Coulomb blockade in silicon nanowire MOSFETs

Max Hofheinz

▶ **To cite this version:**

Max Hofheinz. Coulomb blockade in silicon nanowire MOSFETs. Physics [physics]. Université Joseph-Fourier - Grenoble I, 2006. English. NNT : . tel-00131052

HAL Id: tel-00131052

https://theses.hal.science/tel-00131052v1

Submitted on 14 Feb 2007

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Thèse

pour obtenir le

Doctorat de l'Université Grenoble I – Joseph Fourier
Discipline: Physique et Nanophysique

Coulomb blockade in silicon nanowire MOSFETs

Blocage de Coulomb dans les transistors silicium à base de nanofils

présentée et soutenue publiquement

le 11 décembre 2006

par

Max Hofheinz

du

CEA-Grenoble / DRFMC / SPSMS

devant le **Jury**:

D. Christian Glattli	rapporteur
Dietmar Weinmann	rapporteur
Daniel Estève	examinateur
Laurent Levy	président
Marc Sanquer	directeur de thèse
Xavier Jehl	encadrant

Abstract

We present electrical transport measurements at low temperature on single-electron transistors (SETs) based on silicon nanowire MOSFETs.

The Coulomb island is formed in the wire not by constrictions or oxide barriers but by a modulation of the doping level and a gate electrode covering the central part of the wire. The devices form very stable SETs with well-controlled properties.

When few electrons are on the island, it is in a localized regime with strong fluctuations of the spacing between Coulomb blockade peaks. When more than a few tens of electrons are on the island it becomes diffusive. Then the fluctuations of the peak spacing are small and scale with the single-particle level spacing.

The well-controlled Coulomb blockade allows to investigate the barriers formed by the low-doped parts of the wire. On a small scale, the charging of single dopants in the barriers causes anomalies in the Coulomb blockade spectrum which allow to determine capacitance matrix, approximate position, dynamics and spin of the individual dopants. On a large scale, the increase of the electron density in the barriers with gate voltage leads to a dramatic increase of the dielectric constant in the barriers. We find dielectric constant and conductance of the barriers to be linked as predicted by scaling laws describing the metal–insulator transition.

Keywords

mesoscopic transport, silicon, single-electron transistor, addition spectrum, individual dopants, dielectric constant, capacitance measurements, metal-insulator transition, Coulomb glass.

Résumé

Cette thèse est consacrée à des mesures de transport électronique dans des transistors mono-électroniques de type MOSFET silicium à base de nanofil.

L'îlot de blocage de Coulomb n'est pas formé par des constrictions ou des barrières d'oxyde mais par une modulation du dopage et une grille couvrant la partie centrale du fil. Ces dispositifs sont des transistors mono-électroniques très stables et bien contrôlés.

Quand il ne contient que peu d'électrons, l'îlot est dans un régime localisé où l'espacement entre résonances de Coulomb est très irrégulier. A partir de quelques dizaines d'électrons l'îlot devient diffusif. Dans ce cas les fluctuations de l'espacement entre résonances sont petites et correspondent à l'espacement entre niveaux à une particule.

Le blocage de Coulomb contrôlé permet d'analyser les barrières formées par les parties faiblement dopées du fil. A petite échelle, le remplissage de dopants individuels cause des anomalies dans le spectre de Coulomb qui permettent de remonter à la matrice de capacité, la position approximative, la dynamique et le spin des dopants. A grande échelle l'augmentation de la densité électronique dans les barrières avec la tension de grille entraîne une forte augmentation de la constante diélectrique dans les barrières. Nous observons un bon accord entre constante diélectrique et conductance des barrières via les lois d'échelle de la transition métal–isolant.

Mots clés

transport mésoscopique, silicium, transistor mono-électronique, spectre d'addition, dopants uniques, constante diélectrique, mesures de capacitance, transition métal-isolant, verre de Coulomb.

Contents

Remerciements

Les résultats dans cette thèse sont entièrement bâtis sur les progrès de la microélectronique, en particulier sur le fait qu'elle s'est dépassée elle-même et qu'elle est devenue nano-électronique. J'ai profité de ces progrès au travers des transistors à effet de champ de très petite taille et très bonne qualité qui m'ont été fournis, d'une part par Antoine Cros, Robin Cerutti, Philippe Coronel et Hugues Brut de l'équipe de Thomas Skotnicki à ST, d'autre part par l'équipe de Simon Deleonibus au CEA/Leti, notamment par Bernard Prévitali, David Fraboulet et Denis Mariolle (lot B146). Je suis particulièrement reconnaissant à Gabriel Molas qui, non seulement, m'a fourni les échantillons avec lesquels pratiquement tous les résultats dans ce manuscrit ont été obtenus (lot 9857) mais qui a aussi supporté mes questions sans fin. Mais je veux aussi remercier Maud, notre agent double au Leti et traductrice entre Microélectroniens et Mésoscopiens, ainsi qu'à Romain qui m'a éclairé sur les méthodes de caractérisation en microélectronique.

J'ai beaucoup apprécié ce temps au LCP (ou maintenant le LaTEQS et l'IMAPEC). Merci à tous les membres du labo, qui ont crée une atmosphère très cordiale où il était naturel de s'aider mutuellement. En particulier m'ont aidé les maîtres des clés Jean-Luc et Marie-Jo. Ils ont résolu 1001 petits problèmes, pour la plupart avant même qu'ils n'apparaissent. Je dois aussi beaucoup à Jean-Michel qui m'a aidé pour tous les problèmes de cryogénie. Et aussi à Pierre dont les capacités en électronique sont à l'origine de beaucoup d'améliorations de nos mesures. Il m'a aussi aidé sur pas mal d'autres questions d'électronique — et avec sa collection complète de cartes TOP25 pour mes week-ends. Merci à Emanuele qui m'a donné de bons conseils pour les radio-fréquences. Merci aussi à tous les thésards et postdocs du labo, notamment à Martin, qui a découvert le blocage de Coulomb avec moi, et à Olivier, Benjamin et Julien qui ont veillé à ce que je garde un minimum d'activités sociales et sportives pendant la rédaction.

Je tiens beaucoup à remercier Laurent Levy, Daniel Estève, Dietmar Weinmann et Christian Glattli d'avoir accepté d'être membres du jury et rapporteurs pour ma thèse. Dietmar Weinmann m'a rendu le manuscrit avec des corrections très pertinentes qui m'ont beaucoup aidé à améliorer le manuscrit. Merci aussi à mes parents et ma soeur qui, sans être physiciens, ont réussi à lire le manuscrit entier et m'ont corrigé un grand nombre de fautes d'anglais.

Mais je veux surtout remercier mes deux encadrants, Marc Sanquer et Xavier Jehl. Quasiment tous les remerciements faits s'appliquent encore plus à eux. Ils étaient des encadrants de thèse parfaits, et si cette thèse est une réussite, c'est en

grande partie grâce à eux. Ils ont su me laisser une grande liberté dans le travail et dans l'évolution du sujet tout en me guidant avec de bons conseils et en soutenant fortement mon travail. Ils étaient toujours disponibles pour des questions et des coups de main. J'ai beaucoup apprécié nos longues discussions sur la physique mais aussi sur des aspects plus pratiques. Ils ont beaucoup clarifié mes idées et, plus j'avançais, plus j'appréciais leur clairvoyance et leurs connaissances très larges. C'était toujours un grand plaisir de travailler avec eux et j'ai beaucoup appris à leur côté.

Résumé

1 Introduction

Cette thèse décrit des mesures de transport électronique dans des MOSFETs silicium à base de nanofils gravés. Les mesures sont effectuées à très basse température où les transistors fonctionnent en régime mono-électronique.

2 Équation maîtresse et blocage de Coulomb

Nous considérons d'abord l'énergie électrostatique d'un ensemble de grains métalliques. Elle est décrite par des formules classiques mais nous considérons la charge sur chaque grain comme discrétisée en multiples de la charge élémentaire e. Le même formalisme peut être adopté dans le cas d'états électroniques localisés à différents endroits tant que leur extension est petite comparée à leur distance. C'est le modèle d'interaction constante. De manière plus générale il peut être adopté si l'interaction entre deux électrons ne dépend ni de la présence d'autres électrons ni des états quantiques dans lesquels se trouvent les électrons.

Après avoir calculé l'énergie statique, on peut calculer en théorie de perturbation le transport par effet tunnel entre les grains ou états localisés. Nous nous limitons au premier ordre. Pour cela il faut supposer que la conductance de la barrière tunnel est plus faible que le quantum de conductance $\frac{e^2}{h}$. Les taux tunnel définissent un système d'équations différentielles linéaires qui relient les probabilités statistiques de chaque configuration de charge. La solution stationnaire de ce système permet de calculer les courant moyens du système. Chaque équation correspondant à un état possible, la taille de ce système peut être très grande. Néanmoins elle peut souvent être réduite car seulement une partie des états est énergétiquement accessible. Un cas trivial de ce modèle est le transistor mono-électronique.

3 Les MOSFETs comme transistors mono-électroniques

Les dispositifs étudiés sont des nanofils silicium entourés d'oxyde et couverts dans leur partie centrale d'une grille en silicium poly-cristallin (voir figures 3.1, page 33

à 3.3, page 34). Dans les parties du fil proches de la grille le dopage est en dessous de la valeur critique de la transition métal–isolant: à basse température le fil est un isolant. Dans le reste du fil le dopage est au-dessus de la concentration critique. Une tension de grille positive fait ensuite passer le centre du segment isolant du côté métallique de la transition. Ainsi un îlot est formé sans qu'il n'y ait de barrière d'oxyde ou de constriction dans le nanofil. Ce modèle est confirmé d'un côté par l'observation de transport thermiquement activé (voir figure 3.8, page 41), de l'autre par l'observation d'oscillations de Coulomb très périodiques en fonction de la tension de grille (voir figures 3.7, page 40 et 3.10, page 42.

La période de ces oscillations est $\frac{e}{C_g}$ où e est la value absolue de la charge élémentaire et C_g la capacité de grille. Elle est en très bon accord avec la capacité géométrique calculée numériquement à partir des dimensions lithographiques (voir figure 3.11, page 44).

4 Spectre d'addition de l'îlot de blocage de Coulomb

Bien que petites, les fluctuations dans l'espacement des résonances de blocage de Coulomb sont bien mesurables. Traduite en énergie, leur amplitude est à un facteur 2 près égale à l'écartement entre niveaux à une particule calculé pour un gaz bidimensionnel de taille correspondante. Ces observations sont en bon accord avec des simulations numériques pour des îlots diffusifs [83]. Ces faibles fluctuations sont observées pour la plupart des échantillons à forte concentration d'électrons (pour plus que 50 électrons typiquement). A plus faible concentration les fluctuations peuvent devenir beaucoup plus grandes, de l'ordre de l'écart moyen entre résonances. Nous expliquons ceci par le fait qu'à faible concentration d'électrons l'îlot devient isolant et que les électrons sont localisés dans des minima du potentiel d'impureté. Le spectre d'addition dépend dans ce cas de la distance entre minima, qui fluctue fortement [67]. Cependant, dans la plupart des échantillons ce régime n'est pas observé car les barrières deviennent trop opaques aux faibles tensions de grille nécessaires pour atteindre ce régime.

Les fluctuations de l'espacement entre résonances ne sont pas toujours complètement chaotiques. Souvent nous observons un espacement bimodal où des espacements plus petits et plus grands sont alternés. Un tel effet est attendu à cause de la dégénérescence de spin [120]. Mais elle ne semble pas être à l'origine de la bimodalité observée car celle-ci n'est pas affectée par le champ magnétique (voir figure 4.8, page 61). De plus, l'amplitude de la bimodalité peut être beaucoup plus grande que les fluctuations en régime diffusif. Ceci exclut également la dégénérescence de bande (2 pour en 2D, 6 en 3D) du silicium comme origine de la bimodalité. L'explication suivante semble plus probable. Le champ électrique est plus fort sur les flancs du fil, dû à l'effet de coin. Par conséquent nous supposons que le gaz d'électrons se forme d'abord sur les flancs. Il s'agit ainsi de deux îlot quantiques couplés au lieu d'un seul, d'où l'effet bimodal.

5 Isolants désordonnés comme barrières de blocage de Coulomb

A très basse température les électrons ne franchissent les barrières ni par activation thermique ni par effet tunnel car, étant donné la géométrie et la hauteur des barrières, ces modes de transport donneraient des courants plusieurs ordres de grandeur plus faibles que ceux observés. Le transport se fait alors à travers des états localisés dans les barrières tels que les états des dopants.

Nous observons quelques anomalies dans les oscillations de Coulomb régulières (voir figures 5.1, page 66 et 5.2, page 67). Des simulations (voir figure 5.4, page 70) reproduisant les anomalies indiquent que ce sont en effet des états localisés dans les barrières qui sont à l'origine des anomalies. Quand la tension de grille est augmentée ces pièges à électrons sont poussés en dessous du niveau de Fermi et se peuplent. Une fois occupés ils décalent la signature de blocage de Coulomb dû à l'interaction électrostatique. Aux tensions de grille où ils sont proches du niveau de Fermi leur occupation dépend également de l'occupation de l'îlot, et l'occupation du piège oscille avec la tension de grille. C'est à ce point que des dents caractéristiques apparaissent dans les diamants de Coulomb (voir figures 5.2, page 67 pour l'expérience et 5.4, page 70 pour la simulation). A partir de cette signature il est possible de remonter à la matrice de capacité du piège et ainsi à sa position approximative.

Les pièges peuvent être observés plus directement en regardant leur dynamique. Avec quelques anomalies nous observons du bruit télégraphique. Le bruit télégraphique permet de déterminer le taux d'occupation moyen du piège. Nous observons en effet qu'il oscille en fonction de la tension de grille comme prédit par le modèle. Cependant, avec la plupart des anomalies nous n'observons pas de bruit télégraphique car leur dynamique est plus rapide que la bande passante de la mesure. Dans ce cas nous appliquons des signaux de fréquence variable sur la grille et observons jusqu'où le piège peut suivre. Il en ressort que la vitesse de relaxation de charge de la plupart des pièges est de l'ordre de la nano-seconde. Cet ordre de grandeur confirme qu'il s'agit bien de pièges dans les nanofils mêmes et non dans l'oxyde.

Sous champ magnétique les anomalies se déplacent. Ce déplacement est associé à l'effet Zeeman. Il se fait pratiquement dans tous les cas dans le sens de plus faible énergie ce qui indique que les pièges sont occupés par un seul électron, comme les états dopants.

Les pentes dans les diamants de Coulomb permettent d'accéder aux capacités de couplage entre îlot et réservoirs. Nous observons que la capacité associée à une barrière peut augmenter considérablement quand un piège se charge à l'intérieur de la barrière. De manière plus globale la capacité associée aux barrières croît considérablement avec la tension de grille. Nous expliquons cela par une augmentation de la constante diélectrique dans les barrières. Une telle augmentation à été observée dans du silicium dopé proche de la transition métal-isolant [19]. Dans notre système la tension de grille attire des électrons dans les barrières et les approche ainsi de la transition. L'augmentation de la constante diélectrique a été paramétrisée

par la longueur de localisation [50]. Ce paramètre détermine aussi la conductance de la barrière, et à travers ce paramètre nous observons en effet un bon accord entre constante diélectrique moyenne et conductance moyenne (voir figures 5.20, page 96 et 5.21, page 96).

6 Conclusions

Nous avons mesuré des transistors monoélectroniques très stables et compatibles avec des procédés de la microélectronique. Leurs propriétés sont contrôlées par leur géométrie. Cela et le fait qu'ils ne nécessitent qu'une seule grille les rend très intéressants pour des applications comme la pompe à électrons, une source de courant quantique.

Avec ces transistors nous avons étudié des dopants uniques dans du silicium, qui ont des propriétés très intéressantes pour former des bits quantiques [60].

Chapter 1

Introduction

1.1 Context

The invention of the bipolar transistor in 1947 [6, 15, 107] was the starting point of semiconductor electronics. In the 1960s the first field effect transistors were realized and soon replaced the bipolar transistors in logic circuits. Since then the size of logic devices, in particular the gate length, has decreased by approximately a factor 1000. This shrinking is motivated by the reduction of switching time, energy consumption and cost per transistor it brings about. It allowed to increase constantly complexity and speed while reducing the cost per processor. In 1965 Gordon Moore [88] found the increase in complexity to be exponential in time, the number of transistors in a processor doubling every 18 months. "Moore's law" has been respected since. Yet it is not a mere observation any more; it has become a guideline permitting to synchronize development in different areas of device fabrication and architecture. It has even been institutionalized with the ITRS (International Technology Roadmap for Semiconductors [135]) and the semiconductor industry tries hard to keep pace with its prediction (see for example [134]). Silicon is now one of the best known materials and the continuous efforts have pushed silicon technology to a high level of perfection. In particular the interface between silicon and its oxide is now virtually defect free. Although in recent transistors used in logic circuits the gate length is 30 nm and the gate oxide thickness 1.2 nm, the transistors are described by the same classical laws as the early micrometer-sized transistors. Quantum effects such as tunneling together with electrostatic effects due to the short gate length cause the so-called short-channel effects, degrading the on/off ratio of the drain–source current. One tries to master them by increasing the gate control by using stronger gate dielectrics and by separating the transistors into several wires being controlled by the gate electrode from 3 or even 4 sides [135].

Despite the maturity of silicon technology the single-electron transistor (SET) was first realized in aluminum in 1987 [30]. Only two years later Coulomb blockade was observed in silicon nanowires [103] and GaAs heterostructures [82]. Coulomb blockade was already fairly well understood in 1992 [35]. Rather than subject of research, SETs are often used more as a tool, in particular as a very sensitive charge

detector allowing almost ideal quantum measurements [21]. In this context, SETs have been used to read the charge in a single-electron/Cooper-pair box [71], as tip in near field microscopy [130] or optimized for speed by driving them with a radio-frequency signal (the so-called RF-SET) [101]. This allowed for example to measure current by counting electrons [75, 17, 38]. Similar devices consisting of several Coulomb blockade islands in series, the so-called electron pump, allow to put a certain number of electrons onto a capacitor and thereby build a new capacitance standard [62].

Coulomb blockade in semiconductor quantum dots, where the density of states is lower, is still of interest in itself. This is for one part because the strength of the tunnel barriers can be controlled. For example, a single quantum dot can be continuously transformed into a double quantum dot by pinching it off in the middle [124]. The most interesting point about semiconductor quantum dots is that their density of states is low, in particular for the first electrons put onto the dot. This permits to have the single-particle level spacing higher than the energy equivalent of temperature $k_{\mathrm{B}}T$ and to observe individual quantum mechanical states [116] and phenomena due to phase coherence. Research is particularly active on the Kondo effect [33]. Recently, measurements of quantum coherence are more and more motivated by quantum computing where the quantum bits (or qubits) are quantum mechanical superpositions of "0" and "1" [111]. Such qubits have been realized in superconducting circuits, for example Josephson junction qubits [112] where the bit is coded in the phase difference across the junction or the "quantronium" [122]. Other realizations of solid state qubits are charge qubits where the bit is coded by the position of an electron, for example in a double dot [34]. The most appealing qubits are electron or nuclear spins because they are intrinsic 2-level systems [74]. In this context the relaxation of single electron spins was measured [42, 41, 4]. The observed time constants are typically in the millisecond range. When the singlet–triplet relaxation of 2-electron systems is considered, the time constants are much shorter [99, 65]. These short time constants are due to the host material GaAs used in these experiments. It has strong spin-orbit coupling so that the electron spins are coupled to the environmental degrees of freedom. In addition the nuclear spin is $\frac{3}{2}$ in Ga and As so that the electron spin is coupled through hyperfine interaction to a huge number of degrees of freedom, even if only one orbital state is accessible.

Kane [60, 61] proposed nuclear spins of donors in silicon as quantum bits. In his proposal a dopant charged with two electrons is close to a Si-SiO_2 interface. When the gate voltage is accurately chosen, the spin singlet state is degenerate with the state where one electron is in the trap and the other at the interface so that tunneling back and forth between these states occurs. The spin of dopant states in silicon can have very long relaxation times because spin–orbit interaction is weak and the most abundant isotope $^{28}_{14}$Si has nuclear spin 0 (see appendix C). In purified $^{28}_{14}$Si, Feher *et al.* [24, 25, 127] measured electron and nuclear spin relaxation times of phosphorous dopants exceeding 1 hour at 1.25 K. It has recently become possible to place dopants individually in silicon, [102, 105] but to our knowledge there are no realizations of the Kane proposal so far. There are, however, experiments where the spin state of an single dopant in a quantum dot is read optically [77] or the spin

of individual charge traps near a MOSFET is revealed by their random telegraph signal [129].

Almost all experiments on semiconductor quantum dots cited above were made in 2-dimensional electron gases (2DEGs) in GaAs/AlGaAs heterostructures, shaped by metallic depletion gates. In these structures the dopants are in a different layer than the actual electron gas. They create a very smooth potential in the 2DEG and allow for long elastic mean free paths and coherence lengths so that the quantum dots are regular even for large sizes exceeding 100 nm. Most silicon devices are etched and doped by ion implantation. Both processes create strong random potentials. Therefore transport in silicon devices at low temperature is often dominated by disorder. Without doping and etching the mobilities are much higher [103], but do not reach the values of GaAs, because the 2DEG is at the interface with the amorphous SiO_2 and the density of states is much higher. Therefore excited states [108, 96] or coherent phenomena [34] are only rarely observed in silicon devices.

Coulomb blockade in silicon is mostly studied in views of memory cells or logic devices replacing the traditional transistor [117, 91]. In consequence, focus is on high temperature operability and compatibility with the existing silicon technology [37, 29, 63, 114]. Most of the time the dot is claimed to be defined by lithography. But disorder is always present. It can be used to form very small dots. In very thin wires, disorder easily produces random constrictions and small quantum dots are formed [53, 118, 52]. In very large but short GaAs [69] and silicon [98] field effect transistors small quantum dots created by disorder have also been observed at very low temperature.

Another approach to form small structures is not to tailor them from a substrate but to grow them. Carbon nanotubes are the most famous example [55] but also in grown silicon nanowires Coulomb blockade with clearly resolved excited states has recently been observed [131]. Such structures are difficult to contact. Therefore it is tried to directly grow them between the contacts.

1.2 Outline

In this work we present a very simple single-electron transistor design consisting of a doped silicon nanowire without constrictions or oxide barriers. The quantum dot is defined by a gate electrode covering the center of the wire. Its Coulomb blockade properties are perfectly controlled by the fabrication process. First we study Coulomb blockade in these devices. Then we use the SET to study in more detail the most peculiar part of the devices, the Coulomb blockade barriers formed by small disordered insulating parts of the wire.

In chapter 2 the rate-equation approach to tunneling through networks of metallic grains or localized states is presented. It is used for the simulations made in this work. This chapter also gives a short overview of the single-electron transistor (SET); it is not intended as an introduction, though.

The devices we measure are presented in chapter 3. After considering their behavior at room temperature we explain how these devices become single-electron transistors at low temperature. We show that contrarily to most silicon nanowire SETs, the size of the quantum dot is defined by lithography and not by disorder. [49]

Having shown that Coulomb blockade is well controlled in our devices, in chapter 4 we concentrate on the statistics of the Coulomb blockade peak spacing. At high electron densities, when the dot is diffusive, the fluctuations of the peak spacing are related to the single-particle level spacing and thus are weak. At low electron density, when the dot is in an insulating regime, the fluctuations are of the order of the mean Coulomb-blockade peak spacing. [14, 56]

In our devices the barriers separating dot and leads are not simple tunnel barriers or constrictions as supposed in most descriptions of Coulomb blockade. In chapter 5, we argue that electron transport between dot and leads occurs by resonant tunneling through well connected dopant states. Thus the Thouless energy of the barriers is very low. There are even dopant sites with well quantized occupation number. When they become occupied they cause characteristic anomalies in the Coulomb blockade spectrum which allow to draw conclusions on the capacitance matrix of the dopant sites. By applying a magnetic field or pulse signals the ground state spin and the dynamics of the dopant states are measured. [48]

In a second part of this chapter we study the behavior of the barriers as a whole. Analyzing the Coulomb blockade diamonds, we observe a strong increase of the capacitance of the barriers when the dopants become occupied with increasing gate voltage.

The details on the methods adopted in order to obtain all these results are described as appendix, the experimental setup in appendix A and the data analysis in appendix B.

