-\beta_1)(1-\beta_2)+(\alpha\beta_1)$ approximately.

requirements [17]. Fabrication and characterization of a double quantum dot containing two coupled spins has been achieved, as well as initialization and single-shot read-out of the spin state. The single-spin relaxation time was found to be very long, but the decoherence time is still unknown. We present concrete ideas on how to proceed towards coherent spin operations. Single-spin manipulation relies on a microfabricated wire located close to the quantum dot, and two-spin interactions are controlled via the tunnel barrier connecting the respective quantum dots. To demonstrate superposition and entanglement of spin states, we plan to use a charge detection approach, without relying on transport measurements.

6.1 Qubit

The first of the five DiVincenzo requirements is to have a scalable physical system with well-characterized qubits. We have fabricated double quantum dot devices in which a single electron can be confined in each of the two dots (see Sect. 2). The spin states $|\uparrow\rangle$ and $|\downarrow\rangle$ of the electron, subject to a large magnetic field B, correspond to the two states of the proposed qubit two-level system. The Zeeman splitting, ΔE_Z, between the two states can be tuned with the magnetic field, according to $\Delta E_Z = g\mu_B B$, with $g \approx -0.44$ the electron g-factor in GaAs [54], and μ_B the Bohr magneton.

These one-electron dots can be fully characterized using a QPC as a charge detector, with the techniques developed in Sects. 2 and 3. First of all, we can use the QPC to monitor the charge configuration of the double dot, in order to reach the regime where both dots contain just a single electron. Then we can evaluate and tune the tunnel rate from each dot to the reservoir using the lock-in technique described above. The same technique can be employed to determine the energy spectrum of each of the two dots, i.e. the Zeeman splitting between the two qubit states, as well as the energy of orbital excited states. Furthermore, the QPC can be used to monitor the inter-dot tunnel barrier, both qualitatively (from the curvature of lines in the honeycomb diagram, as shown in Fig. 2.6) and quantitatively (by performing photon-assisted tunnelling spectroscopy to measure the tunnel splitting between the one-electron bonding and anti-bonding state, as in [86]). In principle, it is even possible to use the lock-in technique to measure the exchange splitting J between the delocalized two-electron singlet and triplet spin states. However, in practical situations the splitting might be too small ($<20\,\mu$eV) to be resolved using tunnelling spectroscopy.

We can thus determine all relevant parameters of the two-spin system without performing transport measurements. The essential advantage of the QPC technique is that it works even for a dot that is very weakly coupled to just a *single* reservoir, with a tunnel rate between zero and $\sim$100 kHz (limited by the bandwidth of the current measurement setup). This gives us more freedom to design simpler dots with fewer gates, which could therefore be easier to operate.

6.2 Read-Out

We have achieved single-shot read-out of the spin orientation of an individual electron in a quantum dot (see Sect. 5). Our approach utilizes the Zeeman splitting, induced by a large magnetic field parallel to the 2DEG, to create spin-to-charge conversion (Fig. 35a). This is followed by real-time detection of single-electron tunnelling events using the QPC. The total visibility of the spin measurement is ∼65%, limited mostly by the ∼40 kHz bandwidth of our current measurement setup, and also by thermal excitation of electrons out of the quantum dot, due to the high effective electron temperature of ∼300 mK.

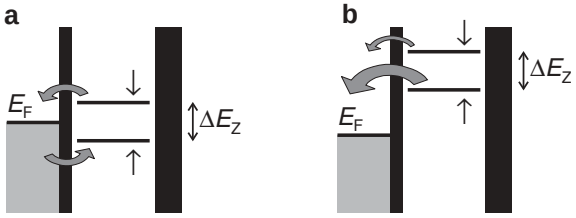

Fig. 35. Schematic energy diagrams depicting spin-to-charge conversion based on a difference in *energy* (**a**) between $|\uparrow\rangle$ and $|\downarrow\rangle$, or on a difference in *tunnel rate* (**b**)

We estimate that we can improve the visibility of the spin read-out technique to more than 90% by lowering the electron temperature below 100 mK, and especially by using a faster way to measure the charge on the dot. This could be possible with a "radio-frequency QPC" (RF-QPC), similar to the well-known RF-SET [57]. In this approach, the QPC is embedded in an LC circuit with a resonant frequency of ∼1 GHz. By measuring the reflection or transmission of a resonant carrier wave, we estimate that it should be possible to read out the charge state of the nearby quantum dot in ∼1 μs, an order of magnitude faster than is currently attainable.

A disadvantage of the read-out technique based on the Zeeman splitting is that it relies on accurate positioning of the dot-levels with respect to the Fermi energy of the reservoir, E_F (see Fig. 35a). This makes the spin read-out very sensitive to charge switches, which can easily push the $|\uparrow\rangle$ level above E_F, or pull $|\downarrow\rangle$ below E_F, resulting in a measurement error. To counteract this effect, a large enough Zeeman splitting is required (in Sect. 5 a magnetic field of more than 8 Tesla was used, although with a more stable sample a lower field might be sufficient). On the other hand, a smaller Zeeman splitting is desirable because it implies a lower and therefore more convenient resonance frequency for coherent spin manipulation. In addition, the spin relaxation time is expected to be longer at smaller ΔE_Z. Therefore, a different spin read-out mechanism that is less sensitive to charge switches and can function at lower fields would be very useful.

A particularly convenient way to perform spin-to-charge conversion could be provided by utilizing not a difference in *energy* between spin-up and spin-down, but a difference in *tunnel rate* (Fig. 35b). To read out the spin orientation of an electron on the dot, we simply raise both dot levels above E_F, so that the electron can leave the dot. If the tunnel rate for spin-up electrons, $\Gamma_\uparrow$, is much larger than that for spin-down electrons, $\Gamma_\downarrow$, then at a suitably chosen time the dot will have a large probability to be already empty if the spin was up, but a large probability to be still occupied if the spin is down. Measuring the charge on the dot within the spin relaxation time can then reveal the spin state.

This scheme is very robust against charge switches, since no precise positioning of the dot levels with respect to the leads is required: both levels simply have to be above E_F. Also, switches have a small influence on the tunnel rates themselves, as they tend to shift the whole potential landscape up or down, which does not change the tunnel barrier for electrons in the dot [87]. Of course, the visibility of this spin measurement scheme depends on the difference in tunnel rate we can achieve.

A difference in tunnel rate for spin-up and spin-down electrons is provided by the magnetic field. From large-bias transport measurements in a magnetic field parallel to the 2DEG [82], we find that the spin-selectivity ($\Gamma_\uparrow/\Gamma_\downarrow$) grows roughly linearly from ∼1.5 at 5 Tesla to ∼5 at 14 Tesla. This is in good agreement with the spin-selectivity of about 3 that was found at 10 Tesla using the single-shot spin measurement technique of Sect. 5.

We believe that this spin-dependence of the tunnel rates is due to exchange interactions in the reservoirs. If ΔE_Z is the same in the dot as in the reservoirs, the tunnel barrier will be the same for $|\uparrow\rangle$ and $|\downarrow\rangle$ electrons, giving $\Gamma_\uparrow = \Gamma_\downarrow$ (Fig. 36a). However, close to the dot there is a region with only $|\uparrow\rangle$ electrons, where an electron that is excited from $|\uparrow\rangle$ to $|\downarrow\rangle$ must overcome not only the single-particle Zeeman energy but also the many-body exchange energy between the reservoir electrons [88]. We can describe this situation with an

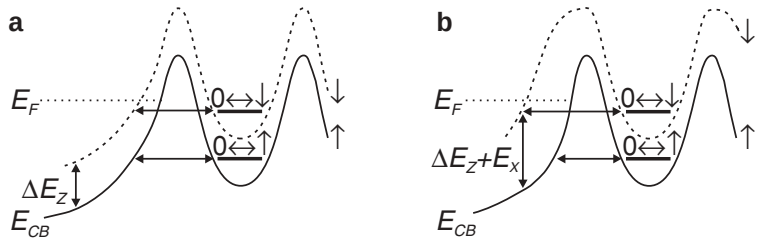

Fig. 36. Exchange interaction in the reservoirs leading to spin-selective tunnel rates. (**a**) Schematic diagram of the conduction band edge E_{CB} near the dot for electrons with spin-up (*solid line*) and spin-down (*dashed line*). If ΔE_Z in the reservoirs is the same as in the dot, the tunnel rates do not depend on spin. (**b**) The exchange energy E_X in the reservoirs close to the dot induces spin-dependent tunnel rates

effective g-factor $g_{\rm eff}$, which can be larger than the bare g-factor (Fig. 36b). In this case, $|\downarrow\rangle$ electrons experience a thicker tunnel barrier than $|\uparrow\rangle$ electrons, resulting in a difference in tunnel rates [43].

In a magnetic field parallel to the 2DEG, the effect only leads to a modest spin-selectivity that does not allow a single-shot measurement. However, a much larger spin-selectivity is possible in a perpendicular magnetic field [88], i.e. in the Quantum Hall regime. Magnetotransport measurements in 2DEGs with odd filling factor have shown that the g-factor can be enhanced by as much as a factor of ten, depending on the field strength. We anticipate that a convenient perpendicular field of $\sim$4 T could already give enough spin-selectivity to allow high-fidelity spin read-out. Therefore, spin read-out should be feasible not only in a large parallel magnetic field, but also in a somewhat smaller perpendicular field.

6.3 Initialization

Initialization of the spin to the pure state $|\uparrow\rangle$ – the desired initial state for most quantum algorithms [1] – has been demonstrated in Sect. 5. There it was shown that by waiting long enough, energy relaxation will cause the the spin on the dot to relax to the $|\uparrow\rangle$ ground state (Fig. 37a). This is a very simple and robust initialization approach, which can be used for any magnetic field orientation (provided that $g\mu_B B > 5k_B T$). However, as it takes about $5T_1$ to reach equilibrium, it is also a very slow procedure ($\geq$10 ms), especially at lower magnetic fields, where the spin relaxation time T_1 might be very long.

A faster initialization method has been used in the "reverse pulse" technique in Sect. 5. By placing the dot in the read-out configuration (Fig. 37b), a spin-up electron will stay on the dot, whereas a spin-down electron will be replaced by a spin-up. After waiting a few times the sum of the typical tunnel times for spin-up and spin-down ($\sim 1/\Gamma_\uparrow + 1/\Gamma_\downarrow$), the spin will be with large probability in the $|\uparrow\rangle$ state. This initialization procedure can therefore be quite fast ($<$1 ms), depending on the tunnel rates.

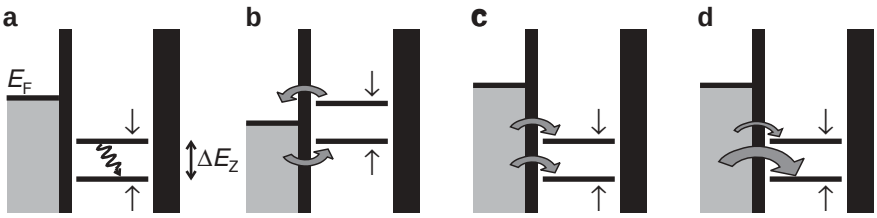

Fig. 37. Schematic energy diagrams depicting initialization procedures in a large parallel or perpendicular magnetic field. (**a**) Spin relaxation to pure state $|\uparrow\rangle$. (**b**) The "read-out" configuration can result in $|\uparrow\rangle$ faster. (**c**) Random spin injection gives a statistical mixture of $|\uparrow\rangle$ and $|\downarrow\rangle$. (**d**) In a large perpendicular field providing a strong spin-selectivity, injection results mostly in $|\uparrow\rangle$

We also have the possibility to initialize the dot to a mixed state, where the spin is probabilistically in $|\uparrow\rangle$ or $|\downarrow\rangle$. In Sect. 5, mixed-state initialization was demonstrated in a parallel field by first emptying the dot, followed by placing both spin levels below E_F during the "injection stage" (Fig. 37c). The dot is then randomly filled with either a spin-up or a spin-down electron. This is very useful, e.g. to test two-spin operations (see paragraph 6.6).

In a large perpendicular field providing a strong spin-selectivity, initialization to the $|\uparrow\rangle$ state is possible via spin relaxation (Fig. 37a) or via direct injection (Fig. 37d). Initialization to a mixed state (or in fact to any state other than $|\uparrow\rangle$) is very difficult due to the spin-selectivity. It probably requires the ability to coherently rotate the spin from $|\uparrow\rangle$ to $|\downarrow\rangle$ (see paragraph 6.5).

6.4 Coherence Times

The long-term potential of GaAs quantum dots as electron spin qubits clearly depends crucially on the spin coherence times T_1 and T_2. In Sect. 5, we have shown that the single-spin relaxation time, T_1, can be very long – on the order of 1 ms at 8 T. This implies that the spin is only very weakly disturbed by the environment. The dominant relaxation mechanism at large magnetic field is believed to be the coupling of the spin to phonons, mediated by the spin-orbit interaction [22].

The fundamental quantity of interest for spin qubits is the decoherence time of a single electron spin in a quantum dot, T_2, which has never been measured. Experiments with electrons in 2DEGs have established an ensemble-averaged decoherence time, T_2^*, of $\sim$100 ns [89]. Recently, a similar lower bound on T_2 has been claimed for a single trapped electron spin, based on the linewidth of the observed electron spin resonance [90]. Theoretically, it has been suggested that the real value of T_2 can be much longer [22], and under certain circumstances could even be given by $T_2 = 2T_1$, limited by the same spin-orbit interactions that limit T_1.

To build a scalable quantum computer, a sufficiently long T_2 (corresponding to more than 10^4 times the gate operation time) is essential in order to reach the "accuracy threshold". However, for experiments in the near future, we only need to perform a few spin rotations within T_2, which might already be possible for much shorter T_2, on the order of a µs. This should also be long enough to perform two-spin operations, which are likely to be much faster. To find the actual value of T_2, the ability to perform coherent spin operations is required. This is discussed in the next paragraphs.

6.5 Coherent Single-Spin Manipulation: ESR

We have not yet satisfied the key requirement for an actual spin qubit: coherent manipulation of one- and two-spin states. To controllably create superpositions of $|\uparrow\rangle$ and $|\downarrow\rangle$, we can use the well-known electron spin resonance

(ESR) effect. A microwave magnetic field $\mathbf{B_{ac}}$ oscillating in the plane perpendicular to $\mathbf{B}$, at a frequency $f = g\mu_B B/h$ (in resonance with the spin precession about $\mathbf{B}$) causes the spin to make transitions between $|\uparrow\rangle$ and $|\downarrow\rangle$. The choice of B strength is a trade-off between reliable initialization and read-out (strong B is better) and experimental convenience (low f is easier). We expect that a perpendicular field of 4 Tesla should be sufficient to provide high-fidelity read-out and initialization, with $f \approx 25\,\mathrm{GHz}$ (for $g = -0.44$). Alternatively, in a parallel field we may have to go up to 8 Tesla, corresponding to $f \approx 45\,\mathrm{GHz}$ [54], for high-fidelity spin measurement. However, since single-shot read-out is not strictly required, a somewhat lower field could also be enough.

Properly timed bursts of microwave power tip the spin state over a controlled angle, e.g. 90° or 180°. In order to observe Rabi oscillations, the Rabi period must be at most of the order of the single-spin decoherence time T_2. For a Rabi period of 150 ns, we need a microwave field strength B_{ac} of $\sim$1 mT. If T_2 is much longer, there is more time to coherently rotate the spin, so a smaller oscillating field is sufficient.

We intend to generate the oscillating magnetic field by sending an alternating current through an on-chip wire running close by the dot (Fig. 38a). If the wire is placed well within one wavelength (which is a few mm at 30 GHz near the surface of a GaAs substrate) from the quantum dot, the dot is in the near-field region and the electric and magnetic field distribution produced by the AC current should be the same as for a DC current [91]. With a wire 200 nm from the dot, a current of $\sim$1 mA should generate a magnetic field of about 1 mT and no electric field at the position of the dot. To minimize

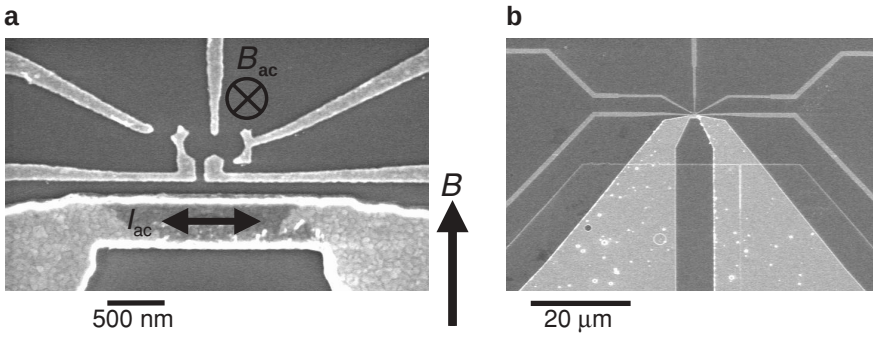

Fig. 38. On-chip wire to apply microwaves to a nearby quantum dot. The device was made by Laurens Willems van Beveren and Jort Wever. (**a**) Scanning electron microscope image of a device consisting of a double quantum dot in close proximity to a gold wire. An AC current through the wire, I_{ac}, generates an oscillating magnetic field, B_{ac}, perpendicular to the plane. If the AC frequency is resonant with the Zeeman splitting due to a large static in-plane magnetic field, B, a spin on the dot will rotate. (**b**) Large-scale view of the wire, designed to be a 50 Ω coplanar stripline

reflection and radiation losses, the wire is designed to be a shorted coplanar stripline (Fig. 38b) with a 50 Ω impedance.

To detect the electron spin resonance (ESR) and obtain a lower bound on T_2 from the linewidth of the resonance signal, various methods have been proposed, either using transport measurements [92] or relying on charge detection [93]. In both cases, the required spin-to-charge conversion is achieved by positioning the dot levels around the Fermi energy of the reservoir (Fig. 39a–b). The ESR-field induces spin flips, exciting $|\uparrow\rangle$ electrons to $|\downarrow\rangle$, which can then tunnel out of the dot. This leads to an average current (Fig. 39a) or to a change in the average occupation of the dot (Fig. 39b). However, in this configuration the dot is particularly sensitive to spurious effects induced by the microwaves, such as $|\uparrow\rangle$ electrons being excited out of the dot via thermal excitation or photon-assisted tunnelling. These processes can completely obscure the spin resonance.

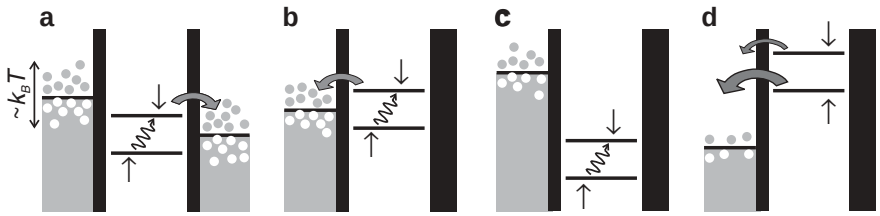

Fig. 39. Detecting ESR. (**a**) To detect ESR in a transport measurement [92], the dot is placed in Coulomb blockade, so that electron spins that are flipped by the ESR field can contribute to a current. (**b**) A similar configuration is used to detect ESR via changes in the occupation of the dot [90], measured using a charge detector. (**c**) If the dot is deep in Coulomb blockade during the spin-flip stage, the electron is not easily excited to the reservoir via thermal excitation or photon-assisted tunnelling. (**d**) The microwaves are off during the spin read-out stage to enhance the measurement fidelity

Such problems can be avoided by combining (pulsed) electron spin resonance with *single-shot* spin measurement. This allows us to separate the spin manipulation stage (during which the microwaves are on) from the spin read-out stage (without microwaves). In this way, excitation out of the dot is prevented by Coulomb blockade (Fig. 39c), until spin read-out is initiated (Fig. 39d). In contrast to the techniques described above – which require a large spin flip rate to generate a *measurable* current or disturbance of the dot occupation – this approach only requires the spin flip rate to be faster than the decoherence rate. Therefore, a longer T_2 allows us to use a smaller B_{ac}, corresponding to (quadratically) smaller microwave power. This should help to suppress heating and photon-assisted tunnelling.

In principle, an ESR experiment can be performed in a parallel or a perpendicular magnetic field. The read-out in a perpendicular field is particularly

suitable for ESR detection, as the dot levels are far above E_F (so are not affected by photon-assisted tunnelling or heating). If $\mathbf{B}$ is perpendicular to the surface, $\mathbf{B_{ac}}$ must run through the dot in a direction parallel to the surface, so we must place the wire above the dot rather than to its side. The wire could be located on top of an insulating dielectric layer that covers the gate electrodes.

6.6 Coherent Spin Interactions: $\sqrt{\text{SWAP}}$

Two electron spins $\mathbf{S_1}$ and $\mathbf{S_2}$ in neighbouring quantum dots are coupled to each other by the exchange interaction, which takes the form $J(t)\mathbf{S_1} \cdot \mathbf{S_2}$. If the double dot is filled with two identical spins, the interaction does not change their orientation. However, if the left electron spin starts out being $|\uparrow\rangle$ and the right one $|\downarrow\rangle$, then the states of the two spins will be swapped after a certain time. An interaction active for half this time performs the $\sqrt{\text{SWAP}}$ gate, which has been shown to be universal for quantum computation when combined with single qubit rotations [94]. In fact, the exchange interaction is even universal by itself when the state of each qubit is encoded in the state of three electron spins [17].

The strength $J(t)$ of the exchange interaction depends on the overlap of the two electron wavefunctions, which varies exponentially with the voltage applied to the gate controlling the inter-dot tunnel barrier. By applying a (positive) voltage pulse with a certain amplitude and duration, we can temporarily turn on the exchange interaction, thereby performing a $\sqrt{\text{SWAP}}$ gate. We expect that J may correspond to a frequency of $\sim$10 GHz, so two-qubit gates could be performed in $\sim$100 ps. A much larger value would not be convenient experimentally, as we would have to control the exact amplitude and duration of the pulse very precisely. On the other hand, a very slow exchange operation would be more sensitive to decoherence resulting from fluctuations in the tunnel rate, due to charge noise. The value of J can in principle be determined in a transport measurement [33], or alternatively by using the QPC tunnel spectroscopy technique developed in Sect. 3. However, in practical situations J might be too small to be resolved.

To explore the operation of the SWAP gate, we only need reliable initialization and read-out, without requiring ESR [2]. Imagine qubit 1 is prepared in a pure state $|\uparrow\rangle$ and qubit 2 is prepared in a statistical mixture of $|\uparrow\rangle$ and $|\downarrow\rangle$. Measurement of qubit 1 should then always give $|\uparrow\rangle$, while measurement of qubit 2 should give probabilistically $|\uparrow\rangle$ or $|\downarrow\rangle$. After application of the SWAP gate, in contrast, measurement of qubit 2 should always give $|\uparrow\rangle$, while measurement of qubit 1 should give a probabilistic outcome. This and other spin-interaction experiments are probably easiest in a parallel magnetic field, where initialization to a statistical mixture is convenient. In addition, a large perpendicular field shrinks the electron wavefunctions, lowering the tunnel coupling and thus the exchange interaction between the two dots.

6.7 Unresolved Issues

Several issues are not yet fully resolved, both experimentally and theoretically. One of these is the question of electron spin resonance in the reservoir. There are indications that the g-factor in the dot is different from that in the reservoir [95] (disregarding enhancement due to exchange interactions, which are not relevant for a "global" excitation such as ESR). However, if the two g-factors are equal, then any coherent operation of the spin on the dot will also influence the spin population outside the dot. This has not been taken into account in this section, but it could lead to complications for the proposed ESR experiments.

Another question is related to the $\sim 10^6$ nuclear spins in the quantum dot that couple to the electron spin via the hyperfine coupling. Through the Overhauser effect they produce an effective magnetic field, which can be very large ($\sim$5 T) for a fully polarized nuclear spin ensemble. Statistical fluctuations in the Overhauser field could lead to changes in the phase of the electron spin. It is not yet clear what the influence will be on spin manipulation experiments. If it turns out to be a problem, we may have to polarize the nuclear system completely in order to suppress the fluctuations.

A more practical consideration is the effect of charge switches in the heterostructure, which make any experiment more difficult. This is particularly true for two-spin interaction experiments, as charge noise can affect the interdot tunnel barrier and therefore the exchange interaction, resulting in decoherence. In collaboration with the group of Prof. Wegscheider in Regensburg, we have started to investigate the possible origin of charge switching, in an effort to produce more quiet heterostructures and devices.

Finally, so far we have used at most two quantum dots, not paying much attention to the scalability of our spin qubit approach. For instance, the ESR-field generated by the big wire runing next to the double dot will also influence other spins in nearby dots. We may therfore have to develop techniques to locally control the g-factor felt by the electron spin in a dot, in order to shift particular dots in or out of resonance.

6.8 Conclusion and Outlook

In summary, we have demonstrated that single electrons trapped in GaAs lateral quantum dots are promising candidates for implementing a spin qubit. We have realized the "hardware" for such a system: a device consisting of two coupled quantum dots that can be filled with one electron spin each, flanked by two quantum point contacts. Using these QPCs as charge detectors, we can determine all relevant parameters of the double dot. In addition, we have developed a technique to measure the spin orientation of an *individual* electron. Now we can proceed to combine all these ingredients with the ability to generate strong microwave magnetic fields close to the dot, and gate voltage

pulses to control the inter-dot coupling, in order to demonstrate superposition and entanglement of spin states.

For such experiments, the QPC is an invaluable tool. It allows us to probe a dot that is nearly isolated from the reservoirs, which is a regime not accessible to conventional transport experiments. Most importantly, it enables us to study a *single* spin or charge, rather than measuring average properties of a large ensemble. The QPC charge and spin detector is therefore essential to achieve the kind of single-particle control that is required for creating a qubit – transport experiments are no longer necessary.

The techniques we have developed are not only suitable for quantum computation. Now that the spin orientation of a single electroncan be measured, we can think of using the spin as a local probe to explore the semiconductor environment. For instance, measuring the spin relaxation time in various situations could reveal details of different mechanisms for spin-orbit coupling. We could vary the orientation of the magnetic field with respect to the crystal axes, or investigate the effect of static or time-varying electric fields. Once we can measure the electron spin resonance frequency, this would allow us to study the polarization of the nuclear spin ensemble via the Overhauser effect. In all these cases, the fact that dot parameters such as the Zeeman splitting or the tunnel coupling to reservoirs can be controlled in situ, makes a lateral quantum dot filled with a single spin a system of great versatility and fundamental importance.

Acknowledgments

We thank T. Fujisawa, T. Hayashi, T. Saku and Y. Hirayama for help with device fabrication, M. Blaauboer, D.P. DiVincenzo, H.A. Engel, C.J.P.M. Harmans, V. Golovach, D. Loss, R. Schoelkopf, K. Schwab and W.G. van der Wiel for helpful discussions, and B. van der Enden and R. Schouten for technical support. We acknowledge financial support from the Specially Promoted Research Grant-in-Aid for Scientific Research from the Japanese Ministry of Education, Culture, Sports, Science and Technology, the DARPA-QUIST program (DAAD19-01-1-0659), the EU-RTN network on spintronics, and the Dutch Organisation for Fundamental Research on Matter (FOM).

References

1. M.A. Nielsen and I.L. Chuang, *Quantum computation and quantum information*, (Cambridge University Press, Cambridge, England, 2000). 25, 86
2. D. Loss and D.P. DiVincenzo, Phys. Rev. A **57**, 120 (1998). 25, 29, 30, 31, 48, 58, 59, 72, 82, 90
3. R.P. Feynman, *The Feynman Lectures on Physics*, Vol. 3 (Addison Wesley, 1970). 26, 27
4. M. Riebe et al., Nature **429**, 734 (2004). 26

5. M.D. Barrett et al., Nature **429**, 737 (2004). 26
6. V.B. Braginsky and F.Y. Khalili, *Quantum Measurement* (Cambridge University Press, 1992). 27
7. S. Singh, *The Code Book* (Anchor Books/Doubleday, 2000). 27
8. R.P. Feynman, *The Feynman Lectures on Computation*, edited by R.W. Allen and T. Hey (Perseus Publishing, 2000). 27
9. D. Deutsch, Proc. R. Soc. Lond. A, 400 (1985). 28
10. P.W. Shor, in *Proceedings of 35th Annual Symposium on Foundations of Computer Science* (IEEE Press, 1994). 28
11. S. Lloyd, Science **273**, 1073 (1996). 28
12. L.K. Grover, Phys. Rev. Lett. **79**, 325 (1997). 28
13. P.W. Shor, in *Proceedings of 37th Annual Symposium on Foundations of Computer Science*, 56 (IEEE Press, 1996). 28
14. A.M. Steane, Phys. Rev. Lett. **77**, 793 (1996). 28
15. A.Y. Kitaev, in *Quantum Communication, Computing, and Measurement*, edited by A.S. Holevo, O. Hirota and C.M. Caves, 181 (Plenum Press, 1997). 28
16. D. Aharonov and M. Ben-Or, quant-ph/9906129 (1999). 28
17. D.P. DiVincenzo, Fortschr. Phys. **48**, 771 (2000). 29, 83, 90
18. L.M.K. Vandersypen et al., Nature **414**, 883 (2001). 29
19. Quantum information science and technology roadmapping project, available at http://qist.lanl.gov/. 29
20. Y. Kato, R.C. Myers, A.C. Gossard and D.D. Awschalom, Science **299**, 1201 (2003). 30
21. A.V. Khaetskii and Yu.V. Nazarov, Phys. Rev. B **64**, 125316 (2001). 31
22. V.N. Golovach, A. Khaetskii and D. Loss, Phys. Rev. Lett. **93**, 016601 (2004) 31, 79, 87
23. L.M. Woods, T.L. Reinecke and Y. Lyanda-Geller, Phys. Rev. B **66**, 161318(R) (2002). 31
24. S.I. Erlingsson and Yu.V. Nazarov, Phys. Rev. B **66**, 155327 (2002). 31
25. Y. Nakamura, Yu.A. Pashkin and J.S. Tsai, Nature **398**, 786 (1999). 31
26. T. Hayashi, T. Fujisawa, H.D. Cheong, Y.H. Jeong, and Y. Hirayama, Phys. Rev. Lett. **91**, 226804 (2003). 31
27. L.P. Kouwenhoven, C.M. Marcus, P.L. McEuen, S. Tarucha, R.M. Westervelt, and N.S. Wingreen, in *Mesoscopic Electron Transport*, edited by L. L. Sohn, L. P. Kouwenhoven, and G. Schön, NATO Advanced Study Institutes, Ser. E, Vol. **345** (Kluwer Academic, Dordrecht, 1997), pp. 105-214. 31, 34, 48, 51, 60
28. N.W. Ashcroft and N.D. Mermin, *Solid state physics* (Brooks/Cole 1976). 38
29. S. Tarucha, D.G. Austing, Y. Tokura, W.G. van der Wiel and L.P. Kouwenhoven, Phys. Rev. Lett. **84**, 2485 (2000). 38
30. W.G. van der Wiel et al., Physica B **256-258**, 173 (1998). 38
31. D. Weinmann, W. Häusler and B. Kramer, Phys. Rev. Lett. **74**, 984 (1995). 39
32. S. Tarucha, D.G. Austing, T. Honda, R.J. van der Hage and L.P. Kouwenhoven, Phys. Rev. Lett. **77**, 3613 (1996). 39
33. V.N. Golovach and D. Loss, Europhys. Lett. **62**, 83 (2003). 40, 90
34. G. Burkard, D. Loss, and D.P. DiVincenzo, Phys. Rev. B **59**, 2070 (1999). 40
35. J.M. Martinis, M.H. Devoret and J. Clarke, Phys. Rev. B **35**, 4682 (1987). 45
36. K. Bladh et al., Rev. Sci. Instr. **74**,1323 (2003). 45
37. J.H. Davies, *The physics of low-dimensional semiconductors* (Cambridge University Press, 1998). 46
38. D. Vion, A. Aassime, A. Cottet, P. Joyez, H. Pothier, C. Urbina, D. Estève, and M.H. Devoret, Science **296**, 886 (2002). 48