Chapter 2

Rate equations and Coulomb blockade

We are interested not only in single-electron transistors (SETs) but also in the interaction of the quantum dot with localized states. Therefore we not only describe Coulomb blockade in a single metallic grain but develop a basic phenomenological description of transport through an array of metallic dots and localized electronic states with interaction. It will serve for numerical calculations. First the dots are considered as isolated i.e. we do not allow electron tunneling and calculate the electrostatic energy of charged metallic dots. We will then see that this description can be maintained if some of the metallic dots are replaced by localized electron states. In a second step we will consider tunneling between the reservoirs, the dots and localized states in the limit of weak coupling and sequential tunneling where only one electron tunnels at a time and calculate the current through the system in function of the applied voltages.

2.1 Electrostatic energy of interacting metallic islands

We call a quantum dot metallic if the electronic states spread over the whole volume of the dot and if the spacing between excitations Δ_1 is much smaller than $k_{\mathrm{B}}T$. The excitation spectrum can then be considered as continuous. This is not necessarily true for the so called addition spectrum, i.e. the spectrum of excitations that consist in adding electrons to the dot. For not too small dots, this addition spectrum is dominated by electron-electron interactions. If the Fermi wave length in the dot is small compared to its size or the mean free path (whichever is smaller), we can suppose in good approximation that the total electron-electron interaction energy does not depend on the exact quantum mechanical states the electrons are in, but

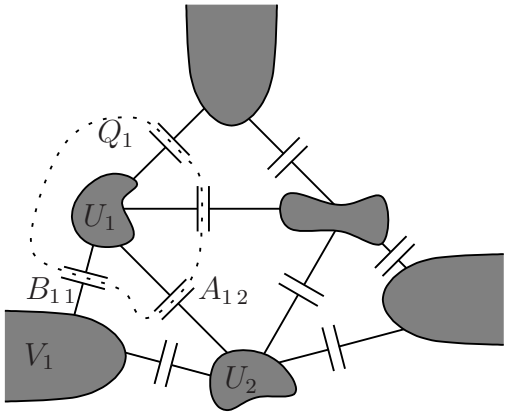

Figure 2.1: Capacitance network between metallic grains.

is proportional to n^2 where n is the number of electrons in the dot. This is the constant interaction model. We obtain the classical electrostatic energy of the dot:

$$W_n = \frac{Q^2}{2C} = \frac{e^2 n^2}{2C} \tag{2.1}$$

In this approximation, the addition energy, i.e. the spacing of the addition spectrum is $\Delta_+ = (W_{n+1} - W_n) - (W_n - W_{n-1}) = \frac{e^2}{C}$. For sufficiently big dots this energy is much higher than the excitation spacing Δ_1 at fixed electron number. This allows to calculate first the classical interaction energy by considering the electrons as free inside the metallic grains and consider only in a second time the excitation spectrum for fixed electron number.

Let us now consider an array of metallic dots as shown in figure 2.1. The interaction between dots i and j can be described by a coupling capacitance $A_{ij} = A_{ji}$. In the same way the interaction between dot i and lead k is described by the capacitance B_{ik}. We call V_k the voltage in lead k. First we calculate the potential U_i of each dot i, given its charge Q_i. This charge Q_i is the sum of the charges on the coupling capacitors of dot i (see figure 2.1).

$$Q_i = \sum_{j \neq i} A_{ij}(U_i - U_j) + \sum_k B_{ik}(U_i - V_k) \tag{2.2}$$

We can write this in matrix form

$$\boldsymbol{Q} + \hat{B}\boldsymbol{V} = \hat{C}\boldsymbol{U} \tag{2.3}$$

where the symmetric matrix $\hat{C}$ is given by

$$C_{ij} = \begin{cases} \sum_{l \neq i} A_{il} + \sum_k B_{ik} & (i = j) \\ -A_{ij} & (i \neq j) \end{cases} \tag{2.4}$$

The electrostatic energy can be calculated by integrating the potential U from the origin, i.e. charge 0 on all dots, to $\boldsymbol{Q}$ along an arbitrary path in the space of

charges; we take a linear one. As we will only be interested in relative values of W as the charges on the dots change we can choose an addition constant.

$$W(\boldsymbol{Q}) + \text{const} = \int_{\boldsymbol{0}}^{\boldsymbol{Q}} \langle \boldsymbol{U}(\boldsymbol{q}) | \mathrm{d}\boldsymbol{q} \rangle = \int_0^1 \langle \hat{C}^{-1}(\lambda \boldsymbol{Q} + \hat{B}\boldsymbol{V}) | \boldsymbol{Q} \mathrm{d}\lambda \rangle \tag{2.5}$$

$$= \frac{1}{2} \langle \boldsymbol{Q} + \hat{B}\boldsymbol{V} | \hat{C}^{-1} | \boldsymbol{Q} + \hat{B}\boldsymbol{V} \rangle - \frac{1}{2} \langle \hat{B}\boldsymbol{V} | \hat{C}^{-1} | \hat{B}\boldsymbol{V} \rangle \tag{2.6}$$

Here $\langle \cdot | \cdot \rangle$ notes the scalar product between classical vectors. We choose the constant such that it cancels the second term of expression (2.6) and define

$$W = \frac{1}{2} \langle \boldsymbol{Q} + \hat{B}\boldsymbol{V} | \hat{C}^{-1} | \boldsymbol{Q} + \hat{B}\boldsymbol{V} \rangle \tag{2.7}$$

Other representations of W are

$$W = \frac{1}{2} \langle \boldsymbol{U} | \hat{C} | \boldsymbol{U} \rangle \tag{2.8}$$

$$= \frac{1}{2} \sum_{i<j} A_{ij} (U_i - U_j)^2 + \frac{1}{2} \sum_{i,k} B_{ik} U_i^2 \tag{2.9}$$

$$= \frac{1}{2} \sum_{i<j} A_{ij} (U_i - U_j)^2 \tag{2.10}$$

$$+ \frac{1}{2} \sum_{i,k} B_{ik} (U_i - V_k)^2 + \sum_{i,k} B_{ik} (U_i - V_k) V_k + \frac{1}{2} \sum_{i,k} B_{ik} V_k^2$$

The representation (2.8) is the equivalent of (2.7) expressed as function of potential instead of charge. For (2.9) we used the definition of $\hat{C}$. The physical meaning of the second term of expr. (2.9) is difficult to see in this form. It is split in expr. (2.10) in order to explain its physical meaning. The first term of expr. (2.10) is the charging energy of the interdot capacitors. The second term the charging energy of the capacitors coupling the dots to the leads. The third term is the energy of the electrons of the exterior plates of the dot-lead capacitors or the work delivered by the voltage sources as the dots are charged. The last term is not important as it does not depend on U_i or Q_i.

We will introduce an additional energy offset ϵ_i for each dot to account for static disorder. The energy writes then

$$W = \frac{e^2}{2} \langle \boldsymbol{n} - \boldsymbol{x} | \hat{C}^{-1} | \boldsymbol{n} - \boldsymbol{x} \rangle + \langle \boldsymbol{\epsilon} | \boldsymbol{n} \rangle \tag{2.11}$$

where the occupation vector $\boldsymbol{n}$ and $\boldsymbol{x}$ are defined by $-e\boldsymbol{n} = \boldsymbol{Q}$ and $e\boldsymbol{x} = \hat{B}\boldsymbol{V}$.

2.2 Discrete levels

We will now make the opposite assumption, namely that the electron states are confined into very small volumes, i.e. that the electrons can be treated as point-like

classical charges. The interaction between them is then

$$W_{\text{discrete}} = \frac{e^2}{2}\langle \boldsymbol{n}|\hat{D}|\boldsymbol{n}\rangle \quad \text{with} \quad D_{ij} = \frac{1}{4\pi\epsilon_0\epsilon_{\text{host}} r_{ij}}, (i \neq j) \quad \text{and} \quad D_{ii} = 0 \qquad (2.12)$$

ϵ_{host} is the dielectric constant of the host material and $r_{ij} = r_{ji}$ the distance between sites i and j. The factor $\frac{1}{2}$ accounts for the fact that due to this symmetry all terms are counted twice. We get again a constant interaction model. Indeed, models (2.11) and (2.12) can be combined and one obtains an interaction matrix of the form

$$\hat{I} = \begin{pmatrix} \hat{C}^{-1} & \hat{C}^{-1}\hat{J}^{\mathrm{T}} \\ \hat{J}\hat{C}^{-1} & \hat{D} \end{pmatrix} \qquad (2.13)$$

which replaces C^{-1} in expression (2.11). $J_{ij} = \frac{\partial U_i}{\partial U_j}$ is the lever-arm factor of the metallic grain j at site i or the absolute value of the mirror charge in grain j (at fixed potential) caused by a unitary charge at i. Now the interaction between localized states $\hat{D}$ is more complex than without metallic grains because it has to include not only the localized charges themselves, but also their virtual images, mirrored at the surface of the metallic grains and electrodes. In particular there is now a negative self-interaction, i.e. negative diagonal elements in $\hat{D}$, due to the attractive interaction between the charge on dot i and its mirror charges.

What happens when the localized states cannot be considered as point-like? The electron-electron interactions not only create offsets to the potential wells into which the electrons are confined but also deform them. When the spatial extensions of the states become comparable to the distances between them, these deformations change the wave functions and in consequence the interactions. Thus the interaction between two states depends on the occupation of the others; The constant interaction description (2.12) is not a good approximation any more. In other words, in order to have constant interactions, the charge distribution and thus the electronic states are required to be "rigid".

Despite being formally very similar, the capacitance description of the last section supposes exactly the opposite: the electrons are free inside the grains and so fully screen the electric field of the other grains by forming surface charges. The charge distributions are supposed to be "soft".

For localized states in different sites with a localization radius comparable to the distance between sites, model (2.12) of rigid quantum mechanical states is as crude an approximation as the capacitance description (2.11) supposing soft states. We will therefore employ the much simpler capacitance model (2.11) also for localized states. The most important error is that the capacitance model has a repulsive self-interaction not present in model (2.12). It can be compensated: as on a localized site i the occupation number n_i can only take the values 0 and 1, $n_i^2 = n_i$. Thus we can cancel the self-interaction $\frac{e^2}{2}\hat{C}_{ii}^{-1} n_i^2$ by adjusting the energy offset ϵ_i. The interaction energies are then calculated in the same way for localized states and metallic grains.

This model of localized states described in terms of capacitance is also useful in order to describe a grain where $\Delta_1 > k_B T$. We consider now states confined into one and the same grain, so they overlap and their coupling capacitance will be very high. In the limit where the inter-state coupling capacitance A_{ij} tends to infinity, the elements of the inverse total capacitance C^{-1} tend to $(C^{-1})_{ij} = (m \sum_k B_k)^{-1}$. m is the size of the matrix C^{-1} (or the number of states) and we supposed that the state–lead coupling capacitances B_{ik} do not depend on the state i anymore. This means that expression (2.11) becomes

$$W = \frac{e^2 n^2}{2C} + \sum_i \epsilon_i n_i \tag{2.14}$$

with $n = \sum_i n_i$. The total capacitance $C = \sum_k B_k$ is now a scalar. In this limit where the localized states overlap, the constant interaction model is not a good approximation. This description is nevertheless useful to phenomenologically describe some aspects of the excitation spectrum. In reality, the correlations beyond constant interactions make that the second term has to be replaced by a more complex many-particle excitation spectrum at fixed total particle number which is different for each n (scrambling).

The energy W is now formally the same for a wide range of physical systems: spatially separated localized states, multiple states of a single quantum dot, metallic quantum dots and any combination of them. But there is a fundamental difference. In the case of localized states W describes the exact energy of the state. For a metallic grain, however, it gives only the ground state energy and in addition there is a continuous excitation spectrum. We will account for this difference when calculating the tunneling rates.

2.3 Tunneling

We now allow electrons to tunnel between the dots. The tunneling coupling is considered as a small perturbation to the Hamiltonian of the isolated dots. This condition is fulfilled if

$$4\pi^2 \rho_\mathrm{i} \rho_\mathrm{f} |M|^2 \ll 1 \tag{2.15}$$

where M is the matrix element of the tunneling Hamiltonian and ρ_i and ρ_f the densities of states[1] in the dot the electron comes from and goes to. For simplicity we suppose them to be constant. We define

$$G_\mathrm{T} = \frac{e^2}{h} 4\pi^2 \rho_\mathrm{i} \rho_\mathrm{f} |M|^2 \tag{2.16}$$

Condition (2.15) reads then

$$G_\mathrm{T} \ll \frac{e^2}{h} \tag{2.17}$$

In the opposite case $G_\mathrm{T} \gg \frac{e^2}{h}$ the levels in the two dots hybridize and the two dots have to be considered as one.

[1] more precisely the mean density Δ_1 of the excitation spectrum at fixed electron number

2.3.1 Tunneling between dots

We will now calculate the tunneling between dots. We have to distinguish 3 cases.

- Tunneling between 2 continuous spectra (metallic dot or lead)
- Tunneling between a continuous spectrum and a discrete level
- Tunneling between 2 discrete levels

Between two continuous spectra the tunneling rate is given by (golden rule):

$$\begin{aligned}\Gamma_{\mathrm{cc}} &= \frac{2\pi}{\hbar}|M|^2 \iint \mathrm{d}\epsilon_{\mathrm{i}}\,\mathrm{d}\epsilon_{\mathrm{f}}\;\delta(\epsilon_{\mathrm{i}}-\epsilon_{\mathrm{f}})\;\rho_{\mathrm{i}}\;\rho_{\mathrm{f}}\;f\left(\frac{\epsilon_{\mathrm{i}}-W(\boldsymbol{n}_{\mathrm{i}})}{k_{\mathrm{B}}T}\right)\left(1-f\left(\frac{\epsilon_{\mathrm{f}}-W(\boldsymbol{n}_{\mathrm{f}})}{k_{\mathrm{B}}T}\right)\right)\\ &= \frac{k_{\mathrm{B}}T}{e^2}G_{\mathrm{T}}f^*\left(\frac{W(\boldsymbol{n}_{\mathrm{f}})-W(\boldsymbol{n}_{\mathrm{i}})}{k_{\mathrm{B}}T}\right)\end{aligned} \tag{2.18}$$

f is the distribution function of kinetic energies in the two dots. We suppose f to be the Fermi-Dirac distribution $f(x)=\frac{1}{\mathrm{e}^x+1}$. $f^*(x)=(f*f)(x)=\frac{x}{\mathrm{e}^x-1}$ is its auto-convolution. Using the Fermi-Dirac distribution supposes that energy relaxation inside the dots is much faster than tunneling: when an electron tunnels we suppose that both dots are at equilibrium, i.e. that the energy transferred by the electrons having tunneled before has been relaxed to the environment. If this assumption is not fulfilled the coupling to the environment has to be taken into account explicitly. This can be done by replacing $\delta(\epsilon_{\mathrm{i}}-\epsilon_{\mathrm{f}})$ in (2.18) by a function $P_{\mathrm{env}}(\epsilon_{\mathrm{i}}-\epsilon_{\mathrm{f}})$ that gives the probability for the tunneling electron to emit an energy $\epsilon_{\mathrm{i}}-\epsilon_{\mathrm{f}}$ to the environment or to absorb the energy $\epsilon_{\mathrm{f}}-\epsilon_{\mathrm{i}}$ from the environment if $\epsilon_{\mathrm{i}}-\epsilon_{\mathrm{f}}<0$. This function is discussed in [51].

If the tunneling electron goes from a continuous spectrum to a discrete level the energy must match exactly the new level and the rate is given by

$$\begin{aligned}\Gamma_{\mathrm{cd}} &= \frac{2\pi}{\hbar}|M|^2 \iint \mathrm{d}\epsilon_{\mathrm{i}}\,\mathrm{d}\epsilon_{\mathrm{f}}\;\delta(\epsilon_{\mathrm{i}}-\epsilon_{\mathrm{f}})\;\rho_{\mathrm{i}}\;\delta(\epsilon_{\mathrm{f}}-W(\boldsymbol{n}_{\mathrm{f}}))\;f\left(\frac{\epsilon_{\mathrm{i}}-W(\boldsymbol{n}_{\mathrm{i}})}{k_{\mathrm{B}}T}\right)(d_{\mathrm{f}}-n_{\mathrm{f}})\\ &= \frac{I_{\mathrm{T}}}{e}f\left(\frac{W(\boldsymbol{n}_{\mathrm{f}})-W(\boldsymbol{n}_{\mathrm{i}})}{k_{\mathrm{B}}T}\right)(d_{\mathrm{f}}-n_{\mathrm{f}})\end{aligned} \tag{2.19}$$

The prefactor I_{T} is defined by $\frac{I_{\mathrm{T}}}{e}=\frac{2\pi}{\hbar}|M|^2\rho_{\mathrm{i}}$ and has the dimensions of a current. d_{f} is the degeneracy of the discrete level. Γ_{cd} is proportional to f whereas Γ_{cc} is proportional to $f*f$.

In the reverse case where the electron tunnels from a discrete level to a continuum the rate is

$$\begin{aligned}\Gamma_{\mathrm{dc}} &= \frac{2\pi}{\hbar}|M|^2 \iint \mathrm{d}\epsilon_{\mathrm{i}}\,\mathrm{d}\epsilon_{\mathrm{f}}\;\delta(\epsilon_{\mathrm{i}}-\epsilon_{\mathrm{f}})\;\delta(\epsilon_{\mathrm{i}}-W(\boldsymbol{n}_{\mathrm{i}}))\;\rho_{\mathrm{f}}\;n_{\mathrm{i}}\left(1-f\left(\frac{\epsilon_{\mathrm{f}}-W(\boldsymbol{n}_{\mathrm{f}})}{k_{\mathrm{B}}T}\right)\right)\\ &= \frac{I_{\mathrm{T}}}{e}\;f\left(\frac{W(\boldsymbol{n}_{\mathrm{f}})-W(\boldsymbol{n}_{\mathrm{i}})}{k_{\mathrm{B}}T}\right)n_{\mathrm{i}}\end{aligned} \tag{2.20}$$

The prefactor I_T is now defined by $\frac{I_\mathrm{T}}{e} = \frac{2\pi}{\hbar}|M|^2\rho_\mathrm{f}$.

At lowest order in the tunneling Hamiltonian the tunneling between two discrete levels is proportional to a delta function:

$$\begin{aligned}\Gamma_\mathrm{dd} &= \frac{2\pi}{\hbar}|M|^2 \iint \mathrm{d}\epsilon_\mathrm{i}\,\mathrm{d}\epsilon_\mathrm{f}\;\delta(\epsilon_\mathrm{i}-\epsilon_\mathrm{f})\;\delta(\epsilon_\mathrm{i}-W(\boldsymbol{n}_\mathrm{i}))\;\delta(\epsilon_\mathrm{f}-W(\boldsymbol{n}_\mathrm{f}))\;n_\mathrm{i}\;(d_\mathrm{f}-n_\mathrm{f}) \\ &= \frac{2\pi|M|^2}{\hbar}\;\delta(W(\boldsymbol{n}_\mathrm{f})-W(\boldsymbol{n}_\mathrm{i}))\end{aligned} \tag{2.21}$$

This tunneling rate does not depend on temperature. To get its energy dependence either the electromagnetic environment [51, 27] has to be taken into account or one has to take into account the life-time broadening (Lorentzian) of the energy levels.

2.3.2 Tunneling between dots and leads

Tunneling rates from dots to leads or from leads to dots can also be described by expressions (2.18), (2.19) and (2.20). However we have to account for the chemical potential $-eV_k$ in lead k. This is done by replacing $W(\boldsymbol{n_\mathrm{f}}) - W(\boldsymbol{n}_\mathrm{i})$ with $W(\boldsymbol{n}_\mathrm{f}) - W(\boldsymbol{n}_\mathrm{i}) + eV_k$ if the electron tunnels from lead k to a dot or by $W(\boldsymbol{n}_\mathrm{f}) - W(\boldsymbol{n}_\mathrm{i}) - eV_k$ if the electron tunnels form a dot to lead k.

2.4 Rate equations and currents

The state of the system without tunneling is described by the vector $\boldsymbol{n}$ giving the occupation number of each dot or localized state. For single electronic states this is straightforward. For metallic dots expression (2.11) gives only the ground state energy for a given occupation vector $\boldsymbol{n}$. But we have supposed the continuous excitations at fixed electron number to be in thermal equilibrium and we fully accounted for them when we calculated the transition rates. Thus, the state of the dot array is described by the occupation vector $\boldsymbol{n}$ alone.

Let $P_{\boldsymbol{n}}$ be the statistical probability to be in state $\boldsymbol{n}$. This probability obeys the following rate equation:

$$\frac{\partial P_{\boldsymbol{n}}}{\partial t} = \sum_{\boldsymbol{m}\neq\boldsymbol{n}} (\Gamma_{\boldsymbol{n}\leftarrow\boldsymbol{m}} P_{\boldsymbol{m}} - \Gamma_{\boldsymbol{m}\leftarrow\boldsymbol{n}} P_{\boldsymbol{n}}) \tag{2.22}$$

$\Gamma_{\boldsymbol{n}\leftarrow\boldsymbol{m}}$ is the tunneling rate from state $\boldsymbol{m}$ to $\boldsymbol{n}$ and it is given by

$$\Gamma_{\boldsymbol{n}\leftarrow\boldsymbol{m}} = \begin{cases} \Gamma^0_{ij} & (\boldsymbol{n} = \boldsymbol{m} + \mathbf{1}_i - \mathbf{1}_j) \\ \sum_k \Gamma^+_{ik} & (\boldsymbol{n} = \boldsymbol{m} + \mathbf{1}_i) \\ \sum_k \Gamma^-_{ik} & (\boldsymbol{n} = \boldsymbol{m} - \mathbf{1}_i) \\ 0 & \text{otherwise} \end{cases} \tag{2.23}$$

In the vector $\mathbf{1}_i$ all elements are 0, only the i-th is 1. Γ^0_{ij} is the tunneling rate from dot j to dot i. Γ^+_{ik} is the transition rate from lead k to dot i and Γ^-_{ik} the tunneling rate from dot i to lead k. These rates are given by one of the expressions (2.18) ... (2.21), depending on whether the initial and final dot (or state or lead) have a continuous or a discrete spectrum. The rate equation formalism is valid if phase coherence is lost after each tunneling event. If one of the dots involved in the tunneling process is metallic, this condition is fulfilled for $k_\mathrm{B}T \gg \hbar\Gamma$. In the case of tunneling between discrete levels one has to take into account the coupling with the environment (see above).

Eq. (2.22) can be written in matrix form

$$\frac{\partial \boldsymbol{P}}{\partial t} = \Gamma \boldsymbol{P} \tag{2.24}$$

where Γ is defined by

$$\Gamma_{\boldsymbol{nm}} = \begin{cases} \Gamma_{\boldsymbol{n}\leftarrow\boldsymbol{m}} & (\boldsymbol{n} \neq \boldsymbol{m}) \\ -\sum_{\boldsymbol{k}} \Gamma_{\boldsymbol{n}\leftarrow\boldsymbol{k}} & (\boldsymbol{n} = \boldsymbol{m}) \end{cases} \tag{2.25}$$

The size s of vector $\boldsymbol{P}$ is not the same as the size of the state vector $\boldsymbol{n}$, i.e. the number of dots, because each element of P corresponds to a specific value of $\boldsymbol{n}$ and therefore $s = \prod_i (d_i + 1)$ where d_i is the degeneracy in case of a discrete level. In the case of a metallic dot d_i is the maximum number of electrons we allow in the dot to get a finite system. s increases exponentially with the number of dots.

As we are interested in the low frequency properties of the system we are searching for stationary solutions of (2.24):

$$\Gamma \boldsymbol{P} = 0 \tag{2.26}$$

This transforms the differential equation system (2.24) into a linear homogeneous equation system. The matrix Γ has rank $s-1$ (unless if, for at least one i, all the Γ^0_{ij}, Γ^+_{ik}, Γ^-_{ik} are zero). Γ cannot have full rank because $\sum_{\boldsymbol{m}} \Gamma_{\boldsymbol{nm}} = 0$ for all $\boldsymbol{n}$. The equation has therefore non-trivial solutions and we can impose the normalization

$$\sum_{\boldsymbol{n}} P_{\boldsymbol{n}} = 1 \tag{2.27}$$

Once the stationary probability vector $\boldsymbol{P}$ is known, the current from dot j to dot i is

$$I_{i\leftarrow j} = -e \sum_{\boldsymbol{n}=\boldsymbol{m}+\mathbf{1}_i-\mathbf{1}_j} (\Gamma^0_{ij} P_{\boldsymbol{m}} - \Gamma^0_{ji} P_{\boldsymbol{n}}) \tag{2.28}$$

and current entering dot i from lead k is

$$I^{\leftarrow}_{ik} = -e \sum_{\boldsymbol{n}=\boldsymbol{m}+\mathbf{1}_i} (\Gamma^+_{ik} P_{\boldsymbol{m}} - \Gamma^-_{ik} P_{\boldsymbol{n}}) \tag{2.29}$$

This alorithm for calculating the current through a system of dots and states is used for all numerical calculations of electron transport in this work. The number of

possible many-particle-states increases exponentially with the number of dots/single particle states, and thus only small systems can be calculated in this way. However, at low bias voltage (i.e. if all electrodes through which electrons can enter or leave the system are within a potential range which is small compared to $k_{\mathrm{B}}T$) it is possible to define a free energy. In that case only the state with the lowest free energy and states with a free energy at most a few[2] $k_{\mathrm{B}}T$ higher have to be considered.