39. P. M. Petroff, A. Lorke, and A. Imamoglu, Phys. Today, 46 (May 2001). 48
40. L.P. Kouwenhoven, D.G. Austing, and S. Tarucha, Rep. Prog. Phys. **64** (6), 701 (2001). 48, 72
41. K. Ono, D.G. Austing, Y. Tokura, and S. Tarucha, Science **297**, 1313 (2002). 48
42. T. Hatano, M. Stopa, T. Yamaguchi, T. Ota, K. Yamada, and S. Tarucha, Phys. Rev. Lett. **3**, 066806 (2004). 48
43. M. Ciorga, A.S. Sachrajda, P. Hawrylak, C. Gould, P. Zawadzki, S. Jullian, Y. Feng, and Z. Wasilewski, Phys. Rev. B **61**, R16315 (2000). 48, 49, 60, 86
44. M. Field, C.G. Smith, M. Pepper, D.A. Ritchie, J.E.F. Frost, G.A.C. Jones, and D.G. Hasko, Phys. Rev. Lett. **70**, 1311 (1993). 52, 60, 66, 73
45. D. Sprinzak, Y. Ji, M. Heiblum, D. Mahalu, and H. Shtrikman, Phys. Rev. Lett. **88**, 176805 (2002). 52, 61
46. H. Pothier, P. Lafarge, C. Urbina, D. Estève, and M.H. Devoret, Europhys. Lett. **17**, 249 (1992). 52
47. W.G. van der Wiel, S. De Franceschi, J.M. Elzerman, T. Fujisawa, S. Tarucha, and L.P. Kouwenhoven, Rev. Mod. Phys. **75**, 1 (2003), see also cond-mat/0205350v2. 52, 55, 58, 59
48. A.W. Rushforth, C.G. Smith, M.D. Godfrey, H.E. Beere, D.A. Ritchie, and M. Pepper, Phys. Rev. B **69**, 113309 (2004). 56
49. L.M.K. Vandersypen, R. Hanson, L.H. Willems van Beveren, J.M. Elzerman, J.S. Greidanus, S. De Franceschi, and L.P. Kouwenhoven, in *Quantum Computing and Quantum Bits in Mesoscopic Systems*, Kluwer Academic/Plenum Publishers, New York 2003), see also quant-ph/0207059. 58, 72
50. A. Aassime, G. Johansson, G. Wendin, R.J. Schoelkopf, and P. Delsing, Phys. Rev. Lett. **86**, 3376 (2001).(1998).
51. P. Lafarge, H. Pothier, E.R. Williams, D. Esteve, C. Urbina, and M.H. Devoret, Zeitschrift für Physik B, **85**, 327 (1991). 60
52. R.C. Ashoori, H.L. Stormer, J.S. Weiner, L.N. Pfeiffer, S.J. Pearton, K.W. Baldwin, and K.W. West, Phys. Rev. Lett. **68**, 3088 (1992). 60
53. K.W. Lehnert, K. Bladh, L.F. Spietz, D. Gunnarsson, D.I. Schuster, P. Delsing, and R.J. Schoelkopf, Phys. Rev. Lett. **90**, 027002 (2003). 60
54. R. Hanson, B. Witkamp, L.M.K. Vandersypen, L.H. Willems van Beveren, J.M. Elzerman, and L.P. Kouwenhoven, Phys. Rev. Lett. **91**, 196802 (2003). 63, 64, 72, 74, 79, 83, 88
55. J.M. Elzerman, R. Hanson, J.S. Greidanus, L.H. Willems van Beveren, S. De Franceschi, L.M.K. Vandersypen, S. Tarucha, and L.P. Kouwenhoven, Phys. Rev. B **67**, R161308 (2003). 61, 67, 72, 73
56. T.A. Fulton and G.J. Dolan, Phys. Rev. Lett. **59**, 109 (1987). 66
57. R.J. Schoelkopf, P. Wahlgren, A.A. Kozhevnikov, P. Delsing, and D.E. Prober, Science **280**, 1238 (1998). 66, 71, 84
58. W. Lu, Z. Ji, L. Pfeiffer, K.W. West, and A.J. Rimberg, Nature **423**, 422 (2003). 66, 71, 72
59. T. Fujisawa, T. Hayashi, Y. Hirayama, H.D Cheong, and Y.H. Jeong, Appl. Phys. Lett. **84**, 2343 (2004). 66, 72
60. J. Cooper, C.G. Smith, D.A. Ritchie, E.H. Linfield, Y. Jin, and H. Launois, Phys. E **6**, 457 (2000). 66, 71
61. R. Schleser, E. Ruh, T. Ihn, K. Ennslin, D.C. Driscoll, and A.C. Gossard, cond-mat/0406568. 66
62. P. Horowitz and W. Hill, The Art of Electronics (Cambridge University Press, Cambridge, UK, 1989). 68

63. A.N. Korotkov, Phys. Rev. B **60**, 5737 (1999) 71
64. A.A. Clerk, S.M. Girvin and A.D. Stone, Phys. Rev. B **67**, 165324 (2003). 71
65. J.J. Sakurai, *Modern Quantum Mechanics*, Addison-Wesley (Reading MA, USA, 1994). 72
66. F.W. Wehrli, Physics Today **6**, 34 (1992). 72
67. S.A. Wolf et al., Science **294**, 1488-1495 (2001). 72
68. R. Blatt and P. Zoller, Eur. J. Phys. **9**, 250-279 (1988). 72
69. H.J. Mamin, R. Budakian, B.W. Chui, and D. Rugar, Phys. Rev. Lett. **91**, 207604 (2003). 72
70. B.E. Kane, Nature **393**, 133-137 (1998). 72
71. M. Xiao, I. Martin, and H.W. Jiang, Phys. Rev. Lett. **91**, 078301 (2003). 72
72. M. Friesen, C. Tahan, R. Joynt, and M.A. Eriksson, Phys. Rev. Lett. **92**, 037901 (2004). 72
73. H.A. Engel, V.N. Golovach, D. Loss, L.M.K. Vandersypen, J.M. Elzerman, R. Hanson, and L.P. Kouwenhoven, Phys. Rev. Lett. **93**, 106804 (2004). 72
74. R. Ionicioiu and A.E. Popescu, http://xxx.lanl.gov/abs/quant-ph/0310047 (2003). 72
75. A.D. Greentree, A.R. Hamilton, L.C.L. Hollenberg, and R.G. Clark, http://xxx.lanl.gov/abs/cond-mat/0403449 (2004). 72
76. J. Weis, R.J. Haug, K. von Klitzing, and K. Ploog, Surf. Sci. **305**, 664 (1994).
77. L.P. Kouwenhoven et al., Science **278**, 1788 (1997). 72
78. M. Ciorga et al., Physica E **11**, 35 (2001). 72
79. T. Fujisawa, D.G. Austing, Y. Tokura, Y. Hirayama, and S. Tarucha, Nature **419**, 278-281 (2002). 72
80. J.A. Folk, R.M. Potok, C.M. Marcus, and V. Umansky, Science **299**, 679 (2003). 72
81. J.M. Elzerman, R. Hanson, L.H. Willems van Beveren, L.M.K. Vandersypen, and L.P. Kouwenhoven, Appl. Phys. Lett. **84**, 4617-4619 (2004). 73
82. R. Hanson et al., http://xxx.lanl.gov/abs/cond-mat/011414 (2003) 74, 85
83. A.V. Khaetskii and Yu.V. Nazarov, Phys. Rev. B **64**, 125316 (2001). 79
84. L.M. Woods, T.L. Reinecke, and Y. Lyanda-Geller, Phys. Rev. B **66**, 161318(R) (2002). 79
85. S.I. Erlingsson and Yu.V. Nazarov, Phys. Rev. B **66**, 155327 (2002). 79
86. J.R. Petta, A.C. Johnson, C.M. Marcus, M.P. Hanson, and A.C. Gossard, cond-mat/0408139 (2004). 83
87. S.W. Jung, T. Fujisawa, Y.H. Jeong and Y. Hirayama, cond-mat (2004). 85
88. T. Englert, D.C. Tsui, A.C. Gossard, and C. Uihlein. Surface Science **113**, 295 (1982). 85, 86
89. J.M. Kikkawa and D.D. Awschalom, Nature **397**, 139 (1999). 87
90. M. Xiao, I. Martin, E. Yablonovitch, and H.W. Jiang, Nature **430**, 435 (2004). 87, 89
91. J.D. Jackson, *Classical electrodynamics*, Wiley, New York (1998). 88
92. H.A. Engel and D. Loss, Phys. Rev. Lett. **86**, 4648 (2001). 89
93. I. Martin, D. Mozyrsky, and H.W. Jiang, Phys. Rev. Lett. textbf90, 018301 (2003). 89
94. G. Burkard, D. Loss, and D.P. DiVincenzo, Phys. Rev. B **59**, 2070 (1999). 90
95. M. Dobers, K. von Klitzing, and G. Weiman. Phys. Rev. B **38**, 5453 (1988). 91

Low-Temperature Conduction of a Quantum Dot

M. Pustilnik[1] and L.I. Glazman[2]

[1] School of Physics, Georgia Institute of Technology, Atlanta, GA 30332, USA
[2] William I. Fine Theoretical Physics Institute, University of Minnesota, Minneapolis, MN 55455, USA

Summary. We review mechanisms of low-temperature electronic transport through a quantum dot weakly coupled to two conducting leads. Transport in this case is dominated by electron-electron interaction. At temperatures moderately lower than the charging energy of the dot, the linear conductance is suppressed by the Coulomb blockade. Upon further lowering of the temperature, however, the conductance may start to increase again due to the Kondo effect. We concentrate on lateral quantum dot systems and discuss the conductance in a broad temperature range, which includes the Kondo regime.

1 Introduction

In quantum dot devices [1] a small droplet of electron liquid, or just a few electrons are confined to a finite region of space. The dot can be attached by tunneling junctions to massive electrodes to allow electronic transport across the system. The conductance of such a device is determined by the number of electrons on the dot N, which in turn is controlled by varying the potential on the gate – an auxiliary electrode capacitively coupled to the dot [1]. At sufficiently low temperatures N is an integer at almost any gate voltage V_g. Exceptions are narrow intervals of V_g in which an addition of a single electron to the dot does not change much the electrostatic energy of the system. Such a degeneracy between different charge states of the dot allows for an activationless electron transfer through it, whereas for all other values of V_g the activation energy for the conductance G across the dot is finite. The resulting oscillatory dependence $G(V_g)$ is the hallmark of the Coulomb blockade phenomenon [1]. The contrast between the low- and high-conductance regions (Coulomb blockade valleys and peaks, respectively) gets sharper at lower temperatures. The pattern of periodic oscillations in the G vs. V_g dependence is observed down to the lowest attainable temperatures in experiments on tunneling through small metallic islands [2].

M. Pustilnik and L.I. Glazman: *Low-temperature Conduction of a Quantum Dot*, Lect. Notes Phys. **667**, 97–130 (2005)
www.springerlink.com

Conductance through quantum dots formed in semiconductor heterostructures exhibits a reacher behavior [1]. In larger dots (in the case of GaAs heterostructures, such dots may contain hundreds of electrons), the fluctuations of the heights of the Coulomb blockade peaks become apparent already at moderately low temperatures. Characteristic for mesoscopic phenomena, the heights are sensitive to the shape of a dot and magnetic flux threading it. The separation in gate voltage between the Coulomb blockade peaks, and the conductance in the valleys also exhibit mesoscopic fluctuations. However, the pattern of sharp conductance peaks separating the low-conductance valleys of the $G(V_g)$ dependence persists. Smaller quantum dots (containing few tens of electrons in the case of GaAs) show yet another feature [3]: in some Coulomb blockade valleys the dependence $G(T)$ is not monotonic and has a minimum at a finite temperature. This minimum is similar in origin [4] to the well-known non-monotonic temperature dependence of the resistivity of a metal containing magnetic impurities [5] – the *Kondo effect*. Typically, the valleys with anomalous temperature dependence correspond to an odd number of electrons in the dot. In an ideal case, the low-temperature conductance in such a valley is of the order of conductance at peaks surrounding it. Thus, at low temperatures the two adjacent peaks merge to form a broad maximum.

The number of electrons on the dot is a well-defined quantity as long the conductances of the junctions connecting the dot to the electrodes is small compared to the conductance quantum e^2/h. In quantum dot devices formed in semiconductor heterostructures the conductances of junctions can be tuned continuously. With the increase of the conductances, the periodic patern in $G(V_g)$ dependence gradually gives way to mesoscopic conductance fluctuations. Yet, electron-electron interaction still affects the transport through the device. A strongly asymmetric quantum dot device with one junction weakly conducting, while another completely open, provides a good example of that. The differential conductance across the device in this case exhibits zero-bias anomaly – suppression at low bias. Clearly, Coulomb blockade is not an isolated phenomenon, but is closely related to interaction-induced anomalies of electronic transport and thermodynamics in higher dimensions [6].

The emphasis of these lectures is on the Kondo effect in quantum dots. We will concentrate on the so-called *lateral quantum dot systems* [1, 3], formed by gate depletion of a two-dimensional electron gas at the interface between two semiconductors. These devices offer the highest degree of tunability, yet allow for relatively simple theoretical treatment. At the same time, many of the results presented below are directly applicable to other systems as well, including vertical quantum dots [7, 8, 9], Coulomb-blockaded carbon nanotubes [9, 10], single-molecule transistors [11], and stand-alone magnetic atoms on metallic surfaces [12].

Kondo effect emerges at relatively low temperature, and we will follow the evolution of the conductance upon the reduction of temperature. On the way to Kondo effect, we encouter also the phenomena of Coulomb blockade and of mesoscopic conductance fluctuations.

2 Model of a Lateral Quantum Dot System

The Hamiltonian of interacting electrons confined to a quantum dot has the following general form,

$$H_{\rm dot} = \sum_s \sum_{ij} h_{ij} d^\dagger_{is} d_{js} + \frac{1}{2} \sum_{ss'} \sum_{ijkl} h_{ijkl} d^\dagger_{is} d^\dagger_{js'} d_{ks'} d_{ls} \; . \tag{1}$$

Here an operator $d^\dagger_{is}$ creates an electron with spin s in the orbital state $\phi_i(\mathbf{r})$ (the wave functions are normalized according to $\int d\mathbf{r} \phi^*_i(\mathbf{r}) \phi_j(\mathbf{r}) = \delta_{ij}$); $h_{ij} = h^*_{ji}$ is an Hermitian matrix describing the single-particle part of the Hamiltonian. The matrix elements h_{ijkl} depend on the potential $U(\mathbf{r} - \mathbf{r}')$ of electron-electron interaction,

$$h_{ijkl} = \int d\mathbf{r}\, d\mathbf{r}' \phi^*_i(\mathbf{r}) \phi^*_j(\mathbf{r}') U(\mathbf{r} - \mathbf{r}') \phi_k(\mathbf{r}') \phi_l(\mathbf{r}) \; . \tag{2}$$

The Hamiltonian (1) can be simplified further provided that the quasiparticle spectrum is not degenerate near the Fermi level, that the Fermi-liquid theory is applicable to the description of the dot, and that the dot is in the metallic conduction regime. The first of these conditions is satisfied if the dot has no spatial symmetries, which implies also that motion of quasiparticles within the dot is chaotic.

The second condition is met if the electron-electron interaction within the dot is not too strong, i.e. the gas parameter r_s is small,

$$r_s = (k_F a_0)^{-1} \lesssim 1 \; , \quad a_0 = \kappa \hbar^2 / e^2 m^* \tag{3}$$

Here k_F is the Fermi wave vector, a_0 is the effective Bohr radius, κ is the dielectric constant of the material, and m^* is the quasiparticle effective mass.

The third condition requires the ratio of the Thouless energy E_T to the mean single-particle level spacing δE to be large [13],

$$g = E_T / \delta E \gg 1 \; . \tag{4}$$

For a ballistic two-dimensional dot of linear size L the Thouless energy E_T is of the order of $\hbar v_F / L$, whereas the level spacing can be estimated as

$$\delta E \sim \hbar v_F k_F / N \sim \hbar^2 / m^* L^2 \; . \tag{5}$$

Here v_F is the Fermi velocity and $N \sim (k_F L)^2$ is the number of electrons in the dot. Therefore,

$$g \sim k_F L \sim \sqrt{N} \; , \tag{6}$$

so that having a large number of electrons $N \gg 1$ in the dot guarantees that the condition (4) is satisfied.

Under the conditions (3), (4) the *Random Matrix Theory* (for a review see, e.g., [14, 15, 16, 17]) is a good starting point for description of non-interacting quasiparticles within the energy strip of the width E_T about the

Fermi level [13]. The matrix elements h_{ij} in (1) belong to a Gaussian ensemble [16, 17]. Since the matrix elements do not depend on spin, each eigenvalue ϵ_n of the matrix h_{ij} represents a spin-degenerate energy level. The spacings $\epsilon_{n+1} - \epsilon_n$ between consecutive levels obey the Wigner-Dyson statistics [16]; the mean level spacing $\langle \epsilon_{n+1} - \epsilon_n \rangle = \delta E$.

We discuss now the second term in the Hamiltonian (1), which describes electron-electron interaction. It turns out [18, 19, 20] that the vast majority of the matrix elements h_{ijkl} are small. Indeed, in the lowest order in $1/g \ll 1$, the wave functions $\phi_i(\mathbf{r})$ are Gaussian random variables with zero mean, statistically independent of each other and of the corresponding energy levels [21]:

$$\overline{\phi_i^*(\mathbf{r})\phi_j(\mathbf{r}')} = \frac{\delta_{ij}}{\mathcal{A}}\, F(|\mathbf{r}-\mathbf{r}'|)\,, \overline{\phi_i(\mathbf{r})\phi_j(\mathbf{r}')} = \frac{\delta_{\beta,1}\delta_{ij}}{\mathcal{A}}\, F(|\mathbf{r}-\mathbf{r}'|)\,. \tag{7}$$

Here $\mathcal{A} \sim L^2$ is the area of the dot, and the function F is given by

$$F(r) \sim \langle \exp(i\mathbf{k}\cdot\mathbf{r})\rangle_{\rm FS}\,. \tag{8}$$

where $\langle \ldots \rangle_{\rm FS}$ stands for the averaging over the Fermi surface $|\mathbf{k}| = k_F$. In two dimensions, the function $F(r)$ decreases with r as $F \propto (k_F r)^{-1/2}$ at $k_F r \gg 1$, and saturates to $F \sim 1$ at $k_F r \ll 1$.

The parameter β in (7) distinguishes between the presence ($\beta = 1$) or absence ($\beta = 2$) of the time-reversal symmetry. The symmetry breaking is driven by the orbital effect of the magnetic field and is characterised by the parameter

$$\chi = (\Phi/\Phi_0)\sqrt{g}\,,$$

where Φ is the magnetic flux threading the dot and $\Phi_0 = hc/e$ is the flux quantum, so that the limits $\chi \ll 1$ and $\chi \gg 1$ correspond to, respectively, $\beta = 1$ and $\beta = 2$. Note that in the case of a magnetic field $H_\perp$ applied perpendicular to the plane of the dot, the crossover (at $\chi \sim 1$) between the two regimes occurs at so weak field that the corresponding Zeeman energy B is negligibly small[1].

After averaging with the help of (7)–(8), the matrix elements (2) take the form

$$\overline{h_{ijkl}} = (2E_C + E_S/2)\,\delta_{il}\delta_{jk} + E_S\delta_{ik}\delta_{jl} + \Lambda\,(2/\beta - 1)\,\delta_{ij}\delta_{kl}\,.$$

We substitute this expression into Hamiltonian (1), and rearrange the sum over the spin indexes with the help of the identity

$$2\,\delta_{s_1 s_2}\delta_{s_1' s_2'} = \delta_{s_1 s_1'}\delta_{s_2' s_2} + \boldsymbol{\sigma}_{s_1 s_1'}\cdot\boldsymbol{\sigma}_{s_2' s_2}\,, \tag{9}$$

where $\boldsymbol{\sigma} = (\sigma^x, \sigma^y, \sigma^z)$ are the Pauli matrices. This results in a remarkably simple form [19, 20]

[1] For example, in the experiments [22] the crossover takes place at $H_\perp \sim 10\ mT$. Zeeman energy in such a field $B \sim 2.5\ mK$, which is by an order of magnitude lower than the base temperture in the measurements.

$$H_{\rm int} = E_C \hat{N}^2 - E_S \hat{\mathbf{S}}^2 + \Lambda\,(2/\beta - 1)\,\hat{T}^\dagger \hat{T} \tag{10}$$

of the interaction part of the Hamiltonian of the dot. Here

$$\hat{N} = \sum_{ns} d^\dagger_{ns} d_{ns}\ , \quad \hat{\mathbf{S}} = \sum_{nss'} d^\dagger_{ns} \frac{\boldsymbol{\sigma}_{ss'}}{2}\, d_{ns'} \quad \hat{T} = \sum_{n} d^\dagger_{n\uparrow} d^\dagger_{n\downarrow} \tag{11}$$

are the operators of the total number of electrons in the dot, of the dot's spin, and the "pair creation" operator corresponding to the interaction in the Cooper channel.

The first term in (10) represents the electrostatic energy. In the conventional equivalent circuit picture, see Fig. 1, the charging energy E_C is related to the total capacitance C of the dot, $E_C = e^2/2C$. For a mesoscopic ($k_F L \gg 1$) conductor, the charging energy is large compared to the mean level spacing δE. Indeed, using the estimates $C \sim \kappa L$ and (3) and (5), we find

$$E_C/\delta E \sim L/a_0 \sim r_s\sqrt{N}\ . \tag{12}$$

Except an exotic case of an extremely weak interaction, this ratio is large for $N \gg 1$; for the smallest quantum dots formed in GaAs heterostructures, $E_C/\delta E \sim 10$ [3]. Note that (4), (6), and (12) imply that

$$E_T/E_C \sim 1/r_s \gtrsim 1\ ,$$

which justifies the use of RMT for the description of single-particle states with energies $|\epsilon_n| \lesssim E_C$, relevant for Coulomb blockade.

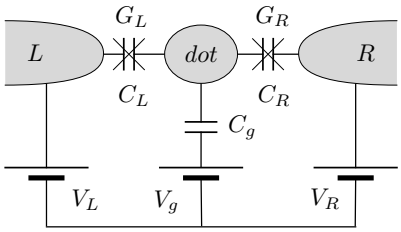

Fig. 1. Equivalent circuit for a quantum dot connected to two leads by tunnel junctions and capacitively coupled to the gate electrode. The total capacitance of the dot $C = C_L + C_R + C_g$

The second term in (10) describes the intra-dot exchange interaction, with the exchange energy E_S given by

$$E_S = \int d\mathbf{r}\, d\mathbf{r}' U(\mathbf{r} - \mathbf{r}') F^2(|\mathbf{r} - \mathbf{r}'|) \tag{13}$$

In the case of a long-range interaction the potential U here should properly account for the screening [20]. For $r_s \ll 1$ the exchange energy can be estimated with logarithmic accuracy by substituting $U(r) = (e^2/\kappa r)\theta(a_0 - r)$ into

(13) (here we took into account that the screening length in two dimensions coincides with the Bohr radius a_0), which yields

$$E_S \sim r_s \ln(1/r_S)\, \delta E \ll \delta E \,. \tag{14}$$

The estimate (14) is valid only for $r_s \ll 1$. However, the ratio $E_S/\delta E$ remains small for experimentally relevant[2] value $r_s \sim 1$ as long as the Stoner criterion for the absence of itinerant magnetism [23] is satisfied. This guarantees the absence of a macroscopic (proportional to N) magnetization of a dot in the ground state [19].

The third term in (10), representing interaction in the Cooper channel, is renormalized by higher-order corrections arising due to virtual transitions to states outside the energy strip of the width E_T about the Fermi level. For attractive interaction ($\Lambda < 0$) the renormalization enhances the interaction, eventually leading to the superconducting instability and formation of a gap Δ_Λ in the electronic spectrum. Properties of very small ($\Delta_\Lambda \sim \delta E$) superconducting grains are reviewed in, e.g., [24]; for properties of larger grains ($\Delta_\Lambda \sim E_C$) see [25]. Here we concentrate on the repulsive interaction ($\Lambda > 0$), in which case Λ is very small,

$$\Lambda \sim \frac{\delta E}{\ln(\epsilon_F/E_T)} \sim \frac{\delta E}{\ln N} \ll \delta E \,.$$

This estimate accounts for the logarithmic renormalization of Λ when the high-energy cutoff is reduced from the Fermi energy ϵ_F down to the Thouless energy E_T [20]. In addition, if the time-reversal symmetry is lifted ($\beta = 2$) then the third term in (10) is zero to start with. Accordingly, hereinafter we neglect this term altogether by setting $\Lambda = 0$.

Obviously, the interaction part of the Hamiltonian (10), is invariant with respect to a change of the basis of single-particle states $\phi_i(\mathbf{r})$. Picking up the basis in which the first term in (1) is diagonal, we arrive at the *universal Hamiltonian* [19, 20],

$$H_{\rm dot} = \sum_{ns} \epsilon_n d^\dagger_{ns} d_{ns} + E_C \left(\hat{N} - N_0\right)^2 - E_S \hat{\mathbf{S}}^2 \,. \tag{15}$$

We included in (15) the effect of the capacitive coupling to the gate electrode: the dimensionless parameter N_0 is proportional to the gate voltage,

$$N_0 = C_g V_g / e \,,$$

where C_g is the capacitance between the dot and the gate, see Fig. 1. The relative magnitude of fully off-diagonal interaction terms in (1) (corresponding

[2] For GaAs ($m^* \approx 0.07 m_e$, $\kappa \approx 13$) the effective Bohr radius $a_0 \approx 10\ nm$, whereas a typical density of the two-dimensional electron gas, $n \sim 10^{11}\ cm^{-2}$ [3], corresponds to $k_F = \sqrt{2\pi n} \sim 10^6\ cm^{-1}$. This gives $k_F a_0 \sim 1$.

to h_{ijkl} with all four indices different), not included in (15), is of the order of $1/g \sim N^{-1/2} \ll 1$. Partially diagonal terms (two out of four indices coincide) are larger, of the order of $\sqrt{1/g} \sim N^{-1/4}$, but still are assumed to be negligible as $N \gg 1$.

As discussed above, in this limit the energy scales involved in (15) form a well-defined hierarchy

$$E_S \ll \delta E \ll E_C \ . \tag{16}$$

If all the single-particle energy levels ϵ_n were equidistant, then the spin S of an even-N state would be zero, while an odd-N state would have $S = 1/2$. However, the level spacings are random. If the spacing between the highest occupied level and the lowest unoccupied one is accidentally small, than the gain in the exchange energy, associated with the formation of a higher-spin state, may be sufficient to overcome the loss of the kinetic energy (cf. the Hund's rule in quantum mechanics). For $E_S \ll \delta E$ such deviations from the simple even-odd periodicity are rare [19, 26, 27]. This is why the last term in (15) is often neglected. Equation (15) then reduces to the Hamiltonian of the *Constant Interaction Model*, widely used in the analysis of experimental data [1]. Finally, it should be emphasized that SU(2)–invariant Hamiltonian (15) describes a dot in the absence of the spin-orbit interaction, which would destroy this symmetry.

Electron transport through the dot occurs via two dot-lead junctions. In a typical geometry, the potential forming a lateral quantum dot varies smoothly on the scale of the Fermi wavelength, see Fig. 2. Hence, the point contacts connecting the quantum dot to the leads act essentially as electronic waveguides. Potentials on the gates control the waveguide width, and, therefore, the number of electronic modes the waveguide support: by making the waveguide narrower one pinches the propagating modes off one-by-one. Each such mode contributes $2e^2/h$ to the conductance of a contact. The Coulomb blockade develops when the conductances of the contacts are small compared to $2e^2/h$, i.e. when the very last propagating mode approaches its pinch-off [28, 29]. Accordingly, in the Coulomb blockade regime each dot-lead junction in a lateral quantum dot system supports only a single electronic mode [30].

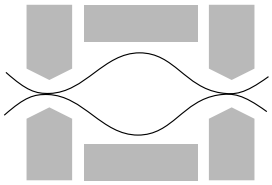

Fig. 2. The confining potential forming a lateral quantum dot varies smoothly on the scale of the de Broglie wavelength at the Fermi energy. Hence, the dot-lead junctions act essentially as electronic waveguides with a well-defined number of propagating modes

As discussed below, for $E_C \gg \delta E$ the characteristic energy scale relevant to the Kondo effect, the Kondo temperature T_K, is small compared to the mean level spacing: $T_K \ll \delta E$. This separation of the energy scales allows us to simplify the problem even further by assuming that the conductances of the dot-lead junctions are small. This assumption will not affect the properties of the system in the Kondo regime. At the same time, it justifies the use of the tunneling Hamiltonian for description of the coupling between the dot and the leads. The microscopic Hamiltonian of the system can then be written as a sum of three distinct terms,

$$H = H_{\text{leads}} + H_{\text{dot}} + H_{\text{tunneling}} \,, \tag{17}$$

which describe free electrons in the leads, isolated quantum dot, and tunneling between the dot and the leads, respectively. The second term in (17), the Hamiltonian of the dot H_{dot}, is given by (15). We treat the leads as reservoirs of free electrons with continuous spectra ξ_k, characterized by constant density of states ν, same for both leads. Moreover, since the typical energies $\omega \lesssim E_C$ of electrons participating in transport through a quantum dot in the Coulomb blockade regime are small compared to the Fermi energy of the electron gas in the leads, the spectra ξ_k can be linearized near the Fermi level, $\xi_k = v_F k$; here k is measured from k_F. With only one electronic mode per junction taken into account, the first and the third terms in (17) have the form

$$H_{\text{leads}} = \sum_{\alpha k s} \xi_k c^\dagger_{\alpha k s} c_{\alpha k s}, \quad \xi_k = -\xi_{-k}, \tag{18}$$

$$H_{\text{tunneling}} = \sum_{\alpha k n s} t_{\alpha n}\, c^\dagger_{\alpha k s} d_{n s} + \text{H.c.} \tag{19}$$

Here $t_{\alpha n}$ are tunneling matrix elements (tunneling amplitudes) "connecting" the state n in the dot with the state k in the lead α ($\alpha = R, L$ for the right/left lead).

Tunneling leads to a broadening of discrete levels in the dot. The width $\Gamma_{\alpha n}$ that level n acquires due to escape of an electron to lead α is given by

$$\Gamma_{\alpha n} = \pi\nu \left| t^2_{\alpha n} \right| \tag{20}$$

Randomness of single-particle states in the dot translates into the randomness of the tunneling amplitudes. Indeed, the amplitudes depend on the values of the electron wave functions at the points $\mathbf{r}_\alpha$ of the contacts, $t_{\alpha n} \propto \phi_n(\mathbf{r}_\alpha)$. For $k_F \left|\mathbf{r}_L - \mathbf{r}_R\right| \sim k_F L \gg 1$ the tunneling amplitudes [and, therefore, the widths (20)] in the left and right junctions are statistically independent of each other. Morover, the amplitudes to different energy levels are also uncorrelated, see (7):

$$\overline{t^*_{\alpha n}\, t_{\alpha' n'}} = \frac{\Gamma_\alpha}{\pi\nu}\, \delta_{\alpha\alpha'}\delta_{nn'} \,, \quad \overline{t_{\alpha n}\, t_{\alpha' n'}} = \frac{\Gamma_\alpha}{\pi\nu}\, \delta_{\beta,1}\delta_{\alpha\alpha'}\delta_{nn'} \,, \tag{21}$$

The average value $\Gamma_\alpha = \overline{\Gamma_{\alpha n}}$ of the width is related to the conductance of the corresponding dot-lead junction

$$G_\alpha = \frac{4e^2}{\hbar}\frac{\Gamma_\alpha}{\delta E}\,. \tag{22}$$

In the regime of strong Coulomb blockade ($G_\alpha \ll e^2/h$), the widths are small compared to the level spacing, $\Gamma_\alpha \ll \delta E$, so that discrete levels in the dot are well defined. Note that statistical fluctuations of the widths $\Gamma_{\alpha n}$ are large, and the corresponding distribution function is not Gaussian. Indeed, using (20) and (21) it is straightforward [16, 17] to show that

$$P(\gamma) = \overline{\delta\left(\gamma - \Gamma_{\alpha n}/\Gamma_\alpha\right)} = \begin{cases} \dfrac{e^{-\gamma/2}}{\sqrt{2\pi\gamma}}\,, \ \beta = 1 \\ \\ e^{-\gamma}, \quad \beta = 2 \end{cases} \tag{23}$$

This expression is known as Porter-Thomas distribution [31].

3 Thermally-Activated Conduction

At high temperatures, $T \gg E_C$, charging energy is negligible compared to the thermal energy of electrons. Therefore the conductance of the device in this regime G_∞ is not affected by charging and, independently of the gate voltage, is given by

$$\frac{1}{G_\infty} = \frac{1}{G_L} + \frac{1}{G_R}\,. \tag{24}$$

Dependence on N_0 develops at lower temperatures, $T \lesssim E_C$. It turns out that the conductance is suppressed for all gate voltages except narrow regions (*Coulomb blockade peaks*) around half-integer values of N_0. We will demonstrate this now using the method of rate [32, 33].

3.1 Onset of the Coulomb Blockade Oscillations

We start with the regime of relatively high temperatures,

$$\delta E \ll T \ll E_C\,, \tag{25}$$

and assume that the gate voltage is tuned sufficiently close to one of the points of charge degeneracy,

$$|N_0 - N_0^*| \lesssim T/E_C \tag{26}$$

(here N_0^* is a half-integer number).

Condition (25) enables us to treat the discrete single-particle levels within the dot as a continuum with the density of states $1/\delta E$. Condition (26), on the other hand, allows us to take into account only two charge states of the dot which are almost degenerate in the vicinity of the Coulomb blockade peak. For N_0 close to N_0^* these are the state $|0\rangle$ with $N = N_0^* - 1/2$ electrons on

the dot, and the state $|1\rangle$ with $N = N_0^* + 1/2$ electrons. According to (15) and (26), the difference of electrostatic energies of these states (the energy cost to add an electron to the dot) is

$$E_+(N_0) = E_{|1\rangle} - E_{|0\rangle} = 2E_C(N_0^* - N_0) \lesssim T \,. \tag{27}$$

In addition to the constraints (25) and (26), we assume here that the inelastic electron relaxation rate within the dot $1/\tau_\epsilon$ is large compared to the escape rates $\Gamma_\alpha/\hbar$. In other words, transitions between discrete levels in the dot occur before the electron escapes to the leads[3]. Under this assumption the tunnelings across the two junctions can be treated independently of each other (this is known as *sequential tunneling* approximation).