2.5 Single electron transistor

A simple case of the model described above is the single electron transistor (SET). The devices studied in these work fall in this category. Nevertheless, as SETs are described in detail in a large number of papers and books on Coulomb blockade (for example [8, 35]), we will make here only the connection to the model described above and summarize the main results needed later in this work.

A SET consists of a dot connected to two electrodes (source and drain) via tunnel junctions. A third electrode, the gate, allows to control the dot potential. No current flows between gate and dot. We consider at first the metallic limit $\Delta_1 \ll k_{\mathrm{B}}T$.

The capacitance matrix defined in (2.4) becomes a scalar $C = C_{\mathrm{g}} + C_{\mathrm{s}} + C_{\mathrm{d}}$ and the electrostatic energy from equation (2.11) simplifies to

$$W_n = \frac{e^2}{2C}\,(n-x)^2 \tag{2.30}$$

with $x = (C_{\mathrm{g}}V_{\mathrm{g}} + C_{\mathrm{s}}V_{\mathrm{s}} + C_{\mathrm{d}}V_{\mathrm{d}})/e$.

We can split this energy in effective one particle levels w_m^* which are filled successively:

$$W_n = W_0 + \sum_{m=1}^{n} w_m^* \tag{2.31}$$

with

$$w_n^* = W_n - W_{n-1} = \frac{e^2}{C}\,(n - x - \tfrac{1}{2}) \tag{2.32}$$

These levels are spaced by $\Delta_+ = \frac{e^2}{C}$, the so called charging energy. Current can pass through the device only if one of the effective one particle levels is in the bias window, i.e. between $-eV_{\mathrm{s}}$ and $-eV_{\mathrm{d}}$. This condition results in the so called Coulomb blockade diamond structure (see figure 2.2(a)). In accordance with the conventions in microelectronics we set $V_{\mathrm{s}} = 0$ throughout the entire work, for calculations as well as measurements. Thus, drain and bias voltage are identical. In this convention the Coulomb blockade diamonds are delimited by lines with slopes

$$\left.\frac{\partial V_{\mathrm{g}}}{\partial V_{\mathrm{d}}}\right.^{+} = 1 + \frac{C_{\mathrm{s}}}{C_{\mathrm{g}}} \tag{2.33}$$

[2] $10\,k_{\mathrm{B}}T$ gives already a very good accuracy of order 10^{-4}.

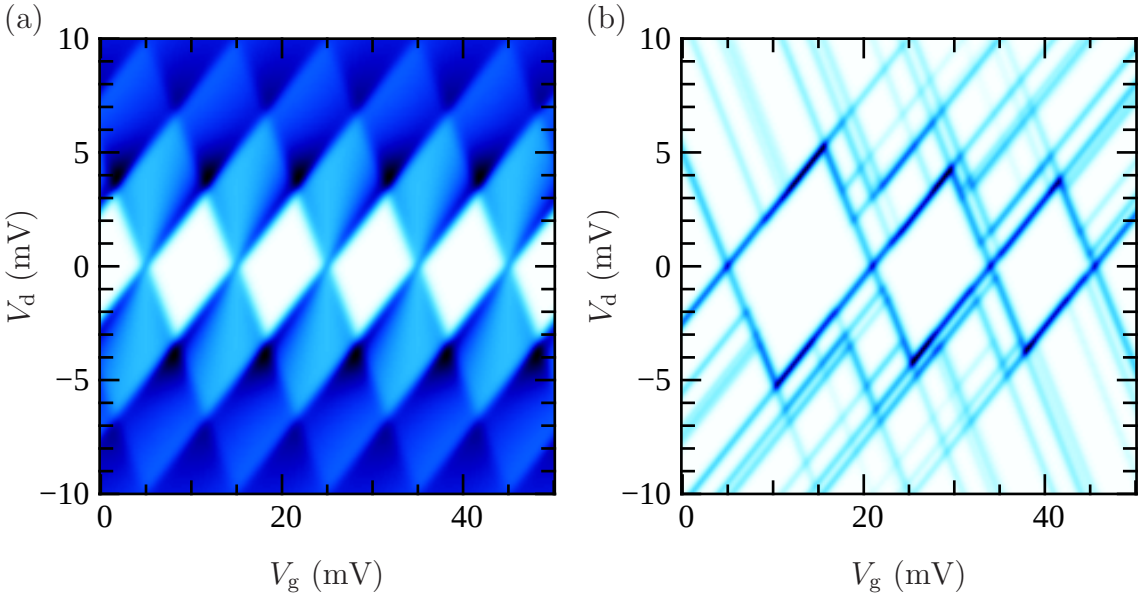

Figure 2.2: Numerically calculated nonlinear conductance. *White is zero differential conductance, dark blue is maximum differential conductance. (a) Metallic limit, negligible single particle level spacing $\Delta_1 \ll k_\mathrm{B}T$. (b) 4 resolved single particle states ($\Delta_1 \gg k_\mathrm{B}T$). The single particle level spacings are from bottom to top* 2 meV, 1 meV *and* 0.5 meV. *The temperature $T = 1\,\mathrm{K}$ and the capacitances $C_\mathrm{g} = C_\mathrm{s} = C_\mathrm{d} = 100\,\frac{e}{\mathrm{V}}$ are the same in (a) and (b).*

and

$$\frac{\partial V_\mathrm{g}}{\partial V_\mathrm{d}}^{-} = -\frac{C_\mathrm{d}}{C_\mathrm{g}} \tag{2.34}$$

The linear response conductance at $|eV_\mathrm{d}|, \Delta_1 \ll k_\mathrm{B}T \ll \frac{e^2}{C}$, can be easily calculated using the tunneling rate (2.18) for the left and the right barrier and the rate equations (2.22). One obtains

$$G_\mathrm{metal} = \frac{1}{2}\frac{G_\mathrm{s}G_\mathrm{d}}{G_\mathrm{s}+G_\mathrm{d}}\frac{\frac{e\alpha V_\mathrm{g}}{k_\mathrm{B}T}}{\sinh\frac{e\alpha V_\mathrm{g}}{k_\mathrm{B}T}} \tag{2.35}$$

where $\alpha = \frac{C_\mathrm{g}}{C}$ is the gate-voltage lever-arm factor and $G_{\mathrm{s/d}}$ the tunneling constant of the source/drain barrier (see eq. (2.18)). The full width at half maximum (FWHM) is $4.41\frac{k_\mathrm{B}T}{e\alpha}$. For simplicity we supposed the resonance to be centered around $V_\mathrm{g} = 0$.

If the temperature increases and $k_\mathrm{B}T \ll \frac{e^2}{C}$ does not hold any more, more than two charge states have to be taken into account at each gate voltage and one obtains [68]:

$$G_\mathrm{hot} = \frac{G_\mathrm{s}G_\mathrm{d}}{G_\mathrm{s}+G_\mathrm{d}}\frac{\sum_{n=-\infty}^{\infty} \mathrm{e}^{-\lambda(n-x+\frac{1}{2})^2}\frac{\lambda(n-x+\frac{1}{2})}{\sinh(\lambda(n-x+\frac{1}{2}))}}{\mathrm{e}^{\frac{\lambda}{4}}\sum_{n=-\infty}^{\infty}\mathrm{e}^{-\lambda(n-x+\frac{1}{2})^2}} \tag{2.36}$$

with $\lambda = \frac{1}{k_B T}\frac{e^2}{2C}$. For low temperatures $k_B T \ll \frac{e^2}{C}$ (2.36) tends to (2.35). At high temperatures it tends to the series conductance of the two barriers:

$$G_\infty = \frac{G_s G_d}{G_s + G_d} \tag{2.37}$$

We will now remove the assumption $\Delta_1 \ll k_B T$. Expression (2.30) then has to be replaced by expression (2.14) which includes the single-particle levels:

$$W_{\boldsymbol{n}} = \frac{e^2}{2C}(n-x)^2 + \sum_i \epsilon_i n_i \tag{2.38}$$

The single particle spacing Δ_1 is the mean spacing between the levels ϵ_i. For the tunneling rates expression (2.18) has to be replaced by (2.19) and (2.20), i.e. the tunneling rates are proportional to the Fermi function and not to its auto-convolution. Then, the Coulomb blockade diamonds persist but inside the conducting regions the conductance increases[3] only when an additional excited state enters the bias window. Therefore there are plateaus with 0 differential conductance (see figure 2.2(b))

The linear response conductance at the transition between n and $n+1$ electrons in a d-fold degenerate level is now

$$G_{\text{degenerate}} = \frac{e}{k_B T}\frac{I_s I_d}{I_s + I_d}\frac{\sqrt{(n+1)(d-n)}}{2\cosh\left(\frac{e\alpha V_g}{k_B T} - \frac{1}{2}\ln\frac{d-n}{n+1}\right) + \frac{1+d}{\sqrt{(n+1)(d-n)}}} \tag{2.39}$$

where $I_{s/d}$ is the tunneling constant of the source/drain barrier (see eqs. (2.19) and (2.20)). In the case of a non-degenerate level ($d = 1$, $n = 0$) this equation simplifies to

$$G_{\text{nondegenerate}} = \frac{1}{4}\frac{e}{k_B T}\frac{I_s I_d}{I_s + I_d}\frac{1}{\cosh^2\frac{e\alpha V_g}{2k_B T}} \tag{2.40}$$

The main difference with respect to the metallic limit is that the peak height is now proportional to T^{-1}. The resonances for nondegenerate discrete levels are also narrower than the resonances in the metallic limit: the FWHM is now $2\,\text{arcosh}(3)\,\frac{k_B T}{e\alpha} \approx 3.53\frac{k_B T}{e\alpha}$ instead of $4.41\frac{k_B T}{e\alpha}$. However, scaled to the same FWHM, the peak shape is almost the same as before because $\frac{x}{\sinh x} \approx \frac{1}{\cosh^2\frac{x}{2.5}}$ holds up to the 4th order in x. The difference between the two expressions is smaller than 10^{-2}.

Now, where exactly occurs the transition between metallic regime and resolved Δ_1? We address the problem numerically by calculating the conductance through a dot where the single particle levels ϵ_i are spaced by Δ_1, using expressions (2.19) and (2.20) for the rates between source and drain and the dot. Then we vary the temperature. The results are shown in figures 2.3 and 2.4. We find that the transition in the peak height occurs rapidly around $\Delta_1 = 2k_B T$. The transition in the peak shape occurs between $\Delta_1 = 2k_B T$ and $\Delta_1 = 5k_B T$.

[3] In case of very different couplings the conductance can also decrease

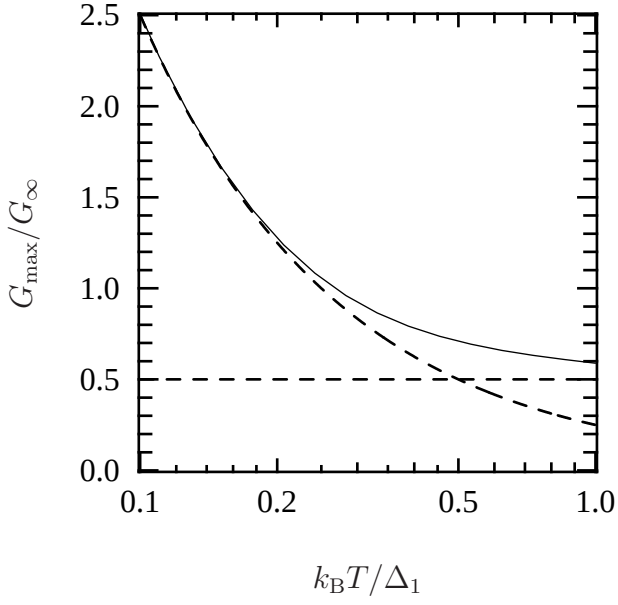

Figure 2.3: Evolution of the peak height with k_BT/Δ_1. *The height of the peaks is plotted in units of $G_\infty = \frac{G_s G_d}{G_s+G_d}$. For this numerical calculation the height of the 6th peak of a 11-level system with uniform spacing Δ_1 has been plotted. The dashed lines are the metallic limit $k_BT \gg \Delta_1$ with peak height $G_{\max} = \frac{1}{2}G_\infty$ and the limit where only one single particle level carries current $k_BT \ll \Delta_1$ with peak height $G_{\max} = \frac{\Delta_1}{4k_BT}G_\infty$.*

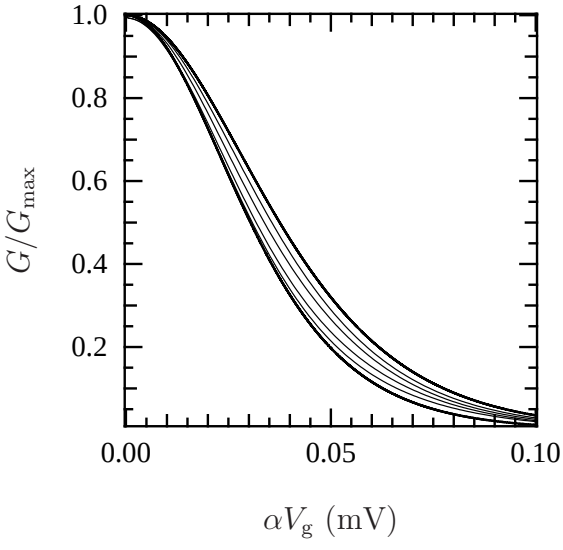

Figure 2.4: Evolution of the peak shape with Δ_1. *$T = 200\,\mathrm{mK}$. The largest peak represents the limit $\Delta_1 \ll k_BT$, the narrowest one the limit $\Delta_1 \gg k_BT$; the peaks in between the cases $\Delta_1 = x \times k_BT$ with $x = 2$ (next to largest one), 3, 4, 5 (next to narrowest one). All peaks are normalized to unitary height. The bold lines are analytical, the fine ones are numerical solutions, taken at the 5th peak of a system with 9 levels and a charging energy of $1\,e$V.*

To summarize, when temperature is decreased one passes from a constant conductance G_∞ to a oscillating conductance when k_BT falls below the charging energy $\frac{e^2}{C}$. The peak conductance goes to $\frac{1}{2}G_\infty$ and the thermally activated valley conductance goes to 0. When k_BT falls below the second characteristic energy, the excitation spacing Δ_1, the peak height begins to increase with T^{-1}.

2.6 Strong coupling

We have considered the coupling between dot and leads only in lowest order perturbation theory and supposed $G_T \ll \frac{e^2}{h}$. This will be sufficient for the major part of the problems addressed in the following chapters.

If the coupling to the leads is increased (but still kept smaller than $\frac{e^2}{h}$), second order effects called co-tunneling, i.e. tunneling through virtual states of the dot, can

cause non-vanishing valley conductance, even at very low temperature. In some cases the valley conductance can even go to $\frac{2e^2}{h}$ at very low temperature. This is due to the Kondo effect [33].

At very low temperature, where our assumption $\hbar\Gamma \ll k_{\mathrm{B}}T$ is not verified anymore, higher order effects play also a role at the peaks because the life-time broadening of the levels has to be taken into account and at very low temperature for $k_{\mathrm{B}}T \ll \Gamma \ll \Delta_1, \frac{e^2}{C}$ the peaks become Lorentzians. At intermediate temperatures the line shapes we have found above and the Lorentzian have to be convoluted.

The persistence of Coulomb blockade for strong coupling has been studied both experimentally, in [92, 59] for example and theoretically in [80, 79, 90, 121] for example. In this case it is not sufficient to know the total conductance of a barrier but it is also important to know how the conductance or total transmission is distributed over the different modes (or channels in the Landauer-Büttiker formalism) of source, drain and dot.

It appears that Coulomb blockade with a reduced Coulomb gap persists until perfect transmission ($G_{\mathrm{T}} = \frac{e^2}{h}$) for point contacts (only one mode) [31] and even for $G_{\mathrm{T}} \gg \frac{e^2}{h}$ in the case of a perfect tunnel barrier (infinite number of modes with each infinitely small transmission). These are the extreme cases, other transmission distributions give intermediate results [90].

Chapter 3

MOSFETs working as single-electron transistors

Coulomb blockade is mostly studied in Al dots where confinement is provided by Al_20_3 tunnel barriers or in 2 dimensional electron gases in GaAs/AlGaAs heterostructures where the barriers consist of pinch-off channels. But the barriers can be of any type provided that they are sufficiently resistive ($G_T < \frac{e^2}{h}$). Coulomb blockade can even be observed in very small standard MOSFETs [98, 44, 45]. As, following the credo of microelectronics, the transistors are scaled down, the doping levels and the thickness of source and drain near the gate have to be reduced to avoid short channel effects. As a consequence their resistance grows. In microelectronics this is an undesired effect because the maximal current drops, which in turn slows down the switching time of a circuit. But the increasing access resistance makes small MOSFETs better and better SETs at low temperature.

In this work SOI nanowire transistors are measured. Their working principle is the same as that of MOSFETs. The main difference is that the width of the wire is comparable to the gate length, while MOSFETs usually are much larger than long. The geometry of our devices is sketched in figure 3.1. SOI (silicon on insulator) means that the transistor is realized in a very thin silicon film which is separated from the substrate by a thick buried silicon oxide layer.

The samples are from two batches, B146 and 9857 that were realized in CEA/Leti. Batch 9857 has been fabricated during the PhD of Gabriel Molas [85].

3.1 Sample fabrication

The sample geometry is illustrated in figures 3.1, 3.2 and 3.3. We will give only a short overview of the fabrication process. More details can be found in [85].

The nanowire transistors are produced on 200 mm SOI wafers (Smartcut process), with a 200 nm thick active layer over 400 nm of buried oxide (BOX) and boron substrate doping of $10^{15}\,\mathrm{cm}^{-3}$. The thickness of the active layer is reduced to 80 nm by oxidation at 1050 °C under O_2+HCl. A second local oxidation further reduces

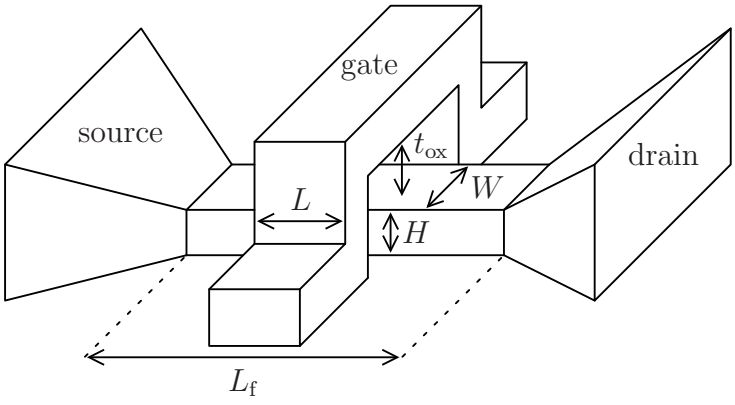

Figure 3.1: The geometry of the nanowires. *Dielectrics have been omitted. In the samples of batch 9857 the sides of the nanowire are inclined by about 45° (see figure 3.4).*

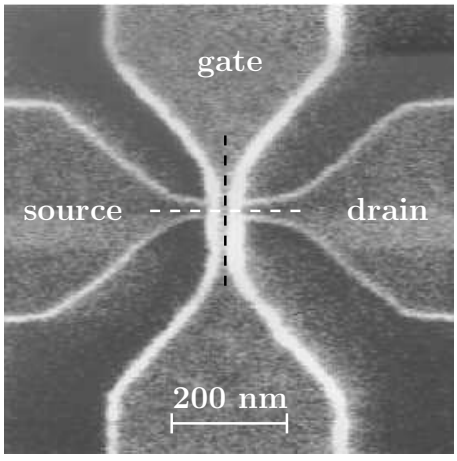

Figure 3.2: Scanning electron micrograph of a nanowire transistor. *The image has been taken before the back-end process, i.e. there are no spacers, no encapsulation oxide and no interconnections yet.*

the thickness of the active zones (where the channel will be) to 15 nm to 25 nm. The 80 nm thick regions will serve to define the contacts. The whole structure is then doped with arsenic (a donor). In batch B146 there are also variants which are doped with boron (an acceptor) or left undoped[1]. This step defines the doping level in the channel and serves to adjust the threshold voltage. Only subsequently a much higher arsenic dose is implanted to form the source and drain regions (see below).

The lithography works with negative photo resist. The active regions are written by an electron beam; the contact regions where lower resolution is needed are exposed to extreme UV. On some wafers the resist has been trimmed to reduce the width of the wire.

The wires are then plasma etched. For batch 9857, this step has been optimized in order to obtain 45° inclined sides. (see figure 3.4). This allows for homogeneous oxidation of the sides of the silicon film.

The film is then heated to 800 ℃ to form a thermal oxide of various thicknesses from 1.5 nm to 4 nm. In batch 9857, a second oxide layer of 8 nm or 20 nm is then deposited (high thermal oxide, HTO) and annealed at 730 ℃.[2]

[1]The active layer is indeed already doped by some $10^{15}\,\mathrm{cm}^{-3}$ of boron. But on average this corresponds to less than one dopant in the active area

[2]The main purpose of batch 9857 was to study permanent memory cells [85]. Therefore, in most samples of batch 9857 nano-crystals have been grown between the two layers of oxide. As

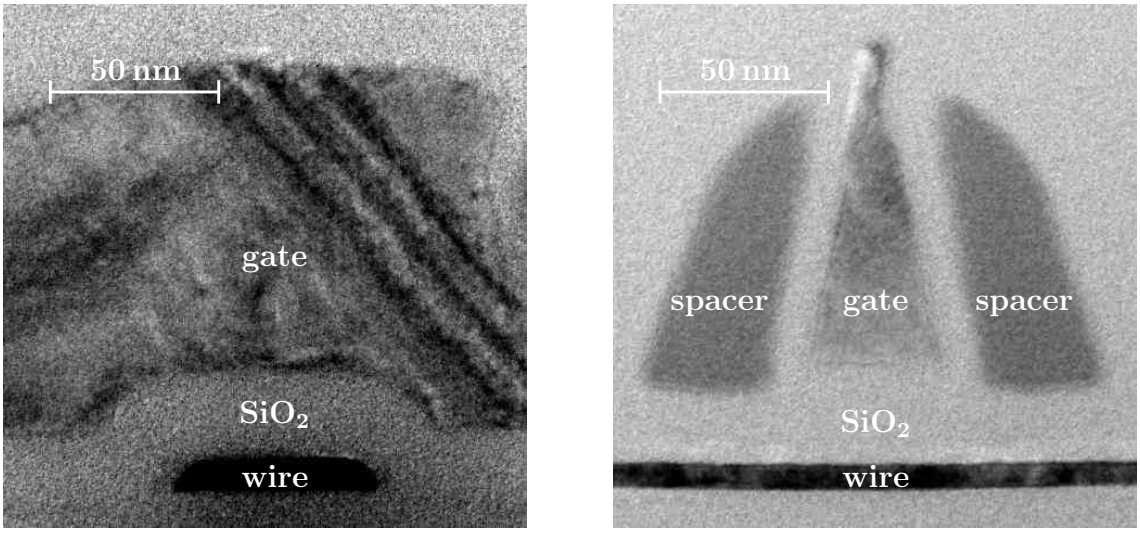

Figure 3.3: Transmission electron micrographs of batch 9857 samples. *The left cut is perpendicular to the wire, along the black dashed line in figure 3.2, the right one parallel to the wire along the white dashed line in figure 3.2. The wires shown here are thinner than the ones measured in this work.*

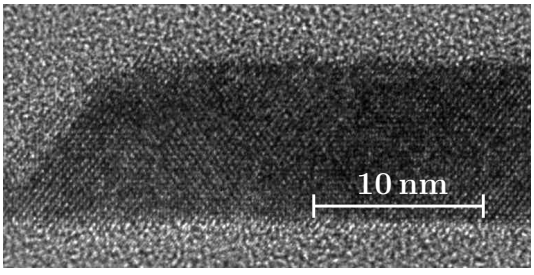

Figure 3.4: Transmission electron micrograph of the wire. *The cut is in the same direction as in the left panel of figure 3.3. The darker region is silicon, the lighter one silicon oxide. Note the atomic planes, visible inside the wire, and the 45 ° inclination of the side of the wire.*

The gate electrode consists of 100 nm of strongly n-doped polycrystalline silicon. It is etched with the same hybrid lithography process used for the wire.

In batch 9857 spacers (10 nm HTO and Si_3N_4) are formed on both sides of the gate (see figure 3.3). They have a total width of 50 nm and serve together with the gate electrode as mask which protects the channel from the following step, the doping of source and drain. This is why in batch B146 only the channel under the gate electrode is low doped while in the batch 9857 also a 50 nm region on either side of the gate preserves the low channel doping.

Doping the source and drain access regions is a very critical process for thin silicon on insulator layers, because on the one hand high doping levels are necessary to obtain sufficient conductance — especially at low temperature where dopants freeze out if the doping is too low. On the other hand, if the defect level after doping exceeds $5 \cdot 10^{21}\,\mathrm{cm}^{-3}$, the active layers become amorphous after annealing rather than recrystallizing. For bulk silicon, this is not the case because the substrate

such samples have not been studied in the present work, this step is not detailed here.

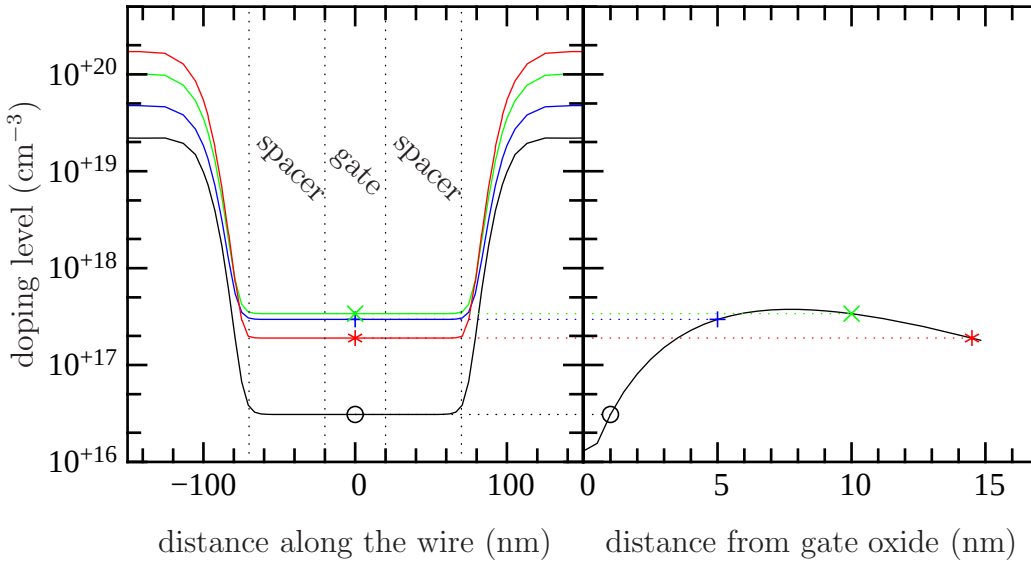

Figure 3.5: Simulated doping profile for batch 9857 for the retained doping process. *The left panel shows the doping level along the wire at* 1 nm *(∘),* 5 nm *(+),* 10 nm *(×) and* 15 nm *(∗) from the top of the film. The right panel shows the doping profile below the center of the gate electrode. A previous estimation gave a slightly higher channel doping of* $10^{18}\,\mathrm{cm}^{-3}$ *[85]. We take* $5 \cdot 10^{17}\,\mathrm{cm}^{-3}$ *throughout the entire work. Simulations: Gabriel Molas*

serves as germ for recrystallization. Indeed, earlier samples with the same geometry became insulating at low temperature despite a very high doping level: the wire was amorphous.

In batch B146 the source and drain regions were implanted with an ion energy of 8 k*e*V through 5 nm of screening oxide. In batch 9857 two methods were tried: doping with a high ion energy of 20 k*e*V through 10 nm of oxide and doping with a lower ion energy of 3 k*e*V through 3 nm of screening oxide, followed by RTP (rapid thermal process) annealing at 1050 ℃. Simulations (see fig. 3.5) and measurements of the saturation current clearly favor the second method, which has been retained. The sample specific doping levels and sizes are summarized in tables 3.1 and 3.2.

3.2 Room temperature characteristics

3.2.1 Below threshold

At low bias voltage a *n*-channel MOSFET can be modeled in the following way. The electrical potential in the channel is controlled by gate voltage. As the gate voltage is swept from negative to positive voltage, the electrons experience first a barrier of decreasing height. When the gate electrode is sufficiently long, electrons

batch	B146-1	B146-2	9857
H (nm)	12, 17, 22		17
L (nm)	various, see table 3.2		
W (nm)			
t_{ox} (nm)	4		10, 24
L_f (nm)	200		
channel doping	—	As: $10^{14}\,cm^{-2}$ $\rightsquigarrow 10^{19}\,cm^{-3}$	As: $5 \cdot 10^{17}\,cm^{-3}$
	B: some $15\,cm^{-3}$		
access doping	As: $2 \cdot 10^{14}\,cm^{-2}$ $\rightsquigarrow 2 \cdot 10^{19}\,cm^{-3}$	As: $10^{14}\,cm^{-2}$ $\rightsquigarrow 10^{19}\,cm^{-3}$	As: $5 \cdot 10^{14}\,cm^{-2}$ $\rightsquigarrow 10^{20}\,cm^{-3}$
	B: some $15\,cm^{-3}$		

Table 3.1: Technological parameters of the studied nanowires. *Note that in batch 9857 there are* 50 nm *wide spacers on both sides of the gate electrode. The doping in the wire below them is the same as in the channel.*

sample	batch	W (nm)	L (nm)	H (nm)	t_{ox} (nm)
1	9857	20	40	17	10
2	9857	30	40	17	10
3	9857	60	30	17	24
4	9857	20	40	17	24
5	9857	30	40	17	24
6	9857	60	40	17	24
7	9857	80	50	17	24
8	B146-1	50	40	17	4
9	B146-1	30	40	22	4
10	B146-1	50	40	22	4
11	B146-2	50	40	22	4
12	B146-2	40	60	22	4

Table 3.2: Sample geometries. *Only samples for which figures are shown are listed here. Others are included in statistical analysis. The widths in batch 9857 refer to the width on top of the wire.*

cannot tunnel through the barrier and transport is thermally activated. The low bias conductance is therefore

$$G_{\text{channel}} \propto \exp \frac{e\alpha V_{\text{g}}}{k_{\text{B}}T} \tag{3.1}$$

where α is the gate-voltage lever arm i.e. the degree of control the gate electrode has over the channel. This factor compares directly to the lever-arm factor defined for the single-electron transistor on page 28.

The upper panel of figure 3.6 shows typical conductance plots on a logarithmic scale. As expected, at low gate voltage the conductance grows exponentially with gate voltage. We extract an average lever-arm factor $\alpha = 0.35$ in batch B146 and 0.55 in batch 9857. The RMS sample-to-sample fluctuations are only 0.02 in batch B146 (despite different geometries) but 0.16 in batch 9857. In long bulk MOSFETs a lever arm factor $\alpha < 1$ is explained by the depletion capacitance coupling the channel to the substrate and thus reducing the gate control. In SOI transistors this capacitance does not exist. The substrate can control the channel only through the very thick BOX. In batch 9857 the BOX is 20 to 40 times thicker than the gate oxide, in batch B146 even 100 times. Therefore the substrate coupling is negligible and the loss of gate control can only be due to electrostatic coupling to the source and drain potential. This is one of the so-called short-channel effects degrading the performances of very short transistors. This effect becomes stronger when the high source and drain doping comes very close to the channel. This is the case in the batch B146 samples, where there are no spacers separating the high source and drain doping and the channel. These samples therefore show a small α despite a thinner oxide. The 9857 samples have very large spacers separating the channel and the source/drain doping by 50 nm. With their thick gate oxides they could not work otherwise.

3.2.2 Above threshold

When the gate voltage crosses the so called threshold voltage V_{t}, the conduction band edge reaches the Fermi level and the density of states increases abruptly. The gate voltage is then screened out and the conduction band edge inside the channel is pinned to the Fermi level while the number of carriers increases linearly with gate voltage. The linear conductance of a MOSFET is then given by

$$G_{\text{channel}} = \frac{\mu_{\text{n}} C_{\text{g}}}{L^2} (V_{\text{g}} - V_{\text{t}}) \tag{3.2}$$

where C_{g} is the total gate capacitance, L the channel length and μ_{n} the electron mobility. In figure 3.6(b) the conductance follows this linear behavior only in a small range above V_{t}; very soon the conductance begins to bend downwards. This saturation is due to series resistances and mobility reduction caused by the strong electric field of the gate electrode. In samples of batch 9857 the saturation should be dominated by the series resistances: the access resistance is high because of the

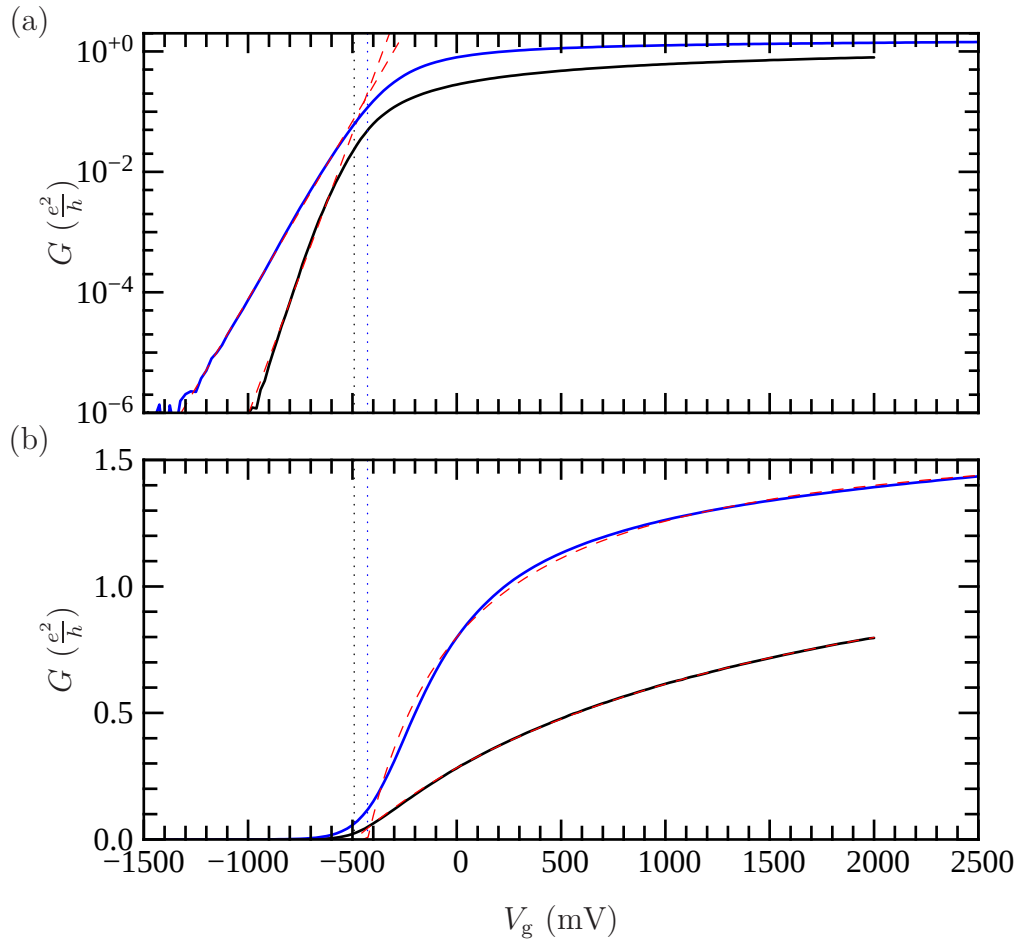

Figure 3.6: Room temperature characteristics of the nanowires. *Black curve: batch 9857 (sample 6); blue curve: batch B146 (sample 10). In the logarithmic plot (a) the exponential characteristic below threshold is fitted (red dashed lines). In the linear plot (b), the "on" regime is fitted (red dashed lines, see text). The dotted lines indicate the threshold voltages.*

large spacers, and the electric field is small because the gate oxide is thick. The conductance of eq. (3.2) in series with a gate-voltage independent conductance fits very well the behavior of the transistors of batch 9857 (see fig. 3.6(b)).

We know the gate capacitance from Coulomb blockade measurements at low temperature (see section 3.4). With the gate length, eq. (3.2) allows then to calculate the channel mobility. Using the lithographic gate length, we obtain extremely low values in the order of $20\,\mathrm{cm^2V^{-1}s^{-1}}$ in batch 9857. However, eq. (3.2) is extremely sensitive to the gate length. The electrical gate length could indeed be longer than the lithographic one. This is due to the very large and low doped parts of the wire below the spacers where the Fermi level is below the conduction band edge and the electron density is low. Therefore the gate electrode can easily shift the potential in the spacer regions. This can have two effects:

- The effective gate length, i.e. the length over which the conduction band edge

falls below the Fermi level and electrons are accumulated is longer than the actual gate length. This can have dramatic effects on the extracted channel mobility because expression (3.2) depends quadratically on the channel length.

- The height of the barriers formed by the regions below the spacers changes with gate voltage, causing the conductance of the regions below the spacers to depend on gate voltage. The access resistance therefore contains a voltage dependent part which is attributed to the channel.

Both effects tend to increase the effective gate length with respect to the actual lithographic gate length. At room temperature these two effects cannot be distinguished as the relevant energy scale, the energy of the dopant levels (45 meV), is of the same order as $k_{\mathrm{B}}T = 26\,\mathrm{m}e\mathrm{V}$. At low temperature however, the second effect is the driving force of most of the phenomena we observe in chapter 5.

The fact that in batch 9857 the electrical channel length is longer than the lithographic one is confirmed when we try to extract mobility from the series resistance. This resistance is dominated by the regions below the spacers: beyond them the doping is 50 times higher and the wire increases in width and thickness. Knowing the geometry, we can estimate the conductivity of the wire below the spacers and in consequence the mobility. We find values in the order of $400\,\mathrm{cm^2V^{-1}s^{-1}}$, much higher than in the channel.

We now choose the effective gate length L_{eff} such that the mobility in eq. (3.2) and the spacer regions becomes equal. The fitting parameters are then: threshold voltage V_{t}, mobility μ_{n}, and effective channel length L_{eff} instead of: threshold voltage, channel mobility, and series resistance. The total conductance is then

$$G = \left(\frac{L_{\mathrm{eff}}^2}{\mu_{\mathrm{n}} C_{\mathrm{g}} \left(V_{\mathrm{g}} - V_{\mathrm{t}}\right)} + \frac{L_{\mathrm{g}} + 2L_{\mathrm{spacer}} - L_{\mathrm{eff}}}{ne\mu_{\mathrm{n}} S} \right)^{-1} \tag{3.3}$$

where n is the doping level in the spacer regions. On average over the measured samples the effective gate length L_{eff} is 70 nm longer than the lithographic gate length L_{g}. The RMS sample to sample fluctuations are 10 nm. This means that the gate controls slightly more than half the spacers. For the total mobility we find $220\,\mathrm{cm^2V^{-1}s^{-1}} \pm 30\,\mathrm{cm^2V^{-1}s^{-1}}$. This value is much smaller than for undoped bulk silicon at room temperature ($1450\,\mathrm{cm^2V^{-1}s^{-1}}$) but expected for doping levels in the order of $10^{18}\,\mathrm{cm^{-3}}$ [54] and quite typical for small silicon devices.

In batch B146 the situation is quite different. There are no spacers and the doping of the access regions extends to the border of the gate electrode and the gate oxide is thinner. Therefore, the gate electrode should control only the region right below the gate. Indeed, with eq. (3.3) we find an effective gate length only $5\,\mathrm{nm} \pm 7\,\mathrm{nm}$ longer than the lithographic length. The mobility is lower: $\mu_{\mathrm{n}} = 70\,\mathrm{cm^2V^{-1}s^{-1}} \pm 20\,\mathrm{cm^2V^{-1}s^{-1}}$. In the access regions this is due to the higher doping level. Below the gate the thinner oxide increases the electric field and the polycrystalline gate electrode induces a stronger disorder in the wire. As there is no doping modulation in source and drain, we took for the length of the access resistor in eq. (3.3) the total wire length minus the effective gate length, neglecting that way the resistance of the triangular parts where the wire widens (see figure 3.2).

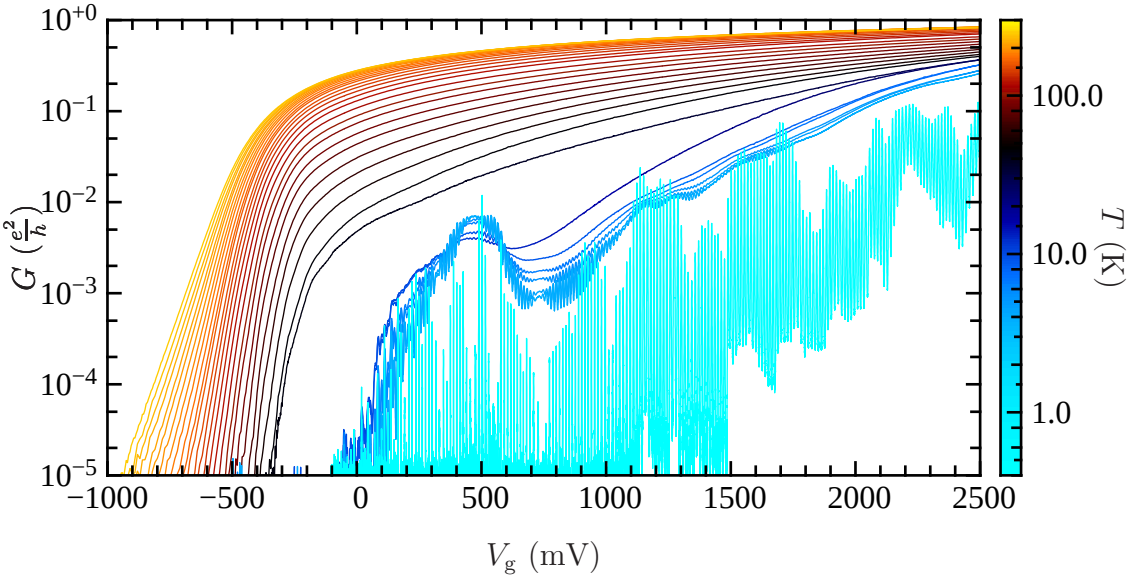

Figure 3.7: Cooling of a nanowire MOSFET from room temperature to 400 mK. *The upper group of curves goes from* 50 K *to* 300 K *with a step of* 10 K. *The cyan curve was taken at* 400 mK *and the light blue curve at* 4.2 K. *Between* 5 K *and* 50 K *the temperature changes significantly within each gate voltage sweep. Sample 3.*

3.3 How the nanowire FET becomes a SET

Figure 3.7 shows the temperature dependence of a batch 9857 nanowire MOSFET. Above approximately 50 K nothing spectacular happens. In accordance with eq. (3.1) the part of the characteristics below threshold becomes steeper as temperature is reduced. But also above threshold the characteristics becomes steeper and the current near threshold decreases drastically. This supports the interpretation we made above that there are potential barriers below the spacers over which the gate electrode has also some control. To confirm this interpretation, in figure 3.8(a) we draw the Arrhenius plot of conductance at different voltages. Below approximately 100 K the conductance depends exponentially on T^{-1}, implying thermally activated transport. At high temperature there are significant deviations from the exponential behavior; at high gate voltage the conductance even increases from 300 K down to 200 K. This effect is often observed in silicon devices. It is due to an increase of mobility as phonon scattering is quenched. Using the slopes in the Arrhenius plot between 50 K and 75 K we calculate the activation energy of conductance (see figure 3.8(b)). Below threshold the activation energy strongly depends on gate voltage. But even above threshold transport is thermally activated with a barrier height that decreases only slowly as gate voltage is increased. This behavior suggests the potential depicted in figure 3.9. Below the threshold voltage the main potential barrier is the part of the wire below the gate electrode which is well controlled by the gate voltage. At the threshold voltage, the potential below source and drain begins to

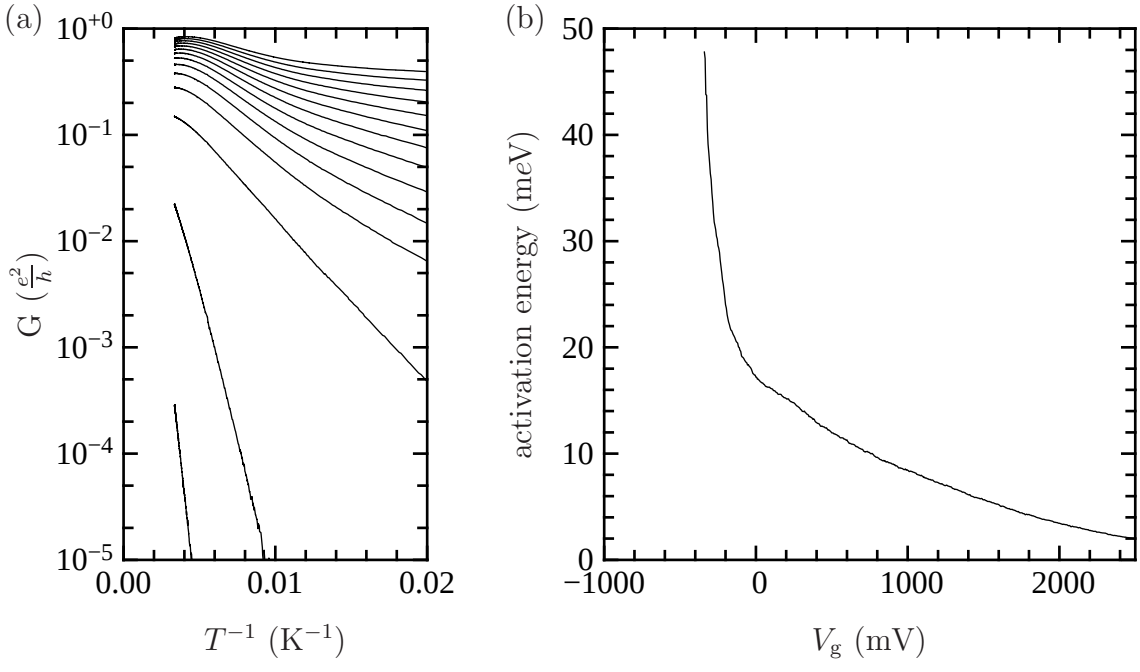

Figure 3.8: Thermal activation of transport. *(a) Arrhenius plot of the drain-source conductance. The curves from bottom to top are taken at* $V_g = -0.75\,V, -0.50\,V \ldots 2.50\,V$. *Transport is thermally activated. Anomalies at high temperature are due to phonon scattering which makes the mobility temperature dependent. (b) Extracted activation energy. There are two regimes: below the threshold voltage* $V_t \approx -100\,mV$ *it is controlled by the part of the wire below the gate, above threshold by the parts below the spacers which are only weakly influenced by the gate voltage. Sample 3.*

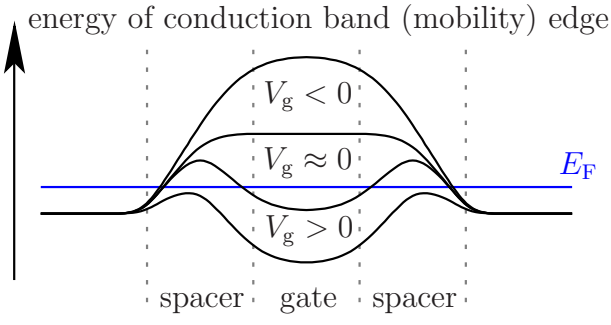

Figure 3.9: Schematic view of the potential along the nanowire for different gate voltages. *The gate has a good control on the nanowire below the gate, but also on the low doped parts below the spacers.*

be higher than the potential below the gate. The height of these potential barriers is of the order of 20 meV and is only slowly lowered by the gate voltage. Between 4 K and 400 mK the conductance does not decrease significantly. Thus below 4 K tunneling through the barriers seems to dominate. We will investigate transport through these barriers in more detail in chapter 5.