With the help of the Fermi golden rule the current I_α from the lead α into the dot can be written as

$$\begin{aligned} I_\alpha = e\,\frac{2\pi}{\hbar}\sum_{kns} \left|t_{\alpha n}^2\right| \delta(\xi_k + eV_\alpha - \epsilon_n - E_+) \\ \times \left\{\mathcal{P}_0 f(\xi_k)[1 - f(\epsilon_n)] - \mathcal{P}_1 f(\epsilon_n)[1 - f(\xi_k)]\right\} \,. \end{aligned} \tag{28}$$

Here $\mathcal{P}_i$ is the probabilitiy to find the dot in the charge states $|i\rangle$ $(i = 0, 1)$, $f(\omega) = [\exp(\omega/T) + 1]^{-1}$ is the Fermi function, and V_α is the electric potential on the lead α, see Fig. 1. In writing (28) we assumed that the distribution functions $f(\xi_k)$ and $f(\epsilon_n)$ are not perturbed. This is well justified provided that the relaxation rate $1/\tau_\epsilon$ exceeds the rate $\sim G_\infty|V_L - V_R|/e$ at which electrons pass through the dot. Replacing the summations over n and k in (28) by integrations over the corresponding continua, and making use of (20) and (22), we find

$$I_\alpha = \frac{G_\alpha}{e}\left[\mathcal{P}_0 F(E_+ - eV_\alpha) - \mathcal{P}_1 F(eV_\alpha - E_+)\right], \quad F(\omega) = \frac{\omega}{e^{\omega/T} - 1} \,. \tag{29}$$

In the steady state, the currents across the two junctions satisfy

$$I = I_L = -I_R \,. \tag{30}$$

Equations (29) and (30), supplemented by the obvious normalization condition $\mathcal{P}_0 + \mathcal{P}_1 = 1$, allow one to find the probabilities $\mathcal{P}_i$ and the current across the dot I in response to the applied bias $V = V_L - V_R$. This yields for the linear conductance across the dot [32]

$$G = \lim_{V\to 0} dI/dV = G_\infty \frac{E_C(N_0 - N_0^*)/T}{\sinh[2E_C(N_0 - N_0^*)/T]} \,. \tag{31}$$

Here $N_0 = N_0^*$ corresponds to the Coulomb blockade peak. At each peak, the conductance equals half of its high-temperature value G_∞, see (24). On

[3] Note that a finite inelastic relaxation rate requires inclusion of mechanisms beyond the model (15), e.g., electron-phonon collisions.

the contrary, in the *Coulomb blockade valleys* ($N_0 \neq N_0^*$), the conductance falls off exponentially with the decrease of temperature, and all the valleys behave exactly the same way. Note that the sequential tunneling approximation disregards any interference phenomena for electrons passing the dot. Accordingly, the result (31) is insensitive to a weak magnetic field.

3.2 Coulomb Blockade Peaks at Low Temperature

At temperatures below the single-particle level spacing in the dot δE, the activation energy for electron transport equals the difference between the ground state energies of the Hamiltonian (15) corresponding to two subsequent (integer) eigenvalues of N. Obviously, this difference includes, in addition to the electrostatic contribution $E_+(N_0)$, see (27), also a finite (and random) level spacing. As a result, the distance in N_0 between adjacent Coulomb blockade peaks is no longer 1, but contains a small fluctuating contribution of the order of $\delta E/E_C$. Mesoscopic fluctuations of spacings between the peaks are still subject of a significant disagreement between theory and experiments. We will not consider these fluctuations here (see [15] for a recent review), and discuss only the heights of the peaks.

We concentrate on the temperature interval

$$\Gamma_\alpha \ll T \ll \delta E \,, \tag{32}$$

which extends to lower temperatures the regime considered in the previous section, see (25), and on the gate voltages tuned to the vicinity of the Coulomb blockade peak, see (26). Just as above, the latter condition allows us to neglect all charge states except the two with the lowest energy, $|0\rangle$ and $|1\rangle$. Due to the second inequality in (32), the thermal broadening of single-particle energy levels in the dot can be neglected, and the states $|0\rangle$ and $|1\rangle$ coincide with the ground states of the Hamiltonian (15) with, respectively, $N = N_0^* - 1/2$ and $N = N_0^* + 1/2$ electrons in the dot. To be definite, consider the case when

$$N_0^* = N + 1/2 \tag{33}$$

with N being an even integer; for simplicity, we also neglect the exchange term in (15). Then $|0\rangle$ (with even number of electrons N) is the state in which all single particle levels below the Fermi level ($n < 0$) are doubly occupied. This state is, obviously, non-degenerate. The state $|1\rangle$ differs from $|0\rangle$ by an addition of a single electron on the Fermi level $n = 0$. The extra electron may be in two possible spin states, hence $|1\rangle$ is doubly degenerate; we denote the two components of $|1\rangle$ by $|s\rangle$ with $s = \uparrow, \downarrow$. As discussed below, the degeneracy eventually gives rise to the Kondo effect. However, at $T \gg \Gamma_\alpha$ [see (32)] the quantum coherence associated with the onset of the Kondo effect is not important, and the rate equations approach can still be used to study the transport across the dot [33].

Applying the Fermi golden rule, we write the contribution of electrons with spin s to the electric current $I_{\alpha s}$ from lead α to the dot as

$$I_{\alpha s} = e\,\frac{2\pi}{\hbar}\,\left|t_{\alpha 0}^2\right| \sum_k \delta\left(\xi_k + eV_\alpha - \epsilon_0 - E_+\right)\Big\{\mathcal{P}_0 f(\xi_k) - \mathcal{P}_s\left[1 - f(\xi_k)\right]\Big\}$$

We now neglect ϵ_0 as it is small compared to E_+ (thereby neglecting the mesoscopic fluctuations of the position of the Coulomb blockade peak) and replace the summation over k by an integration. This yields

$$I_{\alpha s} = \frac{2e}{\hbar}\,\Gamma_{\alpha 0}\Big\{\mathcal{P}_0 f\left(E_+ - eV_\alpha\right) - \mathcal{P}_s f\left(eV_\alpha - E_+\right)\Big\}\,. \tag{34}$$

In the steady state the currents $I_{\alpha s}$ satisfy

$$I_{Ls} = -I_{Rs} = I/2 \tag{35}$$

(here we took into account that both projections of spin contribute equally to the total electric current across the dot I). Solution of (34) and (35) subject to the normalization condition $\mathcal{P}_0 + \mathcal{P}_\uparrow + \mathcal{P}_\downarrow = 1$ results in [20]

$$G = \frac{4e^2}{\hbar}\,\frac{\Gamma_{L0}\Gamma_{R0}}{\Gamma_{L0}+\Gamma_{R0}}\,\left[\frac{-df/d\omega}{1+f(\omega)}\right]_{\omega=E_+(N_0)}\,. \tag{36}$$

The case of odd N in (33) is also described by (36) after replacement $E_+(N_0) \to -E_+(N_0)$.

There are several differences between (36) and the corresponding expression (31) valid in the temperature range (25). First of all, the maximum of the conductance (36) occurs at the gate voltage slightly (by an amount of the order of T/E_C) off the degeneracy point $N_0 = N_0^*$, and, more importantly, the shape of the peak is not symmetric about the maximum. This asymmetry is due to correlations in transport of electrons with opposite spins through a single discrete level in the dot. In the maximum, the function (36) takes value

$$G_{\rm peak} \sim \frac{e^2}{h}\,\frac{\Gamma_{L0}\Gamma_{R0}}{\Gamma_{L0}+\Gamma_{R0}}\,\frac{1}{T}\,. \tag{37}$$

Note that (36) and (37) depend on the widths $\Gamma_{\alpha 0}$ of the energy level $n = 0$ rather then on the averages Γ_α over many levels in the dot, as in (31). As already discussed in Sect. 2, the widths $\Gamma_{\alpha 0}$ are related to the values of the electron wave functions at the position of the dot-lead contacts, and, therefore, are random. Accordingly, the heights $G_{\rm peak}$ of the Coulomb blockade peaks exhibit strong mesoscopic fluctuations. In view of (23), the distribution function of $G_{\rm peak}$, see (37), is expected to be broad and strongly non-Gaussian, as well as very sensitive to the magnetic flux threading the dot. This is indeed confirmed by calculations [34, 35] and agrees with experimental data [22, 36]. The expression for the distribution function is rather combersome and we will

not reproduce it here, referring the reader to the original papers [34, 35] and reviews [15, 17, 20]) instead.

An order-of-magnitude estimate of the average height of the peak can be obtained by replacing $\Gamma_{\alpha 0}$ in (37) by Γ_α, see (22), which yields

$$\overline{G_{\rm peak}} \sim G_\infty \frac{\delta E}{T} \ . \tag{38}$$

This is by a factor $\delta E/T$ larger than the corresponding figure $G_{\rm peak} = G_\infty/2$ for the temperature range (25), and may even approach the unitary limit ($\sim e^2/h$) at the lower end of the temperature interval (32). Interestingly, breaking of time-reversal symmetry results in an increase of the average conductance [20]. This increase is analogous to negative magnetoresistance due to weak localization in bulk systems [37], with the same physics involved.

4 Activationless Transport through a Blockaded Quantum Dot

According to the rate equations theory [32], at low temperatures, $T \ll E_C$, conduction through the dot is exponentially suppressed in the Coulomb blockade valleys. This suppression occurs because the process of electron transport through the dot involves a *real transition* to the state in which the charge of the dot differs by e from the thermodynamically most probable value. The probability of such fluctuation is proportional to $\exp\left(-E_C|N_0 - N_0^*|/T\right)$, which explains the conductance suppression, see (30). Going beyond the lowest-order perturbation theory in conductances of the dot-leads junctions G_α allows one to consider processes in which states of the dot with a "wrong" charge participate in the tunneling process as *virtual states*. The existence of these higher-order contributions to the tunneling conductance was envisioned already in 1968 by Giaever and Zeller [38]. The first quantitative theory of this effect, however, was developed much later [39].

The leading contributions to the activationless transport, according to [39], are provided by the processes of *inelastic and elastic co-tunneling*. Unlike the sequential tunneling, in the co-tunneling mechanism, the events of electron tunneling from one of the leads into the dot, and tunneling from the dot to the other lead occur as a single quantum process.

4.1 Inelastic Co-Tunneling

In the inelastic co-tunneling mechanism, an electron tunnels from a lead into one of the vacant single-particle levels in the dot, while it is an electron occupying some other level that tunnels out of the dot, see Fig. 3(a). As a result, transfer of charge e between the leads is accompanied by a simultaneous creation of an electron-hole pair in the dot.

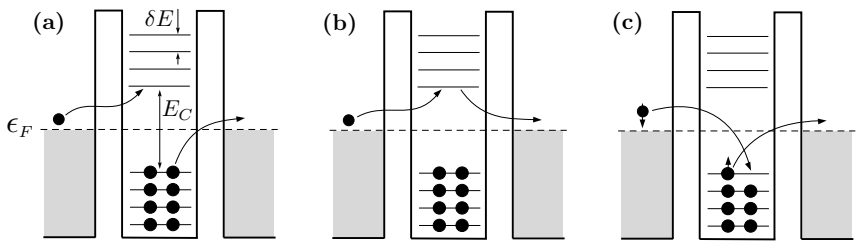

Fig. 3. Examples of the co-tunneling processes. (**a**) Inelastic co-tunneling: transferring of an electron between the leads leaves behind an electron-hole pair in the dot; (**b**) elastic co-tunneling; (**c**) elastic co-tunneling with a flip of spin

Here we will estimate the contribution of the inelastic co-tunneling to the conductance deep in the Coulomb blockade valley, i.e. at almost integer N_0. Consider an electron that tunnels into the dot from the lead L. If energy ω of the electron relative to the Fermi level is small compared to the charging energy, $\omega \ll E_C$, then the energy of the virtual state involved in the co-tunneling process is close to E_C. The amplitude A_{in} of the inelastic transition via this virtual state to the lead R is then given by

$$A_{in} = \frac{t^*_{Ln} t_{Rn'}}{E_C} \,. \tag{39}$$

The initial state of this transition has an extra electron in the single-particle state k in the lead L, while the final state has an extra electron in the state k' in the lead R and an electron-hole pair in the dot (state n is occupied, state n' is empty).

Given the energy of the initial state ω, the number of available final states can be estimated from the phase space argument, familiar from the calculation of the quasiparticle lifetime in the Fermi liquid theory [40]. For $\omega \gg \delta E$ this number is of the order of $(\omega/\delta E)^2$. Since the typical value of ω is T, the inelastic co-tunneling contribution to the conductance can be estimated as

$$G_{in} \sim \frac{e^2}{h} \left(\frac{T}{\delta E}\right)^2 \nu^2 \overline{|A^2_{in}|} \,.$$

Using now (20) and (22), we find [39]

$$G_{in} \sim \frac{G_L G_R}{e^2/h} \left(\frac{T}{E_C}\right)^2 \,. \tag{40}$$

A comparison of (40) with the result of the rate equations theory (31) shows that the inelastic co-tunneling takes over the thermally-activated hopping at moderately low temperatures

$$T \lesssim T_{\rm in} = E_C \left[\ln\left(\frac{e^2/h}{G_L + G_R}\right)\right]^{-1} \,. \tag{41}$$

The smallest energy of the electron-hole pair is of the order of δE. At temperatures below this threshold the inelastic co-tunneling contribution to the conductance becomes exponentially small. It turns out, however, that even at much higher temperatures this mechanism becomes less effective than the elastic co-tunneling.

4.2 Elastic Co-Tunneling

In the process of elastic co-tunneling, transfer of charge between the leads is not accompanied by the creation of an electron-hole pair in the dot. In other words, occupation numbers of single-particle energy levels in the dot in the initial and final states of the co-tunneling process are exactly the same, see Fig. 3(b). Close to the middle of the Coulomb blockade valley (at almost integer N_0) the average number of electrons on the dot, $N \approx N_0$, is also an integer. Both an addition and a removal of a single electron cost E_C in electrostatic energy, see (15). The amplitude of the elastic co-tunneling process in which an electron is transfered from lead L to lead R can then be written as

$$A_{el} = \sum_n t^*_{Ln} t_{Rn} \frac{\mathrm{sign}(\epsilon_n)}{E_C + |\epsilon_n|} \tag{42}$$

The two types of contributions to the amplitude A_{el} are associated with virtual creation of either an electron if the level n is empty ($\epsilon_n > 0$), or of a hole if the level is occupied ($\epsilon_n < 0$); the relative sign difference between the two types of contributions originates in the fermionic commutation relations.

As discussed in Sect. 2, the tunneling amplitudes $t_{\alpha n}$ entering (42) are Gaussian random variables with zero mean and variances given by (21). It is then easy to see that the second moment of the amplitude (42) is given by

$$\overline{|A^2_{el}|} = \frac{\Gamma_L \Gamma_R}{(\pi\nu)^2} \sum_n \left(E_C + |\epsilon_n|\right)^{-2} \ .$$

Since for $E_C \gg \delta E$ the number of terms making significant contribution to the sum over n here is large, and since the sum is converging, one can replace the summation by an integral which yields

$$\overline{|A^2_{el}|} \approx \frac{\Gamma_L \Gamma_R}{(\pi\nu)^2} \frac{1}{E_C \delta E} \ . \tag{43}$$

Substitution of this expression into

$$G_{el} = \frac{4\pi e^2 \nu^2}{\hbar} \left|A^2_{el}\right| \tag{44}$$

and making use of (22) gives [39]

$$\overline{G_{el}} \sim \frac{G_L G_R}{e^2/h} \frac{\delta E}{E_C} \ . \tag{45}$$

for the average value of the elastic co-tunneling contribution to the conductance.

This result is easily generalized to gate voltages tuned away from the middle of the Coulomb blockade valley. The corresponding expression reads

$$\overline{G_{el}} \sim \frac{G_L G_R}{e^2/h}\frac{\delta E}{E_C}\left(\frac{1}{N_0 - N_0^*} + \frac{1}{N_0^* - N_0 + 1}\right) . \tag{46}$$

and is valid when N_0 is not too close to the degeneracy points $N_0 = N_0^*$ and $N_0 = N_0^* + 1$ (N_0^* is a half-integer number):

$$\min\left\{ |N_0 - N_0^*| , |N_0 - N_0^* - 1| \right\} \gg \delta E/E_C$$

Comparison of (45) with (40) shows that the elastic co-tunneling mechanism dominates the electron transport already at temperatures

$$T \lesssim T_{el} = \sqrt{E_C \delta E} , \tag{47}$$

which may exceed significantly the level spacing. However, as we will see shortly below, mesoscopic fluctuations of G_{el} are strong [41], of the order of its average value. Thus, although $\overline{G_{el}}$ is always positive, see (46), the sample-specific value of G_{el} for a given gate voltage may vanish.

The key to understanding the statistical properties of the elastic co-tunneling contribution to the conductance is provided by the observation that there are many ($\sim E_C/\delta E \gg 1$) terms making significant contribution to the amplitude (42). All these terms are random and statisticaly independent of each other. The central limit theorem then suggests that the distribution of A_{el} is Gaussian [20], and, therefore, is completely characterised by the first two statistical moments,

$$\overline{A_{el}} = \overline{A_{el}^*}, \quad \overline{A_{el}A_{el}} = \overline{A_{el}^* A_{el}^*} = \delta_{\beta,1}\,\overline{A_{el}A_{el}^*} \tag{48}$$

with $\overline{A_{el}^* A_{el}}$ given by (43). This can be proven by explicit consideration of higher moments. For example,

$$\overline{|A_{el}^4|} = 2\left(\overline{A_{el}A_{el}^*}\right)^2 + \left|\overline{A_{el}A_{el}}\right|^2 + \delta\,\overline{|A_{el}^4|} . \tag{49}$$

The non-Gaussian correction here, $\delta\,\overline{|A_{el}^4|} \sim \overline{|A_{el}^2|}\,(\delta E/E_C)$, is by a factor of $\delta E/E_C \ll 1$ smaller than the main (Gaussian) contribution.

It follows from (44), (48), and (49) that the fluctuation of the conductance $\delta G_{el} = G_{el} - \overline{G_{el}}$ satisfies

$$\overline{\delta G_{el}^2} = \frac{2}{\beta}\left(\overline{G_{el}}\right)^2 . \tag{50}$$

Note that breaking of time reversal symmetry reduces the fluctuations by a factor of 2, similar to conductance fluctuations in bulk systems, whereas the average conductance (46) is not affected by the magnetic field.

It is clear from (50) that the fluctuations of the conductance are of the order of the conductance itself, despite naive expectations that the large number of the contributing states results in self-averaging. The reason is that one has to add amplitudes, rather than probabilities, in order to compute the conductance. Because the fluctuations of the conductance are large, its distribution function is not Gaussian. Given the statistics (48) of the amplitudes, it is not quite surprising that the distribution of G_{el} normalized to its average coincides with Porter-Thomas distribution (23). This result was obtained first in [41] by a different (and more general) method.

Finally, it is interesting to compare the elastic co-tunneling contribution to the conductance fluctuations with that of inelastic co-tunneling [41],

$$\overline{\delta G_{in}^2} \sim G_{in}^2 \frac{\delta E}{T}$$

with G_{in} given by (40). Even though the inelastic co-tunneling is the main conduction mechanism at $T \gtrsim T_{el}$, see (47), elastic co-tunneling continues to dominate the fluctuations as long as $T \lesssim T_{el}(E_C/\delta E)^{1/6}$.

5 Kondo Regime in Transport through a Quantum Dot

In the above discussion of the elastic co-tunneling we made a tacit assumption that all single-particle levels in the dot are either empty or doubly occupied. This, however, is not the case when the dot has a non-zero spin in the ground state. A dot with odd number of electrons, for example, would necessarily have a half-integer spin S. In the most important case of $S = 1/2$ the top-most occupied single-particle level is filled by a single electron and is spin-degenerate. This opens a possibility of a co-tunneling process in which a transfer of an electron between the leads is accompanied by a flip of electron's spin with simultaneous flip of the spin on the dot, see Fig. 3(c).

The amplitude of such a process, calculated in the fourth order in tunneling matrix elements, diverges logarithmically when the energy ω of an incoming electron approaches 0. Since $\omega \sim T$, the logarithmic singularity in the transmission amplitude translates into a dramatic enhancement of the conductance G across the dot at low temperatures: G may reach values as high as the quantum limit $2e^2/h$ [42, 43]. This conductance enhancement is not really a surprise. Indeed, in the spin-flip co-tunneling process a quantum dot with odd N behaves as $S = 1/2$ magnetic impurity embedded into a tunneling barrier separating two massive conductors [44]. It is known [45] since mid-60's that the presence of such impurities leads to zero-bias anomalies in tunneling conductance [46], which are adequately explained [47, 48] in the context of the Kondo effect [5].

5.1 Effective Low-Energy Hamiltonian

At energies well below the threshold $\Delta \sim \delta E$ for intradot excitations the transitions within the $(2S+1)$-fold degenerate ground state manifold of a dot can be conveniently described by a spin operator $\mathbf{S}$. The form of the *effective Hamiltonian* describing the interaction of the dot with conduction electrons in the leads is then dictated by SU(2) symmetry[4],

$$H_{\rm eff} = \sum_{\alpha k s} \xi_k c^\dagger_{\alpha k s} c_{\alpha k s} + \sum_{\alpha\alpha'} J_{\alpha\alpha'} (\mathbf{s}_{\alpha'\alpha} \cdot \mathbf{S}) \tag{51}$$

with $\mathbf{s}_{\alpha\alpha'} = \sum_{kk'ss'} c^\dagger_{\alpha k s} (\boldsymbol{\sigma}_{ss'}/2)\, c_{\alpha' k' s'}$. The sum over k in (51) is restricted to $|\xi_k| < \Delta$. The exchange amplitudes $J_{\alpha\alpha'}$ form 2×2 Hermitian matrix $\hat{J}$. The matrix has two real eigenvalues, the exchange constants J_1 and J_2 (hereafter we assume that $J_1 \geq J_2$) By an appropriate rotation in the $R-L$ space the Hamiltonian (52) can then be brought into the form

$$H_{\rm eff} = \sum_{\gamma k s} \xi_k \psi^\dagger_{\gamma k s} \psi_{\gamma k s} + \sum_{\gamma} J_\gamma (\mathbf{s}_\gamma \cdot \mathbf{S})\,. \tag{52}$$

Here the operators ψ_γ are certain linear combinations of the original operators $c_{R,L}$ describing electrons in the leads, and

$$\mathbf{s}_\gamma = \sum_{kk'ss'} \psi^\dagger_{\gamma k s} \frac{\boldsymbol{\sigma}_{ss'}}{2} \psi_{\gamma k' s'}$$

is local spin density of itinerant electrons in the "channel" $\gamma = 1, 2$.

The symmetry alone is not sufficient to determine the exchange constants J_γ; their evaluation must rely upon a microscopic model. Here we breifly outline the derivation [30, 49, 50] of (51) for a generic model of a quantum dot system discussed in Sect. 2 above. For simplicity, we will assume that the gate voltage N_0 is tuned to the middle of the Coulomb blockade valley. The tunneling (19) mixes the state with $N = N_0$ electrons on the dot with states having $N \pm 1$ electrons. The electrostatic energies of these states are high ($\sim E_C$), hence the transitions $N \to N \pm 1$ are virtual, and can be taken into account perturbatively in the second order in tunneling amplitudes [51].

For the Hamiltonian (15) the occupations of single-particle energy levels are good quantum numbers. Therefore, the amplitude $J_{\alpha\alpha'}$ can be written as a sum of partial amplitudes,

$$J_{\alpha\alpha'} = \sum_n J^n_{\alpha\alpha'}\,. \tag{53}$$

[4] In writing (51) we omitted the potential scattering terms associated with the usual elastic co-tunneling. This approximation is well justified when the conductances of the dot-lead junctions are small, $G_\alpha \ll e^2/h$, in which case G_{el} is also very small, see (43)

Each term in the sum here corresponds to a process during which an electron or a hole is created virtually on the level n in the dot, cf. (42). For $G_\alpha \ll e^2/h$ and $E_S \ll \delta E$ the main contribution to the sum in (53) comes from singly-occupied energy levels in the dot. A dot with spin S has $2S$ such levels near the Fermi level (hereafter we assign indexes $n = -S, \ldots, n = S$ to these levels), each carrying a spin $\mathbf{S}/2S$, and contributing

$$J^n_{\alpha\alpha'} = \frac{\lambda_n}{E_C}\, t^*_{\alpha n} t_{\alpha' n}\ , \quad \lambda_n = 2/S, \quad |n| \le S \tag{54}$$

to the exchange amplitude in (51). This yields

$$J_{\alpha\alpha'} \approx \sum_{|n|\le S} J^n_{\alpha\alpha'}\ . \tag{55}$$

It follows from (53) and (54) that

$$\mathrm{tr}\hat{J} = \frac{1}{E_C}\sum_n \lambda_n \left(|t^2_{Ln}| + |t^2_{Rn}|\right)\ . \tag{56}$$

By restricting the sum over n here to $|n| \le S$, as in (55), and taking into account that all λ_n in (54) are positive, we find $J_1 + J_2 > 0$. Similarly, from

$$\det \hat{J} = \frac{1}{2E_C^2}\sum_{m,n} \lambda_m \lambda_n |\mathcal{D}^2_{mn}|\ , \quad \mathcal{D}_{mn} = \det \begin{pmatrix} t_{Lm} & t_{Rm} \\ t_{Ln} & t_{Rn} \end{pmatrix} \tag{57}$$

and (54) and (55) follows that $J_1 J_2 > 0$ for $S > 1/2$. Indeed, in this case the sum in (57) contains at least one contribution with $m \ne n$; all such contributions are positive. Thus, both exchange constants $J_{1,2} > 0$ if the dot's spin S exceeds $1/2$ [30]. The pecularities of the Kondo effect in quantum dots with large spin are discussed in [30, 50].

Here we concentrate on the most common situation of $S = 1/2$ on the dot [3]. The ground state of such dot has only one singly-occupied energy level ($n = 0$), so that $\det \hat{J} \approx 0$, see (55) and (57). Accordingly, one of the exchange constants vanishes,

$$J_2 \approx 0\ , \tag{58}$$

while the remaining one, $J_1 = \mathrm{tr}\hat{J}$, is positive. Equation (58) resulted, of course, from the approximation made in (55). For the model (15) the leading correction to (55) originates in the co-tunneling processes with an intermediate state containing an extra electron (or an extra hole) on one of the empty (doubly-occupied) levels. Such contribution arises because the spin on the level n is not conserved by the Hamiltonian (15), unlike the corresponding occupation number. Straightforward calculation [49] yields the partial amplitude in the form of (54), but with

$$\lambda_n = -\frac{2E_C E_S}{(E_C + |\epsilon_n|)^2}\ , \qquad n \ne 0\ .$$

Unless the tunneling amplitudes $t_{\alpha 0}$ to the only singly-occupied level in the dot are anomalously small, the corresponding correction

$$\delta J_{\alpha\alpha'} = \sum_{n\neq 0} J^n_{\alpha\alpha'} \tag{59}$$

to the exchange amplitude (55) is small,

$$\left|\frac{\delta J_{\alpha\alpha'}}{J_{\alpha\alpha'}}\right| \sim \frac{E_S}{\delta E} \ll 1 \,,$$

see (16). To obtain this estimate, we assumed that all tunneling amplitudes $t_{\alpha n}$ are of the same order of magnitude, and replaced the sum over n in (59) by an integral. A similar estimate yields the leading contribution to $\det \hat{J}$,

$$\det \hat{J} \approx \frac{1}{E_C^2} \sum_n \lambda_0 \lambda_n |\mathcal{D}^2_{0n}| \sim -\frac{E_S}{\delta E} \left(\mathrm{tr}\hat{J}\right)^2 \,,$$

or

$$J_2/J_1 \sim -E_S/\delta E \,. \tag{60}$$

According to (60), the exchange constant J_2 is negative [52], and its absolute value is small compared to J_1. Hence (58) is indeed an excellent approximation for large chaotic dots with spin $S = 1/2$ as long as the intradot exchange interaction remains weak, $E_S \ll \delta E$[5]. Note that corrections to the universal Hamiltonian (15) also result in finite values of both exchange constants, $|J_2| \sim J_1 N^{-1/2}$, and become important for small dots with $N \lesssim 10$ [43]. Although this may significantly affect the conductance across the system in the weak coupling regime $T \gtrsim T_K$, it does not lead to qualitative changes in the results for $S = 1/2$ on the dot, as the channel with smaller exchange constant decouples at low energies [54]. With this caveat, we adopt the approximation (58) in our description of the Kondo effect in quantum dots with spin $S = 1/2$. Accordingly, the effective Hamiltonian of the system (52) assumes the"block-diagonal" form

$$H_{\mathrm{eff}} = H_1 + H_2 \tag{61}$$

$$H_1 = \sum_{ks} \xi_k \psi^\dagger_{1ks}\psi_{1ks} + J(\mathbf{s}_1 \cdot \mathbf{S}) \tag{62}$$

$$H_2 = \sum_{ks} \xi_k \psi^\dagger_{2ks}\psi_{2ks} \tag{63}$$

with $J = \mathrm{tr}\hat{J} > 0$.

[5] Equation (58) holds identically for the Anderson impurity model [48] frequently employed to study transport through quantum dots [42, 53]. In that model a quantum dot is described by a single energy level, which formally corresponds to the infinite level spacing limit $\delta E \to \infty$ of the Hamiltonian (15).

To get an idea about the physics of the Kondo model (see [55] for recent reviews), let us first replace the fermion field operator $\mathbf{s}_1$ in (62) by a single-particle spin-1/2 operator $\mathbf{S}_1$. The ground state of the resulting Hamiltonian of two spins

$$\widetilde{H} = J(\mathbf{S}_1 \cdot \mathbf{S})$$

is obviously a singlet. The excited state (a triplet) is separated from the ground state by the energy gap J_1. This separation can be interpreted as the binding energy of the singlet. Unlike $\mathbf{S}_1$ in this simple example, the operator $\mathbf{s}_1$ in (62) is merely a spin density of the conduction electrons at the site of the "magnetic impurity". Because conduction electrons are freely moving in space, it is hard for the impurity to "capture" an electron and form a singlet. Yet, even a weak local exchange interaction suffices to form a singlet ground state [56, 57]. However, the characteristic energy (an analogue of the binding energy) for this singlet is given not by the exchange constant J, but by the so-called Kondo temperature

$$T_K \sim \Delta \exp(-1/\nu J) \; . \tag{64}$$

Using $\Delta \sim \delta E$ and (56) and (22), one obtains from (64) the estimate

$$\ln\left(\frac{\delta E}{T_K}\right) \sim \frac{1}{\nu J} \sim \frac{e^2/h}{G_L + G_R}\frac{E_C}{\delta E} \; . \tag{65}$$

Since $G_\alpha \ll e^2/h$ and $E_C \gg \delta E$, the r.h.s. of (65) is a product of two large parameters. Therefore, the Kondo temperature T_K is small compared to the mean level spacing,

$$T_K \ll \delta E \; . \tag{66}$$

It is this separation of the energy scales that justifies the use of the effective low-energy Hamiltonian (51), (52) for the description of the Kondo effect in a quantum dot system. The inequality (66) remains valid even if the conductances of the dot-leads junctions G_α are of the order of $2e^2/h$. However, in this case the estimate (65) is no longer applicable [58].

In our model, see (61)–(63), one of the channels (ψ_2) of conduction electrons completely decouples from the dot, while the ψ_1-particles are described by the standard single-channel antiferromagnetic Kondo model [5, 55]. Therefore, the thermodynamic properties of a quantum dot in the Kondo regime are identical to those of the conventional Kondo problem for a single magnetic impurity in a bulk metal; thermodynamics of the latter model is fully studied [59]. However, all the experiments addressing the Kondo effect in quantum dots test their transport properties rather then thermodynamics. The electron current operator is not diagonal in the (ψ_1, ψ_2) representation, and the contributions of these two sub-systems to the conductance are not additive. Below we relate the linear conductance and, in some special case, the non-linear differential conductance as well, to the t-matrix of the conventional Kondo problem.