If our description is correct, there is a quantum dot below the gate separated from source and drain by barriers. As the conductance is below the quantum of conductance $\frac{e^2}{h}$ we should be in a single-electron transistor configuration as described

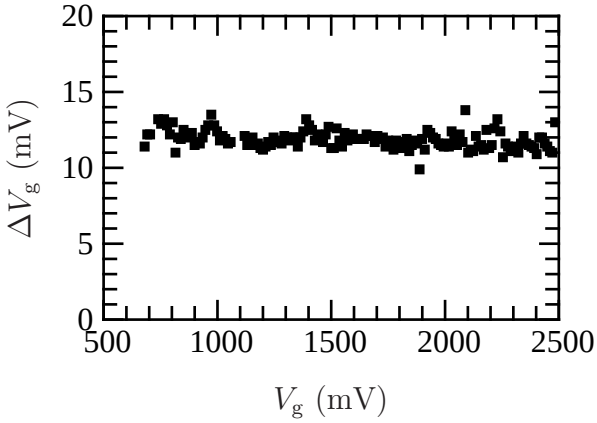

Figure 3.10: Peak spacing vs. gate voltage. *The peak spacing changes only very little. This indicates that the size of the quantum dot is well defined and barely influenced by gate voltage. Sample 6.*

in 2.5. Indeed, the blue curves in figure 3.7 below 50 K show regular oscillations. These oscillations become sharp resonances at very low temperature (400 mK, cyan). The period of these oscillations is 13 mV. With a lever-arm factor (see eq. (3.1)) of 0.3 these resonances correspond to a charging energy of 4 m*e*V which is equivalent to 50 K.

We will now try to confirm this model by comparing the measured period of these oscillations with calculations.

3.4 Size of the quantum dot

As shown in 2.5, the gate-voltage spacing V_+ of the Coulomb blockade resonances of a single-electron transistor is determined by the gate capacitance:

$$V_+ = \frac{e}{C_g} \tag{3.4}$$

Figure 3.10 shows a typical example of the peak spacing versus gate voltage. The peak spacing is obtained by an automatic peak detection described in appendix B.2. Its gate-voltage dependence and fluctuations are very weak. Therefore, we can suppose that there is only one quantum dot rather than multiple impurity quantum dots [103, 100, 70, 69]. We observe the latter case in very short MOSFETs with more classical geometry, i.e. more than 10 times wider than long [98, 44, 45] and in the first peaks in batch B146 samples (in particular split 1, see chapter 4.2).

The gate capacitance can be roughly estimated by an equivalent planar capacitor:

$$C_g = \epsilon_0 \epsilon_{SiO_2} \frac{S}{t_{ox}} \tag{3.5}$$

$\epsilon_{SiO_2} = 3.9$ is the dielectric constant, t_{ox} is the oxide thickness and S the surface area of the gate–wire overlap. We take into account the top and the flanks of the wire. In the case of batch B146 this means $S = L \times (W + 2H)$ (L is the gate length, W is the wire width and H its height, see figure 3.1). In the case of batch 9857 the sides of the wire are inclined by 45° and $S = L \times (W + 2\sqrt{2}H)$.

In figure 3.11 the measured average peak spacing of several samples is plotted versus the calculated peak spacing. For the red markers the gate capacitance has been calculated in planar capacitor approximation (eq. (3.5)). For the batch B146 samples with 4 nm gate oxide the agreement is good. For the batch 9857 samples with thicker gate oxide the gate capacitance corresponding to the measured peak spacing is higher than the estimation (3.5). Yet in this case the gate oxide is almost as thick as the gate and the wire are large, so that we indeed expect the planar capacitor to underestimate the capacitance because it does not take into account the corner charges. We thus have to do numerical calculations.

We do them as follows. The wire and the gate are supposed to be ideal metals. For the gate electrode, the source and drain regions and the dot where the Fermi level is inside the conduction band this is a good approximation. For the parts of the wire below the spacers the approximation is more questionable. In metals charges appear only near the surface. We create a rectangular mesh on each surface and take the charge density of each control point as a variable. A quite lengthy expression from [39] gives the potential in any point of space created by a bilinear charge distribution on a planar rectangle (our mesh cells). It is a linear function of the charge density at the corner points. With this formula the potential at each control point of the mesh[3] is calculated as a linear function of the charge densities at all the control points. Thus the potentials and charge densities are related through a linear equation system. As we are dealing with metals, the potential on each electrode is constant. By fixing the two potentials we can solve this system for the charge densities at the control points. As we need only 2 dimensional meshes, the linear equation system is small (several 1000 points) and can be solved in some seconds on a recent personal computer. The size of the problem can be further reduced because of symmetry: Our devices are symmetric to a vertical plane along the wire and another along the gate electrode. So only one quarter of the geometry has to be calculated. A typical mesh and the charge density are plotted in figure 3.12(a). To calculate the gate-wire capacitance we set the gate voltage to -1 V and the wire potential to 0. The integral of the charge density on the wire then gives the gate–wire capacitance. The integral over the charge density on the gate electrode is not the gate–wire capacitance but the total capacitance of the gate electrode. This is due to the fact that our potential formula implies as boundary condition 0 potential at infinity. In order to calculate the total wire capacitance we would have to set the gate voltage to 0 and the wire voltage to a finite value.

Figure 3.12(b) shows the charge density around the wire below the center of the gate electrode. There are singularities in the corners due to the sharp edges in our geometry. In reality the charge density in the edges should be smoother, firstly because the corners are not perfectly sharp, secondly because the silicon wire is not an ideal metal. Nevertheless, as the gate oxide thickness exceeds the radius of the effective rounding of the corners, the sum over the corner charges should be correct.

[3] In fact, we calculate the potential at points of the surface slightly besides the control points to avoid numerical instabilities.

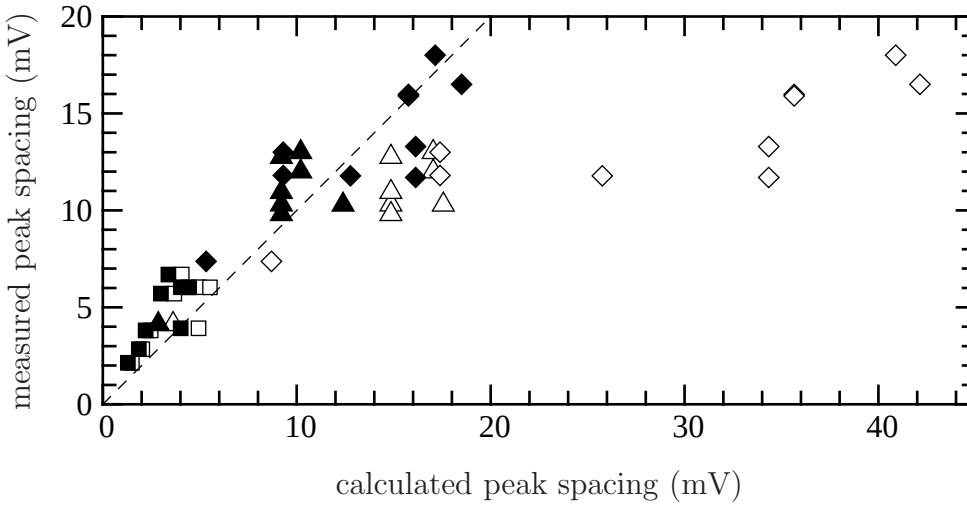

Figure 3.11: Measured vs. calculated peak spacing for various sample geometries. *The oxides thicknesses are $t_{ox} = 4\,nm$ (□), $t_{ox} = 10\,nm$ (△), $t_{ox} = 24\,nm$ (◇). Empty and filled symbols represent the same measurements but the calculations have been done by planar capacitor approximation for the empty symbols and numerically (see text) for the filled ones. The numerical result is much closer to perfect agreement (dashed line).*

More problematic is the direction along the wire, shown in figure 3.12(c). We did not model the vertical sides of the gate electrode and we omitted the spacers which have a higher dielectric constant than silicon oxide (7 versus 3.9) because our model does not allow for inhomogeneous dielectric materials. However, these two omissions should compensate themselves to some extent in the vicinity of the gate electrode. Neglecting the flanks of the gate electrode causes increased corner charges at the edge of the gate. The charge of the flanks of the gate is thus projected to the bottom of the gate electrode and the influence of the gate electrode on the parts of the wire near the corners of the gate electrode is increased. The effect of the spacers would be the same: the higher dielectric constant projects the charge on the side of the gate electrode to the bottom of the spacers.

The charge density integrated over the whole length of the wire yields a much higher gate capacitance than the one related to the measured peak spacing. However, this capacitance includes the gate-source and gate-drain cross capacitances. In consequence, we have to draw a limit between the dot and the reservoirs and stop the integration there. In figure 3.12(c) the charge density falls off most rapidly at approximately 30 nm from the center, i.e. 10 nm beyond the border of the gate electrode for this geometry with 24 nm gate oxide. All measured geometries considered, we obtain the best agreement between measured peak spacing and capacitance calculations for a cut-off half a gate-oxide thickness beyond the gate border, i.e. the length of the quantum dot is $L + t_{ox}$ (L is the gate length and t_{ox} the gate oxide thickness). With this cut-off we find an excellent agreement between measured peak-spacing and calculations (blue markers in figure 3.11). Our findings of a quan-

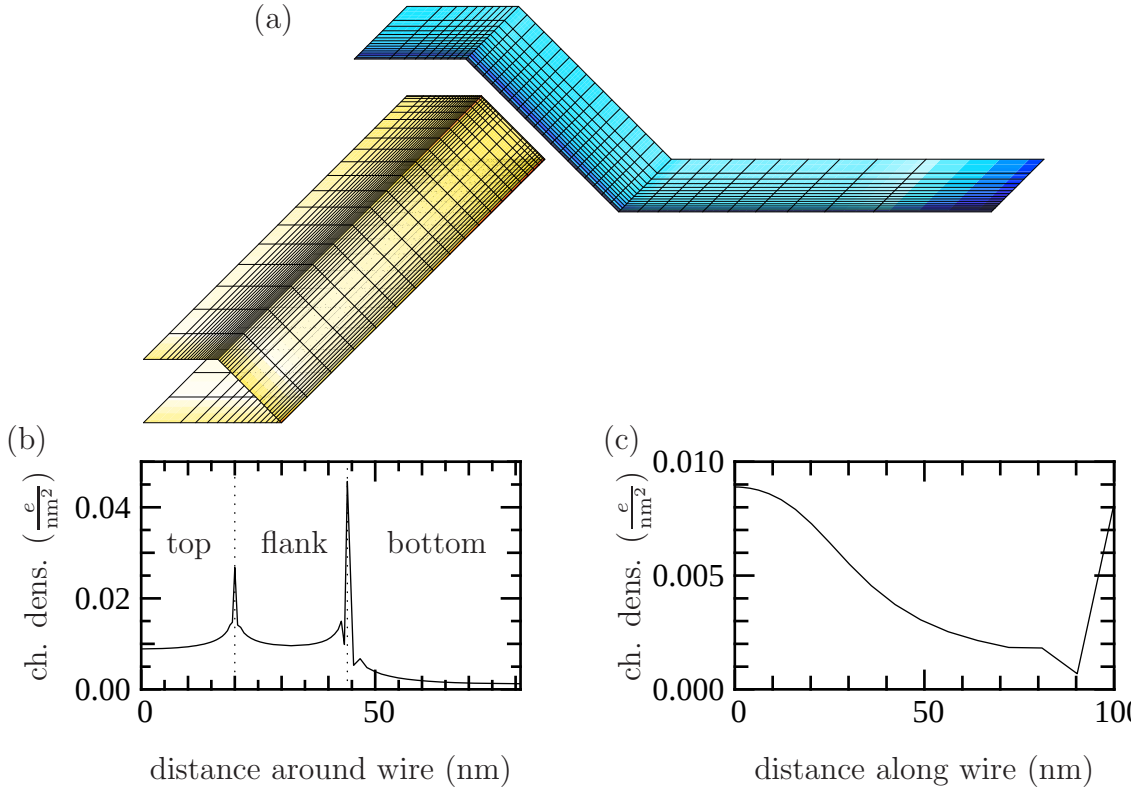

Figure 3.12: Numerical calculations of the gate capacitance. *Both gate and wire are supposed to be ideal metals; thus only their surface is considered. Because of symmetry only one quarter of the geometry has to be calculated. This example is a batch 9857 sample with width* $W = 40\,\mathrm{nm}$, *gate length* $L = 40\,\mathrm{nm}$ *and* $t_{ox} = 24\,\mathrm{nm}$ *gate oxide. (a) The rectangular mesh used for the calculations. (b) Electron density around the wire at gate voltage* $-1\,\mathrm{V}$ *below the gate electrode center. Distance 0 corresponds to the top center of the wire* 80 nm *to the bottom center. (c) Electron density along the wire at gate voltage* $-1\,\mathrm{V}$. *The trace is taken at the top center of the wire, beginning below the gate electrode center. The increase of the charge density at* 100 nm *is an artifact due to the abrupt end of the wire in the simulated geometry.*

tum dot longer than the gate electrode is consistent with the effective gate length obtained by mobility considerations at room temperature (section 3.2).

For the batch B146 samples with 4 nm gate oxide the calculation slightly underestimates the peak spacing. For such thin oxides the retreat of the charge density of approximately 1.5 nm from the surface (see above) becomes important and the effective gate oxide thickness is increased. Thus, our calculations with 4 nm gate oxide overestimate the capacitance and underestimate the peak spacing. The planar capacitor approximation works better in this case because the omission of the corner charges partly compensates the increased effective gate oxide thickness.

Chapter 4

Addition spectrum of the Coulomb-blockade island

In chapter 3.4 the average Coulomb blockade peak spacing has been studied and compared to the simple model of the metallic single-electron transistor (SET, see chapter 2.5). We will now study its fluctuations. The peak spacing represents the addition spectrum, i.e. the energies needed to successively add electrons to the dot.

So far we have considered the quantum dot as perfectly metallic and the addition energy was simply given by the screened electron-electron interactions expressed through the geometric capacitance. Fluctuations of the peak spacing depend on the regime the dot is in (see figure 4.1). We will first consider the case of diffusive or ballistic quantum dots which is closest to the metallic SET. Then, in section 4.2, the case of low electron densities is considered, where disorder and interactions make the electrons behave like classical particles.

4.1 Diffusive dots

4.1.1 Existing theories and experiments

The first correction to the model of a metallic quantum dot as described in chapter 2.5 would be to consider individual quantum mechanical states with energy ϵ_i, but to describe the interactions by a constant capacitance as before. With some precautions we found expressions (2.14) and (2.38) for the energy in this case. Here we can simplify them because we will be interested only in the ground state properties where the energy levels are filled from bottom to top. Thus expression (2.14) becomes:

$$E(n) = \frac{e^2 n^2}{2C} + \sum_{i=1}^{n} \epsilon_i \tag{4.1}$$

For this notation we supposed the levels to be sorted $\epsilon_i \leq \epsilon_{i+1}$ and non-degenerate, i.e. degenerate levels are described by several levels with the same energy in order to simplify notation.

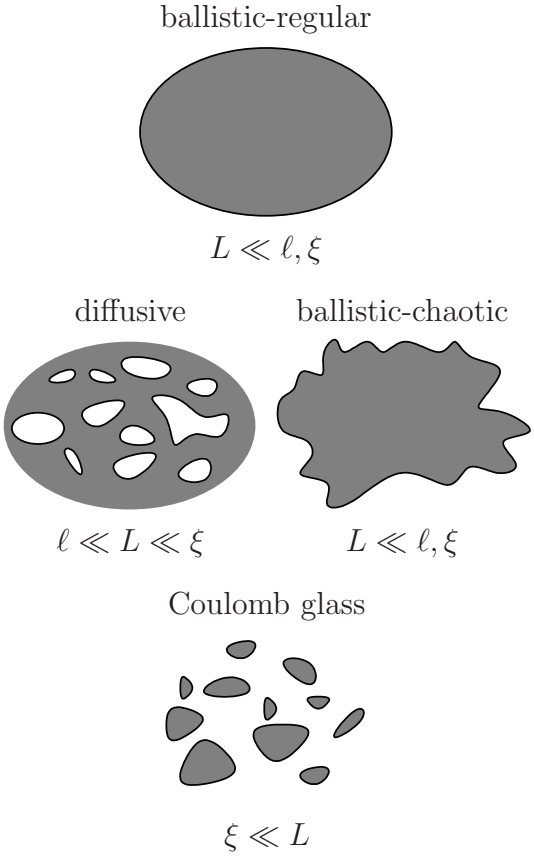

Figure 4.1: Different types of quantum dots. *L is the characteristic size of the dot, ℓ the elastic mean free path and ξ the localization length. In grayed zones the potential lies below the Fermi level. In Coulomb-glass and diffusive dots the disorder determines the properties and the boundary conditions play no important role. In ballistic dots the borders of the dot determine its nature. Diffusive and ballistic-chaotic dots show universal chaotic behavior for $k_F \ell \gg 1$ and $k_F L \gg 1$.*

The difference in chemical potential $\Delta_+(n)$ necessary to put an additional electron onto the dot is then

$$\Delta_+(n) = E(n+1) - 2E(n) + E(n-1) = \frac{e^2}{C} + \epsilon_{n+1} - \epsilon_n = \frac{e^2}{C} + \Delta_1(n) \qquad (4.2)$$

I.e. the charging energy $\frac{e^2}{C}$ and the single-particle level spacing Δ_1 have to be overcome in order to add an electron to the dot.

In this model the distribution of the addition energies is simply the distribution of the single-particle level spacings shifted by the charging energy.

For regular quantum dots (see figure 4.1) the single-particle levels have a shell structure due to the symmetries of the dot just like levels in atoms. However, real quantum dots are rarely perfectly symmetric. Disorder can change the boundaries of the dot; in this case the dot is a chaotic cavity. Or disorder can also be present inside the dot; electrons are then scattered inside the dot and the dot is called diffusive.

In diffusive or chaotic dots the wave functions can change dramatically with very little changes of the potential so that one cannot make predictions on the exact wave function. However, the statistics of the wave functions show generic properties which can be efficiently described by random matrix theory (RMT, see [9, 1]), where the Hamiltonian has to respect the system's space-time symmetries but is otherwise considered as random. The random matrix theory states that the single-particle level

spacing (scaled by the mean single-particle level spacing) follows the Wigner-Dyson distribution

$$w(x) = \tfrac{\pi}{2} x e^{-\frac{\pi}{4}x^2} \quad (x > 0) \tag{4.3}$$
$$w(x) = 0 \quad (x \leq 0) \tag{4.4}$$

with standard deviation

$$\sigma = \sqrt{\tfrac{4}{\pi} - 1} \approx 0.52 \tag{4.5}$$

We give here the result for time-reversal symmetry i.e. in the absence of magnetic field and spin-orbit coupling. The spin and band degeneracies add a δ-peak at 0 to the distribution of the single-particle level spacing. This combination of constant interaction model and random matrix theory is usually labeled CI+RMT.

On the experimental side, only few measurements of the peak spacing statistics of ballistic-chaotic semiconductor quantum dots were reported. They disagree on the width of the peak spacing distribution. In some experiments a peak spacing distribution scaling with $\frac{e^2}{C}$ was found [110, 109, 108], others agreed with RMT in finding a peak spacing distribution with a width of order Δ_1 [93, 81, 76, 96]. While they disagree on the actual width of the peak spacing distribution, they agree in finding more or less Gaussian-like distributions, in any case nothing similar to the shifted Wigner-Dyson-Distribution of the CI+RMT model which goes sharply to 0 for $\Delta_+ \to \frac{e^2}{C}$. References [108] and [96] study silicon quantum dots as we do, the others GaAs heterostructures. In both cases the mobilities are much higher than in our samples.

The fact that the CI+RMT model fails to describe the peak spacing statistics is not surprising. As argued in chapter 2.2, the constant interaction model is not a good approximation for discrete states occupying the same volume. This is so because the CI+RMT does not take into account all effects of order Δ_1, in particular it neglects exchange interaction and scrambling, i.e. the fact that the states of the electrons already in the dot change when an additional electron comes in.

Motivated by the experimental results, more detailed analytical calculations were carried out [13, 119, 120, 3]. The corrections to the RMT result depend on the dimensionless conductance $g_\Box = k_\mathrm{F} \min\{L, \ell\}$, supposed to be large compared to 1. The corrections to the CI+RMT result scale with $\frac{\Delta_1}{\sqrt{g_\Box}}$. Thus, the theories agree with RMT in that the width of the peak spacing distribution is of order Δ_1. The strong even/odd effect of the RMT result (i.e. the δ peak due to spin degeneracy) is significantly reduced when the electron density becomes low ($r_\mathrm{s} \approx 1$) [13] or temperature is increased above approximately $0.3\,\Delta_1$ [120]. The width of the peak-spacing distribution is only little affected by temperature: it reduces by approximately a factor of 2 when $k_\mathrm{B}T$ is increased above Δ_1 [3]. Numerical calculations [2, 57, 58] in chaotic ballistic dots confirm these results and find good agreements with some of the experimental results. Recently the limit of a diffusive dot has also been studied numerically [83]. The peak spacing distributions are found to be significantly larger ($\sigma(\Delta_+) \approx \Delta_1$) than in the ballistic case, and only very weak even/odd effects were found.

4.1.2 Material properties of the dot

The theoretical predictions of the peak spacing statistics depend on the material properties of the quantum dot. Here we will try to roughly estimate these properties. For most of the estimations we suppose the electrons to form a non-interacting degenerate Fermi gas. We calculate the material properties for 2 dimensions and 3 dimensions and only in view of these results we decide whether the dot is actually 2- or 3-dimensional.

mobility: At low doping the mobility as a function of temperature has a maximum. The position of the maximum depends on the doping level. Typically it occurs at approximately 50 K. Above the maximum, mobility is dominated by phonon scattering and the temperature dependence is $T^{-3/2}$ [7] or even stronger when optical phonons begin to play a role. Below, it is dominated by impurity scattering and the temperature dependence is $T^{3/2}$ [20]. This $T^{3/2}$ dependence is due to the fact that, at not too low temperature, electrons in the conduction band can be described as a Boltzmann gas. At very low temperature, however, there are only conduction electrons if the Fermi level is above the conduction band (mobility) edge. It is then the position of the Fermi level that defines mobility and not temperature.

When the doping level is high, conduction band and impurity band begin to hybridize. The fluctuations of the conduction band edge are then approximately 50 meV (the donor binding energy) and the Fermi level lies within this smearing. In this case the low temperature behavior where temperature is not the dominant parameter, is recovered already at room temperature. The transition to this high doping regime occurs around a dopant concentration of $10^{24}\,\mathrm{m}^{-3}$ (see figure 5 in ref. [54]) which is the doping level in the channel regions of our devices. Thus we consider the high temperature mobility (see chapter 3.2) to be relevant also at low temperature – as long as the Fermi level lies within the smearing of the conduction band edge. This is different from devices with very weak doping where the mobility can reach very high values at low temperature (up to $\mu \approx 10^4\,\mathrm{cm^2V^{-1}s^{-1}}$ in silicon and $\mu \approx 10^6\,\mathrm{cm^2V^{-1}s^{-1}}$ in GaAs heterostructures). According to our results in 3.2 we take

$$\mu \approx 100\,\mathrm{cm^2V^{-1}s^{-1}}. \tag{4.6}$$

The mobilities in batch 9857 are approximately twice as high, in batch B146 they are somewhat lower.

electron density: A typical dot shape is $W \times L \times H = 50\,\mathrm{nm} \times 40\,\mathrm{nm} \times 20\,\mathrm{nm}$. The surface area and the volume of the dot are then

$$A = (W + 2H)L = 3600\,\mathrm{nm}^2 \quad \text{(2D)} \tag{4.7}$$

$$V = W\,H\,L \quad = 4 \cdot 10^4\,\mathrm{nm}^2 \quad \text{(3D)} \tag{4.8}$$

For such a dot we count approximately 200 Coulomb blockade resonances between the first observable peaks and the highest gate voltages we can apply.

In batch 9857 there are approximately 20 dopants in the channel, in batch B146 split 2 of the order of 200. This number of dopants is also the number of electrons at $V_g \approx 0$ where we observe the first resonances. We take $n = 200$ as a typical electron number. This implies the following sheet and volume densities.

$$n_S = \frac{n}{A} \approx 5 \cdot 10^{-2}\,\mathrm{nm}^{-2} \quad \text{(2D)} \tag{4.9}$$

$$n_V = \frac{n}{V} \approx 5 \cdot 10^{-3}\,\mathrm{nm}^{-3} \quad \text{(3D)} \tag{4.10}$$

single-particle level spacing:

$$\Delta_1 = \frac{2\pi\hbar^2}{dm^* A} = 170\,\mu\mathrm{eV} \quad \text{(2D)} \tag{4.11}$$

$$\Delta_1 = \frac{2\pi\hbar^2}{d^{2/3}m^* V}(6\pi^2 n_V)^{-1/3} = 11\,\mu\mathrm{eV} \quad \text{(3D)} \tag{4.12}$$

We used the effective masses $m^* = m_t = 0.19\,m_e$ in 2D and $m^* = \sqrt[3]{m_t^2 m_l} = 0.32\,m_e$. The degeneracy is $d = d_{spin}d_{valley}$ with $d_{spin} = 2$ and $d_{valley} = 2$ in 2D [5] and $d_{valley} = 6$ in 3D (see appendix C).