5.2 Linear Response

The linear conductance can be calculated from the Kubo formula

$$G = \lim_{\omega\to 0} \frac{1}{\hbar\,\omega} \int_0^\infty dt\, \mathrm{e}^{i\omega t} \left\langle \left[\hat{I}(t), \hat{I}(0)\right] \right\rangle \ , \tag{67}$$

where the current operator $\hat{I}$ is given by

$$\hat{I} = \frac{d}{dt}\frac{e}{2}\left(\hat{N}_R - \hat{N}_L\right) \ , \quad \hat{N}_\alpha = \sum_{ks} c^\dagger_{\alpha ks} c_{\alpha ks} \tag{68}$$

Here $\hat{N}_\alpha$ is the operator of the total number of electrons in the lead α. Evaluation of the linear conductance proceeds similarly to the calculation of the impurity contribution to the resistivity of dilute magnetic alloys (see, e.g., [60]). In order to take the full advantage of the decomposition (61)–(63), we rewrite $\hat{I}$ in terms of the operators $\psi_{1,2}$. These operators are related to the original operators $c_{R,L}$ representing the electrons in the right and left leads via

$$\begin{pmatrix} \psi_{1ks} \\ \psi_{2ks} \end{pmatrix} = \begin{pmatrix} \cos\theta_0 & \sin\theta_0 \\ -\sin\theta_0 & \cos\theta_0 \end{pmatrix} \begin{pmatrix} c_{Rks} \\ c_{Lks} \end{pmatrix} \ . \tag{69}$$

The rotation matrix here is the same one that diagonalizes matrix $\hat{J}$ of the exchange amplitudes in (51); the rotation angle θ_0 satisfies the equation $\tan\theta_0 = |t_{L0}/t_{R0}|$. With the help of (69) we obtain

$$\hat{N}_R - \hat{N}_L = \cos(2\theta_0)\left(\hat{N}_1 - \hat{N}_2\right) - \sin(2\theta_0)\sum_{ks}\left(\psi^\dagger_{1ks}\psi_{2ks} + \text{H.c.}\right) \tag{70}$$

The current operator $\hat{I}$ entering the Kubo formula (67) is to be calculated with the equilibrium Hamiltonian (61)–(63). Since both $\hat{N}_1$ and $\hat{N}_2$ commute with H_{eff}, the first term in (70) makes no contribution to $\hat{I}$. When the second term in (70) is substituted into (68) and then into the Kubo formula (67), the result, after integration by parts, can be expressed via 2-particle correlation functions such as $\langle\psi^\dagger_1(t)\psi_2(t)\psi^\dagger_2(0)\psi_1(0)\rangle$ (see Appendix B of [61] for further details about this calculation). Due to the block-diagonal structure of H_{eff}, see (61), these correlation functions factorize into products of the single-particle correlation functions describing the (free) ψ_2-particles and the (interacting) ψ_1-particles. The result of the evaluation of the Kubo formula can then be written as

$$G = G_0 \int d\omega \left(-\frac{df}{d\omega}\right) \frac{1}{2}\sum_s \left[-\pi\nu\,\mathrm{Im}\,T_s(\omega)\right] \ . \tag{71}$$

Here

$$G_0 = \frac{2e^2}{h}\sin^2(2\theta_0) = \frac{2e^2}{h}\frac{4|t^2_{L0}t^2_{R0}|}{\left(|t^2_{L0}| + |t^2_{R0}|\right)^2} \ , \tag{72}$$

$f(\omega)$ is the Fermi function, and $T_s(\omega)$ is the t-matrix for the Kondo model (62). The t-matrix is related to the exact retarded Green function of the ψ_1-particles in the conventional way,

$$G_{ks,k's}(\omega) = G^0_k(\omega) + G^0_k(\omega) T_s(\omega) G^0_{k'}(\omega) \ , \quad G^0_k = (\omega - \xi_k + i0)^{-1} \ .$$

Here $G_{ks,k's}(\omega)$ is the Fourier transform of $G_{ks,k's}(t) = -i\theta(t)\langle\{\psi_{1ks}(t), \psi^\dagger_{1k's}\}\rangle$, where $\langle \ldots \rangle$ stands for the thermodynamic averaging with the Hamiltonian (62). In writing (71) we took into account the conservation of the total spin (which implies that $G_{ks,k's'} = \delta_{ss'} G_{ks,k's}$, and that the interaction in (62) is local (which in turn means that the t-matrix is independent of k and k').

5.3 Weak Coupling Regime: $T_K \ll T \ll \delta E$

When the exchange term in the Hamiltonian (62) is treated perturbatively, the main contribution to the t-matrix comes from the transitions of the type [62]

$$|ks, \sigma\rangle \to |k's', \sigma'\rangle \ . \tag{73}$$

Here the state $|ks, \sigma\rangle$ has an extra electron with spin s in the orbital state k whereas the dot is in the spin state σ. By SU(2) symmetry, the amplitude of the transition (73) satisfies

$$A_{|k's',\sigma'\rangle \leftarrow |ks,\sigma\rangle} = A(\omega) \frac{1}{4} \left(\sigma_{\mathbf{s's}} \cdot \sigma_{\sigma'\sigma}\right) \tag{74}$$

Note that the amplitude is independent of k, k' because the interaction is local. However, it may depend on ω due to retardation effects.

The transition (73) is *elastic* in the sense that the number of quasiparticles in the final state of the transition is the same as that in the initial state (in other words, the transition (73) is not accompanied by the production of electron-hole pairs). Therefore, the imaginary part of the t-matrix can be calculated with the help of the optical theorem [63], which yields

$$-\pi\nu \operatorname{Im} T_s = \frac{1}{2} \sum_{\sigma} \sum_{s'\sigma'} \left| \pi\nu \, A^2_{|k's',\sigma'\rangle \leftarrow |ks,\sigma\rangle} \right| \ . \tag{75}$$

The factor $1/2$ here accounts for the probability to have spin σ on the dot in the initial state of the transition. Substitution of the tunneling amplitude in the form (74) into (75), and summation over the spin indexes with the help of the identity (9) result in

$$-\pi\nu \operatorname{Im} T_s = \frac{3\pi^2}{16} \nu^2 \left| A^2(\omega) \right| \ . \tag{76}$$

In the leading (first) order in J one readily obtains $A^{(1)} = J$, independently of ω. However, as discovered by Kondo [5], the second-order contribution $A^{(2)}$ not only depends on ω, but is logarithmically divergent as $\omega \to 0$:

$$A^{(2)}(\omega) = \nu J^2 \ln|\Delta/\omega| \ .$$

Here Δ is the high-energy cutoff in the Hamiltonian (62). It turns out [62] that similar logarithmically divergent contributions appear in all orders of perturbation theory,

$$\nu A^{(n)}(\omega) = (\nu J)^n \left[\ln|\Delta/\omega|\right]^{n-1} \ ,$$

resulting in a geometric series

$$\nu A(\omega) = \sum_{n=1}^{\infty} \nu A^{(n)} = \nu J \sum_{n=0}^{\infty} \left[\nu J \ln|\Delta/\omega|\right]^n = \frac{\nu J}{1 - \nu J \ln|\Delta/\omega|} \ .$$

With the help of the definition of the Kondo temperature (64), this can be written as

$$\nu A(\omega) = \frac{1}{\ln|\omega/T_K|} \ . \tag{77}$$

Substitution of (77) into (76) and then into (71), and evaluation of the integral over ω with logarithmic accuracy yield for the conductance across the dot

$$G = G_0 \frac{3\pi^2/16}{\ln^2(T/T_K)} \ , \quad T \gg T_K. \tag{78}$$

Equation (78) is the leading term of the asymptotic expansion in powers of $1/\ln(T/T_K)$, and represents the conductance in the *leading logarithmic approximation*.

Equation (78) resulted from summing up the most-diverging contributions in all orders of perturbation theory. It is instructive to re-derive it now in the framework of *renormalization group* [64]. The idea of this approach rests on the observation that the electronic states that give a significant contribution to observable quantities, such as conductance, are states within an interval of energies of the width $\omega \sim T$ about the Fermi level, see (71). At temperatures of the order of T_K, when the Kondo effect becomes important, this interval is narrow compared to the width of the band $D = \Delta$ in which the Hamiltonian (62) is defined.

Consider a narrow strip of energies of the width $\delta D \ll D$ near the edge of the band. Any transition (73) between a state near the Fermi level and one of the states in the strip is associated with high ($\sim D$) energy deficit, and, therefore, can only occur virtually. Obviously, virtual transitions via each of the states in the strip result in the second-order correction $\sim J^2/D$ to the amplitude $A(\omega)$ of the transition between states in the vicinity of the Fermi level. Since the strip contains $\nu\delta D$ electronic states, the total correction is [64]

$$\delta A \sim \nu J^2 \delta D/D \ .$$

This correction can be accounted for by modifying the exchange constant in the effective Hamiltonian $\widetilde{H}_{\rm eff}$ which is defined for states within a narrower energy band of the width $D - \delta D$ [64],

$$\widetilde{H}_{\rm eff} = \sum_{ks} \xi_k \psi^\dagger_{1ks} \psi_{1ks} + J_{D-\delta D}(\mathbf{s}_1 \cdot \mathbf{S}) \ , \quad |\xi_k| < D - \delta D \ , \tag{79}$$

$$J_{D-\delta D} = J_D + \nu J_D^2 \frac{\delta D}{D} \ . \tag{80}$$

Here J_D is the exchange constant in the original Hamiltonian. Note that the $\widetilde{H}_{\rm eff}$ has the same form as (62). This is not merely a conjecture, but can be shown rigorously [57, 65].

The reduction of the bandwidth can be considered to be a result of a unitary transformation that decouples the states near the band edges from the rest of the band. In principle, any such transformation should also affect the operators that describe the observable quantities. Fortunately, this is not the case for the problem at hand. Indeed, the angle θ_0 in (69) is not modified by the transformation. Therefore, the current operator and the expression for the conductance (71) retain their form.

Successive reductions of the high-energy cutoff D by small steps δD can be viewed as a continuous process during which the initial Hamiltonian (62) with $D = \Delta$ is transformed to an effective Hamiltonian of the same form that acts within the band of the reduced width $D \ll \Delta$. It follows from (80) that the dependence of the effective exchange constant on D is described by the differential equation [64, 65]

$$\frac{dJ_D}{d\zeta} = \nu J_D^2 \ , \quad \zeta = \ln\left(\Delta/D\right) \ . \tag{81}$$

With the help of (64), the solution of the RG (81) subject to the initial condition $J_\Delta = J$ can be cast into the form

$$\nu J_D = \frac{1}{\ln(D/T_K)} \ .$$

The renormalization described by (81) can be continued until the bandwidth D becomes of the order of the typical energy $|\omega| \sim T$ for real transitions. After this limit has been reached, the transition amplitude $A(\omega)$ is calculated in lowest (first) order of perturbation theory in the effective exchange constant (higher order contributions are negligibly small for $D \sim \omega$),

$$\nu A(\omega) = \nu J_{D\sim|\omega|} = \frac{1}{\ln|\omega/T_K|}$$

Using now (76) and (71), we recover (78).

5.4 Strong Coupling Regime: $T \ll T_K$

As temperature approaches T_K, the leading logarithmic approximation result (78) diverges. This divergence signals the failure of the approximation. Indeed, we are considering a model with single-mode junctions between the

dot and the leads. The maximal possible conductance in this case is $2e^2/h$. To obtain a more precise bound, we discuss in this section the conductance in the strong coupling regime $T \ll T_K$.

We start with the zero-temperature limit $T = 0$. As discussed above, the ground state of the Kondo model (62) is a singlet [56], and, accordingly, is not degenerate. Therefore, the t-matrix of the conduction electrons interacting with the localized spin is completely characterized by the scattering phase shifts δ_s for electrons with spin s at the Fermi level. The t-matrix is then given by the standard scattering theory expression [63]

$$-\pi\nu\, T_s(0) = \frac{1}{2i}\,(\mathbb{S}_s - 1)\,, \quad \mathbb{S}_s = \mathrm{e}^{2i\delta_s}\,, \tag{82}$$

where $\mathbb{S}_s$ is the scattering matrix for electrons with spin s, which for a single channel case reduces to its eignevalue. Substitution of (82) into (71) yields

$$G(0) = G_0 \frac{1}{2} \sum_s \sin^2 \delta_s \tag{83}$$

for the conductance, see (71). The phase shifts in (82), (83) are obviously defined only mod π (that is, δ_s and $\delta_s + \pi$ are equivalent). This ambiguity can be removed if we set to zero the values of the phase shifts at $J = 0$ in (62).

In order to find the two phase shifts δ_s, we need two independent relations. The first one follows from the invariance of the Kondo Hamiltonian (62) under the particle-hole transformation $\psi_{ks} \to s\psi^\dagger_{-k,-s}$ (here $s = \pm 1$ for spin-up/down electrons). The particle-hole symmetry implies the relation for the t-matrix

$$T_s(\omega) = -T^*_{-s}(-\omega)\,, \tag{84}$$

valid at all ω and T. In view of (82), it translates into the relation for the phase shifts at the Fermi level ($\omega = 0$) [66],

$$\delta_\uparrow + \delta_\downarrow = 0\,. \tag{85}$$

The second relation follows from the requirement that the ground state of the Hamiltonian (62) is a singlet [66]. In the absence of exchange ($J = 0$) and at $T = 0$, an infinitesimally weak ($B \to +0$) magnetic field acting on the dot's spin,

$$H_B = BS^z, \tag{86}$$

would polarize it; here B is the Zeeman energy. Since free electron gas has zero spin in the ground state, the total spin in a very large but finite region of space $\mathcal{V}$ surrounding the dot coincides with the spin of the dot, $\langle S^z\rangle_{J=0} = -1/2$. If the exchange with electron gas is now turned on, $J > 0$, a very weak field will not prevent the formation of a singlet ground state. In this state, the total spin within $\mathcal{V}$ is zero. Such change of the spin is possible if the numbers of spin-up and spin-down electrons within $\mathcal{V}$ have changed to compensate for the

dot's spin: $\delta N_\uparrow - \delta N_\downarrow = 1$. By the Friedel sum rule, δN_s are related to the scattering phase shifts at the Fermi level, $\delta N_s = \delta_s/\pi$, which gives

$$\delta_\uparrow - \delta_\downarrow = \pi \ . \tag{87}$$

Combining (85) and (87), we find $|\delta_s| = \pi/2$. Equation (83) then yields for zero-temperature and zero-field conductance across the dot [42]

$$G(0) = G_0 \ . \tag{88}$$

Thus, the growth of the conductance with lowering the temperature is limited only by the value of G_0. This value, see (72), depends only on the ratio of the tunneling amplitudes $|t_{L0}/t_{R0}|$. If $|t_{L0}| = |t_{R0}|$, then the conductance at $T = 0$ will reach the maximal value $G = 2e^2/h$ allowed by quantum mechanics [42].

The maximal conductance, (88), is reached when a a singlet state is formed by the itinerant electrons interacting with the local spin, as described by the Kondo Hamiltonian (62). Perturbation of this singlet [66] by a magnetic field B or temperature T leads to a decrease of the conductance. This decrease is small as long as B and T are small compared to the singlet "binding energy" T_K. The reader is referred to the original papers [66] for the details. Here we only quote the result [60] for the imaginary part of the t-matrix at $|\omega|$ and T small compared to the Kondo temperature T_K,

$$-\pi\nu \,\mathrm{Im}\, T_s(\omega) = 1 - \frac{3\omega^2 + \pi^2 T^2}{2\, T_K^2} \ . \tag{89}$$

Substitution of (89) into (71) yields

$$G = G_0 \left[1 - (\pi T/T_K)^2\right] \ , \quad T \ll T_K \ . \tag{90}$$

Accordingly, corrections to the conductance are quadratic in temperature – a typical Fermi liquid result [66]. The weak-coupling ($T \gg T_K$) and the strong-coupling ($T \ll T_K$) asymptotes of the conductance have a strikingly different structure. Nevertheless, since the Kondo effect is a crossover phenomenon rather than a phase transition [55, 56, 57, 59], the dependence $G(T)$ is a smooth and featureless [67] function throughout the crossover region $T \sim T_K$.

Finally, note that both (78) and (90) have been obtained here for the particle-hole symmetric model (62). This approximation is equivalent to neglecting the elastic co-tunneling contribution to the conductance G_{el}. The asymptotes (78), (90) remain valid [30] as long as $G_{el}/G_0 \ll 1$. The overall temperature dependence of the linear conductance in the middle of the Coulomb blockade valley is sketched in Fig. 4.

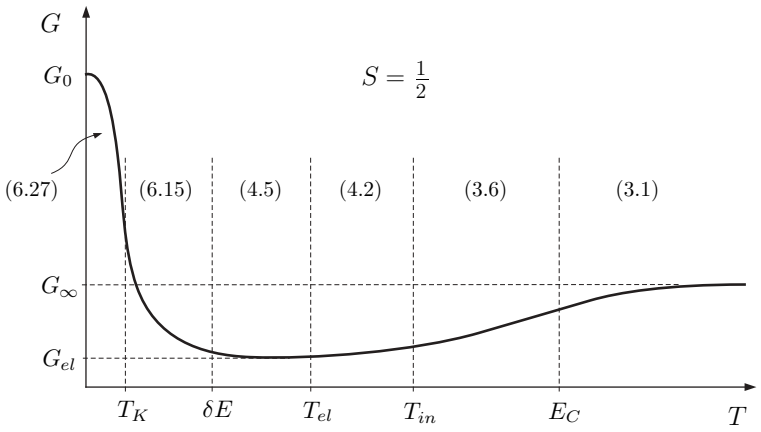

Fig. 4. Sketch of the temperature dependence of the conductance in the middle of the Coulomb blockade valley with $S = 1/2$ on the dot. The numbers in brackets refer to the corresponding equations in the text

6 Discussion

In the simplest form of the Kondo effect considered in these notes, a quantum dot behaves essentially as an artificial "magnetic impurity" with spin S, coupled via exchange interaction to two conducting leads. The details of the temperature dependence $G(T)$ of the linear conductance across the dot depend on the dot's spin S. In the most common case of $S = 1/2$ the conductance in the Kondo regime monotonically increases with the decrease of temperature, potentially up to the quantum limit $2e^2/h$. Qualitatively (although not quantitatively), this increase can be understood from the Anderson impurity model in which the dot is described by a single energy level. On the contrary, when spin on the dot exceeds $1/2$ [68, 69, 70], the evolution of the conductance proceeds in two stages: the conductance first raises, and then drops again when the temperature is lowered [30, 50]. The two-stage Kondo effect was observed recently in a quantum dot tuned to the vicinity of the singlet-triplet transition in its ground state [69]. In GaAs-based lateral quantum dot systems such tuning is achieved by applying a weak magnetic field perpendicular to the plane of the dot [69]. Pecularities of the Kondo effect in lateral quantum dots in the vicinity of the singlet-triplet transition are discussed in [71, 72, 73] (theory of the singlet-triplet transition in vertical dots [8, 9] was developed in [74], see also [61]).

In a typical experiment [3], one measures the dependence of the differential conductance on temperature T, Zeeman energy B, and dc voltage bias V. When one of these parameters is much larger than the other two, and is also large compared to the Kondo temperature T_K, the differential conductance exhibits a logarithmic dependence

$$\frac{1}{G_0}\frac{dI}{dV} \propto \left[\ln \frac{\max\{T, B, eV\}}{T_K}\right]^{-2}, \tag{91}$$

characteristic for the weak coupling regime of the Kondo system. Consider now a zero-temperature transport through a quantum dot with $S = 1/2$ in the presence of a strong field $B \gg T_K$. In accordance with (91), the differential conductance is small compared to G_0 both for $eV \ll B$ and for $eV \gg B$. However, the calculation in the third order of perturbation theory in the exchange constant yields a contribution that diverges logarithmically at $eV = B$ [47]. The divergence arises because of the partial restoration of the coherence associated with the formation of the Kondo singlet: at $eV = B$ the scattered electron has just the right amount of energy to allow for a real transition with a flip of spin. However, the full development of resonance is inhibited by a finite lifetime of the excited spin state of the dot [53, 75]. As a result, the peak in the differential conductance at $eV \sim B$ is broader and lower [53] then the corresponding peak at zero bias in the absence of the field. Even though for $eV \sim B \gg T_K$ the system is clearly in the weak coupling regime, a resummation of the perturbation series turns out to be a prohibitively difficult task, and the expression for the shape of the peak is still unknown. This problem remains to be a subject of active research, see e.g. [76] and references therein.

One encounters similar difficulties in studies of the effect of a weak ac signal of frequency $\Omega \gtrsim T_K$ applied to the gate electrode [77] on transport across the dot. In close analogy with the usual photon-assisted tunneling [78], such perturbation is expected to result in the formation of satellites [79] at $eV = n\hbar\Omega$ (here n is an integer) to the zero-bias peak in the differential conductance. Again, the restoration of coherence responsible for the formation of the satellite peaks is limited by the finite lifetime effects [80].

The spin degeneracy is not the only possible source of the Kondo effect in quantum dots. Consider, for example, a large dot connected by a single-mode junction to a conducting lead and tuned to the vicinity of the Coulomb blockade peak [28]. If one neglects the finite level spacing in the dot, then the two almost degenerate charge state of the dot can be labeled by a pseudospin, while real spin plays the part of the channel index [28, 81]. This setup turns out to be a robust realization [28, 81] of the symmetric (i.e. having equal exchange constants) two-channel $S = 1/2$ Kondo model [54]. The model results in a peculiar temperature dependence of the observable quantities, which at low temperatures follow power laws with manifestly non-Fermi-liquid fractional values of the exponents [82].

It should be emphasized that in the usual geometry consisting of two leads attached to a Coulomb-blockaded quantum dot with $S = 1/2$, only the conventional Fermi-liquid behavior can be observed at low temperatures. Indeed, in this case the two exchange constants in the effective exchange Hamiltonian (52) are vastly different, and their ratio is not tunable by conventional means, see the discussion in Sect. 5.1 above. A way around this

difficulty was proposed recently in [83]. The key idea is to replace one of the leads in the standard configuration by a very large quantum dot, characterized by a level spacing $\delta E'$ and a charging energy E'_C. At $T \gg \delta E'$, particle-hole excitations within this dot are allowed, and electrons in it participate in the screening of the smaller dot's spin. At the same time, as long as $T \ll E'_C$, the number of electrons in the large dot is fixed. Therefore, the large dot provides for a separate screening channel which does not mix with that supplied by the remaining lead. In this system, the two exchange constants are controlled by the conductances of the dot-lead and dot-dot junctions. A strategy for tuning the device parameters to the critical point characterized by the two-channel Kondo physics is discussed in [84].

Finally, we should mention that the description based on the universal Hamiltonian (15) is not applicable to large quantum dots subjected to a *quantizing* magnetic field $H_\perp$ [85, 86]. Such field changes drastically the way the screening occurs in a confined droplet of a two-dimensional electron gas [87]. The droplet is divided into alternating domains containing compressible and incompressible electron liquids. In the metal-like compressible regions, the screening is almost perfect. On the contrary, the incompressible regions behave very much like insulators. In the case of lateral quantum dots, a large compressible domain is formed near the center of the dot. The domain is surrounded by a narrow incompressible region separating it from another compressible ring-shaped domain formed along the edges of the dot [88]. This system can be viewed as two concentric capacitively coupled quantum "dots" – the core dot and the edge dot [85, 88]. When the leads are attached to the edge dot, the measured conductance is sensitive to its spin state: when the number of electrons in the edge dot is odd, the conductance becomes large due to the Kondo effect [85]. Changing the field causes redistribution of electrons between the core and the edge, resulting in a striking checkerboard-like pattern of high- and low-conductance regions [85, 86]. This behavior persists as long as the Zeeman energy remains small compared to the Kondo temperature. Note that compressible regions are also formed around an *antidot* – a potential hill in a two-dimensional electron gas in the quantum Hall regime [89]. Both Coulomb blockade oscillations and Kondo-like behavior were observed in these systems [90].

7 Summary

Kondo effect arises whenever a coupling to a Fermi gas induces transitions within otherwise degenerate ground state multiplet of an interacting system. Both coupling to a Fermi gas and interactions are naturally present in a nanoscale transport experiment. At the same time, many nanostructures can be easily tuned to the vicinity of a degeneracy point. This is why the Kondo effect in its various forms often influences the low temperature transport in meso- and nanoscale systems.

We reviewed in these notes the theory of the Kondo effect in transport through quantum dots. A Coulomb-blockaded quantum dot behaves in many aspects as an artificial "magnetic impurity" coupled via exchange interaction to two conducting leads. Kondo effect in transport through such "impurity" manifests itself in the lifting of the Coulomb blockade at low temperatures, and, therefore, can be unambiguously identified. Quantum dot systems not only offer a direct access to transport properties of an artificial impurity, but also provide one with a broad arsenal of tools to tweak the impurity properties, unmatched in conventional systems. The characteristic energy scale for the intra-dot excitations is much smaller than the corresponding scale for natural magnetic impurities. This allows one to induce degeneracies in the ground state of a dot which are more exotic than just the spin degeneracy. This is only one out of many possible extensions of the simple model discussed here.

Acknowledgements

The research at the University of Minnesota was supported by NSF grants DMR02-37296, and EIA02-10736.

References

1. L.P. Kouwenhoven et al: In: *Mesoscopic Electron Transport*, ed by L.L. Sohn et al (Kluwer, Dordrecht 1997) pp 105–214
   M.A. Kastner: Rev. Mod. Phys. **64**, 849 (1992)
   U. Meirav, E.B. Foxman: Semicond. Sci. Technol. **11**, 255 (1996)
   L.P. Kouwenhoven, C.M. Marcus: Phys. World **11**, 35 (1998) 97, 98, 103
2. P. Joyez et al: Phys. Rev. Lett. **79**, 1349 (1997)
   M. Devoret, C. Glattli: Phys. World **11**, 29 (1998) 97
3. D. Goldhaber-Gordon et al: Nature **391**, 156 (1998)
   S.M. Cronenwett, T.H. Oosterkamp, L.P. Kouwenhoven: Science **281**, 540 (1998)
   J. Schmid et al: Physica B **256–258**, 182 (1998) 98, 101, 102, 115, 124
4. L. Kouwenhoven, L. Glazman: Physics World **14**, 33 (2001) 98
5. J. Kondo: Prog. Theor. Phys. **32**, 37 (1964) 98, 113, 117, 119
6. B.L. Altshuler, A.G. Aronov: In: *Electron-Electron Interactions in Disordered Systems*, ed by A.L. Efros and M. Pollak (North-Holland, Amsterdam 1985) p 1 98
7. S. Tarucha et al: Phys. Rev. Lett. **84**, 2485 (2000)
   L.P. Kouwenhoven, D.G. Austing, S. Tarucha: Rep. Prog. Phys. **64**, 701 (2001) 98
8. S. Sasaki et al: Nature **405**, 764 (2000)
   S. Sasaki et al: Phys. Rev. Lett. **93**, 017205 (2004) 98, 124
9. M. Pustilnik et al: Lecture Notes in Physics **579**, 3 (2001) 98, 124
10. J. Nygård, D.H. Cobden, P.E. Lindelof: Nature **408**, 342 (2000)
    W. Liang, M. Bockrath, H. Park: Phys. Rev. Lett. **88**, 126801 (2002)
    B. Babić, T. Kontos, C. Schönenberger: *Preprint* cond-mat/0407193 98
11. J. Park et al: Nature **417**, 722 (2002)
    W. Liang et al: Nature **417**, 725 (2002)
    L.H. Yu, D. Natelson: Nano Lett. **4**, 79 (2004) 98

12. L.T. Li et al: Phys. Rev. Lett. **80**, 2893 (1998)
V. Madhavan et al: Science **280**, 567 (1998)
H.C. Manoharan et al: Nature **403**, 512 (2000) 98
13. M.V. Berry: Proc. R. Soc. A **400**, 229 (1985)
B.L. Altshuler, B.I. Shklovskii: Sov. Phys.–JETP **64**, 127 (1986) 99, 100
14. C.W.J. Beenakker: Rev. Mod. Phys. **69**, 731 (1997) 99
15. Y. Alhassid, Rev. Mod. Phys. **72**, 895 (2000) 99, 107, 109
16. F. Haake: *Quantum Signatures of Chaos* (Springer, New York 2001) 99, 100, 105
17. K. Efetov: *Supersymmetry in Disorder and Chaos* (Cambridge University Press, Cambridge 1997) 99, 100, 105, 109
18. B.L. Altshuler et al: Phys. Rev. Lett. **78**, 2803 (1997)
O. Agam et al: Phys. Rev. Lett. **78**, 1956 (1997)
Ya.M. Blanter: Phys. Rev. B **54**, 12807 (1996)
Ya.M. Blanter, A.D. Mirlin: Phys. Rev. B **57**, 4566 (1998)
Ya.M. Blanter, A.D. Mirlin, B.A. Muzykantskii: Phys. Rev. Lett. **78**, 2449 (1997)
I.L. Aleiner, L.I. Glazman: Phys. Rev. B **57**, 9608 (1998) 100
19. I.L. Kurland, I.L. Aleiner, B.L. Altshuler: Phys. Rev. B **62**, 14886 (2000) 100, 102, 103
20. I.L. Aleiner, P.W. Brouwer, L.I. Glazman: Phys. Rep. **358**, 309 (2002) 100, 101, 102, 108, 109, 112
21. M.V. Berry: Journ. of Phys. A **10**, 2083 (1977)
A.D. Mirlin: Phys. Rep. **326**, 259 (2000) 100
22. J.A. Folk et al: Phys. Rev. Lett. **76**, 1699 (1996) 100, 108
23. J.M. Ziman: *Principles of the Theory of Solids* (Cambridge University Press, Cambridge 1972) p 339 102
24. J. von Delft, D.C. Ralph: Phys. Rep. **345**, 61 (2001) 102
25. K.A. Matveev, L.I. Glazman, R.I. Shekhter: Mod. Phys. Lett. B **8**, 1007 (1994) 102
26. P.W. Brouwer, Y. Oreg, B.I. Halperin: Phys. Rev. B **60**, R13977 (1999)
H.U. Baranger, D. Ullmo, L.I. Glazman: Phys. Rev. B **61**, R2425 (2000) 103
27. R.M. Potok et al: Phys. Rev. Lett. **91**, 016802 (2003)
J.A. Folk et al: Phys. Scripta **T90**, 26 (2001)
S. Lindemann et al: Phys. Rev. B **66**, 195314 (2002) 103
28. K.A. Matveev: Phys. Rev. B **51**, 1743 (1995) 103, 125
29. K. Flensberg: Phys. Rev. B **48**, 11156 (1993) 103
30. M. Pustilnik, L.I. Glazman: Phys. Rev. Lett. **87**, 216601 (2001) 103, 114, 115, 123, 124
31. C.E. Porter, R.G. Thomas: Phys. Rev. **104**, 483 (1956) 105
32. I.O. Kulik, R.I. Shekhter: Sov. Phys.–JETP **41**, 308 (1975)
L.I. Glazman, R.I. Shekhter: Journ. of Phys.: Condens. Matter **1**, 5811 (1989) 105, 106, 109
33. L.I. Glazman, K.A. Matveev: JETP Lett. **48**, 445 (1988)
C.W.J. Beenakker: Phys. Rev. B **44**, 1646 (1991)
D.V. Averin, A.N. Korotkov, K.K. Likharev: Phys. Rev. B **44**, 6199 (1991) 105, 107
34. R.A. Jalabert, A.D. Stone, Y. Alhassid: Phys. Rev. Lett. **68**, 3468 (1992) 108, 109
35. V.N. Prigodin, K.B. Efetov, S. Iida: Phys. Rev. Lett. **71**, 1230 (1993) 108, 109
36. A.M. Chang et al: Phys. Rev. Lett. **76**, 1695 (1996) 108
37. B.L. Altshuler et al: Phys. Rev. B **22**, 5142 (1980)
S. Hikami, A.I. Larkin, Y. Nagaoka: Progr. Theor. Phys. **63**, 707 (1980) 109
38. I. Giaever, H.R. Zeller: Phys. Rev. Lett. **20**, 1504 (1968)
H.R. Zeller, I. Giaever: Phys. Rev. **181**, 789 (1969) 109
39. D.V. Averin, Yu.V. Nazarov: Phys. Rev. Lett. **65**, 2446 (1990) 109, 110, 111
40. A.A. Abrikosov: *Fundamentals of the Theory of Metals* (North-Holland, Amsterdam 1988) p 620 110

41. I.L. Aleiner, L.I. Glazman: Phys. Rev. Lett. **77**, 2057 (1996) 112, 113
42. L.I. Glazman, M.E. Raikh: JETP Lett. **47**, 452 (1988)
T.K. Ng, P.A. Lee: Phys. Rev. Lett. **61**, 1768 (1988) 113, 116, 123
43. W.G. van der Wiel et al: Science **289**, 2105 (2000)
Y. Ji, M. Heiblum, H. Shtrikman: Phys. Rev. Lett. **88**, 076601 (2002) 113, 116
44. L.I. Glazman, M. Pustilnik: In: *New Directions in Mesoscopic Physics (Towards Nanoscience)* ed by R. Fazio et al (Kluwer, Dordrecht 2003) p 93 113
45. C.B. Duke: *Tunneling in Solids* (Academic Press, New York 1969)
J.M. Rowell: In: *Tunneling Phenomena in Solids* ed by E. Burstein, S. Lundqvist (Plenum Press, New York 1969) p 385 113
46. A.F.G. Wyatt: Phys. Rev. Lett. **13**, 401 (1964)
R.A. Logan, J.M. Rowell: Phys. Rev. Lett. **13**, 404 (1964) 113
47. J. Appelbaum: Phys. Rev. Lett. **17**, 91 (1966)
J.A. Appelbaum: Phys. Rev. **154**, 633 (1967) 113, 125
48. P.W. Anderson: Phys. Rev. Lett. **17**, 95 (1966) 113, 116
49. G.A. Fiete et al: Phys. Rev. B **66**, 024431 (2002) 114, 115
50. M. Pustilnik, L.I. Glazman: Journ. of Physics: Condens. Matter **16**, R513 (2004) 114, 115, 124
51. J.R. Schrieffer, P.A. Wolff: Phys. Rev. **149**, 491 (1966) 114
52. P.G. Silvestrov, Y. Imry: Phys. Rev. Lett. **85**, 2565 (2000) 116
53. Y. Meir, N.S. Wingreen, P.A. Lee: Phys. Rev. Lett. **70**, 2601 (1993) 116, 125
54. P. Nozières, A. Blandin: J. Physique **41**, 193 (1980) 116, 125
55. P. Coleman: In: *Lectures on the Physics of Highly Correlated Electron Systems VI*, ed. F. Mancini (American Institute of Physics, New York 2002) pp 79–160 (cond-mat/0206003)
A.S. Hewson: *The Kondo Problem to Heavy Fermions* (Cambridge University Press, Cambridge 1997) 117, 123
56. P.W. Anderson: *Basic Notions of Condensed Matter Physics* (Addison-Wesley, Reading 1997) 117, 122, 123
57. K.G. Wilson: Rev. Mod. Phys. **47**, 773 (1975) 117, 121, 123
58. L.I. Glazman, F.W.J. Hekking, A.I. Larkin: Phys. Rev. Lett. **83**, 1830 (1999) 117
59. A.M. Tsvelick, P.B. Wiegmann: Adv. Phys. **32**, 453 (1983)
N. Andrei, K. Furuya, J.H. Lowenstein: Rev. Mod. Phys. **55**, 331 (1983) 117, 123
60. I. Affleck, A.W.W. Ludwig: Phys. Rev. B **48**, 7297 (1993) 118, 123
61. M. Pustilnik, L.I. Glazman: Phys. Rev. B **64**, 045328 (2001) 118, 124
62. A.A. Abrikosov: Physics **2**, 5 (1965)
A.A. Abrikosov: Sov. Phys.–Uspekhi **12**, 168 (1969) 119, 120
63. R.G. Newton: *Scattering Theory of Waves and Particles* (Dover, Mineola 2002) 119, 122
64. P.W. Anderson: Journ. of Physics C **3**, 2436 (1970) 120, 121
65. P.W. Anderson, G. Yuval, D.R. Hamann: Phys. Rev. B **1**, 4464 (1970) 121
66. P. Nozières: J. Low Temp. Phys. **17**, 31 (1974)
P. Nozières: J. Physique **39**, 1117 (1978) 122, 123
67. T.A. Costi, A.C. Hewson, V. Zlatić: Journ. of Physics: Condens. Matter **6**, 2519 (1994) 123
68. J. Schmid et al: Phys. Rev. Lett. **84**, 5824 (2000) 72, 124
69. W.G. van der Wiel et al: Phys. Rev. Lett. **88**, 126803 (2002) 124
70. A. Kogan et al: Phys. Rev. B **67**, 113309 (2003) 124
71. M. Pustilnik, Y. Avishai, K. Kikoin: Phys. Rev. Lett. **84**, 1756 (2000) 124
72. V.N. Golovach, D. Loss: Europhys. Lett. **62**, 83 (2003)
M. Pustilnik, L.I. Glazman, W. Hofstetter: Phys. Rev. B **68**, 161303(R) (2003) 124