Fermi wave number:

$$k_F = \sqrt{\frac{4\pi n_S}{d_{spin}d_{valley}}} \approx 0.4\,\mathrm{nm}^{-1} \quad \text{(2D)} \tag{4.13}$$

$$k_F = \sqrt[3]{\frac{6\pi^2 n_V}{d_{spin}d_{valley}}} \approx 0.3\,\mathrm{nm}^{-1} \quad \text{(3D)} \tag{4.14}$$

mean free path:

$$\ell = \frac{\hbar k_F \mu}{e} \approx 2\,\mathrm{nm} \tag{4.15}$$

The mean free path is much smaller than the typical device sizes of several 10 nm. The samples are diffusive.

disorder parameter:

$$k_F\ell \approx \begin{cases} 1.0 & \text{(2D)} \\ 0.6 & \text{(3D)} \end{cases} \tag{4.16}$$

As ℓ is smaller than the device size, $k_F\ell$ is the dimensionless or Thouless conductance $g_\square$.

gas parameter: The gas parameter is the Wigner Seitz radius, i.e. the radius of a hypersphere with volume n^{-1}, normalized to the effective Bohr radius. It measures how strong the Coulomb interaction is compared to the kinetic Fermi energy.

$$r_s = (a_B\sqrt{\pi n_S})^{-1} \approx 0.8 \quad \text{(2D)} \tag{4.17}$$

$$r_s = \left(a_B\sqrt[3]{\tfrac{4}{3}\pi n_V}\right)^{-1} \approx 1.1 \quad \text{(3D)} \tag{4.18}$$

$a_B \approx 3.3\,\mathrm{nm}$ is the effective Bohr radius in silicon. It is derived in chapter 5.2. For a 2DEG very close to the Si/SiO_2 interface $a_B \approx 2.1\,\mathrm{nm}$ due to the lower dielectric constant in SiO_2 [5]. With this value $r_s \approx 1.2$ in the 2D case.

inverse screening length: At low electron densities ($r_s \gg 1$) the screening length is the distance between electrons

$$\lambda = \frac{1}{a_B r_s} = \begin{cases} \sqrt{\pi n_S} & = 0.4\,\mathrm{nm}^{-1} \quad (2D) \\ \sqrt[3]{\frac{4}{3}\pi n_V} & = 0.3\,\mathrm{nm}^{-1} \quad (3D) \end{cases} \tag{4.19}$$

At high electron densities ($r_s \ll 1$) the screening length is given by the density of states (see expression (5.24)).

$$\lambda = \frac{e^2}{2\epsilon_0 \epsilon_{Si} \Delta_1 A} = \frac{d}{a_B} = 1.2\,\mathrm{nm}^{-1} \quad (2D) \tag{4.20}$$

$$\lambda = \sqrt{\frac{e^2}{\epsilon_0 \epsilon_{Si} \Delta_1 V}} \qquad = 1.8\,\mathrm{nm}^{-1} \quad (3D) \tag{4.21}$$

In 2D the two results are equal for $r_s = d^{-1} = 0.25$. Therefore we rather take the screening length $\lambda \approx 0.4\,\mathrm{nm}^{-1}$ for high r_s. Screening due to the gate electrode is weaker. The gate oxide thickness can be as low as 4 nm but the dielectric constant in the oxide is 3 times lower than in silicon. The effective distance between gate electrode and channel electrons is therefore 12 nm (or more in batch 9857). This is larger than $a_B r_s \approx 3\,\mathrm{nm}$, the mean distance between electrons.

dimension and degeneracy: The As doped channels in our devices are only weakly compensated: on average there are approximately 0.1 acceptors in the channel. The thickness of the electron gas is therefore the screening length of the electrons in the disorder potential of the donors. This is different from a standard n-channel MOS structure working in inversion where the space charge due to charged acceptors defines the band bending and in consequence the thickness of the inversion layer [5].

In batch B146 split 2, the doping level is comparable to the number of electrons attracted by the gate electrode. The screening length therefore does not vary considerably as a function of gate voltage. Taking the screening length $\lambda^{-1} \approx 2.5\,\mathrm{nm}$ as thickness H in the 3D level spacing (4.12), we obtain the same result as in the 2D case, i.e. $\Delta_1 \approx 170\,\mu\mathrm{eV}$. In batch 9857 the doping is lower so that at low electron numbers the screening length could be considerably higher. Another important parameter is the doping profile. In batch B146-2 we do not know it precisely. For batch 9857 however, the doping profile has been simulated and it actually seems that the doping is low close to the gate electrode (see figure 3.5). This gradient could compensate the electric field created by the gate electrode and make the bottom of the potential well flatter than it would be with homogeneous doping. Therefore we can not

determine precisely the thickness of the electron gas. We use the 2D result. The resulting single-particle level spacing Δ_1 must then be seen as an upper limit, in particular in batch 9857. The decrease of Δ_1 due to the thickness of the electron gas can be interpreted as multiple 2D sub-bands, increasing the degeneracy factor d.

The gas parameter in our samples is comparable to the ones in the existing experiments. The disorder parameter, however, is very different. While the samples in previous experiments are ballistic, our samples are very diffusive ($\ell \ll L$). The parameter $k_F\ell \approx 1$ indicates that our samples are at the limit between localized Coulomb glass regime and diffusive regime (see figure 4.1).

4.1.3 Experimental results

For this analysis of the peak-spacing statistics we use the samples of batch 9857 and of batch B146 split 2 at high gate voltage where the dots are in the diffusive regime. In most of the samples of batch B146 split 1 the access conductances are above the quantum of conductance because their doping is high (see chapter 3.1) so that Coulomb blockade can only be observed when only few electrons are in the dot (see section 4.2).

First, the positions of the Coulomb blockade peaks are detected with the algorithm described in appendix B.2. The peak spacings plotted in figure 4.2 are the differences between the detected peak positions.

As figure 4.2 testifies, long-range evolutions of the peak spacing are smaller than the short range fluctuations. This allows us to analyze the peak spacing distribution directly. In most other experimental studies of the peak spacing statistics only the differences of successive peak spacings could be analyzed because the average peak spacing changed with gate voltage.[1] The result of such an analysis depends on the correlation of successive peak spacings. While at low temperature theory predicts an anti-correlation (i.e. an even/odd effect due to spin degeneracy (see below)) this effect is washed out for $k_BT \gtrapprox 0.3\Delta_1$ and could actually turn into a positive correlation when due to thermal smearing each state contributes to several Coulomb blockade peaks.

In batch B146-2 there are many samples where the conductance decreases considerably between 4 K and base temperature so that Coulomb blockade peaks are not visible at very low temperature. In addition strong anomalies appear which create tails in the peak spacing distribution. We explain this by the fact that the access regions are very close to the metal–insulator transition. (The estimated doping level ($10^{19}\,\mathrm{cm}^{-3}$) in the access regions of batch B146-2 is very close to the critical concentration for the metal–insulator transition in doped silicon ($6.4 \cdot 10^{18}\,\mathrm{cm}^{-3}$ for bulk Si:As [19])). At low temperature the charges become localized: the conductance becomes very weak and the localized charges in the barriers form offset charges.

[1] The size and capacitance of dots in 2DEGs is defined by the voltage on the depletion gates whereas our dot is limited on 4 sides by the Si–SiO_2 interface.

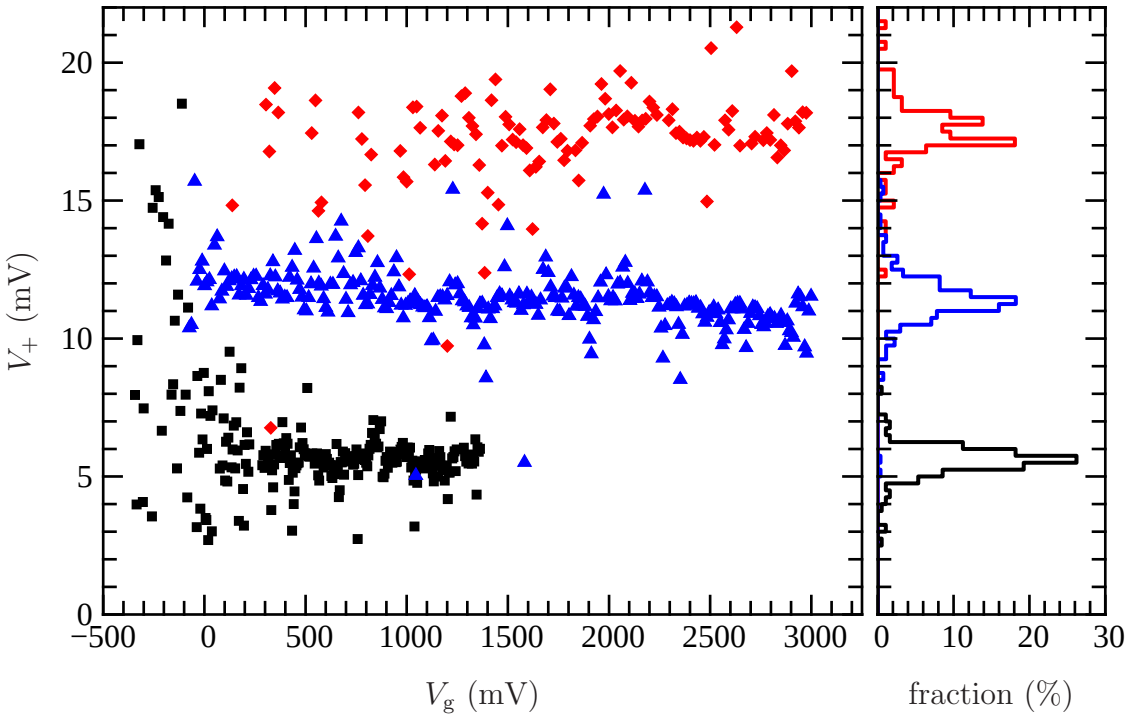

Figure 4.2: Peak-spacing distributions at 4.2 K. *The left panel shows the evolution of the peak spacing with gate voltage for 3 different samples of comparable surface area* $\approx 6000\,\mathrm{nm}^2$ *but different gate oxide thicknesses:* $t_{ox} = 4\,\mathrm{nm}$ *(sample 12,* ■*),* $t_{ox} = 10\,\mathrm{nm}$ *(sample 1,* ▲*),* $t_{ox} = 24\,\mathrm{nm}$ *(sample 4,* ♦*). The right panel shows the corresponding histograms.*

This low doping concerns not only the wire but the whole system up to the metallic contacts. The spacer regions in batch 9857 play the same role. Their doping significantly below the critical concentration but they are much shorter: only the zones below the spacers have low doping. Therefore the effect is better controlled and the temperature dependence is weak below 4 K. We will discuss these phenomena in chapter 5.

Due to these effects the peak spacing can in many samples only be analyzed at 4 K or 1 K but not at base temperature. When we are able to go to very low temperature, we find no significant reduction of the width of the distribution between base temperature ($T \approx 300\,\mathrm{mK}$ effective electronic temperature[2]) and 4.2 K. If we take the 2D result for Δ_1 we get at base temperature $k_B T \approx 0.2\Delta_1$. Under this condition ripples due to excited states should be visible inside the Coulomb blockade diamonds. Usually we do not observe them at high electron density. This could have two explanations.

- The $I(V)$ characteristics of the access regions in batch B146 are strongly nonlinear so that the high conductance at high bias voltage and dissipative pro-

[2]In other types of samples we observe rather 100 mK effective electronic temperature. See appendix A.2.5

cesses could smear out the excited states. Or the strong variations of the drain-source conductance could simply mask the ripples.

- The 2D mean single-particle level spacing is an upper limit and the actual mean single-particle level spacing could be smaller. So we could actually not get electronic temperatures as low as $k_BT \approx 0.2\Delta_1$. This would mean that we never reach the limit where only one quantum mechanical state is occupied. Indeed, as soon as only one state is occupied, the electrons cannot evacuate energy from the dot any more and cooling becomes difficult. Therefore the electronic temperature could be pinned to $k_BT \approx \Delta_1$. An indication for this pinning could be that the effective electronic temperature seems highest in the smallest dots where Δ_1 is largest.

The second explanation is supported by the fact that Patel *et al.* [93] find a stronger increase of the peak spacing fluctuations when k_BT falls below Δ_1. Thus, even at base temperature we probably stay in the regime $k_BT > \Delta_1$.

The fact that our samples are fabricated on 200 mm wavers in an industrial process allows to analyze the peak spacing statistics of many samples having exactly the same fabrication parameters. We have analyzed the width of the peak spacing distribution by fitting its histogram with a Gaussian. We do so rather than calculating directly the variance of the distribution because of the tails in the distribution. They are caused by occasional anomalous regions with very large or small spacings due to charging events in the barriers (see above). These tails in the distribution would significantly increase the estimated variance but affect much less the histogram fit.

In order to translate the peak spacing into addition energy one has to multiply by $e\alpha$. In batch B146, the factor α extracted from Coulomb blockade diamonds (see appendix B.3) or the width of the Coulomb blockade resonances is close to 0.4. This is in good agreement with what we found at high temperature (see page 28). There are some exceptions, however, and the α factor is higher for the first resonances where the dot decomposes into several "puddles" not covering the whole surface (see below). In our analysis we exclude both, unusual samples and the first resonances.

In batch 9857 the determination of the lever-arm factor α is more involved because, as figure 4.3 shows, it has strong sample-to-sample fluctuations and depends on gate voltage (we will analyze the origin of this dependence in chapter 5.4). We do not take into account the evolution of the lever arm with gate voltage but rather take a constant lever arm factor 0.1 for all 9857 samples. As figure 4.3 shows, the actual lever arm factors beyond 1000 mV vary between approximately 0.05 and 0.2.

Finally we normalize the fluctuations of the addition energy by the 2D single-particle level spacing (4.11) with degeneracy $d = d_{\text{spin}}d_{\text{valley}} = 4$ and the specific dimensions of each sample. The result of this analysis is shown in figure 4.4. The width of this addition-energy distribution $\sigma(\Delta_+)$ is in good approximation given by the 2D single-particle level spacing. The deviations are within $\pm 50\,\%$ of the level spacing, whereas the surface, and in consequence the level spacing, are varied by a

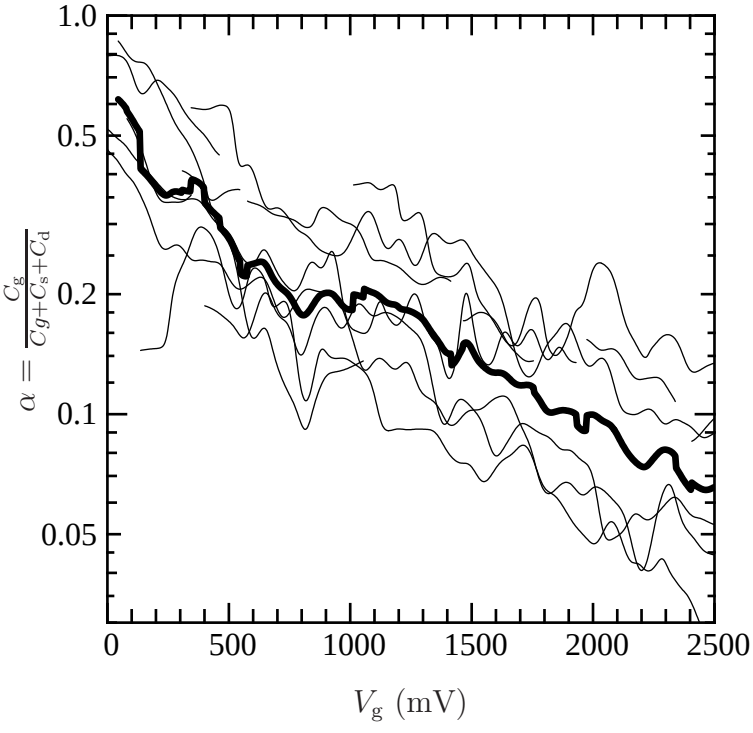

Figure 4.3: Evolution of gate control in batch 9857 with gate voltage. *The gate voltage lever arm factor* $\alpha = \frac{C_g}{C_g+C_s+C_d}$ *has been extracted by analyzing the slopes of the Coulomb blockade diamonds (see appendix B.3) for different samples. The detected slopes have been filtered with a* 30 mV *wide Gaussian filter to obtain the slopes as a continuous function of gate voltage (thin lines). The thick line is the geometric sample average of the detected lever arm factors.*

factor of almost 5. The result $\sigma(\Delta_+) \approx \Delta_1$ is in good agreement with the recent theoretical results cited above, especially [83].

Despite this good agreement however, the scaling of the fluctuations with the inverse surface area does not necessarily exclude scaling with the charging energy $\frac{e^2}{C}$: as we showed in chapter 3.4, the (gate) capacitance scales also with the surface area, i.e. $C \sim A$. In order to exclude this possibility, the capacitance has to be varied without changing the size of the dot. The increase of the gate oxide thickness between batch B146 (4 nm) and batch 9857 (10 nm and 24 nm) could provide this. Unfortunately, only the gate capacitance changes by a factor 6. The change in total capacitance is not significant because the total capacitance is dominated by the source and drain capacitances (the lever arm factor is small). If we call $C_{g,0}$ the gate capacitance in a batch B146 sample, its total capacitance is $\alpha^{-1}C_{g,0} = 2.5\,C_{g,0}$. In a batch 9857 sample with corresponding gate–wire overlap the gate capacitance is 6 times smaller due to the thicker oxide but the lever arm factor is only 0.1 instead of 0.4, so that the total capacitance $\alpha^{-1}C_{g,0}/6 = 1.7\,C_{g,0}$ is very similar to that in batch B146.

In addition, the fabrication process is not exactly the same in batch B146 and batch 9857. In particular the lower channel doping in batch 9857 could lead to a thicker electron layer in batch 9857 than in batch B146. Therefore, one must be careful when interpreting differences between the two batches.

Actually there are long-range fluctuations of the addition energies in batch 9857, and there are even zones where they dominate the short-range fluctuations (for example in the red curve in figure 4.2 between $V_g = 2000$ mV and 2500 mV). This

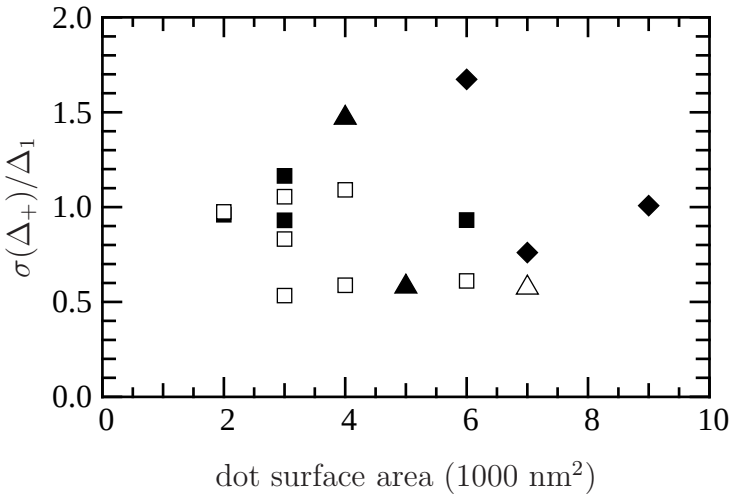

Figure 4.4: Normalized addition-energy fluctuations. *Fluctuations of the addition energy are close to the mean single-particle level spacing* Δ_1*. Data taken at base temperature (filled symbols) and* 4.2 K *(empty symbols). Oxide thickness* 4 nm *(batch B146,* □*),* 10 nm *(batch 9857,* △*) and* 24 nm *(batch 9857,* ◊*).*

could indicate that the main contribution to the fluctuations is always caused by the barriers (see chapter 5) and that the peak spacing fluctuations caused by the dot itself are very small. This would be the case when the dot is 3-dimensional as suspected for batch 9857 in section 4.1.2. In batch B146 the short range fluctuations are always dominant.

4.2 Coulomb-glass dots

In the analysis above we only considered samples of batch 9857 and batch B146, split 2. We left out the first resonances because the peak-spacing fluctuations are very strong there (see figure 4.2). In batch B146 split 1, this effect is even more pronounced (see figure 4.5). Indeed, in this example the fluctuations are of the order of the mean level spacing which in turn is 2.5 times higher than expected from geometry (see chapter 3.4).

In this split the doping of the access region is high and the channel, i.e. the part of the wire below the gate electrode, is not doped (see chapter 3.1). So the first electrons entering the dot can be observed (see figure 4.6). The parameters we calculated in section 4.1.2 are not valid in this regime. The electron density is very weak so that $k_\mathrm{F}\ell$ is much smaller than above, so that $k_\mathrm{F}\ell \ll 1$ and $r_\mathrm{s} \gg 1$. Both figures indicate that the dot is in the localized regime (see figure 4.1) where electron wave functions do not spread over the whole dot but are localized in parts of the dot where the disorder potential is particularly low.

This interpretation is supported by the fact that in batch B146 split 1 the doping of the access regions is high and there are not, as in batch 9857, spacer regions that form barriers. Consistently, at high gate voltage the conductance is of the order of $G_\mathrm{peak} \approx \frac{e^2}{h}$ or above so that often universal conductance fluctuations rather than Coulomb blockade oscillations are observed. This indicates that the access regions do not have sufficient resistance to allow for Coulomb blockade and that it is the resistance of the dot itself that dominates. A conductance $G < \frac{e^2}{h}$ then indicates

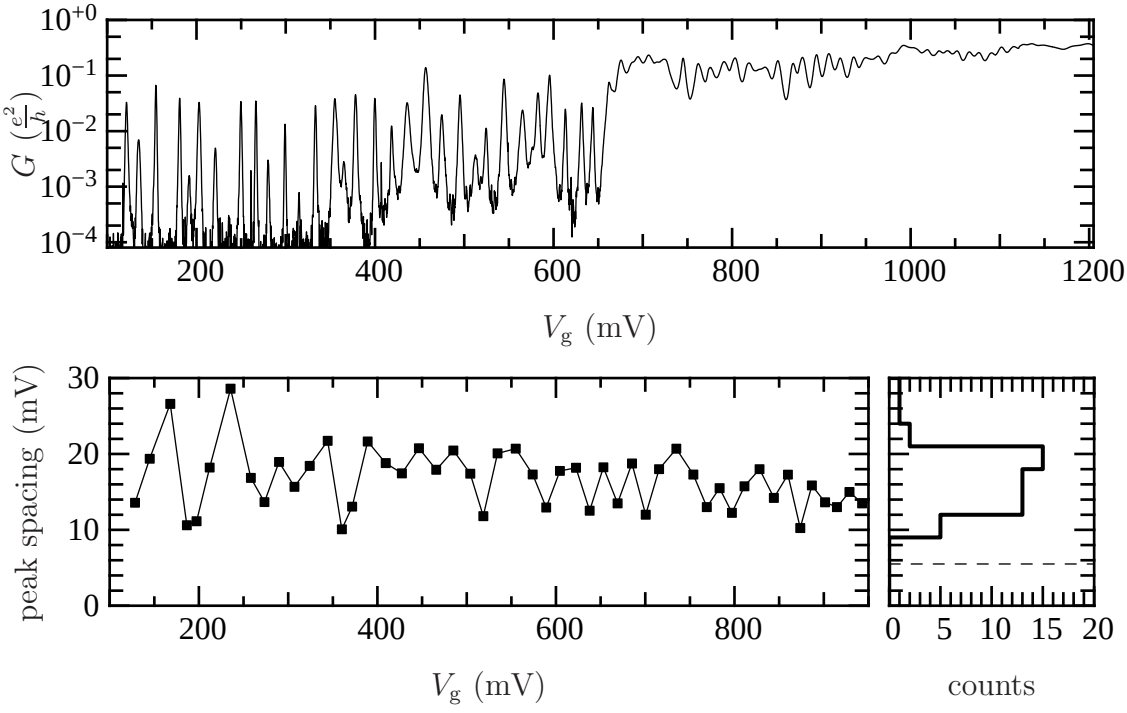

Figure 4.5: Peak-spacing fluctuations when the dot is in the Coulomb glass regime. *The upper panel shows the Coulomb blockade spectrum from which the peak-spacing distribution in the lower panel has been extracted. The fluctuations of the peak spacing have the same order of magnitude as the mean peak spacing. The average peak spacing is strongly increased with respect to a diffusive dot of the same size: the dashed line in the histogram panel (bottom-right) is the peak spacing calculated from the sample geometry. Sample 9.*

that the localization length ξ in the dot is shorter than its length L, i.e. that the dot is in the Coulomb glass regime (see figure 4.1).

The peak spacing of a quantum dot in the localized regime has been considered theoretically by Koulakov, Pikus and Shklovskii [66]. They predict peak-spacing fluctuations of the order of the mean peak spacing when electrons are completely localized (localization length $\xi \ll L$). The crossover to the metallic regime is explained by the increase of the dielectric constant when ξ begins to increase (see chapter 5.4). As long as $\xi < L$, the peak-spacing fluctuations decrease: $\sigma(\Delta_+) \propto \frac{1}{\epsilon} \propto \frac{1}{\lambda^2\xi^2}$ until $\xi = L$ where they saturate at $\sigma(\Delta_+) \propto \frac{1}{\lambda^2 L^2} \propto \Delta_1$.

In the Coulomb-glass regime the evolution of the Coulomb-blockade peaks under magnetic field is spectacular, as shown in figure 4.7. The peaks shift significantly with magnetic field applied in parallel to the nanowire. These shifts differ in direction and amplitude. Very similar effects have been observed in silicon nanowires with comparable properties [95]. The shifts were attributed to the Zeeman effect, the amplitude of the shift V_+ being

$$e\alpha V_+ = g\mu_B \Delta S B \tag{4.22}$$

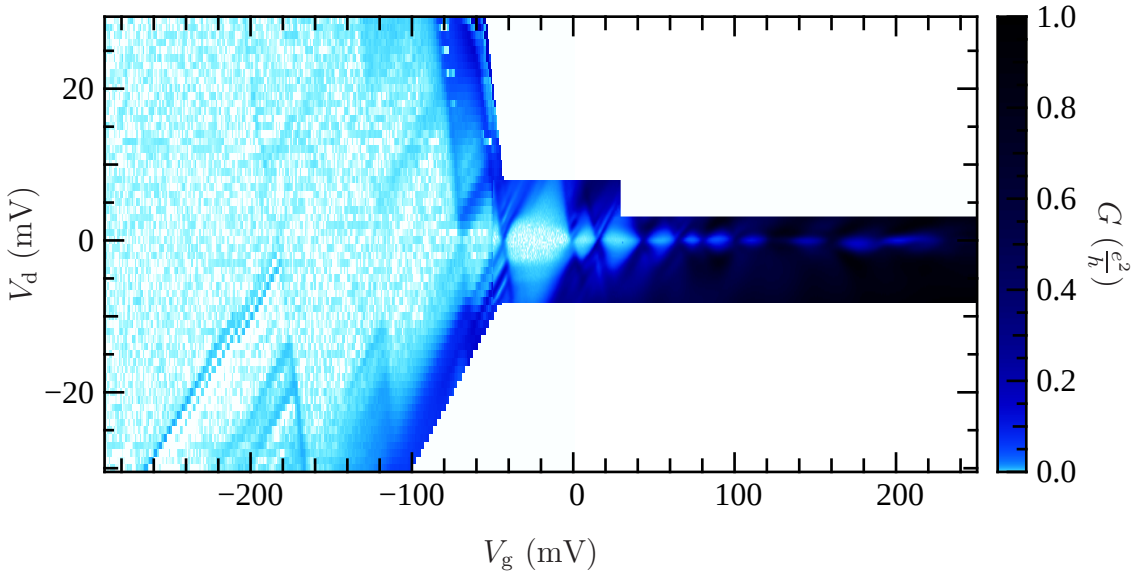

Figure 4.6: A batch B146-1 sample at low gate voltage. *The excited states of the first resonance can be well observed and indicate that it is due to one of the first electrons in the dot. The conductance soon becomes very strong (inelastic co-tunneling is clearly visible). Sample 8.*

where ΔS is the change in spin polarization due to the electron added to the dot, μ_B the Bohr magneton and g the Landé factor, which is very close to 2 for electron spins in silicon [24]. ΔS is $\pm\frac{1}{2}$ when the spins of the electrons already in the dot are not rearranged. In [95] different amplitudes of the Zeeman shifts were attributed to higher spin shifts $\Delta S = \pm\frac{3}{2}$. Surprisingly, at large electron numbers the corresponding resonances were found to be as strong as the others. Yet, they should be strongly suppressed because the rearrangement of the spins is a slow process and should therefore cause a "spin blockade" [125, 126]. In our dots all shifts seem to agree with $\Delta S = \pm\frac{1}{2}$. We explain the different amplitudes of the shifts by a fluctuating α factor. This fluctuation is caused by the source and drain capacitances. In chapter 5 we will study them in detail. We observe that resonances showing a strong shift are systematically wider and according to equations (2.35) or (2.40), the width of the resonances is inversely proportional to α so that wide resonances also show a strong Zeeman shift. For the first peaks we can also obtain the α-factor by analyzing the slopes of the Coulomb blockade diamonds (see appendix B.3). At least up to $V_g \approx 600\,\mathrm{mV}$ we verify that all resonances are compatible with $\Delta S = \pm\frac{1}{2}$. Above, the diamonds are not sufficiently sharp to extract the slopes, and the life-time broadening of the lines becomes important so that we cannot determine α from the peak width.

In the Coulomb glass regime it is not surprising to observe a strongly fluctuating lever-arm factor because electron states are localized in different parts of the dot. Parts of the dot closer to the source and drain regions will have a much weaker gate control and α factor than the central parts. In the dots of batch 9857 we also observe

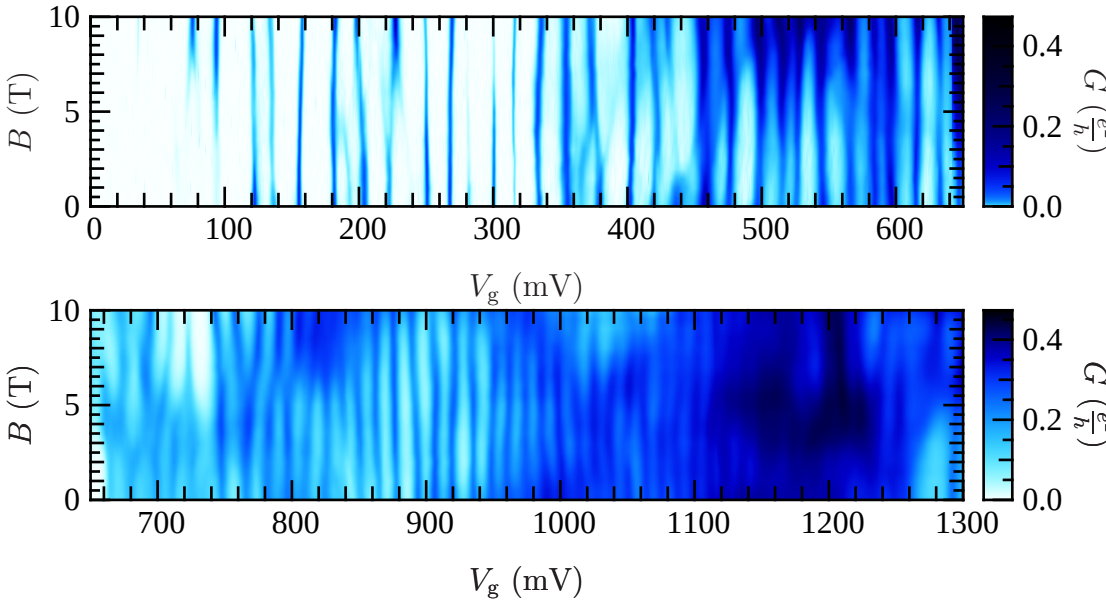

Figure 4.7: Evolution of the conductance peaks under magnetic field. *Coulomb blockade peaks split or merge under magnetic field (e.g. $V_g = 770\,\mathrm{mV}$). This can be explained by a Coulomb glass island (see text). Sample 9.*

shifts of the Coulomb blockade resonances with magnetic field with an amplitude compatible with the Zeeman effect. However, the shifts are correlated over a large number of peaks so that we rather attribute them to effects of the barriers discussed in the first part of chapter 5.

The second, even more surprising feature of figure 4.7 is that some resonances split or merge with magnetic field. This occurs for example at $V_g = 70\,\mathrm{mV}$, 180 mV, 220 mV, 370 mV, 760 mV. Such features could indicate double resonances, i.e. that, depending on magnetic field, it is energetically favorable to pass directly from a state with n electrons to a state with $n+2$ electrons as gate voltage is increased. Indeed, Raikh *et al.* [94] determine with purely electrostatic arguments that in a Coulomb glass dot approximately 5 % of the resonances should be double resonances. We find at least 5 out of 40 resonances up to $V_g = 800\,\mathrm{mV}$ to be double resonances. This is somewhat more than predicted. However, as Raikh *et al.* argue, double resonances should be suppressed because they are second order processes. What we observe is the contrary: The resonances become stronger once they have joined. And this effect is even particularly strong when conductance is low. This indicates that the joined line is the first order process and not the split lines. Thus the following interpretation.[3] Suppose there are at least two electron puddles A and B in the dot, A having the lower charging energy but the last electron in B being very close to

[3]The argument is the same for the trap states in the barriers which we discuss in chapter 5. As the effect can be more clearly observed for them, we give here only a sketch of the argument and refer to chapter 5 for details. There puddle A is replaced by the entire quantum dot and puddle B by a dopant site in the barrier.

the Fermi level. Due to the weaker charging energy, the next electron is added to A but then the Coulomb interaction pushes out the last electron of B. This can either be an internal transition with an electron transferred from B to A or a second order effect with a coherent transfer of an electron from the reservoirs to A and from B to the reservoirs. The former process does not cause any current in the leads, and the latter is slow so that the corresponding resonance is indeed weak. The second split line would be due to the electron of puddle B being re-attracted. The lines join when, due to magnetic field, the last level of puddle B is lowered or raised with respect to the last level of A so that an electron can enter puddle A without changing the occupation of puddle B.

In batch B146 split 2, the signature of the first Coulomb blockade peaks (in some samples even of all peaks) is similar to what we presented here. For example the fluctuations of the black trace in figure 4.2 increase dramatically at low gate voltage. This is why we did not include them in our analysis of diffusive dots.

4.3 Even/odd effects

According to theory, the peak spacing distribution should show an even/odd effect due to spin degeneracy. The corrections beyond CI+RMT weaken them but the even/odd effect should still be important at low temperature. However, to our knowledge they have not yet been clearly observed in quantum dots with many electrons (~ 100). This is most probably due to a too high electron temperature. According to [120], $k_\mathrm{B}T$ has to be lower than $0.3\,\Delta_1$ for a pronounced even/odd effect. To our knowledge, this condition has not yet been fulfilled experimentally for large electron numbers. In silicon there is, in addition to the spin degeneracy, a band degeneracy of 2 in 2D or 6 in 3D. This can cause an additional even/odd effect in the peak-spacing statistics [40].

Some of our samples have a bimodal peak spacing. For example, the addition spectrum of the Coulomb glass dot presented in figure 4.5 shows a pronounced even/odd effect. Figure 4.8 shows a striking example of a sample in the diffusive regime with strong even/odd effect over the whole gate voltage range. Only some anomalous regions, where the peak spacing is increased or reduced, do not show it. In most other samples, if there are any at all, even/odd effects are only visible on the first peaks or in certain gate-voltage ranges.

What is the origin of the even/odd effect? As the lower panel of figure 4.8 shows, it is not affected by magnetic field. This excludes the spin degeneracy as origin. We think that the valley degeneracy in silicon is not the origin either, because our samples are very disordered and the orbital degeneracy should therefore be lifted. Even if an even/odd effect remained, it seems unlikely to have perfect even/odd sequences over several peaks. This is supported by a recent simulation of diffusive dots with short range disorder (Anderson Hamiltonian) showing that the even/odd effect due to the spin degeneracy is surprisingly weak in strongly diffusive dots [83]. This tendency could also apply to the band degeneracy in silicon.

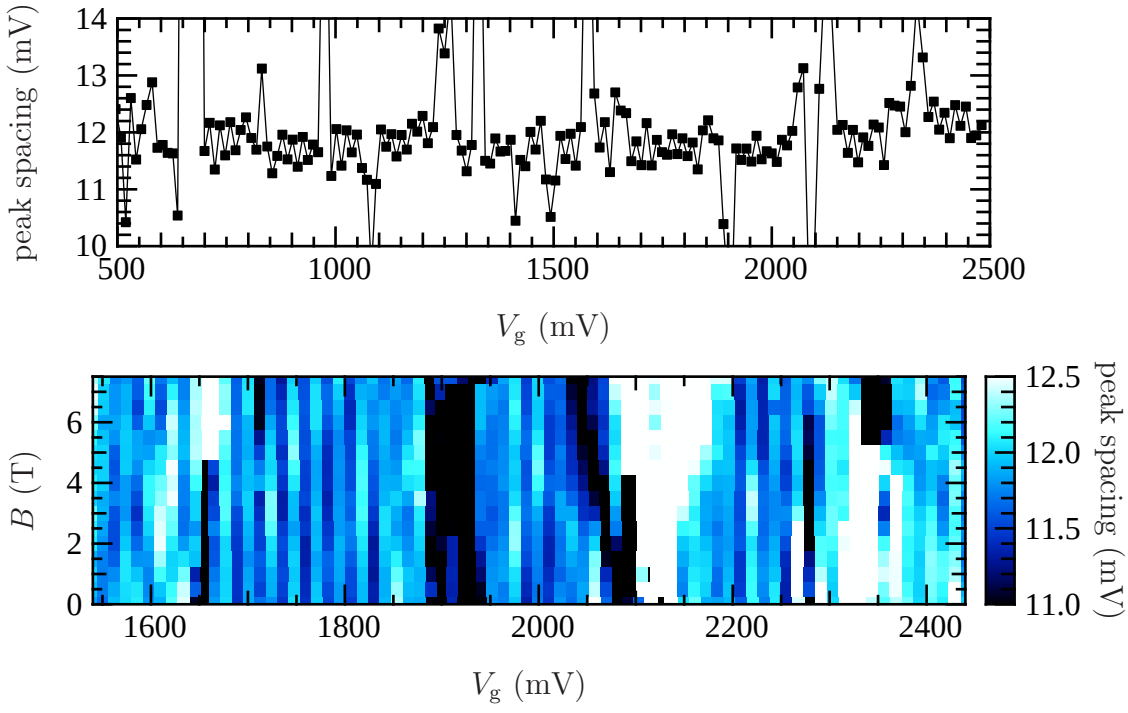

Figure 4.8: Color plot of the peak spacing under magnetic field. *The peak spacing is bimodal (best visible near $V_g = 1800\,\mathrm{mV}$) The even/odd effect is unchanged under magnetic field. The strong fluctuations where the color plot saturates are explained in chapter 5. Sample 7.*

Spin and valley degeneracies cause even/odd effects in the peak spacing of the order of $\frac{\Delta_1}{e\alpha}$. So they cannot explain the even/odd effect in the Coulomb glass sample where the fluctuations and the even/odd effect are of the order of the mean peak spacing (see figure 4.5).

We attribute the even/odd effect rather to the geometry of our samples. Our wires have a rectangular or trapezoidal cross section with the gate electrode covering the three upper sides (see chapter 3.1). The electric field created by the gate electrode and thus the electron density are highest in the corners. In the simulation with a metallic wire (see figure 3.12) there are even singularities in the corners. The screening length is between 2.5 nm and 10 nm so that the charge density should not fall off significantly between the upper and lower corner of each side. However,if the wire is sufficiently wide, the electron density could be considerably higher in the corners and on the sides of the wire than on top of it. At low gate voltages the sides of the wire could therefore already be in the diffusive regime whereas the top and the core of the wire are still in the localized Coulomb glass regime. The wire then forms a double dot system where two quantum dots with similar properties are in parallel. Double dots where the tunneling strength can be tuned showed such an even/odd effect [124]. When the tunnel-coupling between the dots is switched on, the Coulomb blockade resonances split. As the strength of tunneling is increased the splitting increases until eventually the splitting is half the original Coulomb

blockade peak spacing: the two dots have turned into one single dot with a doubled capacitance and a twice smaller peak spacing. For intermediate coupling such a system shows a robust even/odd effect of the order of the mean peak spacing.

The fact that we observe the even/odd effect in the Coulomb glass regime and in a particularly large sample is consistent with this explanation. However, it is not clear why the even/odd effect persists up to so high gate voltages in the large diffusive sample. Perhaps there is a strong defect in the center of the channel that reduces the coupling between the two sides of the dot even at high gate voltage.

4.4 Conclusion

We have analyzed the distribution of the Coulomb blockade peak spacing in very disordered samples. We identified two regimes. When few electrons are in the dot, the dot is in a glassy state and the peak spacing fluctuations are of the order of the mean Coulomb blockade peak spacing. When the electron number is high, our dots are in the diffusive regime and we find fluctuations of the addition energy between 0.5 and 1.5 times the single-particle level spacing, in agreement with recent numerical simulations. Our results are complementary to existing experiments for ballistic chaotic dots. Some strong fluctuations of the peak spacing in diffusive dots are not explained by the properties of the dot. We attribute them to traps in the barriers as discussed in the next chapter.

In several samples we find even/odd effects in the spacing of the first Coulomb blockade peaks, in some samples it persists even at high gate voltage. We do not attribute this effect to spin or valley degeneracy in silicon, but to our device geometry where the dot could be decomposed into two interacting dots.

Chapter 5

Disordered insulators as Coulomb-blockade barriers

In chapter 3 we found the quantum dot in our system to be separated from the source and drain reservoirs by weakly doped parts of the wire with low conductance. In this chapter we will investigate these access regions. First we shall consider individual dopant states. We will see that they are the most plausible explanation for anomalies in the Coulomb blockade spectrum. Afterwords we will analyze the barriers as a whole and investigate the relation between their conductance and their capacitance.

5.1 Type of the conductance through the barriers at low temperature

As described in chapter 3, there are barriers below the spacers that separate the dot from source and drain — at least in batch 9857 samples. We have estimated their height to 20 meV near threshold. How do electrons pass these barriers at low temperature? To answer this question we will first estimate the number of conduction channels in the central part of the wire. At high gate voltage the quantum dot contains typically up to 200 electrons for a size of $L \times W \times H = 40\,\mathrm{nm} \times 50\,\mathrm{nm} \times 20\,\mathrm{nm}$. This corresponds to an electron density $n_\mathrm{V} = 5 \cdot 10^{24}\,\mathrm{m}^{-3}$, approximately one order of magnitude higher than the doping level in the access regions in batch 9857. This electron density opens a number of channels

$$n_\mathrm{ch} = \frac{k_\mathrm{F}^2}{4\pi^2} WH = (d_\mathrm{spin} d_\mathrm{valley})^{1/3} \left(\frac{3n_\mathrm{V}}{4\pi}\right)^{2/3} WH \approx 0.88\, WH n_\mathrm{V}^{2/3} \approx 26 \qquad (5.1)$$

We use $d_\mathrm{spin} = 2$ and $d_\mathrm{valley} = 6$ for the spin and valley degeneracy. As we evaluate n_ch at the highest gate voltage, this number has to be seen as an upper limit.

Below $T = 4\,\mathrm{K}$ the transmission for thermally activated transport over the barriers is too weak to be measurable:

$$\mathcal{T}_\mathrm{th} = \int_{E_\mathrm{barrier}}^{\infty} f\left(\frac{E}{k_\mathrm{B}T}\right) \frac{\mathrm{d}n_\mathrm{ch}}{\mathrm{d}E} \mathrm{d}E \approx \int_{E_\mathrm{barrier}}^{\infty} \exp\left(-\frac{E}{k_\mathrm{B}T}\right) \frac{m^* WH}{2\pi\hbar^2} \mathrm{d}E < 10^{-25} \qquad (5.2)$$

We used $m^* = m_e$ for the effective mass and $E_{barrier} = 20\,meV$ for the barrier height relative to the Fermi level. The fact that thermally activated transport plays no role in batch 9857 is supported by the fact that between 4 K and base temperature the peak height does not evolve significantly[1] (see figure 3.7).

As electrons cannot be activated over the barriers they must tunnel through them. For direct tunneling, the tunneling transmission of each channel is

$$\mathcal{T} \approx e^{-2\kappa L} \tag{5.3}$$

where L is the thickness of the barrier and $\kappa = \sqrt{2m^* E_{barrier}}/\hbar$. For the effective mass we take the smallest value $m^* = m_t = 0.19\,m_e$ resulting in the highest transmissions. At high gate voltage we observe peak conductances above $0.1\,\frac{e^2}{h}$. According to expression (2.35) this implies barrier conductances above $0.4\,\frac{e^2}{h}$, or with 26 channels a transmission per channel $\mathcal{T} > 0.015$. To obtain such high transparencies the barriers must be as short as $L = 10\,nm$ even for barrier heights as low as $E_{barrier} = 1\,meV$.

Yet, as we found in chapter 3, the barriers are rather 35 nm long and have a height in the order of 20 meV. So direct tunneling is excluded as well.

Thus transport must occur by resonant tunneling through the barriers. As possible localized states for resonant tunneling, there are some tens (batch 9857) or even hundreds (batch B146) of dopants in each barrier. We will elaborate this model of resonant tunneling through localized dopant states in the remainder of this chapter.

5.2 Donor states in silicon

Shallow donors in group IV semiconductors, such as silicon, are elements of group V replacing a host atom in the crystal. For silicon arsenic and phosphorus are most frequently used. Only 4 of the 5 outer electrons in the neutral state of such donor atoms participate in the crystal bonds and inside the crystal the donor becomes singly positively charged. The potential seen from the conduction electrons beyond a few lattice constants is

$$U(r) = -\frac{1}{4\pi\epsilon_0\epsilon_{Si}}\frac{e^2}{r} \tag{5.4}$$

where r is the distance from the dopant center and $\epsilon_{Si} = 11.9$ the dielectric constant of silicon. In the effective mass approximation dopants can thus be considered as hydrogen atoms with renormalized dielectric constant and electron mass and the binding energy is

$$E_1 = -\frac{e^4 m^*}{8\epsilon_0^2\epsilon_{Si}^2 h^2} \approx -40\,meV \tag{5.5}$$

For this rough estimation we used an isotropic effective mass of $m^* = \sqrt[3]{m_l m_t^2} = 0.33 m_e$.

[1] This is different in most samples of batch B146-2 where the peak height strongly decreases between 4 K and base temperature: the activation energy is lower but the barrier is longer.

However, the short range potential is not screened by the host dielectric constant and depends on the actual element used as dopant. It causes a specific correction to the binding energy for each chemical element. Silicon has 6-fold degenerate conduction band minima (see appendix C). This degeneracy splits the hydrogen ground state in one singlet, one doublet and one triplet, the singlet state having the lowest energy. The experimental values for arsenic in silicon are $-53.7\,\mathrm{m}e\mathrm{V}$ for the 1s singlet, $-32.6\,\mathrm{m}e\mathrm{V}$ for the 1s doublet, $-31.2\,\mathrm{m}e\mathrm{V}$ for the 1s triplet and $-11.49\,\mathrm{m}e\mathrm{V}$ for the $2p_0$ state [106].

In addition these binding energies can fluctuate for several reasons:

- due to an external potential, different for each site.
- due to their different distance from the Si–SiO_2 interface: close to the interface the average dielectric constant decreases and the effective mass approximation cannot be maintained; the energy levels shift.
- due to hybridization: strongly coupled dopants form a bound state lower in energy than that of isolated dopants and an anti-bound state higher in energy.
- due to electrostatic interaction with other dopants: compensation (i.e. both, donors and acceptors are present) or external fields make that some dopants are charged, even at very low temperature

The first two points are specific to small devices. The others also occur in bulk semiconductors and are described in detail in [106].

The spatial extension of the donor states is given by the renormalized Bohr radius

$$a_\mathrm{B} = \frac{4\pi\epsilon_0\epsilon_\mathrm{Si}\hbar^2}{m_\mathrm{t}e^2} = 3.3\,\mathrm{nm} \tag{5.6}$$

We use the Bohr radius in particular for the exponential tail of the wave function. In each direction, the wave functions of the conduction band minima having the smallest effective mass in this direction will dominate. This is why it is a good approximation to use the transversal effective mass here, which is smaller than the longitudinal effective mass [106].

Another property of dopants we will be interested in is their polarizability α (the ratio of induced electric dipole moment to driving local electric field). For a hydrogen atom one calculates [72]:

$$\alpha = 18\pi\epsilon_0 a_\mathrm{B}^3 \tag{5.7}$$

With the renormalized dielectric constant and Bohr radius in silicon one finds a polarizability of a donor of $\alpha = 2 \cdot 10^{-34}\,\mathrm{Fm}^2$. The experimental value [19, 18] is much smaller:

$$\alpha = 1 \cdot 10^{-35}\,\mathrm{Fm}^2 \tag{5.8}$$

The difference is explained by the fact that we have chosen the smallest possible effective mass in (5.6) in order to describe the long range part of the wave function.