73. W. Izumida, O. Sakai,Y. Shimizu: J. Phys. Soc. Jpn. **67**, 2444 (1998) 124
74. M. Eto, Yu.V. Nazarov: Phys. Rev. Lett. **85**, 1306 (2000)
M. Pustilnik, L.I. Glazman: Phys. Rev. Lett. **85**, 2993 (2000) 124
75. D.L. Losee, E.L. Wolf: Phys. Rev. Lett. **23**, 1457 (1969) 125
76. A. Rosch et al: Phys. Rev. B **68**, 014430 (2003) 125
77. J.M. Elzerman et al: J. Low Temp. Phys. **118**, 375 (2000) 125
78. P.K. Tien, J.P. Gordon: Phys. Rev. **129**, 647 (1963) 125
79. M.H. Hettler and H. Schoeller: Phys. Rev. Lett. **74**, 4907 (1995) 125
80. A. Kaminski, Yu.V. Nazarov, L.I. Glazman: Phys. Rev. Lett. **83**, 384 (1999)
A. Kaminski, Yu.V. Nazarov, L.I. Glazman: Phys. Rev. B **62**, 8154 (2000) 125
81. K.A. Matveev: Sov. Phys.–JETP **72**, 892 (1991) 125
82. D.L. Cox, A. Zawadowski: Adv. Phys. **47**, 599 (1998) 125
83. Y. Oreg, D. Goldhaber-Gordon: Phys. Rev. Lett. **90**, 136602 (2003) 126
84. M. Pustilnik et al: Phys. Rev. B **69**, 115316 (2004) 126
85. M. Keller et al: Phys. Rev. B **64**, 033302 (2001)
M. Stopa et al: Phys. Rev. Lett. **91**, 046601 (2003) 126
86. S.M. Maurer et al: Phys. Rev. Lett. **83**, 1403 (1999)
D. Sprinzak et al: Phys. Rev. Lett. **88**, 176805 (2002)
C. Fühner et al: Phys. Rev. B **66**, 161305(R) (2002) 126
87. C.W.J. Beenakker: Phys. Rev. Lett. **64**, 216 (1990)
A.M. Chang: Solid State Commun. **74**, 871 (1990)
D.B. Chklovskii, B.I. Shklovskii, L.I. Glazman: Phys. Rev. B **46**, 4026 (1992) 126
88. P.L. McEuen et al: Phys. Rev. B **45**, 11419 (1992) A.K. Evans, L.I. Glazman, B.I. Shklovskii: Phys. Rev. B **48**, 11120 (1993) 126
89. V.J. Goldman, B. Su: Science **267**, 1010 (1995) 126
90. M. Kataoka et al: Phys. Rev. Lett. **83**, 160 (1999)
M. Kataoka et al: Phys. Rev. Lett. **89**, 226803 (2002) 126

Andreev Billiards

C.W.J. Beenakker

Instituut-Lorentz, Universiteit Leiden, P.O. Box 9506, 2300 RA Leiden,
The Netherlands

Summary. This is a review of recent advances in our understanding of how Andreev reflection at a superconductor modifies the excitation spectrum of a quantum dot. The emphasis is on two-dimensional impurity-free structures in which the classical dynamics is chaotic. Such Andreev billiards differ in a fundamental way from their non-superconducting counterparts. Most notably, the difference between chaotic and integrable classical dynamics shows up already in the level density, instead of only in the level–level correlations. A chaotic billiard has a gap in the spectrum around the Fermi energy, while integrable billiards have a linearly vanishing density of states. The excitation gap $E_{\rm gap}$ corresponds to a time scale $\hbar/E_{\rm gap}$ which is classical ($\hbar$-independent, equal to the mean time $\tau_{\rm dwell}$ between Andreev reflections) if $\tau_{\rm dwell}$ is sufficiently large. There is a competing quantum time scale, the Ehrenfest time τ_E, which depends logarithmically on $\hbar$. Two phenomenological theories provide a consistent description of the τ_E-dependence of the gap, given qualitatively by $E_{\rm gap} \simeq \min(\hbar/\tau_{\rm dwell}, \hbar/\tau_E)$. The analytical predictions have been tested by computer simulations but not yet experimentally.

1 Introduction

Forty years ago, Andreev discovered a peculiar property of superconducting mirrors [1]. As illustrated in Fig. 1, an electron that tries to enter a superconductor coming from the Fermi level of a normal metal is forced to retrace its path, as if time is reversed. Also the charge of the particle is reversed, since the negatively charged electron is converted into a positively charged hole. The velocity of a hole is opposite to its momentum, so the superconducting mirror conserves the momentum of the reflected particle. In contrast, reflection at an ordinary mirror (an insulator) conserves charge but not momentum. The unusual scattering process at the interface between a normal metal (N) and a superconductor (S) is called Andreev reflection.

Andreev reflection is the key concept needed to understand the properties of nanostructures with NS interfaces [2]. Most of the research has concentrated on transport properties of open structures, see [3, 4] for reviews. There

C.W.J. Beenakker: *Andreev Billiards*, Lect. Notes Phys. **667**, 131–174 (2005)
www.springerlink.com

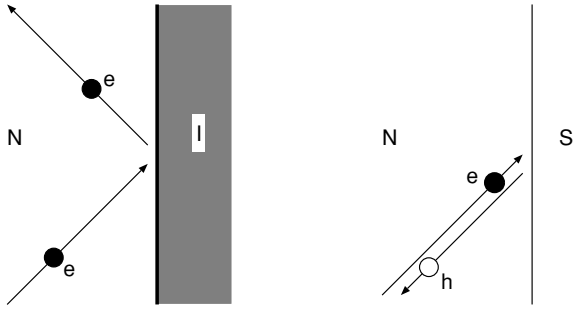

Fig. 1. Normal reflection by an insulator (I) versus Andreev reflection by a superconductor (S) of an electron excitation in a normal metal (N) near the Fermi energy E_F. Normal reflection (*left*) conserves charge but does not conserve momentum. Andreev reflection (*right*) conserves momentum but does not conserve charge: The electron (e) is reflected as a hole (h) with the same momentum and opposite velocity (*retroreflection*). The missing charge of $2e$ is absorbed as a Cooper pair by the superconducting condensate. The electron-hole symmetry is exact at the Fermi level. If the electron is at a finite energy E above E_F, then the hole is at an energy E below E_F. The energy difference of $2E$ breaks the electron-hole symmetry. From [3]

experiment and theory have reached a comparable level of maturity. In the present review we focus on spectral properties of closed structures, such as the quantum dot with superconducting contacts shown in Fig. 2. The theoretical understanding of these systems, gained from the combination of analytical theory and computer simulations, has reached the stage that a comprehensive review is called for – even though an experimental test of the theoretical predictions is still lacking.

An impurity-free quantum dot in contact with a superconductor has been called an "Andreev billiard" [5].[1] The name is appropriate, and we will use it too, because it makes a connection with the literature on quantum chaos [7, 8]. A billiard (in the sense of a bounded two-dimensional region in which all scattering occurs at the boundaries) is the simplest system in which to search for quantum mechanical signatures of chaotic classical dynamics. That is the basic theme of the field of quantum chaos. By introducing a superconducting segment in the boundary of a billiard one has the possibility of unraveling the chaotic dynamics, so to say by making time flow backwards. Andreev billiards therefore reveal features of the chaotic dynamics that are obscured in their normal (non-superconducting) counterparts.

The presence of even the smallest superconducting segment suppresses the quantum mechanical level density at sufficiently low excitation energies. This suppression may take the form of an excitation gap, at an energy $E_{\rm gap}$ well

[1] Open structures containing an antidot lattice have also been called "Andreev billiards" [6], but in this review we restrict ourselves to closed systems.

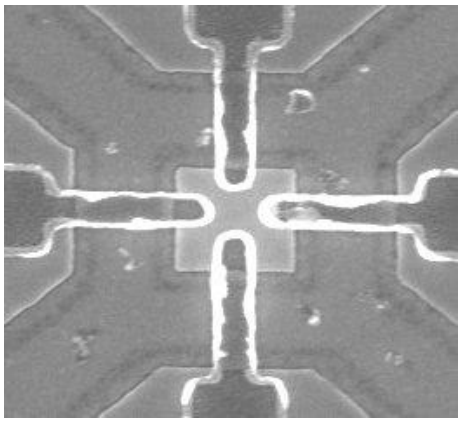

Fig. 2. Quantum dot (central square of dimensions 500 nm × 500 nm) fabricated in a high-mobility InAs/AlSb heterostructure and contacted by four superconducting Nb electrodes. Device made by A. T. Filip, Groningen University (unpublished figure)

below the gap Δ in the bulk superconductor (hence the name "minigap"). It may also take the form of a level density that vanishes smoothly (typically linearly) upon approaching the Fermi level, without an actual gap. The presence or absence of a gap is a quantum signature of chaos. That is a fundamental difference between normal billiards and Andreev billiards, since in a normal billiard the level density can not distinguish chaotic from integrable classical dynamics. (It depends only on the area of the billiard, not on its shape.)

A powerful technique to determine the spectrum of a chaotic system is random-matrix theory (RMT) [3, 9, 10]. Although the appearance of an excitation gap is a quantum mechanical effect, the corresponding time scale $\hbar/E_{\rm gap}$ as it follows from RMT is a classical (meaning $\hbar$-independent) quantity: It is the mean time $\tau_{\rm dwell}$ that an electron or hole excitation dwells in the billiard between two subsequent Andreev reflections. A major development of the last few years has been the discovery of a competing quantum mechanical time scale $\tau_E \propto |\ln\hbar|$. (The subscript E stands for Ehrenfest.) RMT breaks down if $\tau_E \gtrsim \tau_{\rm dwell}$ and a new theory is needed to determine the excitation gap in this regime. Two different phenomenological approaches have now reached a consistent description of the τ_E-dependence of the gap, although some disagreement remains.

The plan of this review is as follows. The next four sections contain background material on Andreev reflection (Sect. 2), on the minigap in NS junctions (Sect. 3), on the scattering theory of Andreev billiards (Sect. 4), and on a stroboscopic model used in computer simulations (Sect. 5). The regime of RMT (when $\tau_E \ll \tau_{\rm dwell}$) is described in Sect. 6 and the quasiclassical regime (when $\tau_E \gg \tau_{\rm dwell}$) is described in Sect. 7. The crossover from $E_{\rm gap} \simeq \hbar/\tau_{\rm dwell}$ to $E_{\rm gap} \simeq \hbar/\tau_E$ is the topic of Sect. 8. We conclude in Sect. 9.

2 Andreev Reflection

The quantum mechanical description of Andreev reflection starts from a pair of Schrödinger equations for electron and hole wave functions $u(\mathbf{r})$ and $v(\mathbf{r})$, coupled by the pair potential $\Delta(\mathbf{r})$. These socalled Bogoliubov-De Gennes (BdG) equations [11] take the form

$$\mathcal{H}_{\rm BG}\begin{pmatrix} u \\ v \end{pmatrix} = E\begin{pmatrix} u \\ v \end{pmatrix} , \tag{1}$$

$$\mathcal{H}_{\rm BG} = \begin{pmatrix} H & \Delta(\mathbf{r}) \\ \Delta^*(\mathbf{r}) & -H^* \end{pmatrix} . \tag{2}$$

The Hamiltonian $H = (\mathbf{p} + e\mathbf{A})^2/2m + V - E_F$ is the single-electron Hamiltonian in the field of a vector potential $\mathbf{A}(\mathbf{r})$ and electrostatic potential $V(\mathbf{r})$. The excitation energy E is measured relative to the Fermi energy E_F. If (u, v) is an eigenfunction with eigenvalue E, then $(-v^*, u^*)$ is also an eigenfunction, with eigenvalue $-E$. The complete set of eigenvalues thus lies symmetrically around zero. The quasiparticle excitation spectrum consists of all positive E.

In a uniform system with $\Delta(\mathbf{r}) \equiv \Delta$, $\mathbf{A}(\mathbf{r}) \equiv 0$, $V(\mathbf{r}) \equiv 0$, the solution of the BdG equations is

$$E = \left[(\hbar^2k^2/2m - E_F)^2 + \Delta^2\right]^{1/2} , \tag{3}$$

$$u(\mathbf{r}) = (2E)^{-1/2}\left(E + \hbar^2k^2/2m - E_F\right)^{1/2} e^{i\mathbf{k}\cdot\mathbf{r}} , \tag{4}$$

$$v(\mathbf{r}) = (2E)^{-1/2}\left(E - \hbar^2k^2/2m + E_F\right)^{1/2} e^{i\mathbf{k}\cdot\mathbf{r}} . \tag{5}$$

The excitation spectrum is continuous, with excitation gap Δ. The eigenfunctions (u, v) are plane waves characterized by a wavevector $\mathbf{k}$. The coefficients of the plane waves are the two coherence factors of the BCS (Bardeen-Cooper-Schrieffer) theory.

At an interface between a normal metal and a superconductor the pairing interaction drops to zero over atomic distances at the normal side. (We assume non-interacting electrons in the normal region.) Therefore, $\Delta(\mathbf{r}) \equiv 0$ in the normal region. At the superconducting side of the NS interface, $\Delta(\mathbf{r})$ recovers its bulk value Δ only at some distance from the interface. This suppression of $\Delta(\mathbf{r})$ is neglected in the step-function model

$$\Delta(\mathbf{r}) = \begin{cases} \Delta & \text{if } \mathbf{r} \in S , \\ 0 & \text{if } \mathbf{r} \in N . \end{cases} \tag{6}$$

The step-function pair potential is also referred to in the literature as a "rigid boundary condition" [12]. It greatly simplifies the analysis of the problem without changing the results in any qualitative way.

Since we will only be considering a single superconductor, the phase of the superconducting order parameter is irrelevant and we may take Δ real. We refer to [13] for a tutorial introduction to mesoscopic Josephson junctions, such as a quantum dot connected to two superconductors, and to [14] for a comprehensive review of the current-phase relation in Josephson junctions.

3 Minigap in NS Junctions

The presence of a normal metal interface changes the excitation spectrum (proximity effect). The continuous spectrum above the bulk gap Δ differs from the BCS form (4) and in addition there may appear discrete energy levels $E_n < \Delta$.

The wave function of the lowest level contains electron and hole components u_0, v_0 of equal magnitude, mixed by Andreev reflection. The mean time $\tau_{\rm dwell}$ between Andreev reflections (corresponding to the mean life time of an electron or hole excitation) sets the scale $E_0 \equiv E_{\rm gap} \simeq \hbar/\tau_{\rm dwell}$ for the energy of this lowest level [15]. This "minigap" is smaller than the bulk gap by a factor $\xi_0/v_F\tau_{\rm dwell}$, with $\xi_0 = \hbar v_F/\Delta$ the superconducting coherence length and v_F the Fermi velocity. The energy $\hbar/\tau_{\rm dwell}$ is called the Thouless energy E_T, because of the role it plays in Thouless's theory of localization [2].

The simplest NS junction, which can be analyzed exactly [16], consists of an impurity-free normal metal layer (thickness d) on top of a bulk superconductor.[2]

Because of translational invariance parallel to the NS interface, the parallel component $p_\parallel$ of the momentum is a good quantum number. The lowest excitation energy

$$E_0(p_\parallel) = \frac{\pi\hbar}{2T\left(p_\parallel\right)}\ ,\quad T\left(p_\parallel\right) = \frac{2md}{\left(p_F^2 - p_\parallel^2\right)^{1/2}}\ , \tag{7}$$

is the reciprocal of the time $T(p_\parallel)$ between two subsequent Andreev reflections. This time diverges when $p_\parallel$ approaches the Fermi momentum $p_F = \hbar k_F = \sqrt{2mE_F}$, so E_0 can come microscopically close to zero. The lower limit $E_0 \gtrsim \hbar^2/md^2$ is set by the quantization of the momentum perpendicular to the layer.

Impurities in the normal metal layer (with mean free path l) prevent the time between Andreev reflections to grow much larger than $\tau_{\rm dwell} \simeq \max(l/v_F, d^2/v_F l)$. The excitation gap [18, 19, 20]

$$E_{\rm gap} \simeq \hbar/\tau_{\rm dwell} \simeq (\hbar v_F l/d^2)\min(1, d^2/l^2) \tag{8}$$

is now a factor $k_F l \min(1, d^2/l^2)$ larger than in the absence of impurities. A precise calculation using disorder-averaged Green functions (reviewed in [21]) gives the curve shown in Fig. 3. The two asymptotes are [20]

[2] The paper by De Gennes and Saint-James [16] first noticed the mixed electron-hole character of excitations below Δ. The property of retroreflection at the NS interface was first noticed by Andreev [1], which in the opinion of the present author justifies the name "Andreev reflection". For a different opinion, see [17].

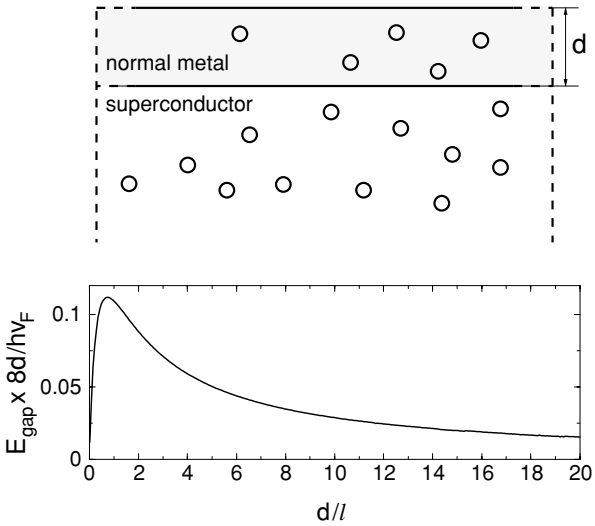

Fig. 3. Excitation gap $E_{\rm gap}$ of a disordered NS junction, as a function of the ratio of the thickness d of the normal metal layer and the mean free path l. The curve in the *bottom* panel is calculated from the disorder-averaged Green function (for $\xi_0 \ll d, l$). The *top* panel illustrates the geometry. The normal metal layer has a specularly reflecting upper surface and an ideally transmitting lower surface. Adapted from [20]

$$E_{\rm gap} = \begin{cases} 0.43\,\hbar v_F/l\,, & \text{if } \ d/l \ll 1\,, \\ 0.78\,\hbar D/d^2\,, & \text{if } \ d/l \gg 1\,, \end{cases} \tag{9}$$

with $D = v_F l/3$ the diffusion constant in the normal metal.

The minigap in a ballistic quantum dot (Andreev billiard) differs from that in a disordered NS junction in two qualitative ways:

1. The opening of an excitation gap depends on the shape of the boundary, rather than on the degree of disorder [22]. A chaotic billiard has a gap at the Thouless energy $E_T \simeq \hbar/\tau_{\rm dwell}$, like a disordered NS junction. An integrable billiard has a linearly vanishing density of states, like a ballistic NS junction.
2. In a chaotic billiard a new time scale appears, the Ehrenfest time τ_E, which competes with $\tau_{\rm dwell}$ in setting the scale for the excitation gap [23]. While $\tau_{\rm dwell}$ is a classical $\hbar$-independent time scale, $\tau_E \propto |\ln\hbar|$ has a quantum mechanical origin.

Because one can not perform a disorder average in Andreev billiards, the Green function formulation is less useful than in disordered NS junctions. Instead, we will make extensive use of the scattering matrix formulation, explained in the next section.

4 Scattering Formulation

In the step-function model (6) the excitation spectrum of the coupled electron-hole quasiparticles can be expressed entirely in terms of the scattering matrix of normal electrons [24].

The scattering geometry is illustrated in Fig. 4. It consists of a finite normal-metal region N adjacent to a semi-infinite superconducting region S. The metal region represents the Andreev billiard. To obtain a well-defined scattering problem we insert an ideal (impurity-free) normal lead between N and S. We assume that the only scattering in the superconductor consists of Andreev reflection at the NS interface (no disorder in S). The superconductor may then also be represented by an ideal lead. We choose a coordinate system so that the normal and superconducting leads lie along the x-axis, with the interface at $x = 0$.

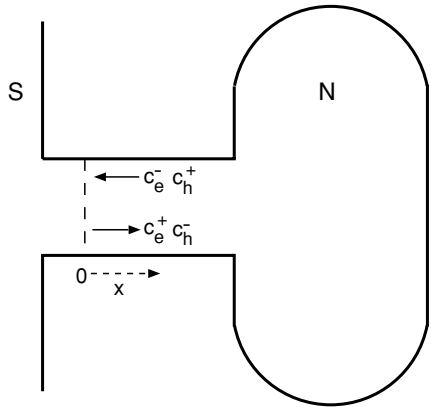

Fig. 4. Normal metal (N) containing an Andreev billiard, coupled to a superconductor (S) by an ideal lead. The *dashed* line represents the NS interface. Scattering states $c^{\rm in} = (c_{\rm e}^+, c_{\rm h}^-)$ and $c^{\rm out} = (c_{\rm e}^-, c_{\rm h}^+)$ are indicated schematically

We first construct a basis for the scattering matrix. In the normal lead N the eigenfunctions of the BdG equation (1) can be written in the form

$$\Psi_{n,{\rm e}}^{\pm}({\rm N}) = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \frac{1}{\sqrt{k_n^{\rm e}}} \Phi_n(y,z) \exp(\pm {\rm i} k_n^{\rm e} x) \;, \tag{10a}$$

$$\Psi_{n,{\rm h}}^{\pm}({\rm N}) = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \frac{1}{\sqrt{k_n^{\rm h}}} \Phi_n(y,z) \exp(\pm {\rm i} k_n^{\rm h} x) \;, \tag{10b}$$

where the wavenumbers $k_n^{\rm e}$ and $k_n^{\rm h}$ are given by

$$k_n^{\rm e,h} = \frac{\sqrt{2m}}{\hbar} (E_F - E_n + \sigma^{\rm e,h} E)^{1/2} \;, \tag{11}$$

and we have defined $\sigma^{\rm e}\equiv 1$, $\sigma^{\rm h}\equiv -1$. The labels e and h indicate the electron or hole character of the wave function. The index n labels the modes, $\Phi_n(y,z)$ is the transverse wave function of the n-th mode, and E_n its threshold energy:

$$\left[\left(p_y^2+p_z^2\right)/2m+V(y,z)\right]\Phi_n(y,z)=E_n\Phi_n(y,z)\;. \tag{12}$$

The eigenfunction Φ_n is normalized to unity, $\int dy\int dz\,|\Phi_n|^2=1$.

In the superconducting lead S the eigenfunctions are

$$\begin{aligned}\Psi^{\pm}_{n,{\rm e}}({\rm S})&=\begin{pmatrix}{\rm e}^{{\rm i}\eta^{\rm e}/2}\\ {\rm e}^{-{\rm i}\eta^{\rm e}/2}\end{pmatrix}\frac{1}{\sqrt{2q_n^{\rm e}}}(E^2/\Delta^2-1)^{-1/4}\\ &\quad\times\Phi_n(y,z)\exp(\pm{\rm i}q_n^{\rm e}x)\;, \qquad (13a)\\ \Psi^{\pm}_{n,{\rm h}}({\rm S})&=\begin{pmatrix}{\rm e}^{{\rm i}\eta^{\rm h}/2}\\ {\rm e}^{-{\rm i}\eta^{\rm h}/2}\end{pmatrix}\frac{1}{\sqrt{2q_n^{\rm h}}}(E^2/\Delta^2-1)^{-1/4}\\ &\quad\times\Phi_n(y,z)\exp(\pm{\rm i}q_n^{\rm h}x)\;. \qquad (13b)\end{aligned}$$

We have defined

$$q_n^{\rm e,h}=\frac{\sqrt{2m}}{\hbar}[E_F-E_n+\sigma^{\rm e,h}(E^2-\Delta^2)^{1/2}]^{1/2}, \tag{14}$$

$$\eta^{\rm e,h}=\sigma^{\rm e,h}\arccos(E/\Delta)\;. \tag{15}$$

The wave functions (10) and (13) have been normalized to carry the same amount of quasiparticle current, because we want to use them as the basis for a unitary scattering matrix. The direction of the velocity is the same as the wave vector for the electron and opposite for the hole.

A wave incident on the Andreev billiard is described in the basis (10) by a vector of coefficients

$$c^{\rm in}=(c_{\rm e}^{+},c_{\rm h}^{-})\;, \tag{16}$$

as shown schematically in Fig. 4. (The mode index n has been suppressed for simplicity of notation.) The reflected wave has vector of coefficients

$$c^{\rm out}=(c_{\rm e}^{-},c_{\rm h}^{+})\;. \tag{17}$$

The scattering matrix $S_{\rm N}$ of the normal region relates these two vectors, $c_{\rm N}^{\rm out}=S_{\rm N}c_{\rm N}^{\rm in}$. Because the normal region does not couple electrons and holes, this matrix has the block-diagonal form

$$S_{\rm N}(E)=\begin{pmatrix}S(E) & 0\\ 0 & S(-E)^{*}\end{pmatrix}. \tag{18}$$

Here $S(E)$ is the unitary scattering matrix associated with the single-electron Hamiltonian H. It is an $N\times N$ matrix, with $N(E)$ the number of propagating modes at energy E. The dimension of $S_{\rm N}(E)$ is $N(E)+N(-E)$.

For energies $0 < E < \Delta$ there are no propagating modes in the superconducting lead S. Restricting ourselves to that energy range, we can define a scattering matrix $S_{\rm A}$ for Andreev reflection at the NS interface by $c^{\rm in} = S_{\rm A} c^{\rm out}$. The elements of $S_{\rm A}$ are obtained by matching the wave function (10) at $x = 0$ to the decaying wave function (13). Since $\Delta \ll E_F$ one may ignore normal reflections at the NS interface and neglect the difference between $N(E)$ and $N(-E)$. This is known as the Andreev approximation [1]. The result is

$$S_{\rm A}(E) = \begin{pmatrix} 0 & \alpha(E) \\ \alpha(E) & 0 \end{pmatrix} , \tag{19}$$

$$\alpha(E) = e^{-i \arccos(E/\Delta)} = \frac{E}{\Delta} - i\sqrt{1 - \frac{E^2}{\Delta^2}} \, . \tag{20}$$

Andreev reflection transforms an electron mode into a hole mode, without change of mode index. The transformation is accompanied by a phase shift $-\arccos(E/\Delta)$ due to the penetration of the wave function into the superconductor.

We are now ready to relate the excitation spectrum of the Andreev billiard to the scattering matrix of the normal region. We restrict ourselves to the discrete spectrum (see [24] for the continuous spectrum). The condition $c_{\rm in} = S_{\rm A} S_{\rm N} c_{\rm in}$ for a bound state implies $\mathrm{Det}\,(1 - S_{\rm A} S_{\rm N}) = 0$. Using (18), (19), and the identity

$$\mathrm{Det} \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \mathrm{Det}\,(ad - aca^{-1}b) \tag{21}$$

one obtains the equation [24]

$$\mathrm{Det}\,\left[1 - \alpha(E)^2 S(E) S(-E)^*\right] = 0 \, . \tag{22}$$

The roots E_p of this determinantal equation constitute the discrete spectrum of the Andreev billiard.

5 Stroboscopic Model

Although the phase space of the Andreev billiard is four-dimensional, like for any billiard it can be reduced to two dimensions on a Poincaré surface of section [7, 8]. This amounts to a stroboscopic description of the classical dynamics, because the position and momentum are only recorded when the particle crosses the surface of section. Quantum mechanically, the stroboscopic evolution of the wave function is described by a compact unitary map rather than by a noncompact Hermitian operator [25, 26]. What one loses by the stroboscopic description is information on time scales below the time of flight across the billiard. What one gains is an enormous increase in computational efficiency.

A stroboscopic model of an Andreev billiard was constructed by Jacquod et al. [27], building on an existing model for open normal billiards called the open kicked rotator [28]. The Andreev kicked rotator possesses the same phenomenology as the Andreev billiard, but is much more tractable numerically.[3] In this subsection we discuss how it is formulated. Some results obtained by this numerical method will be compared in subsequent sections with results obtained by analytical means.

A compact unitary map is represented in quantum mechanics by the Floquet operator F, which gives the stroboscopic time evolution $u(p\tau_0) = F^p u(0)$ of an initial wave function $u(0)$. (We set the stroboscopic period $\tau_0 = 1$ in most equations.) The unitary $M \times M$ matrix F has eigenvalues $\exp(-i\varepsilon_m)$, with the quasi-energies $\varepsilon_m \in (-\pi, \pi)$ (measured in units of $\hbar/\tau_0$). This describes the electron excitations above the Fermi level. Hole excitations below the Fermi level have Floquet operator F^* and wave function $v(p) = (F^*)^p v(0)$. The mean level spacing of electrons and holes separately is $\delta = 2\pi/M$.

An electron is converted into a hole by Andreev reflection at the NS interface, with phase shift $-i$ for $\varepsilon \ll \tau_0\Delta/\hbar$ [cf. (20)]. In the stroboscopic description one assumes that Andreev reflection occurs only at times which are multiples of τ_0. The $N \times M$ matrix P projects onto the NS interface. Its elements are $P_{nm} = 1$ if $m = n \in \{n_1, n_2, \ldots n_N\}$ and $P_{nm} = 0$ otherwise. The dwell time of a quasiparticle excitation in the normal metal is $\tau_{\rm dwell} = M/N$, equal to the mean time between Andreev reflections.

Putting all this together one constructs the quantum Andreev map from the matrix product

$$\mathcal{F} = \mathcal{P}\begin{pmatrix} F & 0 \\ 0 & F^* \end{pmatrix}, \quad \mathcal{P} = \begin{pmatrix} 1 - P^{\rm T}P & -iP^{\rm T}P \\ -iP^{\rm T}P & 1 - P^{\rm T}P \end{pmatrix}. \tag{23}$$

(The superscript "T" indicates the transpose of a matrix.) The particle-hole wave function $\Psi = (u, v)$ evolves in time as $\Psi(p) = \mathcal{F}^p\Psi(0)$. The Floquet operator can be symmetrized (without changing its eigenvalues) by the unitary transformation $\mathcal{F} \to \mathcal{P}^{-1/2}\mathcal{F}\mathcal{P}^{1/2}$, with

$$\mathcal{P}^{1/2} = \begin{pmatrix} 1 - (1 - \frac{1}{2}\sqrt{2})P^{\rm T}P & -i\frac{1}{2}\sqrt{2}P^{\rm T}P \\ -i\frac{1}{2}\sqrt{2}P^{\rm T}P & 1 - (1 - \frac{1}{2}\sqrt{2})P^{\rm T}P \end{pmatrix}. \tag{24}$$

The quantization condition $\det(\mathcal{F} - e^{-i\varepsilon}) = 0$ can be written equivalently as [27]

$$\mathrm{Det}\,[1 + S(\varepsilon)S(-\varepsilon)^*] = 0\,, \tag{25}$$

in terms of the $N \times N$ scattering matrix [28, 29]

$$S(\varepsilon) = P[e^{-i\varepsilon} - F(1 - P^{\rm T}P)]^{-1}FP^{\rm T}\,. \tag{26}$$

[3] The largest simulation to date of a two-dimensional Andreev billiard has $N = 30$, while for the Andreev kicked rotator $N = 10^5$ is within reach, cf. Fig. 24.

Equation (25) for the Andreev map has the same form as (22) for the Andreev billiard (with $\alpha \to -i$). In particular, both equations have roots that lie symmetrically around zero.

A specific realization of the Andreev map is the Andreev kicked rotator. (See [30] for a different realization, based on the kicked Harper model.) The normal kicked rotator has Floquet operator [31]

$$F = \exp\left(i\frac{\hbar\tau_0}{4I_0}\frac{\partial^2}{\partial\theta^2}\right)\exp\left(-i\frac{KI_0}{\hbar\tau_0}\cos\theta\right) \times \exp\left(i\frac{\hbar\tau_0}{4I_0}\frac{\partial^2}{\partial\theta^2}\right) . \tag{27}$$

It describes a particle that moves freely along the unit circle $(\cos\theta, \sin\theta)$ with moment of inertia I_0 for half a period τ_0, is then kicked with a strength $K\cos\theta$, and proceeds freely for another half period. Upon increasing K the classical dynamics varies from fully integrable ($K = 0$) to fully chaotic [$K \gtrsim 7$, with Lyapunov exponent $\alpha \approx \ln(K/2)$]. For $K < 7$ stable and unstable motion coexist (mixed phase space). If needed, a magnetic field can be introduced into the model as described in [32].

The transition from classical to quantum behavior is governed by the effective Planck constant $h_{\rm eff} \equiv \hbar\tau_0/2\pi I_0$. For $1/h_{\rm eff} \equiv M$ an even integer, F can be represented by an $M \times M$ unitary symmetric matrix. The angular coordinate and momentum eigenvalues are $\theta_m = 2\pi m/M$ and $p_m = \hbar m$, with $m = 1, 2, \ldots M$, so phase space has the topology of a torus. The NS interface is an annulus around the torus, either in the θ-direction or in the p-direction. (The two configurations give equivalent results.) The construction (23) produces a $2M \times 2M$ Floquet operator $\mathcal{F}$, which can be diagonalized efficiently in $\mathcal{O}(M^2 \ln M)$ operations [rather than $\mathcal{O}(M^3)$] by combining the Lanczos technique with the fast-Fourier-transform algorithm [33].

6 Random-Matrix Theory

An ensemble of isolated chaotic billiards, constructed by varying the shape at constant area, corresponds to an ensemble of Hamiltonians H with a particular distribution function $P(H)$. It is convenient to think of the Hamiltonian as a random $M \times M$ Hermitian matrix, eventually sending M to infinity. The basic postulate of random-matrix theory (RMT) [9] is that the distribution is invariant under the unitary transformation $H \to UHU^\dagger$, with U an arbitrary unitary matrix. This implies a distribution of the form

$$P(H) \propto \exp[-\mathrm{Tr}\, V(H)] . \tag{28}$$

If $V(H) \propto H^2$, the ensemble is called Gaussian. This choice simplifies some of the calculations but is not essential, because the spectral correlations become

largely independent of V in the limit $M \to \infty$. More generally, the ensemble of the form (28) is called the Wigner-Dyson ensemble, after the founding fathers of RMT.

By computing the Jacobian from the space of matrix elements to the space of eigenvalues $E_n (n = 1, 2, \ldots M)$, one obtains the eigenvalue probability distribution [9]

$$P(\{E_n\}) \propto \prod_{i<j} |E_i - E_j|^\beta \prod_k e^{-V(E_k)}. \tag{29}$$

The symmetry index β counts the number of degrees of freedom in the matrix elements. These are real ($\beta = 1$) in the presence of time-reversal symmetry or complex ($\beta = 2$) in its absence. (A third possibility, $\beta = 4$, applies to time-reversally symmetric systems with strong spin-orbit scattering, which we will not consider here.) Since the unitary transformation $H \to UHU^\dagger$ requires an orthogonal U to keep a real Hamiltonian, one speaks of the Gaussian orthogonal ensemble (GOE) when $\beta = 1$. The name Gaussian unitary ensemble (GUE) refers to $\beta = 2$.

There is overwhelming numerical evidence that chaotic billiards are well described by the Wigner-Dyson ensemble [8]. (This is known as the Bohigas-Giannoni-Schmit conjecture [34].) A complete theoretical justification is still lacking, but much progress has been made in that direction [35]. In this section we will take (28) for the ensemble of isolated billiards as our starting point and deduce what properties it implies for the ensemble of Andreev billiards.

The isolated billiard becomes an Andreev billiard when it is connected by a point contact to a superconductor, cf. Fig. 5. In the isolated billiard RMT breaks down on energy scales greater than $\hbar/\tau_{\rm erg}$, with the ergodic time $\tau_{\rm erg} \simeq A^{1/2}/v_F$ set by the time of flight across the billiard (of area A, at Fermi velocity v_F). On larger energy scales, hence on shorter time scales, non-chaotic dynamics appears which is beyond RMT. The superconductor affects the billiard in an energy range around the Fermi level that is set by

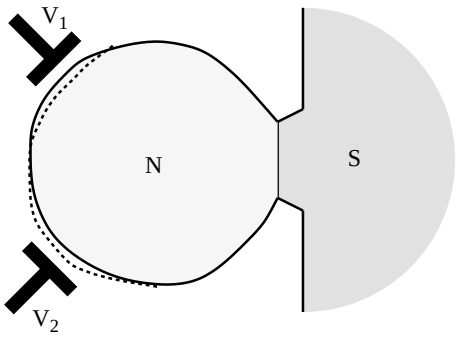

Fig. 5. A quantum dot (N) connected to a superconductor (S). The voltages on the gates V_1 and V_2 change the shape of the dot. Different values of the applied voltages create different samples within the same ensemble. From [36]

the Thouless energy $E_T \simeq \hbar/\tau_{\rm dwell}$. (We assume that E_T is less than the gap Δ in the bulk superconductor.) In this context the dwell time $\tau_{\rm dwell}$ is the mean time between Andreev reflections (being the life time of an electron or hole quasiparticle). The condition $\tau_{\rm erg} \ll \tau_{\rm dwell}$ of weak coupling is therefore sufficient to be able to apply RMT to the entire relevant energy range.

6.1 Effective Hamiltonian

The excitation energies E_p of the Andreev billiard in the discrete part of the spectrum are the solutions of the determinantal (22), given in terms of the scattering matrix $S(E)$ in the normal state (i.e. when the superconductor is replaced by a normal metal). This equation can alternatively be written in terms of the Hamiltonian H of the isolated billiard and the $M \times N$ coupling matrix W that describes the N-mode point contact. The relation between S and H, W is [3, 10]

$$S(E) = 1 - 2\pi i W^T \Big(E - H + i\pi W W^T \Big)^{-1} W \; . \tag{30}$$

The $N \times N$ matrix $W^T W$ has eigenvalues w_n given by

$$w_n = \frac{M\delta}{\pi^2 \Gamma_n} \left(2 - \Gamma_n - 2\sqrt{1-\Gamma_n} \right) \; , \tag{31}$$

where δ is the mean level spacing in the isolated billiard and $\Gamma_n \in [0,1]$ is the transmission probability of mode $n = 1, 2, \ldots N$ in the point contact. For a ballistic contact, $\Gamma_n = 1$, while $\Gamma_n \ll 1$ for a tunneling contact. Both the number of modes N and the level spacing δ refer to a single spin direction.

Substituting (30) into (22), one arrives at an alternative determinantal equation for the discrete spectrum [37]:

$$\mathrm{Det}\,[E - \mathcal{H} + \mathcal{W}(E)] = 0 \; , \tag{32}$$

$$\mathcal{H} = \begin{pmatrix} H & 0 \\ 0 & -H^* \end{pmatrix} \; , \tag{33}$$

$$\mathcal{W}(E) = \frac{\pi}{\sqrt{\Delta^2 - E^2}} \begin{pmatrix} EWW^T & \Delta WW^T \\ \Delta WW^T & EWW^T \end{pmatrix} \; . \tag{34}$$

The density of states follows from

$$\rho(E) = -\frac{1}{\pi} \mathrm{Im}\,\mathrm{Tr}\,(1 + d\mathcal{W}/dE)(E + i0^+ - \mathcal{H} + \mathcal{W})^{-1} \; . \tag{35}$$

In the relevant energy range $E \lesssim E_T \ll \Delta$ the matrix $\mathcal{W}(E)$ becomes energy independent. The excitation energies can then be obtained as the eigenvalues of the effective Hamiltonian [38]

$$\mathcal{H}_{\rm eff} = \begin{pmatrix} H & -\pi WW^T \\ -\pi WW^T & -H^* \end{pmatrix} . \tag{36}$$

The effective Hamiltonian $\mathcal{H}_{\rm eff}$ should not be confused with the Bogoliubov-de Gennes Hamiltonian $\mathcal{H}_{\rm BG}$, which contains the superconducting order parameter in the off-diagonal blocks [cf. (2)]. The Hamiltonian $\mathcal{H}_{\rm BG}$ determines the entire excitation spectrum (both the discrete part below Δ and the continuous part above Δ), while the effective Hamiltonian $\mathcal{H}_{\rm eff}$ determines only the low-lying excitations $E_p \ll \Delta$.

The Hermitian matrix $\mathcal{H}_{\rm eff}$ (like $\mathcal{H}_{\rm BG}$) is antisymmetric under the combined operation of charge conjugation ($\mathcal{C}$) and time inversion ($\mathcal{T}$) [39]:

$$\mathcal{H}_{\rm eff} = -\sigma_y \mathcal{H}_{\rm eff}^T \sigma_y, \quad \sigma_y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} . \tag{37}$$

(An $M \times M$ unit matrix in each of the four blocks of σ_y is implicit.) The $\mathcal{CT}$-antisymmetry ensures that the eigenvalues lie symmetrically around $E = 0$. Only the positive eigenvalues are retained in the excitation spectrum, but the presence of the negative eigenvalues is felt as a level repulsion near $E = 0$.

6.2 Excitation Gap

In zero magnetic field the suppression of the density of states $\rho(E)$ around $E = 0$ extends over an energy range E_T that may contain many level spacings δ of the isolated billiard. The ratio $g \simeq E_T/\delta$ is the conductance of the point contact in units of the conductance quantum e^2/h. For $g \gg 1$ the excitation gap $E_{\rm gap} \simeq g\delta$ is a mesoscopic quantity, because it is intermediate between the microscopic energy scale δ and the macroscopic energy scale Δ. One can use perturbation theory in the small parameter $1/g$ to calculate $\rho(E)$. The analysis presented here follows the RMT of Melsen et al. [22]. An alternative derivation [40], using the disorder-averaged Green function, is discussed in the next sub-section.

In the presence of time-reversal symmetry the Hamiltonian H of the isolated billiard is a real symmetric matrix. The appropriate RMT ensemble is the GOE, with distribution [9]

$$P(H) \propto \exp\left(-\frac{\pi^2}{4M\delta^2} \mathrm{Tr}\, H^2 \right) . \tag{38}$$

The ensemble average $\langle \cdots \rangle$ is an average over H in the GOE at fixed coupling matrix W. Because of the block structure of $\mathcal{H}_{\rm eff}$, the ensemble averaged Green function $\mathcal{G}(E) = \langle (E - \mathcal{H}_{\rm eff})^{-1} \rangle$ consists of four $M \times M$ blocks $\mathcal{G}_{11}$, $\mathcal{G}_{12}$, $\mathcal{G}_{21}$, $\mathcal{G}_{22}$. By taking the trace of each block separately, one arrives at a 2×2 matrix Green function

$$G = \begin{pmatrix} G_{11} & G_{12} \\ G_{21} & G_{22} \end{pmatrix} = \frac{\delta}{\pi} \begin{pmatrix} \mathrm{Tr}\, \mathcal{G}_{11} & \mathrm{Tr}\, \mathcal{G}_{12} \\ \mathrm{Tr}\, \mathcal{G}_{21} & \mathrm{Tr}\, \mathcal{G}_{22} \end{pmatrix} . \tag{39}$$

(The factor δ/π is inserted for later convenience.)

The average over the distribution (38) can be done diagrammatically [41, 42]. To leading order in $1/M$ and for $E \gg \delta$ only simple (planar) diagrams need to be considered. Resummation of these diagrams leads to the selfconsistency equation [22, 37]

$$\mathcal{G} = [E + \mathcal{W} - (M\delta/\pi)\sigma_z G \sigma_z]^{-1}, \quad \sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} . \tag{40}$$

This is a matrix-generalization of Pastur's equation in the RMT of normal systems [43].

The matrices in (40) have four $M \times M$ blocks. By taking the trace of each block one obtains an equation for a 4×4 matrix,

$$G = \frac{1}{M} \sum_{m=1}^{M} \begin{pmatrix} \pi E/M\delta - G_{11} & \tilde{w}_m + G_{12} \\ \tilde{w}_m + G_{21} & \pi E/M\delta - G_{22} \end{pmatrix}^{-1} , \tag{41}$$

$$\tilde{w}_m = \begin{cases} \pi^2 w_m/M\delta & \text{if } \ m = 1, 2, \ldots N \ , \\ 0 & \text{if } \ m = N+1, \ldots M \ . \end{cases} \tag{42}$$

Since $G_{22} = G_{11}$ and $G_{21} = G_{12}$ there are two unknown functions to determine. For $M \gg N$ these satisfy

$$G_{12}^2 = 1 + G_{11}^2 \ , \tag{43a}$$

$$\frac{2\pi E}{\delta} G_{12} = G_{11} \sum_{n=1}^{N} (-G_{12} + 1 - 2/\Gamma_n)^{-1} \ , \tag{43b}$$

where we have used the relation (31) between the parameters w_n and the transmission probabilities Γ_n. Equation (43) has multiple solutions. The physical solution satisfies $\lim_{E\to\infty} \langle \rho(E) \rangle = 2/\delta$, when substituted into

$$\langle \rho(E) \rangle = -(2/\delta) \, \mathrm{Im} \, G_{11}(E) \ . \tag{44}$$

In Fig. 6 we plot the density of states in the mode-independent case $\Gamma_n \equiv \Gamma$, for several vaues of Γ. It vanishes as a square root near the excitation gap. The value of E_{gap} can be determined directly by solving (43) jointly with $dE/dG_{11} = 0$. The result is

$$\frac{k^6 - k^4}{(1-k)^6} x^6 - \frac{3k^4 - 20k^2 + 16}{(1-k)^4} x^4 + \frac{3k^2 + 8}{(1-k)^2} x^2 = 1 \ ,$$
$$x = E_{\mathrm{gap}}/E_T \ , \ k = 1 - 2/\Gamma \ , \ E_T = N\Gamma\delta/4\pi \ . \tag{45}$$

For later use we parametrize the square-root dependence near the gap as

$$\langle \rho(E) \rangle \to \frac{1}{\pi} \sqrt{\frac{E - E_{\mathrm{gap}}}{\Delta_{\mathrm{gap}}^3}} \ , \quad E \to E_{\mathrm{gap}} \ . \tag{46}$$

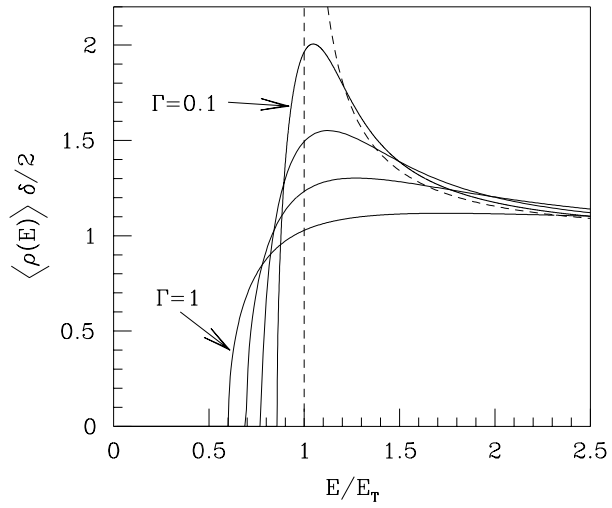

Fig. 6. Ensemble averaged density of states of a chaotic billiard coupled by a point contact to a superconductor, for several values of the transmission probability through the point contact. The energy is in units of the Thouless energy $E_T = N\Gamma\delta/4\pi$. The *solid* curves are computed from (43) and (44), for mode-independent transmission probabilities $\Gamma = 1$, 0.5, 0.25, 0.1. The *dashed* curve is the asymptotic result (51) for $\Gamma \ll 1$ Adapted from [22] (The definition of δ used in that paper differs from the one used here by a factor of two)

When $E \gg E_{\rm gap}$ the density of states approaches the value $2/\delta$ from above, twice the value in the isolated billiard. The doubling of the density of states occurs because electron and hole excitations are combined in the excitation spectrum of the Andreev billiard, while in an isolated billiard electron and hole excitations are considered separately.

A rather simple closed-form expression for $\langle\rho(E)\rangle$ exists in two limiting cases [22]. In the case $\Gamma = 1$ of a ballistic point contact one has

$$\langle\rho(E)\rangle = \frac{E_T\sqrt{3}}{3E\delta}\Big[Q_+(E/E_T) - Q_-(E/E_T)\Big]\,, \tag{47}$$

$$Q_\pm(x) = \Big[8 - 36x^2 \pm 3x\sqrt{3x^4 + 132x^2 - 48}\Big]^{1/3}\,, \tag{48}$$

$$E > E_{\rm gap} = 2\gamma^{5/2}E_T = 0.60\,E_T = \frac{0.30\,\hbar}{\tau_{\rm dwell}} = 0.048\,N\delta\,, \tag{49}$$

where $\gamma = \frac{1}{2}(\sqrt{5}-1)$ is the golden number. In this case the parameter $\Delta_{\rm gap}$ in (46) is given by

$$\Delta_{\rm gap} = [(5-2\sqrt{5})\delta^2 E_{\rm gap}/8\pi^2]^{1/3} = 0.068\,N^{1/3}\delta\,. \tag{50}$$

In the opposite tunneling limit $\Gamma \ll 1$ one finds

$$\langle \rho(E) \rangle = \frac{2E}{\delta}(E^2 - E_T^2)^{-1/2},\ E > E_{\rm gap} = E_T\ . \tag{51}$$

In this limit the density of states of the Andreev billiard has the same form as in the BCS theory for a bulk superconductor [44], with a reduced value of the gap ("minigap"). The inverse square-root singularity at the gap is cut off for any finite Γ, cf. Fig. 6.

6.3 Effect of Impurity Scattering

Impurity scattering in a chaotic Andreev billiard reduces the magnitude of the excitation gap by increasing the mean time $\tau_{\rm dwell}$ between Andreev reflections. This effect was calculated by Vavilov and Larkin [40] using the method of impurity-averaged Green functions [21]. The minigap in a disordered quantum dot is qualitatively similar to that in a disordered NS junction, cf. Sect. 3. The main parameter is the ratio of the mean free path l and the width of the contact W. (We assume that there is no barrier in the point contact, otherwise the tunnel probability Γ would enter as well.)

For $l \gg W$ the mean dwell time saturates at the ballistic value

$$\tau_{\rm dwell} = \frac{2\pi\hbar}{N\delta} = \frac{\pi A}{v_F W}\ ,\ \text{ if }\ l \gg W\ . \tag{52}$$

In the opposite limite $l \ll W$ the mean dwell time is determined by the two-dimensional diffusion equation. Up to a geometry-dependent coefficient c of order unity, one has

$$\tau_{\rm dwell} = c\frac{A}{v_F l}\ln(A/W^2)\ ,\ \text{ if }\ l \ll W\ . \tag{53}$$

The density of states in the two limits is shown in Fig. 7. There is little difference, once the energy is scaled by $\tau_{\rm dwell}$. For $l \gg W$ the excitation gap is given by the RMT result $E_{\rm gap} = 0.300\,\hbar/\tau_{\rm dwell}$, cf. (49). For $l \ll W$ Vavilov and Larkin find $E_{\rm gap} = 0.331\,\hbar/\tau_{\rm dwell}$.

6.4 Magnetic Field Dependence

A magnetic field B, perpendicular to the billiard, breaks time-reversal symmetry, thereby suppressing the excitation gap. A perturbative treatment remains possible as long as $E_{\rm gap}(B)$ remains large compared to δ [45].

The appropriate RMT ensemble for the isolated billiard is described by the Pandey-Mehta distribution [9, 46]

$$P(H) \propto \exp\left(-\frac{\pi^2(1+b^2)}{4M\delta^2} \times \sum_{i,j=1}^{M} \left[(\mathrm{Re}\, H_{ij})^2 + b^{-2}(\mathrm{Im}\, H_{ij})^2 \right] \right)\ . \tag{54}$$

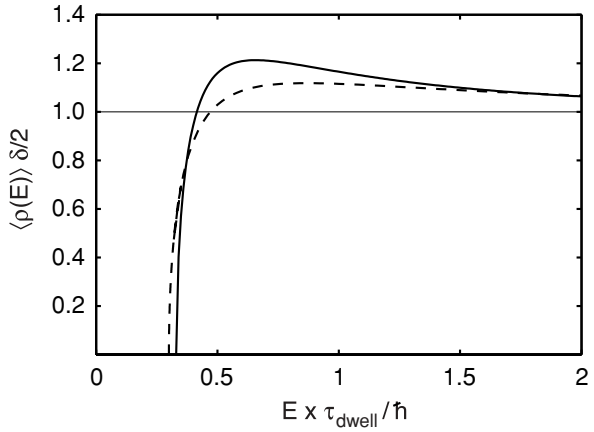

Fig. 7. The dashed curve is the $\Gamma = 1$ result of Fig. 6, corresponding to a quantum dot with weak impurity scattering (mean free path l much larger than the width W of the point contact). The *solid* curve is the corresponding result for strong impurity scattering ($l \ll W$). The line shape is almost the same, but the energy scale is different (given by (52) and (53), respectively). Adapted from [40]

The parameter $b \in [0, 1]$ measures the strength of the time-reversal symmetry breaking. The invariance of $P(H)$ under unitary transformations is broken if $b \neq 0, 1$. The relation between b and the magnetic flux Φ through the billiard is [3]

$$Mb^2 = c(\Phi e/h)^2 \frac{\hbar v_F}{\delta\sqrt{A}} , \tag{55}$$

with c a numerical coefficient that depends only on the shape of the billiard. Time-reversal symmetry is effectively broken when $Mb^2 \simeq g$, which occurs for $\Phi \simeq (h/e)\sqrt{\tau_{\rm erg}/\tau_{\rm dwell}} \ll h/e$. The effect of such weak magnetic fields on the bulk superconductor can be ignored.

The selfconsistency equation for the Green function is the same as (41), with one difference: On the right-hand-side the terms G_{12} and G_{21} are multiplied by the factor $(1 - b^2)/(1 + b^2)$. In the limit $M \to \infty$, $b \to 0$, Mb^2 finite, the first (43a) still holds, but the second (43b) is replaced by

$$\begin{aligned}(2\pi E/\delta - 4Mb^2 G_{11})G_{12} &= G_{11} \\ &\times \sum_{n=1}^{N} (-G_{12} + 1 - 2/\Gamma_n)^{-1} .\end{aligned} \tag{56}$$

The resulting magnetic field dependence of the average density of states is plotted in Fig. 8, for the case $\Gamma_n \equiv 1$ of a ballistic point contact. The gap closes when $Mb^2 = N\Gamma/8$. The corresponding critical flux Φ_c follows from (55).

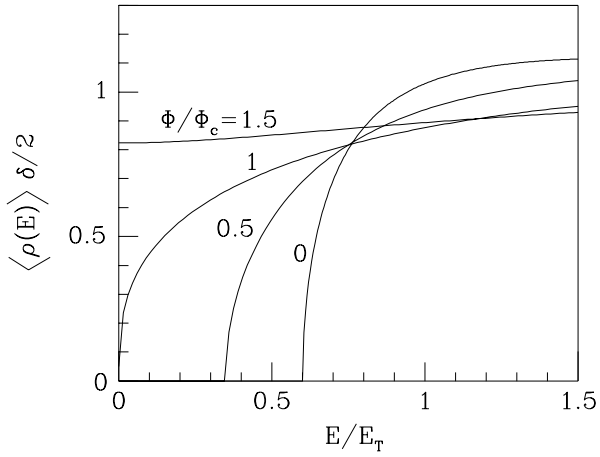

Fig. 8. Magnetic field dependence of the density of states for the case of a ballistic point contact ($\Gamma_n \equiv 1$), computed from (43a), (44), and (56). The microscopic gap of order δ which persists when $\Phi > \Phi_c$ is not resolved in this calculation. Adapted from [45]

6.5 Broken Time-Reversal Symmetry

A microscopic suppression of the density of states around $E = 0$, on an energy scale of the order of the level spacing, persists even if time-reversal symmetry is fully broken. The suppression is a consequence of the level repulsion between the lowest excitation energy E_1 and its mirror image $-E_1$, which itself follows from the $\mathcal{CT}$-antisymmetry (37) of the Hamiltonian. Because of this mirror symmetry, the effective Hamiltonian $\mathcal{H}_{\rm eff}$ of the Andreev billiard can be factorized as

$$\mathcal{H}_{\rm eff} = U \begin{pmatrix} \mathcal{E} & 0 \\ 0 & -\mathcal{E} \end{pmatrix} U^\dagger \,, \tag{57}$$

with U a $2M \times 2M$ unitary matrix and $\mathcal{E} = \mathrm{diag}(E_1, E_2, \ldots E_M)$ a diagonal matrix containing the positive excitation energies.

Altland and Zirnbauer [39] have surmised that an ensemble of Andreev billiards in a strong magnetic field would have a distribution of Hamiltonians of the Wigner-Dyson form (28), constrained by (57). This constraint changes the Jacobian from the space of matrix elements to the space of eigenvalues, so that the eigenvalue probability distribution is changed from the form (29) (with $\beta = 2$) into

$$P(\{E_n\}) \propto \prod_{i<j} \left(E_i^2 - E_j^2\right)^2 \prod_k E_k^2 e^{-V(E_k)-V(-E_k)} \,. \tag{58}$$

The distribution (58) with $V(E) \propto E^2$ is related to the Laguerre unitary ensemble (LUE) of RMT [9] by a change of variables. The average density of states vanishes quadratically near zero energy [47],

$$\langle\rho(E)\rangle = \frac{2}{\delta}\left(1 - \frac{\sin(4\pi E/\delta)}{4\pi E/\delta}\right) . \tag{59}$$

All of this is qualitatively different from the "folded GUE" that one would obtain by simply combining two independent GUE's of electrons and holes [48].

A derivation of Altland and Zirnbauer's surmise has been given by Frahm et al. [38], who showed that the LUE for the effective Hamiltonian $\mathcal{H}_{\rm eff}$ of the Andreev billiard follows from the GUE for the Hamiltonian H of the isolated billiard, provided that the coupling to the superconductor is sufficiently strong. To compute the spectral statistics on the scale of the level spacing, a non-perturbative technique is needed. This is provided by the supersymmetric method [49]. (See [50] for an alternative approach using quantum graphs.)

The resulting average density of states is [38]

$$\langle\rho(E)\rangle = \frac{2}{\delta} - \frac{\sin(2\pi E/\delta)}{\pi E}\int_0^\infty ds\; e^{-s} \times \cos\left(\frac{2\pi E}{\delta}\sqrt{1+\frac{4s}{g_A}}\right) , \tag{60}$$

$$g_A = \sum_{n=1}^{N}\frac{2\Gamma_n^2}{(2-\Gamma_n)^2} . \tag{61}$$

The parameter g_A is the Andreev conductance of the point contact that couples the billiard to the superconductor [51]. The Andreev conductance can be much smaller than the normal-state conductance $g = \sum_{n=1}^{N}\Gamma_n$. (Both conductances are in units of $2e^2/h$.) In the tunneling limit $\Gamma_n \equiv \Gamma \ll 1$ one has $g = N\Gamma$ while $g_A = \frac{1}{2}N\Gamma^2$.

Equation (9) describes the crossover from the GUE result $\rho(E) = 2/\delta$ for $g_{\rm A} \ll 1$ to the LUE result (59) for $g_{\rm A} \gg 1$. The opening of the gap as the coupling to the superconductor is increased is plotted in Fig. 9. The $\mathcal{CT}$-antisymmetry becomes effective at an energy E for $g_{\rm A} \gtrsim E/\delta$. For small energies $E \ll \delta\,\min(\sqrt{g_{\rm A}}, 1)$ the density of states vanishes quadratically, regardless of how weak the coupling is.

6.6 Mesoscopic Fluctuations of the Gap

The smallest excitation energy E_1 in the Andreev billiard fluctuates from one member of the ensemble to the other. Vavilov et al. [36] have surmised that the distribution of these fluctuations is identical upon rescaling to the known distribution [52] of the lowest eigenvalue in the Gaussian ensembles of RMT. This surmise was proven using the supersymmetry technique by Ostrovsky, Skvortsov, and Feigelman [53] and by Lamacraft and Simons [54]. Rescaling amounts to a change of variables from E_1 to $x = (E_1 - E_{\rm gap})/\Delta_{\rm gap}$, where $E_{\rm gap}$ and $\Delta_{\rm gap}$ parameterize the square-root dependence (46) of the mean density

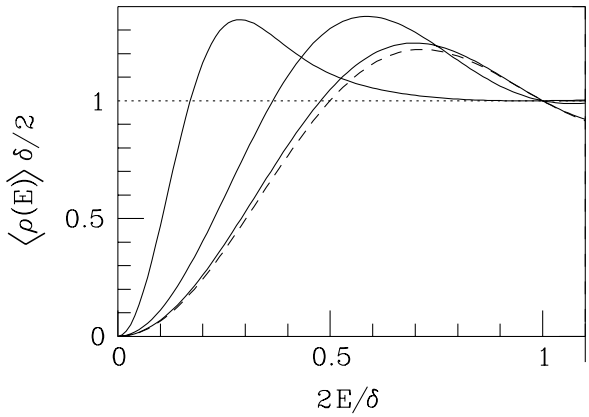

Fig. 9. Density of states of the Andreev billiard in a strong magnetic field for three different values of the Andreev conductance of the point contact: $g_A = 0.4,\ 4,\ 40$. The *solid* curves are calculated from (60). The *dashed* line is the LUE result (59), corresponding to the limit $g_A \to \infty$. The *dotted* line is the GUE limit $g_A \to 0$. Adapted from [38]

of states near the gap in perturbation theory. The gap fluctuations are a mesoscopic, rather than a microscopic effect, because the typical magnitude $\Delta_{\rm gap} \simeq E_{\rm gap}^{1/3}\delta^{2/3}$ of the fluctuations is $\gg \delta$ for $E_{\rm gap} \gg \delta$. Still, the fluctuations are small on the scale of the gap itself.

Following [36], in zero magnetic field the gap distribution is obtained by rescaling the GOE result of Tracy and Widom [52],

$$P(E_1) = \frac{d}{dE_1} F_1\left[(E_1 - E_{\rm gap})/\Delta_{\rm gap}\right] \ , \tag{62}$$

$$F_1(x) = \exp\left(-\tfrac{1}{2}\int_{-\infty}^{x} [q(x') + (x-x')q^2(x')]dx'\right) \ . \tag{63}$$

The function $q(x)$ is the solution of

$$q''(x) = -xq(x) + 2q^3(x) \ , \tag{64}$$

with asymptotic behavior $q(x) \to {\rm Ai}(-x)$ as $x \to -\infty$ [${\rm Ai}(x)$ being the Airy function]. For small x there is a tail of the form

$$P(x) \approx \frac{1}{4\sqrt{\pi}|x|^{1/4}} \exp\left(-\tfrac{2}{3}|x|^{3/2}\right), \quad x \ll -1 \ . \tag{65}$$

The distribution (62) is shown in Fig. 10 (solid curve). The mean and standard deviation are

$$\langle E_1\rangle = E_{\rm gap} + 1.21\,\Delta_{\rm gap}, \quad \langle (E_1 - \langle E_1\rangle)^2\rangle^{1/2} = 1.27\,\Delta_{\rm gap} \ . \tag{66}$$

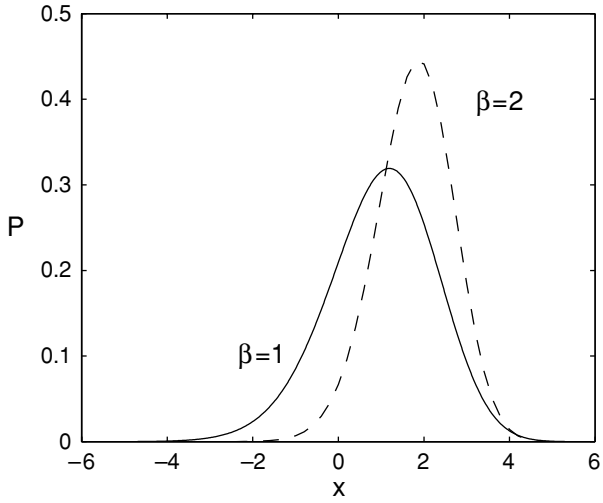

Fig. 10. Probability distribution of the rescaled excitation gap $x = (E_1 - E_{\rm gap})/\Delta_{\rm gap}$, in the presence [$\beta = 1$, (62)] and absence [$\beta = 2$, (67)] of time-reversal symmetry. Adapted from [36]

Because the mesoscopic fluctuations in the gap occur on a much smaller energy scale $\Delta_{\rm gap}$ than $E_{\rm gap}$, there exists a range of magnetic fields that break time-reversal symmetry of the gap fluctuations without significantly reducing $E_{\rm gap}$ [36]. In this field range, specified in Table 1, the distribution of the lowest excitation is given by the GUE result [52]

$$P(E_1) = \frac{d}{dE_1} F_2[(E_1 - E_{\rm gap})/\Delta_{\rm gap}] \,, \tag{67}$$

$$F_2(x) \exp\left(- \int_{-\infty}^{x} (x - x') q^2(x') dx' \right) \,. \tag{68}$$

This curve is shown dashed in Fig. 10. The tail for small x is now given by

Table 1. Characteristic energy and magnetic flux scales for the spectral statistics in the bulk and at the edge of the spectrum and for the size of the gap. The $\beta = 2$ distribution (67) applies to the flux range $(h/e)\tau_{\rm erg}^{1/2}\delta^{1/6}E_{\rm gap}^{1/3}/\hbar^{1/2} \ll \Phi \ll (h/e)\tau_{\rm erg}^{1/2}E_{\rm gap}^{1/2}/\hbar^{1/2}$

	Energy Scale	Flux Scale
Bulk statistics	δ	$(h/e)\tau_{\rm erg}^{1/2}\delta^{1/2}/\hbar^{1/2}$
Edge statistics	$E_{\rm gap}^{1/3}\delta^{2/3}$	$(h/e)\tau_{\rm erg}^{1/2}\delta^{1/6}E_{\rm gap}^{1/3}/\hbar^{1/2}$
Gap size	$E_{\rm gap}$	$(h/e)\tau_{\rm erg}^{1/2}E_{\rm gap}^{1/2}/\hbar^{1/2}$

$$P(x) \approx \frac{1}{8\pi|x|} \exp\left(-\tfrac{4}{3}|x|^{3/2}\right), \quad x \ll -1 \,. \tag{69}$$

The mesoscopic gap fluctuations induce a tail in the ensemble averaged density of states $\langle\rho(E)\rangle$ for $E < E_{\rm gap}$. In the same rescaled variable x the tail is given by [36]

$$\langle\rho(x)\rangle = -x\mathrm{Ai}^2(x) + [\mathrm{Ai}'(x)]^2 + \tfrac{1}{2}\delta_{\beta,1}\mathrm{Ai}(x)\left[1 - \int_x^\infty \mathrm{Ai}(y)dy\right] \,. \tag{70}$$

Asymptotically, $\langle\rho(x)\rangle \propto \exp(-\frac{2}{3}\beta|x|^{3/2})$ for $x \ll -1$. The tail in $\langle\rho(E)\rangle$ is the same as the tail in $P(E)$, as it should be, since both tails are due to the lowest eigenvalue. In Fig. 11 we compare these two functions in zero magnetic field, together with the square-root density of states from perturbation theory.

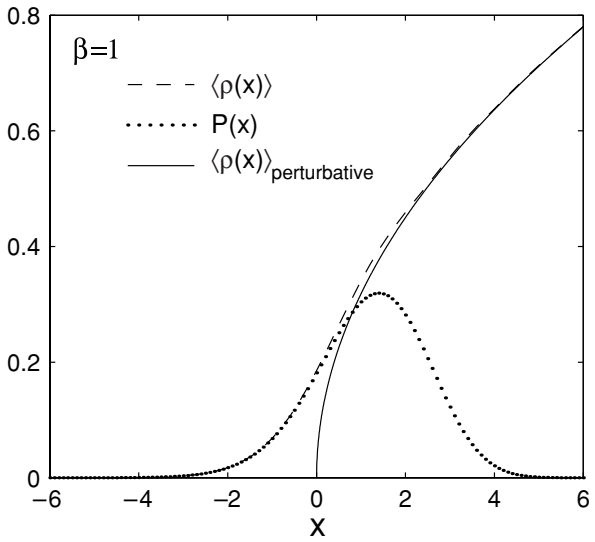

Fig. 11. Ensemble averaged density of states $\langle\rho\rangle$ together with the probability distribution P of the excitation gap, as a function of the rescaled energy $x = (E - E_{\rm gap})/\Delta_{\rm gap}$. The *dotted* and *dashed* curves are the universal results (62) and (70) of RMT in the presence of time-reversal symmetry ($\beta = 1$). The *solid* curve is the mean density of states (46) in perturbation theory. Adapted from [36]

A numerical simulation of the stroboscopic model of Sect. 5 provides a test of these predictions [27]. Results are shown in Fig. 12, for the case $\beta = 1$ and deep in the chaotic regime (kicking strength $K \gg 1$). The agreement with RMT is very good – without any adjustable parameters.

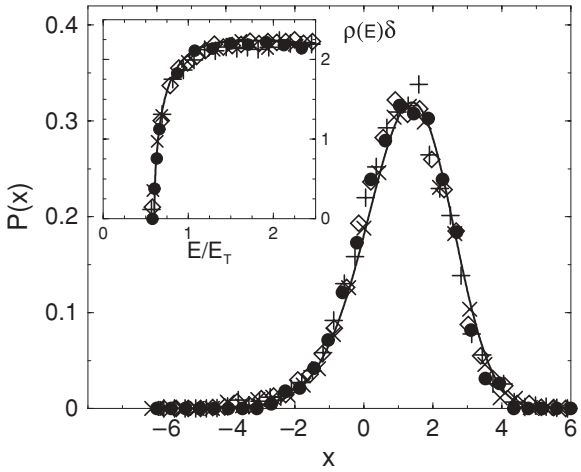

Fig. 12. Main plot: Gap distribution for the Andreev kicked rotator with parameters $M = 2\pi/\delta = 8192$, kicking strength $K = 45$, and $M/N = \tau_{\rm dwell} = 10$ ($\diamond$), 20 ($\bullet$), 40 ($+$), and 50 ($\times$). There is no magnetic field. The *solid* line is the RMT prediction (62). Inset: Average density of states for the same system. The *solid* line is the RMT prediction (49). (Deviations from perturbation theory are not visible on the scale of the inset.) Adapted from [27]

6.7 Coulomb Blockade

Coulomb interactions between electron and hole quasiparticles break the charge-conjugation invariance (37) of the Hamiltonian. Since Andreev reflection changes the charge on the billiard by $2e$, this scattering process becomes energetically unfavorable if the charging energy E_C exceeds the superconducting condensation energy (Josephson energy) E_J. For $E_C \gtrsim E_J$ one obtains the Coulomb blockade of the proximity effect studied by Ostrovsky, Skvortsov, and Feigelman [55].

The charging energy $E_C = e^2/2C$ is determined by the capacitance C of the billiard. The Josephson energy is determined by the change in free energy of the billiard resulting from the coupling to the superconductor,

$$E_J = -\int_0^\infty [\rho(E) - 2/\delta]\, E dE \; . \tag{71}$$

The discrete spectrum $E < E_{\rm gap}$ contributes an amount of order $E_{\rm gap}^2/\delta$ to E_J. In the continuous spectrum $E > E_{\rm gap}$ the density of states $\rho(E)$, calculated by RMT, decays $\propto 1/E^2$ to its asymptotic value $2/\delta$. This leads to a logarithmic divergence of the Josephson energy [37, 56], with a cutoff set by $\min(\Delta, \hbar/\tau_{\rm erg})$:

$$E_J = \frac{E_{\rm gap}^2}{\delta} \ln\left(\frac{\min(\Delta, \hbar/\tau_{\rm erg})}{E_{\rm gap}}\right) \; . \tag{72}$$

The suppression of the excitation gap with increasing E_C is plotted in Fig. 13, for the case $\Gamma \ll 1$, $\Delta \ll \hbar/\tau_{\rm erg}$ [55]. The initial decay is a square root,

$$1 - \Delta_{\rm eff}/E_{\rm gap} = \tfrac{1}{2}\left(\frac{E_C\delta}{E_{\rm gap}^2 \ln(2\Delta/E_{\rm gap})}\right)^{1/2} \ll 1\,, \tag{73}$$

and the final decay is exponential,

$$\Delta_{\rm eff}/\Delta = 2\exp\left(-2E_C\delta/E_{\rm gap}^2\right) \ll 1\,. \tag{74}$$

Here $\Delta_{\rm eff}$ refers to the gap in the presence of Coulomb interactions and $E_{\rm gap} = N\Gamma\delta/4\pi$ is the noninteracting value (51).

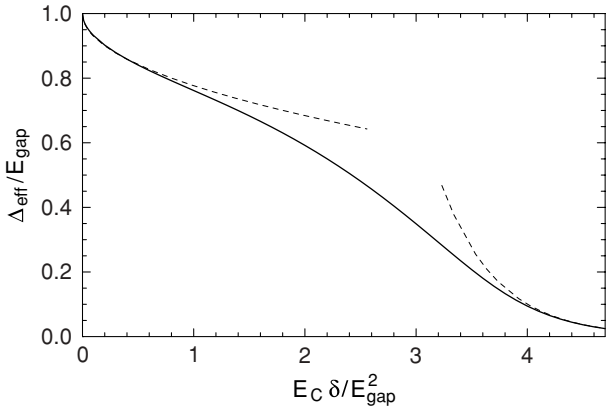

Fig. 13. Suppression due to Coulomb interactions of the gap $\Delta_{\rm eff}$ in the density of states of an Andreev billiard coupled by a tunnel junction to a superconductor, relative to the noninteracting gap $E_{\rm gap} = N\Gamma\delta/4\pi$ (with $\Gamma \ll 1 \ll N\Gamma$). The plot is for the case $\Delta = e^5 E_{\rm gap} \ll \hbar/\tau_{\rm erg}$. The *dashed* lines are the asymptotes (73) and (74). Adapted from [55]

The gap $\Delta_{\rm eff}$ governs the thermodynamic properties of the Andreev billiard, most importantly the critical current. It is not, however, the relevant energy scale for transport properties. Injection of charge into the billiard via a separate tunnel contact measures the tunneling density of states $\rho_{\rm tunnel}$, which differs in the presence of Coulomb interactions from the thermodynamic density of states ρ considered so far. The gap $\Delta_{\rm tunnel}$ in $\rho_{\rm tunnel}$ crosses over from the proximity gap $E_{\rm gap}$ when $E_C \ll E_J$ to the Coulomb gap E_C when $E_C \gg E_J$, see Fig. 14. The single peak in $\rho_{\rm tunnel}$ at $\Delta_{\rm tunnel}$ splits into two peaks when E_C and E_J are of comparable magnitude [55]. This peak splitting happens because two states of charge $+e$ and $-e$ having the same charging energy are mixed by Andreev reflection into symmetric and antisymmetric linear combinations with a slightly different energy.

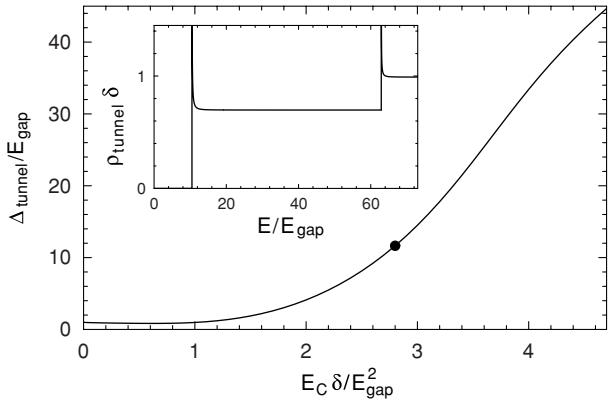

Fig. 14. Main plot: gap $\Delta_{\rm tunnel}$ in the tunneling density of states as a function of the charging energy (for $\Delta = e^5 E_{\rm gap}$ and $N\Gamma = 40\,\pi$). The initial decay (barely visible on the scale of the plot) follows (73) and crosses over to an increase ($\Delta_{\rm tunnel} \to E_C$). Inset: tunneling density of states at $E_C\delta/E_{\rm gap}^2 = 2.8$ (corresponding to the dot in the main plot). Adapted from [55]

7 Quasiclassical Theory

It was noticed by Kosztin, Maslov, and Goldbart [5] that the classical dynamics at the Fermi energy in an Andreev billiard is *integrable* – even if the dynamics in the isolated billiard is chaotic. Andreev reflection suppresses chaotic dynamics because it introduces a periodicity into the orbits: The trajectory of an electron is retraced by the Andreev reflected hole. At the Fermi energy the hole is precisely the time reverse of the electron, so that the motion is strictly periodic. For finite excitation energy or in a non-zero magnetic field the electron and the hole follow slightly different trajectories, so the orbit does not quite close and drifts around in phase space [5, 57, 58, 59, 60].

The near-periodicity of the orbits implies the existence of an adiabatic invariant. Quantization of this invariant leads to the quasiclassical theory of Silvestrov et al. [61].

7.1 Adiabatic Quantization

Figures 15 and 16 illustrate the nearly periodic motion in a particular Andreev billiard. Figure 15 shows a trajectory in real space while Fig. 16 is a section of phase space at the interface with the superconductor ($y = 0$). The tangential component p_x of the electron momentum is plotted as a function of the coordinate x along the interface. Each point in this Poincaré map corresponds to one collision of an electron with the interface. (The collisions of holes are not plotted.) The electron is retroreflected as a hole with the same p_x. At the Fermi level ($E = 0$) the component p_y is also the same, and so the hole retraces the path of the electron (the hole velocity being opposite to

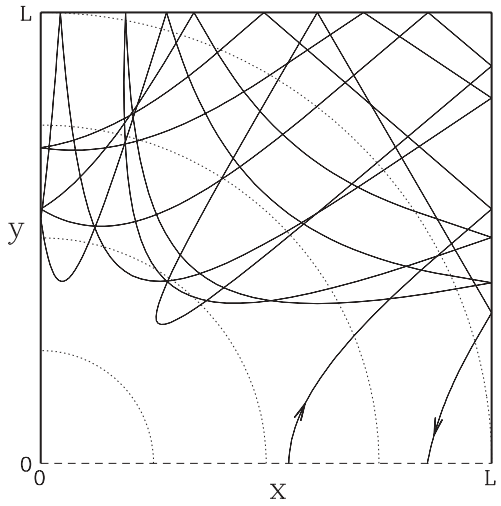

Fig. 15. Classical trajectory in an Andreev billiard. Particles in a two-dimensional electron gas are deflected by an electrostatic potential. (The *dotted* circles are equipotentials.) There is specular reflection at the boundaries with an insulator (*thick solid lines*) and Andreev reflection at the boundary with a superconductor (*dashed line*). The trajectory follows the motion between two Andreev reflections of an electron near the Fermi energy. The Andreev reflected hole retraces this trajectory in opposite direction. From [61]

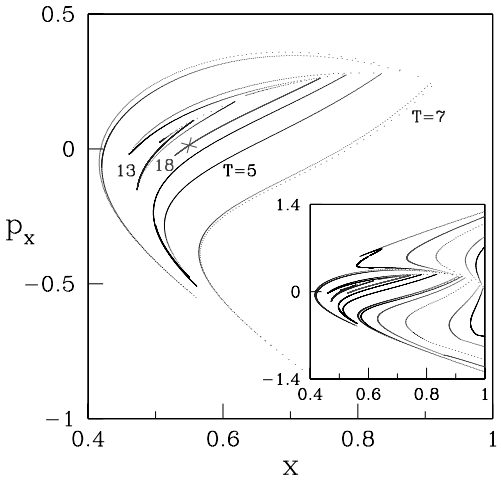

Fig. 16. Poincaré map for the Andreev billiard of Fig. 15. Each dot represents a starting point of an electron trajectory, at position x and with tangential momentum p_x (in dimensionless units). The inset shows the full surface of section, while the main plot is an enlargement of the central region. The drifting nearly periodic motion follows contours of constant time T between Andreev reflections. The cross marks the starting point of the trajectory shown in the previous figure. From [61]

its momentum). The Poincaré map would then consist of a single point. At non-zero excitation energy E the retroreflection occurs with a slight change in p_y, because of the difference $2E$ in the kinetic energy of electrons (at energy $E_F + E$) and holes (at energy $E_F - E$).

The resulting slow drift of the periodic trajectory traces out a contour in the surface of section. These are *isochronous* contours [61], meaning that the time T between Andreev reflections is the same for each point x, p_x on the contour. The adiabatic invariance of T follows from the adiabatic invariance of the action integral I over the nearly periodic motion from electron to hole and back to electron:

$$I = \oint p dq = 2ET \, . \tag{75}$$

Since E is a constant of the motion, adiabatic invariance of I implies adiabatic invariance of the time T between Andreev reflections.

Adiabatic invariance is defined in the limit $E \to 0$ and is therefore distinct from invariance in the sense of Kolmogorov-Arnold-Moser (KAM) [7], which would require a critical E^* such that a contour is exactly invariant for $E < E^*$. There is numerical evidence [5] that the KAM theorem does not apply to a chaotic Andreev billiard.

It is evident from Fig. 16 that contours of large T enclose a very small area in a chaotic system. To estimate the area, it is convenient to measure x in units of the width W of the constriction to the superconductor. Similarly, p_x is conveniently measured in units of the range Δp of transverse momenta inside the constriction.[4] The highly elongated shape evident in Fig. 16 is a consequence of the exponential divergence in time of nearby trajectories, characteristic of chaotic dynamics. The rate of divergence is the Lyapunov exponent α. Since the Hamiltonian flow is area preserving, a stretching $\ell_+(t) = \ell_+(0)e^{\alpha t}$ of the dimension in one direction needs to be compensated by a squeezing $\ell_-(t) = \ell_-(0)e^{-\alpha t}$ of the dimension in the other direction. The area $A \simeq \ell_+\ell_-$ is then time independent. Initially, $\ell_\pm(0) < 1$. The constriction at the superconductor acts as a bottleneck, enforcing $\ell_\pm(T) < 1$. These two inequalities imply $\ell_+(t) < e^{\alpha(t-T)}$, $\ell_- < e^{-\alpha t}$. The enclosed area, therefore, has upper bound

$$A_{\max} \simeq W \Delta p \, e^{-\alpha T} \simeq \hbar N e^{-\alpha T} \, , \tag{76}$$

where $N \simeq W\Delta p/\hbar \gg 1$ is the number of channels in the point contact.

The two invariants E and T define a two-dimensional torus in the four-dimensional phase space. Quantization of this adiabatically invariant torus proceeds following Einstein-Brillouin-Keller [7, 63], by quantizing the area

[4] We consider in this estimate the symmetric case $W/L \simeq \Delta p/p_F \ll 1$, typical for the smooth confining potential of Fig. 15. In the asymmetric case $W/L \ll \Delta p/p_F \simeq 1$, typical for the computer simulations using the kicked rotator, the maximal area $A_{\max}$ is smaller by a factor W/L, cf. [62]. Consequently, the factor $\ln N$ in (81) should be replaced by $\ln(NW/L)$.

$$\oint p dq = 2\pi\hbar(m + \nu/4), \quad m = 0, 1, 2, \ldots \tag{77}$$

enclosed by each of the two topologically independent contours on the torus. Equation (77) ensures that the wave functions are single valued. The integer ν counts the number of caustics (Maslov index) and in this case should also include the number of Andreev reflections.

The first contour follows the quasiperiodic orbit of (75), leading to

$$ET = (m + \tfrac{1}{2})\pi\hbar, \quad m = 0, 1, 2, \ldots \tag{78}$$

The quantization condition (78) is sufficient to determine the smoothed (or ensemble averaged) density of states

$$\langle \rho(E) \rangle = N \int_0^\infty dT\, P(T) \sum_{m=0}^{\infty} \delta\left(E - (m + \tfrac{1}{2})\pi\hbar/T\right) , \tag{79}$$

using the classical probability distribution $P(T)$ for the time between Andreev reflections. (The distribution $P(T)$ is defined with a uniform measure in the surface of section (x, p_x) at the interface with the superconductor.)

Equation (79) is the "Bohr-Sommerfeld rule" of Melsen et al. [22]. It generalizes the familiar Bohr-Sommerfeld quantization rule for translationally invariant geometries [cf. (7)] to arbitrary geometries. The quantization rule refers to classical periodic motion with period $2T$ and phase increment per period of $2ET/\hbar - \pi$, consisting of a part $2ET/\hbar$ because of the energy difference $2E$ between electron and hole, plus a phase shift of $-\pi$ from two Andreev reflections. If E is not $\ll \Delta$, this latter phase shift should be replaced by $-2\arccos(E/\Delta)$ [64, 65, 66], cf. (20). In the presence of a magnetic field an extra phase increment proportional to the enclosed flux should be included [67]. Equation (79) can also be derived from the Eilenberger equation for the quasiclassical Green function [23].

To find the location of individual energy levels a second quantization condition is needed [61]. It is provided by the area $\oint_T p_x dx$ enclosed by the isochronous contours,

$$\oint_T p_x dx = 2\pi\hbar(n + \nu/4), \quad n = 0, 1, 2, \ldots \tag{80}$$

Equation (80) amounts to a quantization of the period T, which together with (78) leads to a quantization of E. For each T_n there is a ladder of Andreev levels $E_{nm} = (m + \frac{1}{2})\pi\hbar/T_n$.

While the classical T can become arbitrarily large, the quantized T_n has a cutoff. The cutoff follows from the maximal area (76) enclosed by an isochronous contour. Since (80) requires $A_{\max} \gtrsim h$, the longest quantized period is $T_0 = \alpha^{-1}[\ln N + \mathcal{O}(1)]$. The lowest Andreev level associated with an adiabatically invariant torus is therefore

$$E_{00} = \frac{\pi\hbar}{2T_0} = \frac{\pi\hbar\alpha}{2\ln N} \,. \tag{81}$$

The time scale $T_0 \propto |\ln\hbar|$ is the Ehrenfest time τ_E of the Andreev billiard, to which we will return in Sect. 8.

The range of validity of adiabatic quantization is determined by the requirement that the drift δx, δp_x upon one iteration of the Poincaré map should be small compared to the characteristic values W, p_F. An estimate is [61]

$$\frac{\delta x}{W} \simeq \frac{\delta p_x}{p_F} \simeq \frac{E_{nm}}{\hbar\alpha N} e^{\alpha T_n} \simeq (m+\tfrac{1}{2})\frac{e^{-\alpha(T_0-T_n)}}{\alpha T_n} \,. \tag{82}$$

For low-lying levels ($m \sim 1$) the dimensionless drift is $\ll 1$ for $T_n < T_0$. Even for $T_n = T_0$ one has $\delta x/W \simeq 1/\ln N \ll 1$.

7.2 Integrable Dynamics

Unlike RMT, the quasiclassical theory is not restricted to systems with a chaotic classical dynamics. Melsen et al. [22, 45] have used the Bohr-Sommerfeld rule (79) to argue that Andreev billiards with an integrable classical dynamics have a smoothly vanishing density of states – without an actual excitation gap. The presence or absence of an excitation gap is therefore a "quantum signature of chaos". This is a unique property of Andreev billiards. In normal, not-superconducting billiards, it is impossible to distinguish chaotic from integrable dynamics by looking at the density of states. One needs to measure density-density correlation functions for that purpose [8].

The difference between chaotic and integrable Andreev billiards is illustrated in Fig. 17. As expected, the chaotic Sinai billiard follows closely the prediction from RMT. (The agreement is less precise than for the kicked rotator of Fig. 12, because the number of modes $N = 20$ is necessarily much smaller in this simulation.) The density of states of the integrable circular billiard is suppressed on the same mesoscopic energy scale E_T as the chaotic billiard, but the suppression is smooth rather than abrupt. Any remaining gap is microscopic, on the scale of the level spacing, and therefore invisible in the smoothed density of states.

That the absence of an excitation gap is generic for integrable billiards can be understood from the Bohr-Sommerfeld rule [22]. Generically, an integrable billiard has a power-law distribution of dwell times, $P(T) \propto T^{-p}$ for $T \to \infty$, with $p \approx 3$ [68, 69]. Equation (79) then implies a power-law density of states, $\langle\rho(E)\rangle \propto E^{p-2}$ for $E \to 0$. The value $p = 3$ corresponds to a linearly vanishing density of states. An analytical calculation [70] of $P(T)$ for a rectangular billiard gives the long-time limit $P(T) \propto T^{-3}\ln T$, corresponding to the low-energy asymptote $\langle\rho(E)\rangle \propto E\ln(E_T/E)$. The weak logarithmic correction to the linear density of states is consistent with exact quantum mechanical calculations [22, 67].

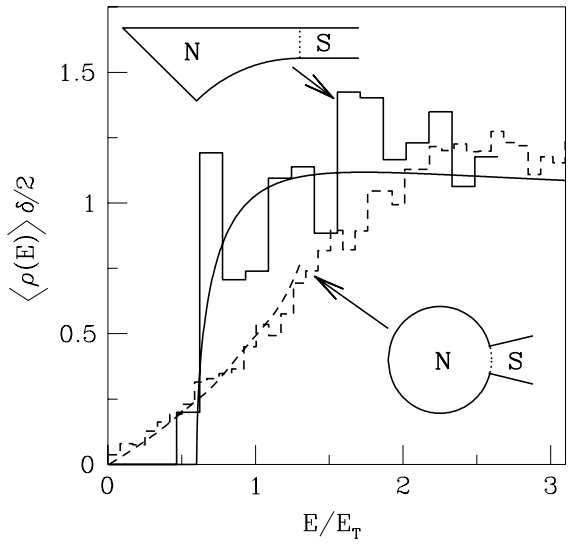

Fig. 17. Histograms: smoothed density of states of a billiard coupled by a ballistic N-mode lead to a superconductor, determined by (22) and averaged over a range of Fermi energies at fixed N. The scattering matrix is computed numerically by matching wave functions in the billiard to transverse modes in the lead. A chaotic Sinai billiard (top inset, *solid* histogram, $N = 20$) is contrasted with an integrable circular billiard (bottom inset, *dashed* histogram, $N = 30$). The *solid* curve is the prediction (49) from RMT for a chaotic system and the *dashed* curve is the Bohr-Sommerfeld result (79), with dwell time distribution $P(T)$ calculated from classical trajectories in the circular billiard. Adapted from [45]

7.3 Chaotic Dynamics

A chaotic billiard has an exponential dwell time distribution, $P(T) \propto e^{-T/\tau_{\rm dwell}}$, instead of a power law [68]. (The mean dwell time is $\tau_{\rm dwell} = 2\pi\hbar/N\delta \equiv \hbar/2E_T$.) Substitution into the Bohr-Sommerfeld rule (79) gives the density of states [71]

$$\langle\rho(E)\rangle = \frac{2}{\delta}\,\frac{(\pi E_T/E)^2\cosh(\pi E_T/E)}{\sinh^2(\pi E_T/E)}\,, \tag{83}$$

which vanishes $\propto e^{-\pi E_T/E}$ as $E \to 0$. This is a much more rapid decay than for integrable systems, but not quite the hard gap predicted by RMT [22]. The two densities of states are compared in Fig. 18.

When the qualitative difference between the random-matrix and Bohr-Sommerfeld theories was discovered [22], it was believed to be a short-coming of the quasiclassical approximation underlying the latter theory. Lodder and Nazarov [23] realized that the two theoretical predictions are actually both correct, in different limits. As the ratio $\tau_E/\tau_{\rm dwell}$ of Ehrenfest time and dwell time is increased, the density of states crosses over from the RMT form (49) to

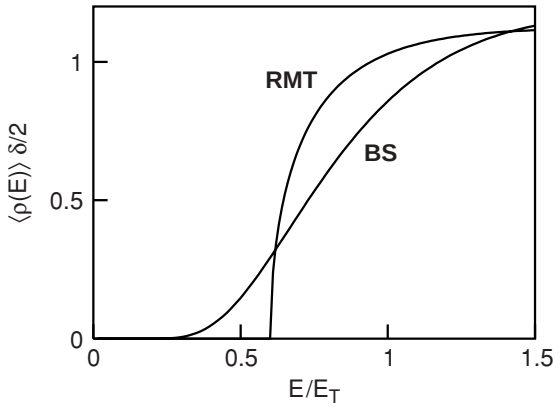

Fig. 18. Comparison of the smoothed density of states in a chaotic Andreev billiard as it follows from RMT ((49), with a hard gap) and as it follows from the Bohr-Sommerfeld (BS) rule ((83), without a hard gap). These are the two limiting distributions when the Ehrenfest time τ_E is, respectively, much smaller or much larger than the mean dwell time $\tau_{\rm dwell}$

the Bohr-Sommerfeld form (83). We investigate this crossover in the following section.

8 Quantum-To-Classical Crossover

8.1 Thouless Versus Ehrenfest

According to Ehrenfest's theorem, the propagation of a quantum mechanical wave packet is described for short times by classical equations of motion. The time scale at which this correspondence between quantum and classical dynamics breaks down in a chaotic system is called the Ehrenfest time τ_E [72].[5] As explained in Fig. 19, it depends logarithmically on Planck's constant: $\tau_E = \alpha^{-1}\ln(S_{\rm cl}/h)$, with $S_{\rm cl}$ the characteristic classical action of the dynamical system and α the Lyapunov exponent.

This logarithmic h-dependence distinguishes the Ehrenfest time from other characteristic time scales of a chaotic system, which are either h-independent (dwell time, ergodic time) or algebraically dependent on h (Heisenberg time $\propto 1/\delta$). That the quasiclassical theory of superconductivity breaks down on time scales greater than τ_E was noticed already in 1968 by Larkin and Ovchinnikov [74].

The choice of $S_{\rm cl}$ depends on the physical quantity which one is studying. For the density of states of the Andreev billiard (area A, opening of width $W \ll A^{1/2}$, range of transverse momenta $\Delta p \simeq p_F$ inside the constriction)

[5] The name "Ehrenfest time" was coined in [73].

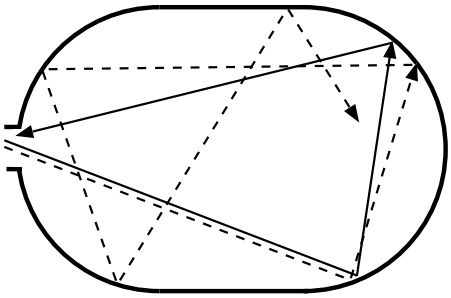

Fig. 19. Two trajectories entering a chaotic billiard at a small separation $\delta x(0)$ diverge exponentially in time, $\delta x(t) = \delta x(0)e^{\alpha t}$. The rate of divergence α is the Lyapunov exponent. An initial microscopic separation λ_F becomes macroscopic at the Ehrenfest time $\tau_E = \alpha^{-1} \ln(L^*/\lambda_F)$. The macroscopic length L^* is determined by the size and shape of the billiard. The Ehrenfest time depends logarithmically on Planck's constant: $\tau_E = \alpha^{-1} \ln(S_{\rm cl}/h)$, with $S_{\rm cl} = mv_F L^*$ the characteristic classical action. The evolution of a quantum mechanical wave packet is well described by a classical trajectory only for times less than τ_E

the characteristic classical action is[6] $S_{\rm cl} = mv_F W^2/A^{1/2}$ [40]. The Ehrenfest time then takes the form

$$\tau_E = \alpha^{-1}[\ln(N^2/M) + \mathcal{O}(1)] \; . \tag{84}$$

Here $M = k_F A^{1/2}/\pi$ and $N = k_F W/\pi$ are, respectively, the number of modes in a cross-section of the billiard and in the point contact. Equation (84) holds for $N \gtrsim \sqrt{M}$. For $N \lesssim \sqrt{M}$ the Ehrenfest time may be set to zero, because the wave packet then spreads over the entire billiard within the ergodic time [62].

Chaotic dynamics requires $\alpha^{-1} \ll \tau_{\rm dwell}$. The relative magnitude of τ_E and $\tau_{\rm dwell}$ thus depends on whether the ratio N^2/M is large or small compared to the exponentially large number $e^{\alpha\tau_{\rm dwell}}$.

The result of RMT [22], cf. Sect. 6.2, is that the excitation gap in an Andreev billiard is of the order of the Thouless energy $E_T \simeq \hbar/\tau_{\rm dwell}$. It was realized by Lodder and Nazarov [23] that this result requires $\tau_E \ll \tau_{\rm dwell}$. More generally, the excitation gap $E_{\rm gap} \simeq \min(E_T, \hbar/\tau_E)$ is determined by the smallest of the Thouless and Ehrenfest energy scales. The Bohr-Sommerfeld theory [22], cf. Sect. 7.3, holds in the limit $\tau_E \to \infty$ and therefore produces a gapless density of states.

8.2 Effective RMT

A phenomenological description of the crossover from the Thouless to the Ehrenfest regime is provided by the "effective RMT" of Silvestrov et al. [61].

[6] The simpler expression $S_{\rm cl} = mv_F W$ of [61] applies to the symmetric case $W/A^{1/2} \simeq \Delta p/p_F \ll 1$.

As described in Sect. 7.1, the quasiclassical adiabatic quantization allows to quantize only the trajectories with periods $T \leq T_0 \equiv \tau_E$. The excitation gap of the Andreev billiard is determined by the part of phase space with periods longer than τ_E. Effective RMT is based on the hypothesis that this part of phase space can be quantized by a scattering matrix $S_{\rm eff}$ in the circular ensemble of RMT, with a reduced dimensionality

$$N_{\rm eff} = N \int_{\tau_E}^{\infty} P(T)\, dT = N e^{-\tau_E/\tau_{\rm dwell}} \,. \tag{85}$$

The energy dependence of $S_{\rm eff}(E)$ is that of a chaotic cavity with mean level spacing $\delta_{\rm eff}$, coupled to the superconductor by a long lead with $N_{\rm eff}$ propagating modes. (See Fig. 20.) The lead introduces a mode-independent delay time τ_E between Andreev reflections, to ensure that $P(T)$ is cut off for $T < \tau_E$. Because $P(T)$ is exponential $\propto \exp(-T/\tau_{\rm dwell})$, the mean time $\langle T\rangle_*$ between Andreev reflections in the accessible part of phase space is simply $\tau_E + \tau_{\rm dwell}$. The effective level spacing in the chaotic cavity by itself (without the lead) is then determined by

$$\frac{2\pi\hbar}{N_{\rm eff}\delta_{\rm eff}} = \langle T\rangle_* - \tau_E = \tau_{\rm dwell} \,. \tag{86}$$

It is convenient to separate the energy dependence due to the lead from that due to the cavity, by writing $S_{\rm eff}(E) = \exp(iE\tau_E/\hbar)S_0(E)$, where $S_0(E)$ represents only the cavity and has an energy dependence of the usual RMT

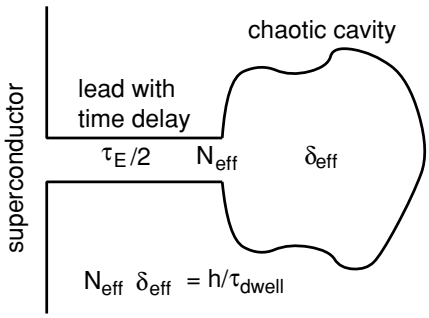

Fig. 20. Pictorial representation of the effective RMT of an Andreev billiard. The part of phase space with time $T > \tau_E$ between Andreev reflections is represented by a chaotic cavity (mean level spacing $\delta_{\rm eff}$), connected to the superconductor by a long lead ($N_{\rm eff}$ propagating modes, one-way delay time $\tau_E/2$ for each mode). Between two Andreev reflections an electron or hole spends, on average, a time τ_E in the lead and a time $\tau_{\rm dwell}$ in the cavity. The scattering matrix of lead plus cavity is $\exp(iE\tau_E/\hbar)S_0(E)$, with $S_0(E)$ distributed according to the circular ensemble of RMT (with effective parameters $N_{\rm eff}$, $\delta_{\rm eff}$). The complete excitation spectrum of the Andreev billiard consists of the levels of the effective RMT (periods $> \tau_E$) plus the levels obtained by adiabatic quantization (periods $<\tau_E$)

form (30) – with effective parameters $N_{\rm eff}$ and $\delta_{\rm eff}$. The determinant (22) for the excitation spectrum then takes the form

$$\mathrm{Det}\left[1-\alpha(E)^2 e^{2iE\tau_E/\hbar} S_0(E) S_0(-E)^*\right] = 0\ . \tag{87}$$

We can safely replace $\alpha(E) \equiv \exp[-i\arccos(E/\Delta)] \to -i$ (since $E \ll \Delta$), but the energy dependence of the phase factor $e^{2iE\tau_E/\hbar}$ can not be omitted.

In App. A we calculate the smallest positive E that solves (87), which is the excitation gap $E_{\rm gap}$ of the effective RMT. The result is plotted in Fig. 21 (solid curve), as a function of $\tau_E/\tau_{\rm dwell}$. The two asymptotes (dotted lines) are

$$E_{\rm gap} = \frac{\gamma^{5/2}\hbar}{\tau_{\rm dwell}}\left(1-(2\gamma-1)\frac{\tau_E}{\tau_{\rm dwell}}\right),\quad \tau_E \ll \tau_{\rm dwell}\ , \tag{88}$$

$$E_{\rm gap} = \frac{\pi\hbar}{2\tau_E}\left(1-(3+\sqrt{8})\frac{\tau_{\rm dwell}}{\tau_E}\right),\quad \tau_E \gg \tau_{\rm dwell}\ , \tag{89}$$

with $\gamma = \frac{1}{2}(\sqrt{5}-1)$ the golden number.

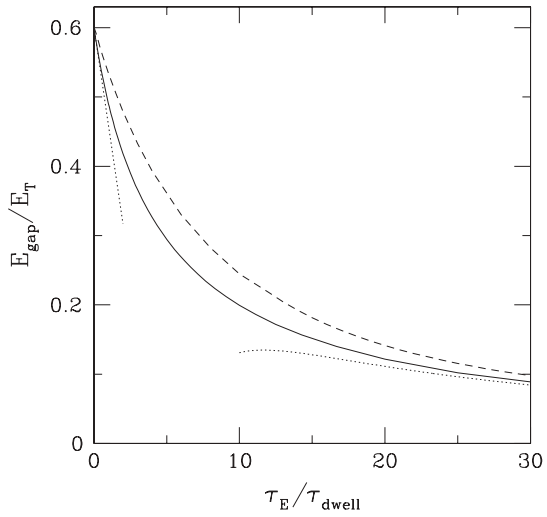

Fig. 21. Excitation gap of the Andreev billiard in the crossover from Thouless to Ehrenfest regimes. The *solid* curve is the solution of the effective RMT of [61], derived in App. A. The *dotted* lines are the two asymptotes (88) and (89). The *dashed* curve is the result of the stochastic model of [40], discussed in Sect. 8.3

The τ_E time delay characteristic of the effective RMT was introduced in [61], but its effect on the excitation gap was not evaluated properly.[7] As a consequence the formula for the gap given in that paper,

[7] I am indebted to P. W. Brouwer for spotting the error.

$$E_{\rm gap} = \frac{0.30\,\hbar}{\langle T\rangle_*} = \frac{0.30\,\hbar}{\tau_E + \tau_{\rm dwell}}\,, \tag{90}$$

provides only a qualitative description of the actual crossover.

The inverse correlation (90) between the gap and the dwell time of long trajectories was observed in a computer simulation of the Andreev kicked rotator [75]. The data points in Fig. 22 track the excitation gap as the location in phase space of the NS interface is varied. The solid curve is a plot of

$$\frac{1}{\langle T\rangle_*} = \frac{\int_{T^*}^\infty P(T)dT}{\int_{T^*}^\infty TP(T)dT}\,, \tag{91}$$

with $P(T)$ the classical dwell time distribution and $T^* = 7$. We see that the sample-to-sample fluctuations in the gap correlate very well with the fluctuations in the mean dwell time of long trajectories. The correlation is not sensitive to the choice of T^*, as long as it is greater than $\tau_E = 4.4$.

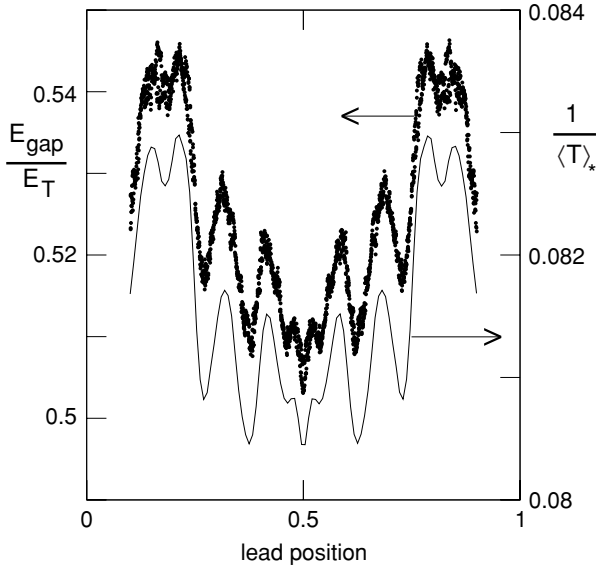

Fig. 22. The data points (*left axis*) are the quantum mechanical gap values $E_{\rm gap}$ of the Andreev kicked rotator as a function of the location of the NS interface, for parameter values $M = 131072$, $\tau_{\rm dwell} = M/N = 5$, $K = 14$. The *solid* curve (*right axis*) is the reciprocal of the mean dwell time (91) of classical trajectories longer than $T^* = 7$. Adapted from [75]

8.3 Stochastic Model

Small-angle scattering by a smooth disorder potential provides a stochastic model for the quantum diffraction of a wave packet in a chaotic billiard [76].

The scattering time of the stochastic model plays the role of the Ehrenfest time in the deterministic chaotic dynamics. The advantage of a stochastic description is that one can average over different realizations of the disorder potential. This provides for an established set of analytical techniques. The disadvantage is that one does not know how well stochastic scattering mimics quantum diffraction.

Vavilov and Larkin [40] have used the stochastic model to study the crossover from the Thouless regime to the Ehrenfest regime in an Andreev billiard. They discovered that the rapid turn-on of quantum diffraction at $\tau_E \gtrsim \tau_{\rm dwell}$ not only causes an excitation gap to open at $\hbar/\tau_E$, but that it also causes oscillations with period $\hbar/\tau_E$ in the ensemble-averaged density of states $\langle\rho(E)\rangle$ at high energies $E \gtrsim E_T$. In normal billiards oscillations with this periodicity appear in the level-level correlation function [77], but not in the level density itself.

The predicted oscillatory high-energy tail of $\langle\rho(E)\rangle$ is plotted in Fig. 23, for the case $\tau_E/\tau_{\rm dwell} = 3$, together with the smooth results of RMT ($\tau_E/\tau_{\rm dwell} \to 0$) and Bohr-Sommerfeld (BS) theory ($\tau_E/\tau_{\rm dwell} \to \infty$).

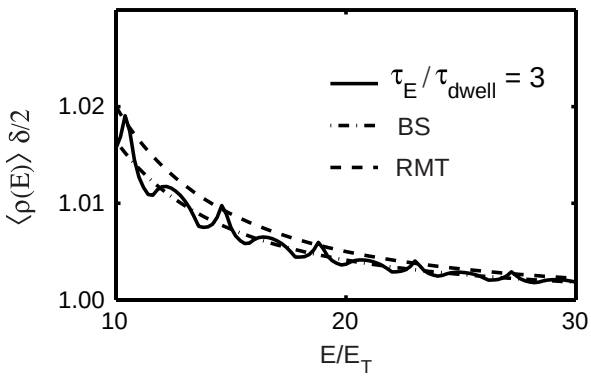

Fig. 23. Oscillatory density of states at finite Ehrenfest time (*solid curve*), compared with the smooth limits of zero (RMT) and infinite (BS) Ehrenfest times. The *solid* curve is the result of the stochastic model of Vavilov and Larkin, for $\tau_E = 3\,\tau_{\rm dwell} = 3\hbar/2E_T$. (The definition (84) of the Ehrenfest time used here differs by a factor of two from that used by those authors.) Adapted from [40]

Independent analytical support for the existence of oscillations in the density of states with period $\hbar/\tau_E$ comes from the singular perturbation theory of [78]. Support from numerical simulations is still lacking. Jacquod et al. [27] did find pronounced oscillations for $E \gtrsim E_T$ in the level density of the Andreev kicked rotator. However, since these could be described by the Bohr-Sommerfeld theory they can not be the result of quantum diffraction, but must be due to nonergodic trajectories [79].

The τ_E-dependence of the gap obtained by Vavilov and Larkin is plotted in Fig. 21 (dashed curve). It is close to the result of the effective RMT (solid curve). The two theories predict the same limit $E_{\rm gap} \to \pi\hbar/2\tau_E$ for $\tau_E/\tau_{\rm dwell} \to \infty$. The asymptotes given in [40] are

$$E_{\rm gap} = \frac{\gamma^{5/2}\hbar}{\tau_{\rm dwell}}\left(1 - 0.23\,\frac{\tau_E}{2\tau_{\rm dwell}}\right), \quad \tau_E \ll \tau_{\rm dwell}\ , \tag{92}$$

$$E_{\rm gap} = \frac{\pi\hbar}{2\tau_E}\left(1 - \frac{2\tau_{\rm dwell}}{\tau_E}\right), \quad \tau_E \gg \tau_{\rm dwell}\ . \tag{93}$$

Both are different from the results (88) and (89) of the effective RMT.[8]

8.4 Numerical Simulations

Because the Ehrenfest time grows only logarithmically with the size of the system, it is exceedingly difficult to do numerical simulations deep in the Ehrenfest regime. Two simulations [27, 80] have been able to probe the initial decay of the excitation gap, when $\tau_E \lesssim \tau_{\rm dwell}$. We show the results of both simulations in Fig. 24 (closed and open circles), together with the full decay as predicted by the effective RMT of Sect. 8.2 (solid curve) and by the stochastic model of Sect. 8.3 (dashed curve).

The closed circles were obtained by Jacquod et al. [27] using the stroboscopic model of Sect. 5 (the Andreev kicked rotator). The number of modes N in the contact to the superconductor was increased from 10^2 to 10^5 at fixed dwell time $\tau_{\rm dwell} = M/N = 5$ and kicking strength $K = 14$ (corresponding to a Lyapunov exponent $\alpha \approx \ln(K/2) = 1.95$). In this way all classical properties of the billiard remain the same while the effective Planck constant $h_{\rm eff} = 1/M = 1/N\tau_{\rm dwell}$ is reduced by three orders of magnitude. To plot the data as a function of $\tau_E/\tau_{\rm dwell}$, (84) was used for the Ehrenfest time. The unspecified terms of order unity in that equation were treated as a single fit parameter. (This amounts to a horizontal shift by -0.286 of the data points in Fig. 24.)

The open circles were obtained by Kormányos et al. [80] for the chaotic Sinai billiard shown in the inset. The number of modes N was varied from 18 to 30 by varying the width of the contact to the superconductor. The Lyapunov exponent $\alpha \approx 1.7$ was fixed, but $\tau_{\rm dwell}$ was not kept constant in this simulation. The Ehrenfest time was computed by means of the same formula (84), with $M = 2L_c k_F/\pi$ and L_c the average length of a trajectory between two consecutive bounces at the curved boundary segment.

The data points from both simulations have substantial error bars (up to 10%). Because of that and because of their limited range, we can not conclude that the simulations clearly favor one theory over the other.

[8] Since $2\gamma - 1 = 0.236$, the small-τ_E asymptote of Vavilov and Larkin differs by a factor of two from that of the effective RMT.

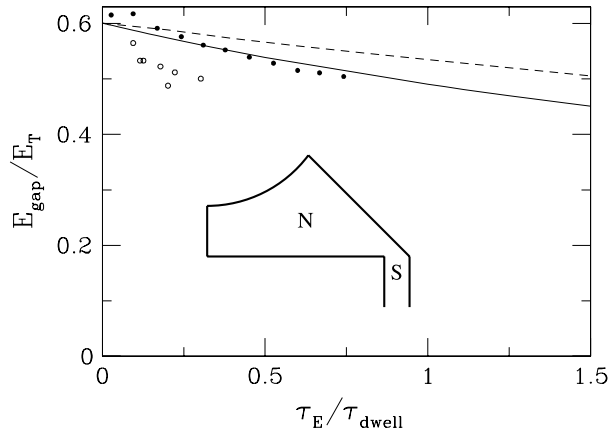

Fig. 24. Ehrenfest-time dependence of the excitation gap in an Andreev billiard, according to the effective RMT (*solid* curve, calculated in App. A) and according to the stochastic model (*dashed* curve, calculated in [40]). The data points result from the simulation of the Andreev kicked rotator [27] (*closed circles*, in the range $N = 10^2$–10^5) and of the Sinai billiard shown in the inset [80] (*open circles*, in the range $N = 18$–30)

9 Conclusion

Looking back at what we have learned from the study of Andreev billiards, we would single out the breakdown of random-matrix theory as the most unexpected discovery and the one with the most far-reaching implications for the field of quantum chaos. In an isolated chaotic billiard RMT provides an accurate description of the spectral statistics on energy scales below $\hbar/\tau_{\rm erg}$ (the inverse ergodic time). The weak coupling to a superconductor causes RMT to fail at a much smaller energy scale of $\hbar/\tau_{\rm dwell}$ (the inverse of the mean time between Andreev reflections), once the Ehrenfest time τ_E becomes greater than $\tau_{\rm dwell}$.

In the limit $\tau_E \to \infty$, the quasiclassical Bohr-Sommerfeld theory takes over from RMT. While in isolated billiards such an approach can only be used for integrable dynamics, the Bohr-Sommerfeld theory of Andreev billiards applies regardless of whether the classical motion is integrable or chaotic. This is a demonstration of how the time-reversing property of Andreev reflection unravels chaotic dynamics.

What is lacking is a conclusive theory for finite $\tau_E \gtrsim \tau_{\rm dwell}$. The two phenomenological approaches of Sects. 8.2 and 8.3 agree on the asymptotic behavior

$$\lim_{\hbar\to 0} E_{\rm gap} = \frac{\pi\hbar\alpha}{2|\ln\hbar| + \text{constant}}\ , \tag{94}$$

in the classical $\hbar \to 0$ limit (understood as $N \to \infty$ at fixed $\tau_{\rm dwell}$). There is still some disagreement on how this limit is approached. We would hope

that a fully microscopic approach, for example based on the ballistic σ-model [81, 82], could provide a conclusive answer. At present technical difficulties still stand in the way of a solution along those lines [83].

A new direction of research is to investigate the effects of a nonisotropic superconducting order parameter on the Andreev billiard. The case of d-wave symmetry is most interesting because of its relevance for high-temperature superconductors. The key ingredients needed for a theoretical description exist, notably RMT [84], quasiclassics [85], and a numerically efficient Andreev map [86].

Acknowledgments

While writing this review, I benefitted from correspondence and discussions with W. Belzig, P. W. Brouwer, J. Cserti, P. M. Ostrovsky, P. G. Silvestrov, and M. G. Vavilov. The work was supported by the Dutch Science Foundation NWO/FOM.

A Excitation Gap in Effective RMT and Relationship with Delay Times

We seek the edge of the excitation spectrum as it follows from the determinant (87), which in zero magnetic field and for $E \ll \Delta$ takes the form

$$\mathrm{Det}\left[1 + e^{2iE\tau_E/\hbar} S_0(E) S_0(-E)^\dagger\right] = 0\,. \tag{95}$$

The unitary symmetric matrix S_0 has the RMT distribution of a chaotic cavity with effective parameters $N_{\rm eff}$ and $\delta_{\rm eff}$ given by (85) and (86). The mean dwell time associated with S_0 is $\tau_{\rm dwell}$. The calculation for $N_{\rm eff} \gg 1$ follows the method described in Sects. 6.1 and 6.2, modified as in [37] to account for the energy dependent phase factor in the determinant.

Since S_0 is of the RMT form (30), we can write (95) in the Hamiltonian form (32). The extra phase factor $\exp(2iE\tau_E/\hbar)$ introduces an energy dependence of the coupling matrix,

$$\mathcal{W}_0(E) = \frac{\pi}{\cos u}\begin{pmatrix} W_0 W_0^T \sin u & W_0 W_0^T \\ W_0 W_0^T & W_0 W_0^T \sin u \end{pmatrix}\,, \tag{96}$$

where we have abbreviated $u = E\tau_E/\hbar$. The subscript 0 reminds us that the coupling matrix refers to the reduced set of $N_{\rm eff}$ channels in the effective RMT. Since there is no tunnel barrier in this case, the matrix W_0 is determined by (31) with $\Gamma_n \equiv 1$. The Hamiltonian

$$\mathcal{H}_0 = \begin{pmatrix} H_0 & 0 \\ 0 & -H_0 \end{pmatrix} \tag{97}$$

is that of a chaotic cavity with mean level spacing $\delta_{\rm eff}$. We seek the gap in the density of states

$$\rho(E) = -\frac{1}{\pi}\mathrm{Im}\,\mathrm{Tr}\left(1+\frac{d\mathcal{W}_0}{dE}\right)\left(E+i0^+-\mathcal{H}_0+\mathcal{W}_0\right)^{-1}, \tag{98}$$

cf. (35).

The selfconsistency equation for the ensemble-averaged Green function,

$$\mathcal{G} = [E+\mathcal{W}_0-(M\delta_{\rm eff}/\pi)\sigma_z G\sigma_z]^{-1}, \tag{99}$$

still leads to (43a), but (43b) should be replaced by

$$\begin{aligned} G_{11}+G_{12}\sin u = &-(\tau_{\rm dwell}/\tau_E)uG_{12} \\ &\times(G_{12}+\cos u+G_{11}\sin u)\,. \end{aligned} \tag{100}$$

(We have used that $N_{\rm eff}\delta_{\rm eff} = 2\pi\hbar/\tau_{\rm dwell}$.) Because of the energy dependence of the coupling matrix, the (44) for the ensemble averaged density of states should be replaced by

$$\langle\rho(E)\rangle = -\frac{2}{\delta}\,\mathrm{Im}\left(G_{11}-\frac{u}{\cos u}G_{12}\right). \tag{101}$$

The excitation gap corresponds to a square root singularity in $\langle\rho(E)\rangle$, which can be obtained by solving (43a) and (100) jointly with $dE/dG_{11}=0$ for $u\in(0,\pi/2)$. The result is plotted in Fig. 21. The small- and large-τ_E asymptotes are given by (88) and (89).

The large-τ_E asymptote is determined by the largest eigenvalue of the time-delay matrix. To see this relationship, note that for $\tau_E \gg \tau_{\rm dwell}$ we may replace the determinant (95) by

$$\mathrm{Det}\left[1+\exp[2iE\tau_E/\hbar+2iEQ(0)]+\mathcal{O}\left(\frac{\tau_{\rm dwell}}{\tau_E}\right)^2\right]=0\,. \tag{102}$$

The Hermitian matrix

$$Q(E) = \frac{1}{i}S_0(E)^\dagger\frac{d}{dE}S_0(E) \tag{103}$$

is known in RMT as the Wigner-Smith or time-delay matrix. The roots E_{np} of (102) satisfy

$$2E_{np}(\tau_E+\tau_n) = (2p+1)\pi\hbar,\quad p=0,1,2,\ldots\,. \tag{104}$$

The eigenvalues $\tau_1,\tau_2,\ldots\tau_{N_{\rm eff}}$ of $\hbar Q(0)$ are the delay times. They are all positive, distributed according to a generalized Laguerre ensemble [87]. In the limit $N_{\rm eff}\to\infty$ the distribution of the τ_n's is nonzero only in the interval

(τ_-, τ_+), with $\tau_\pm = \tau_{\rm dwell}(3 \pm \sqrt{8})$. By taking $p = 0$, $\tau_n = \tau_+$ we arrive at the asymptote (89).

The precise one-to-one correspondence (104) between the spectrum of low-lying energy levels of the Andreev billiard and the spectrum of delay times is a special property of the classical limit $\tau_E \to \infty$. For $\tau_E \lesssim \tau_{\rm dwell}$ there is only a qualitative similarity of the two spectral densities [88].

References

1. A. F. Andreev, Zh. Eksp. Teor. Fiz. **46**, 1823 (1964) [Sov. Phys. JETP **19**, 1228 (1964)]. 131, 135, 139
2. Y. Imry, *Introduction to Mesoscopic Physics* (Oxford University, Oxford, 2002). 131, 135
3. C. W. J. Beenakker, Rev. Mod. Phys. **69**, 731 (1997). 131, 132, 133, 143, 148
4. B. J. van Wees and H. Takayanagi, in *Mesoscopic Electron Transport*, edited by L. L. Sohn, L. P. Kouwenhoven, and G. Schön, NATO ASI Series E345 (Kluwer, Dordrecht, 1997). 131
5. I. Kosztin, D. L. Maslov, and P. M. Goldbart, Phys. Rev. Lett. **75**, 1735 (1995). 132, 156, 158
6. J. Eroms, M. Tolkiehn, D. Weiss, U. Rössler, J. De Boeck, and G. Borghs, Europhys. Lett. **58**, 569 (2002). 132
7. M. C. Gutzwiller, *Chaos in Classical and Quantum Mechanics* (Springer, New York, 1990). 132, 139, 158
8. F. Haake, *Quantum Signatures of Chaos* (Springer, Berlin, 2001). 132, 139, 142, 160
9. M. L. Mehta, *Random Matrices* (Academic, New York, 1991). 133, 141, 142, 144, 147, 149
10. T. Guhr, A. Müller-Groeling, and H. A. Weidenmüller, Phys. Rep. **299**, 189 (1998). 133, 143
11. P. G. de Gennes, *Superconductivity of Metals and Alloys* (Benjamin, New York, 1966). 134
12. K. K. Likharev, Rev. Mod. Phys. **51**, 101 (1979). 134
13. C. W. J. Beenakker, in: *Transport Phenomena in Mesoscopic Systems*, edited by H. Fukuyama and T. Ando (Springer, Berlin, 1992); `cond-mat/0406127`. 134
14. A. A. Golubov, M. Yu. Kupriyanov, and E. Ilichev, Rev. Mod. Phys. **76**, 411 (2004). 134
15. W. L. McMillan, Phys. Rev. **175**, 537 (1968). 135
16. P. G. de Gennes and D. Saint-James, Phys. Lett. **4**, 151 (1963); D. Saint-James, J. Physique **25** 899 (1964). 135
17. G. Deutscher, `cond-mat/0409225`. 135
18. A. A. Golubov and M. Yu. Kuprianov, J. Low Temp. Phys. **70**, 83 (1988). 135
19. W. Belzig, C. Bruder, and G. Schön, Phys. Rev. B **54**, 9443 (1996). 135
20. S. Pilgram, W. Belzig, and C. Bruder, Phys. Rev. B **62**, 12462 (2000). 135, 136
21. W. Belzig, F. K. Wilhelm, C. Bruder, G. Schön, and A. D. Zaikin, Superlatt. Microstruct. **25**, 1251 (1999). 135, 147
22. J. A. Melsen, P. W. Brouwer, K. M. Frahm, and C. W. J. Beenakker, Europhys. Lett. **35**, 7 (1996). 136, 144, 145, 146, 159, 160, 161, 163
23. A. Lodder and Yu. V. Nazarov, Phys. Rev. B **58**, 5783 (1998). 136, 159, 161, 163
24. C. W. J. Beenakker, Phys. Rev. Lett. **67**, 3836 (1991); **68**, 1442(E) (1992). 137, 139
25. E. B. Bogomolny, Nonlinearity **5**, 805 (1992). 139
26. R. E. Prange, Phys. Rev. Lett. **90**, 070401 (2003). 139

27. Ph. Jacquod, H. Schomerus, and C. W. J. Beenakker, Phys. Rev. Lett. **90**, 207004 (2003). 140, 153, 154, 167, 168, 169
28. A. Ossipov, T. Kottos, and T. Geisel, Europhys. Lett. **62**, 719 (2003). 140
29. Y. V. Fyodorov and H.-J. Sommers, JETP Lett. **72**, 422 (2000). 140
30. A. Ossipov and T. Kottos, Phys. Rev. Lett. **92**, 017004 (2004). 141
31. F. M. Izrailev, Phys. Rep. **196**, 299 (1990). 141
32. J. Tworzydło, A. Tajic, and C. W. J. Beenakker, `cond-mat/0405122`. 141
33. R. Ketzmerick, K. Kruse, and T. Geisel, Physica D **131**, 247 (1999). 141
34. O. Bohigas, M.-J. Giannoni, and C. Schmit, Phys. Rev. Lett. **52**, 1 (1984). 142
35. S. Müller, S. Heusler, P. Braun, F. Haake, and A. Altland, Phys. Rev. Lett. **93**, 014103 (2004). 142
36. M. G. Vavilov, P. W. Brouwer, V. Ambegaokar, and C. W. J. Beenakker, Phys. Rev. Lett. **86**, 874 (2001). 142, 150, 151, 152, 153
37. P. W. Brouwer and C. W. J. Beenakker, Chaos, Solitons & Fractals **8**, 1249 (1997). Several misprints have been corrected in `cond-mat/9611162`. 143, 145, 154, 170
38. K. M. Frahm, P. W. Brouwer, J. A. Melsen, and C. W. J. Beenakker, Phys. Rev. Lett. **76**, 2981 (1996). 143, 150, 151
39. A. Altland and M. R. Zirnbauer, Phys. Rev. Lett. **76**, 3420 (1996); Phys. Rev. B **55**, 1142 (1997). 144, 149
40. M. G. Vavilov and A. I. Larkin, Phys. Rev. B **67**, 115335 (2003). 144, 147, 148, 163, 165, 167, 168, 169
41. A. Pandey, Ann. Phys. (N.Y.) **134**, 110 (1981). 145
42. E. Brézin and A. Zee, Phys. Rev. E **49**, 2588 (1994). 145
43. L. A. Pastur, Teoret. Mat. Fiz. **10**, 102 (1972) [Theoret. Math. Phys. **10**, 67 (1972)]. 145
44. M. Tinkham, *Introduction to Superconductivity* (McGraw-Hill, New York, 1995). 147
45. J. A. Melsen, P. W. Brouwer, K. M. Frahm, and C. W. J. Beenakker, Physica Scripta **T69**, 223 (1997). 147, 149, 160, 161
46. A. Pandey and M. L. Mehta, Commun. Math. Phys. **87**, 449 (1983). 147
47. T. Nagao and K. Slevin, J. Math. Phys. **34**, 2075 (1993). 149
48. J. T. Bruun, S. N. Evangelou, and C. J. Lambert, J. Phys. Condens. Matt. **7**, 4033 (1995). 150
49. K. Efetov, *Supersymmetry in Disorder and Chaos* (Cambridge University, Cambridge, 1997). 150
50. S. Gnutzmann, B. Seif, F. von Oppen, and M. R. Zirnbauer, Phys. Rev. E **67**, 046225 (2003). 150
51. C. W. J. Beenakker, Phys. Rev. B **46**, 12841 (1992). 150
52. C. A. Tracy and H. Widom, Commun. Math. Phys. **159**, 151 (1994); **177**, 727 (1996). 150, 151, 152
53. P. M. Ostrovsky, M. A. Skvortsov, and M. V. Feigelman, Phys. Rev. Lett. **87**, 027002 (2001); JETP Lett. **75**, 336 (2002). 150
54. A. Lamacraft and B. D. Simons, Phys. Rev. B **64**, 014514 (2001). 150
55. P. M. Ostrovksy, M. A. Skvortsov, and M. V. Feigelman, Phys. Rev. Lett. **92**, 176805 (2004). 154, 155, 156
56. L. G. Aslamasov, A. I. Larkin, and Yu. N. Ovchinnikov, Zh. Eksp. Teor. Fiz. **55**, 323 (1968) [Sov. Phys. JETP **28**, 171 (1969)]. 154
57. A. V. Shytov, P. A. Lee, and L. S. Levitov, Phys. Uspekhi **41**, 207 (1998). 156
58. J. Wiersig, Phys. Rev. E **65**, 036221 (2002). 156
59. I. Adagideli and P. M. Goldbart, Phys. Rev. B **65**, 201306 (2002). 156
60. J. Cserti, P. Polinák, G. Palla, U. Zülicke, and C. J. Lambert, Phys. Rev. B **69**, 134514 (2004). 156

61. P. G. Silvestrov, M. C. Goorden, and C. W. J. Beenakker, Phys. Rev. Lett. **90**, 116801 (2003). 156, 157, 158, 159, 160, 163, 165
62. P. G. Silvestrov, M. C. Goorden, and C. W. J. Beenakker, Phys. Rev. B **67**, 241301 (2003). 158, 163
63. K. P. Duncan and B. L. Györffy, Ann. Phys. (N.Y.) **298**, 273 (2002). 158
64. J. Cserti, A. Kormányos, Z. Kaufmann, J Koltai, and C. J. Lambert, Phys. Rev. Lett. **89**, 057001 (2002). 159
65. J. Cserti, A. Bodor, J. Koltai, and G. Vattay, Phys. Rev. B **66**, 064528 (2002). 159
66. J. Cserti, B. Béri, A. Kormányos, P. Pollner, and Z. Kaufmann, J. Phys. Condens. Matt. **16**, 6737 (2004). 159
67. W. Ihra, M. Leadbeater, J. L. Vega, and K. Richter, Europhys. J. B **21**, 425 (2001). 159, 160
68. W. Bauer and G. F. Bertsch, Phys. Rev. Lett. **65**, 2213 (1990). 160, 161
69. H. U. Baranger, R. A. Jalabert, and A. D. Stone, Chaos **3**, 665 (1993). 160
70. A. Kormányos, Z. Kaufmann, J. Cserti, and C. J. Lambert, Phys. Rev. B **67**, 172506 (2003). 160
71. H. Schomerus and C. W. J. Beenakker, Phys. Rev. Lett. **82**, 2951 (1999). 161
72. G. P. Berman and G. M. Zaslavsky, Physica A **91**, 450 (1978). 162
73. B. V. Chirikov, F. M. Izrailev, and D. L. Shepelyansky, Physica D **33**, 77 (1988). 162
74. A. I. Larkin and Yu. N. Ovchinnikov, Zh. Eksp. Teor. Fiz. **55**, 2262 (1968) [Sov. Phys. JETP **28**, 1200 (1969)]. 162
75. M. C. Goorden, Ph. Jacquod, and C. W. J. Beenakker, Phys. Rev. B **68**, 220501 (2003). 166
76. I. L. Aleiner and A. I. Larkin, Phys. Rev. B **54**, 14423 (1996). 166
77. I. L. Aleiner and A. I. Larkin, Phys. Rev. E **55**, 1243 (1997). 167
78. I. Adagideli and C. W. J. Beenakker, Phys. Rev. Lett. **89**, 237002 (2002). 167
79. W. Ihra and K. Richter, Physica E **9**, 362 (2001). 167
80. A. Kormányos, Z. Kaufmann, C. J. Lambert, and J. Cserti, Phys. Rev. B **70**, 052512 (2004). 168, 169
81. B. A. Muzykantskii and D. E. Khmelnitskii, JETP Lett. **62**, 76 (1995). 170
82. A. V. Andreev, B. D. Simons, O. Agam, and B. L. Altshuler, Nucl. Phys. B **482**, 536 (1996). 170
83. D. Taras-Semchuk and A. Altland, Phys. Rev. B **64**, 014512 (2001). 170
84. A. Altland, B. D. Simons, and M. R. Zirnbauer, Phys. Rep. **359**, 283 (2002). 170
85. I. Adagideli, D. E. Sheehy, and P. M. Goldbart, Phys. Rev. B **66**, 140512 (2002). 170
86. I. Adagideli and Ph. Jacquod, Phys. Rev. B **69**, 020503 (2004). 170
87. P. W. Brouwer, K. M. Frahm, and C. W. J. Beenakker, Phys. Rev. Lett. **78**, 4737 (1997); Waves in Random Media, **9**, 91 (1999). 171
88. M. G. A. Crawford and P. W. Brouwer, Phys. Rev. E **65**, 026221 (2002). 172