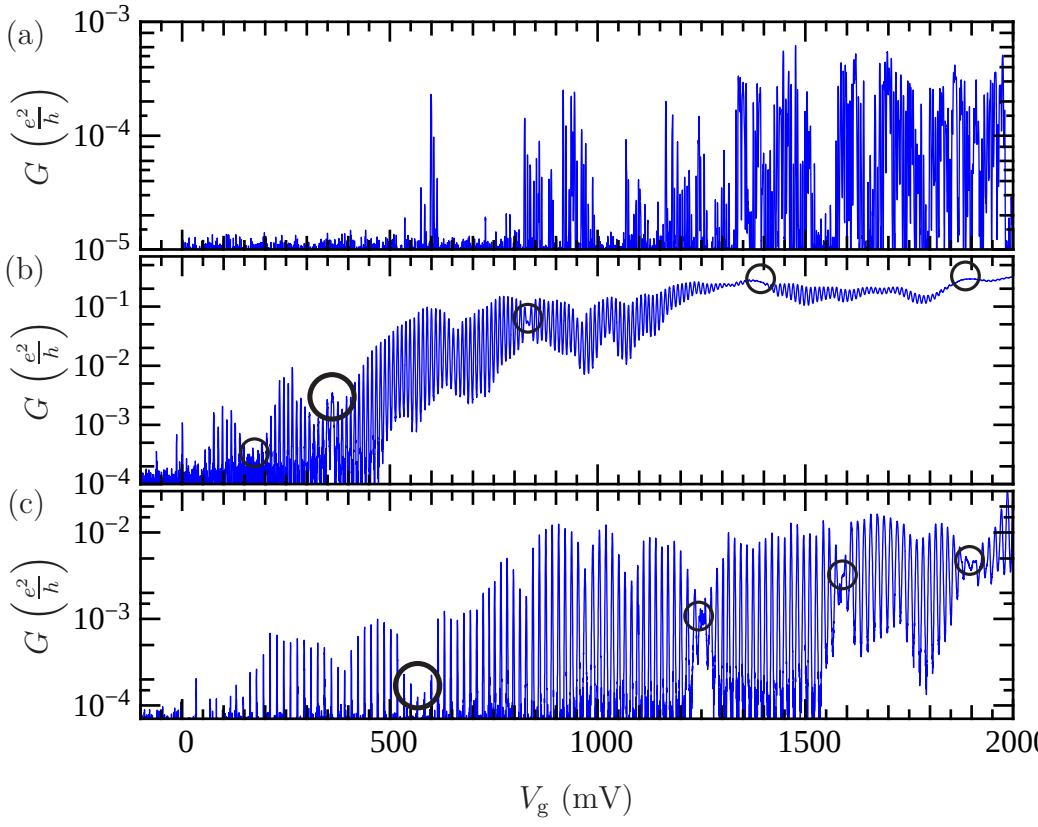

Figure 5.1: Drain-source conductance versus gate voltage. *(a) sample 11. (b) sample 2. (c) sample 5. The Coulomb blockade oscillations in (a) are irregular as compared to (b) and (c) where only a few anomalies perturb the otherwise very regular spectrum. These anomalous regions with reduced contrast and fluctuating peak spacing are highlighted with circles. The anomalies marked with big circles will be studied in detail.*

For the polarizability, which involves the central part of the wave function, a higher effective mass should be chosen. The experimental value corresponds to an effective mass $0.53\,m_e$ in expr. (5.6), a value between longitudinal and transversal effective mass (see appendix C).

5.3 Anomalies in the Coulomb blockade spectrum

Figure 5.1 shows typical $G(V_g)$ plots. In chapter 4 we studied the peak spacing statistics and attributed it to the properties of the quantum dot. But there were tails in the peak spacing distribution which we did not take into account for the analysis. Here we focus on their origin: anomalous regions where the conductance contrast is markedly reduced and a phase shift of the Coulomb blockade oscillations occurs. Such perturbations to the periodic pattern are marked with circles in figure 5.1. In batch 9857 samples with low doping, these perturbations occur only rarely (we count typically 3 to 5 well marked anomalies per sample). In the unperturbed

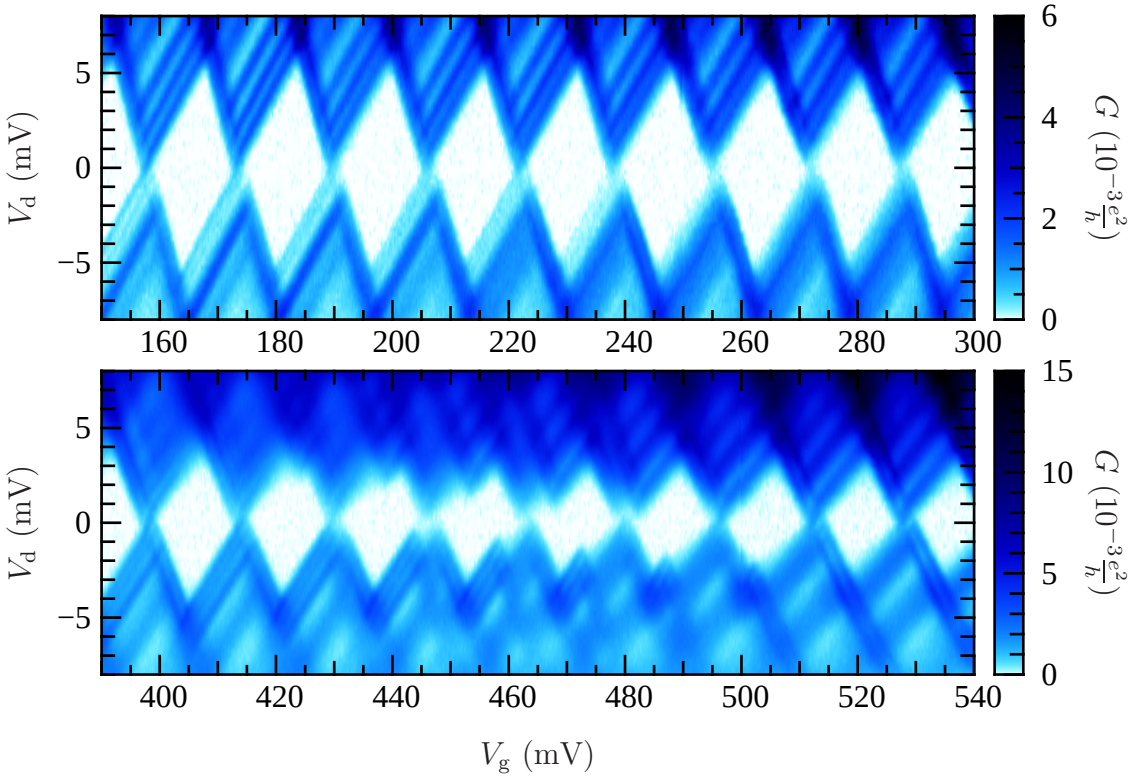

Figure 5.2: 2D-plots of the measured drain-source conductance versus gate and drain voltages. *Upper panel: in an unperturbed, very periodic gate voltage range. Lower panel: in an anomalous region where a charge trap is observed. White areas correspond to Coulomb blockaded regions (no detectable current). The lines inside the conducting regions are not the excited states of the dot (see section 5.3.3). Compared to the lower panel of figure 5.1, the anomalous region has shifted by* 50 mV *in gate voltage after thermal cycling between base and room temperature. We do not observe such shifts as long as the sample is kept cold. Sample 5.*

regions, the height of the Coulomb blockade peaks shows long-range correlations. In batch B146 samples with high doping level the perturbations are more frequent at very low temperature, and as a result the whole spectrum looks irregular (see top panel in figure 5.1). This suggests that the perturbations are related to the doping.

In the $G(V_g, V_d)$ plot they are even more visible (see figure 5.2). In the perturbed regions additional teeth appear in the Coulomb diamonds.

We develop a simple model based on a trap state located in the vicinity of the quantum dot, and compare the simulation with the experimental data.

5.3.1 Model

The dopants in the barriers are randomly distributed and the coupling between them depends exponentially on their distance. Thus, the strength of this coupling

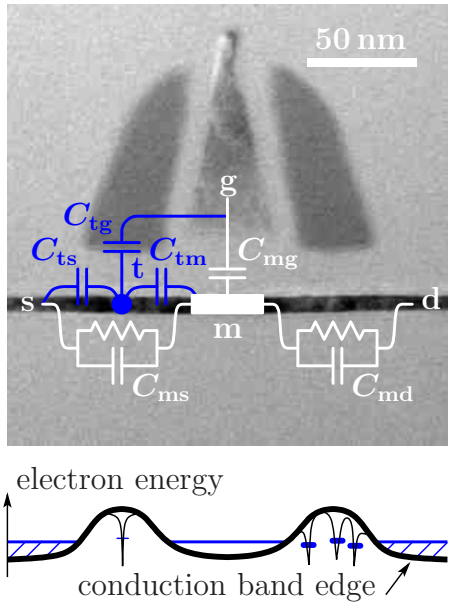

Figure 5.3: Electrical model. *Below the transmission electron micrograph of a batch 9857 sample a schematic energy diagram is drawn. The reduced doping level below the spacers and the gate electrode creates a potential barrier, in the middle of which a well is created by a positive gate voltage. Conductance through the barriers separating the well from source and drain occurs by tunneling through a chain of well connected dopants (plotted in the right barrier) [100]. In more isolated dopants (plotted in the left barrier) the number of charges is well quantized. Such traps are the main concern of this section. Their interaction with the quantum well is mainly electrostatic. We describe it with the lumped network superimposed to the micrograph.*

is distributed over a wide range. As a consequence, transport takes place mainly through a percolation path formed by well connected dopants [106] while other dopant states are only weakly connected and their occupation is a good quantum number (see figure 5.3). We attribute the anomalies in the Coulomb blockade spectrum to electrostatic interaction of the quantum dot with such a charge trap formed by a weakly connected dopant site.

The corresponding lumped network is shown in figure 5.3. Similar models have been considered in [36] and [10]. A small trap (t) is capacitively coupled to source (s) or drain (d), gate (g) and to the main dot (m). We note

$$C_i = C_{is} + C_{id} + C_{ig} \qquad (i = \mathrm{m}, \mathrm{t}) \tag{5.9}$$

$$-ex_i = C_{is}V_\mathrm{s} + C_{id}V_\mathrm{d} + C_{ig}V_\mathrm{g} \qquad (i = \mathrm{m}, \mathrm{t}) \tag{5.10}$$

$$\beta_\mathrm{t} = \frac{C_\mathrm{tm}}{C_\mathrm{t} + C_\mathrm{tm}} \tag{5.11}$$

With the capacitance matrix for the network in figure 5.3 (see equation (2.4)) and its LDU decomposition

$$\hat{C} = \begin{pmatrix} C_\mathrm{t}+C_\mathrm{tm} & -C_\mathrm{tm} \\ -C_\mathrm{tm} & C_\mathrm{m}+C_\mathrm{tm} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ -\beta_\mathrm{t} & 1 \end{pmatrix} \begin{pmatrix} C_\mathrm{t}+C_\mathrm{tm} & 0 \\ 0 & C_\mathrm{m}+\beta_\mathrm{t}C_\mathrm{t} \end{pmatrix} \begin{pmatrix} 1 & -\beta_\mathrm{t} \\ 0 & 1 \end{pmatrix} \tag{5.12}$$

we calculate the electrostatic energy for n_m electrons in the main dot and n_t electrons in the trap (see equation (2.7)):

$$
\begin{aligned}
W(n_\mathrm{m}, n_\mathrm{t}) &= \frac{e^2}{2}\langle \begin{pmatrix} n_\mathrm{t}-x_\mathrm{t} \\ n_\mathrm{m}-x_\mathrm{m} \end{pmatrix} |\hat{C}^{-1}| \begin{pmatrix} n_\mathrm{t}-x_\mathrm{t} \\ n_\mathrm{m}-x_\mathrm{m} \end{pmatrix} \rangle \\
&= \frac{e^2}{2}\langle \begin{pmatrix} n_\mathrm{t}-x_\mathrm{t} \\ n_\mathrm{m}-x_\mathrm{m} \end{pmatrix} | \begin{pmatrix} 1 & \beta_\mathrm{t} \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \frac{1}{C_\mathrm{t}+C_\mathrm{tm}} & 0 \\ 0 & \frac{1}{C_\mathrm{m}+\beta_\mathrm{t}C_\mathrm{t}} \end{pmatrix} \begin{pmatrix} 1 & 0 \\ \beta_\mathrm{t} & 1 \end{pmatrix} | \begin{pmatrix} n_\mathrm{t}-x_\mathrm{t} \\ n_\mathrm{m}-x_\mathrm{m} \end{pmatrix} \rangle \\
&= \underbrace{\frac{e^2\left(n_\mathrm{m}+\beta_\mathrm{t}n_\mathrm{t}-x\right)^2}{2C}}_{M(n_\mathrm{m},n_\mathrm{t})} + \underbrace{\frac{e^2\left(n_\mathrm{t}-x_\mathrm{t}\right)^2}{2\left(C_\mathrm{t}+C_\mathrm{tm}\right)}}_{D(n_\mathrm{t})}
\end{aligned}
\tag{5.13}
$$

where $C = C_\mathrm{m} + \beta_\mathrm{t}C_\mathrm{t}$ and $x = x_\mathrm{m} + \beta_\mathrm{t}x_\mathrm{t}$. For a small trap ($C_\mathrm{m} \gg C_\mathrm{t}$) these renormalizations are weak: $C \approx C_\mathrm{m}$ and $x \approx x_\mathrm{m}$. Note that the model is symmetric under exchange of main dot and trap. Expressions M and D have different forms because we attributed the interaction term entirely to M to ease further analysis. W is plotted in the top panel of figure 5.4.

The current through the system is calculated numerically using the algorithm described in chapter 2. We focus on the structure of the conductance through the system as fixed by eq. (5.13), not the exact values of the conductance. This is why we choose very simplistic models for the following parameters that affect only the absolute value of the conductance or smooth variations of it. We consider the main dot as metallic (negligible single-particle level spacing Δ_1, i.e. $\Delta_1 \ll k_\mathrm{B}T$) and we consider only one non-degenerate energy level for the trap. The dot–lead tunneling rates are then given by expression (2.18) and the trap–dot and trap–lead tunneling rates by expressions (2.19) and (2.20). We suppose symmetric prefactors for the tunneling rates and we suppose the rates to or from the trap to be very small:

$$1000\, eI_\mathrm{ts} = 1000\, eI_\mathrm{tm} = G_\mathrm{ms}\Delta_1 = G_\mathrm{md}\Delta_1 \tag{5.14}$$

These are the rates for a trap on the source side of the dot; for a trap on the drain side I_ts is replaced with I_td. The small tunneling rates for the trap make that electrons can be added or removed from the trap, while their contribution to the total current through the device is negligible. This contrasts with models of stochastic Coulomb blockade [97] or in-series quantum dots [124, 96] where the current has to pass through both dots.

Results of such a numerical study are presented in figure 5.4. The middle panel shows the mean occupation of the trap. On a large scale, the trap becomes occupied with increasing gate voltage. In the central region of the figure however, whenever an electron is added onto the main dot, the electron in the trap is repelled and only later it is re-attracted by the gate electrode. Inversely, the trap charge repels the charges on the main dot and the Coulomb blockade structure of the main dot is shifted to higher gate voltage when the trap is occupied (see lower panel). The two Coulomb blockade structures for unoccupied and occupied trap are respectively indicated by dotted and dashed lines in the middle and lower panel of figure 5.4.

This explanation is illustrated in terms of energy in the top panel of figure 5.4, which shows the energies for the different charge configurations. The crossings of the

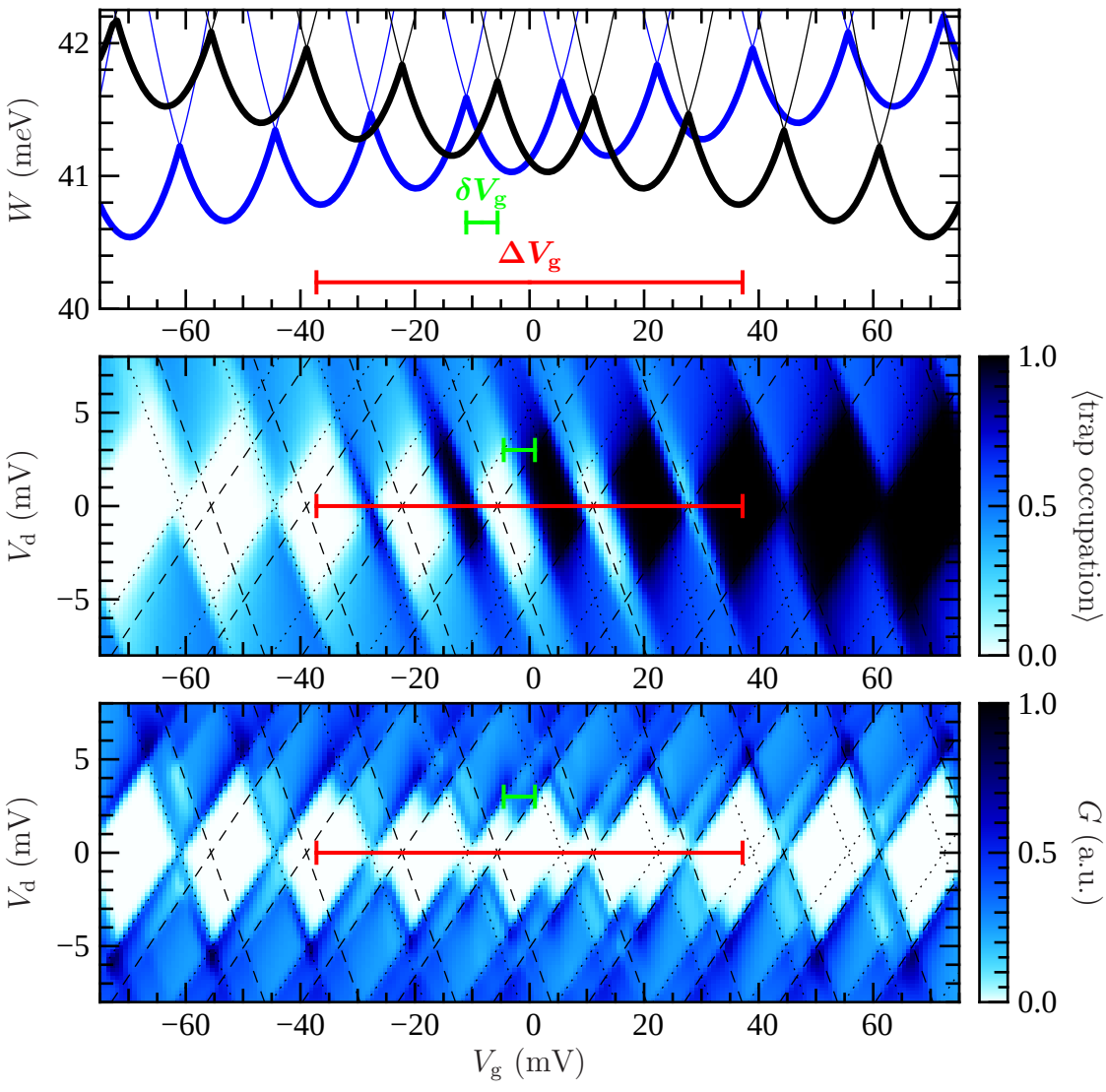

Figure 5.4: Numerical study of a trap. *The trap is coupled to the source and to the main quantum dot as sketched in figure 5.3. Parameters: temperature:* $D = 1\,\mathrm{K}$*; main dot:* $C_{\mathrm{mg}} = 60\,\frac{e}{\mathrm{V}}$, $C_{\mathrm{md}} = C_{\mathrm{ms}} = 70\,\frac{e}{\mathrm{V}}$*; trap:* $C_{\mathrm{tg}} = 0.045\,\frac{e}{\mathrm{V}}$, $C_{\mathrm{ts}} = 2\,\frac{e}{\mathrm{V}}$, $C_{\mathrm{td}} = 0$, $C_{\mathrm{tm}} = 1\,\frac{e}{\mathrm{V}}$*. The trap can either be empty or charged with one electron. The upper panel shows the energy for the different charge states at zero bias in function of gate voltage. The blue parabolas are for empty trap, the black ones for occupied trap. The thick blue and black lines indicate the ground state of the main dot for respectively empty and occupied trap. The middle panel shows the self-consistent mean occupation number of the trap, the lower panel the resulting conductance through the dot. When occupied, the charge trap shifts the Coulomb blockade diamonds of the main dot by* δV_{g}*. The dotted (dashed) lines indicate the position of the diamonds when the trap is empty (occupied). The interaction between trap and dot cause a suppression of the Coulomb peaks over a range of* ΔV_{g} *as well as characteristic teeth on the Coulomb diamonds. This result is in very good agreement with the experimental data in the lower panel of figure 5.2.*

blue (black) parabolas give the positions of the Coulomb blockade peaks for empty (occupied) trap. The shift of the crossings of the black parabolas with respect to the crossings of the blue parabolas is due to the term $\beta_t n_t$ in $M(n_m, n_t)$. One Coulomb blockade oscillation corresponds to a change of 1 in $\beta_t n_t - x$. Thus the shift due to $\Delta n_t = 1$ is

$$\delta V_g = \beta_t V_+ \tag{5.15}$$

where V_+ is the Coulomb blockade peak spacing of the main dot.

We will now determine the width of the anomaly in the Coulomb blockade spectrum at low bias voltage, i.e. the gate voltage range where the Coulomb peaks are suppressed. The reason of this suppression can be seen in the top panel of figure 5.4: when the trap occupation oscillates in function of gate voltage, the degeneracy points of the blue (black) parabolas, which cause the Coulomb peaks, are not reached. The transitions between the black and blue parabolas carry no first order current because they either imply an electron exchange between dot and trap or only a change of the electron number of the trap. So no first order current can flow through the dot.

The width of the anomaly is thus the range between the first and the last crossing of the thick black line and the thick blue line in the top panel of figure 5.4. First we calculate ΔM, the difference of the ground state energies for empty and occupied trap arising from the term M in eq. (5.13). Then we calculate the change in gate voltage necessary for $D(1) - D(0)$ to exceed this difference.

ΔM reaches its extreme values when, for one state of the trap, the main dot is at a degeneracy point (the kinks in the thick lines), where $M = \frac{(e/2)^2}{2C}$. For the other state of the trap the main dot is then a fraction β_t of a Coulomb blockade period away from the degeneracy point and $M = \frac{e^2(1/2-\beta_t)^2}{2C}$. The extrema of ΔM are therefore $\pm\frac{e^2}{2C}\beta_t(1-\beta_t)$.

The gate voltage dependence of term D is given by $\alpha_t = \frac{1}{-e}\frac{\mathrm{d}}{\mathrm{d}V_g}(D(1) - D(0)) = \frac{C_{tg}}{C_t + C_{tm}}$. Note that α_t is the long-range gate voltage lever arm of the trap over several Coulomb blockade oscillations, where the charge of the main dot has to be considered as relaxed with the source and drain Fermi levels. $D(1) - D(0)$ has to change from $+\frac{e^2}{2C}\beta_t(1-\beta_t)$ to $-\frac{e^2}{2C}\beta_t(1-\beta_t)$ in order to toggle the trap definitively. Therefore the width ΔV_g of the anomaly is given by $e\alpha_t \Delta V_g = 2\frac{e^2}{2C}\beta_t(1-\beta_t)$ or

$$\Delta V_g = \frac{\beta_t(1-\beta_t)}{\alpha_t}\alpha_m V_+ \tag{5.16}$$

$\alpha_m = \frac{C_g}{C}$ with $C_g = C_{mg} + \beta_t C_{tg}$ is the gate voltage lever arm of the main dot.

We have identified $\alpha_t = \frac{C_{tg}}{C_t + C_{tm}}$ and $\beta_t = \frac{C_{tm}}{C_t + C_{tm}}$ as parameters determining the structure of the trap signature. Neither depends on the absolute value of the trap capacitances. Indeed, if one allows only 0 or 1 electron in the trap, the absolute value of the trap capacitances enters the problem only indirectly by modifying slightly the capacitance matrix of the main dot and, in the limit of a small trap, does not enter the problem at all. This is why our model only contains these 2 effective

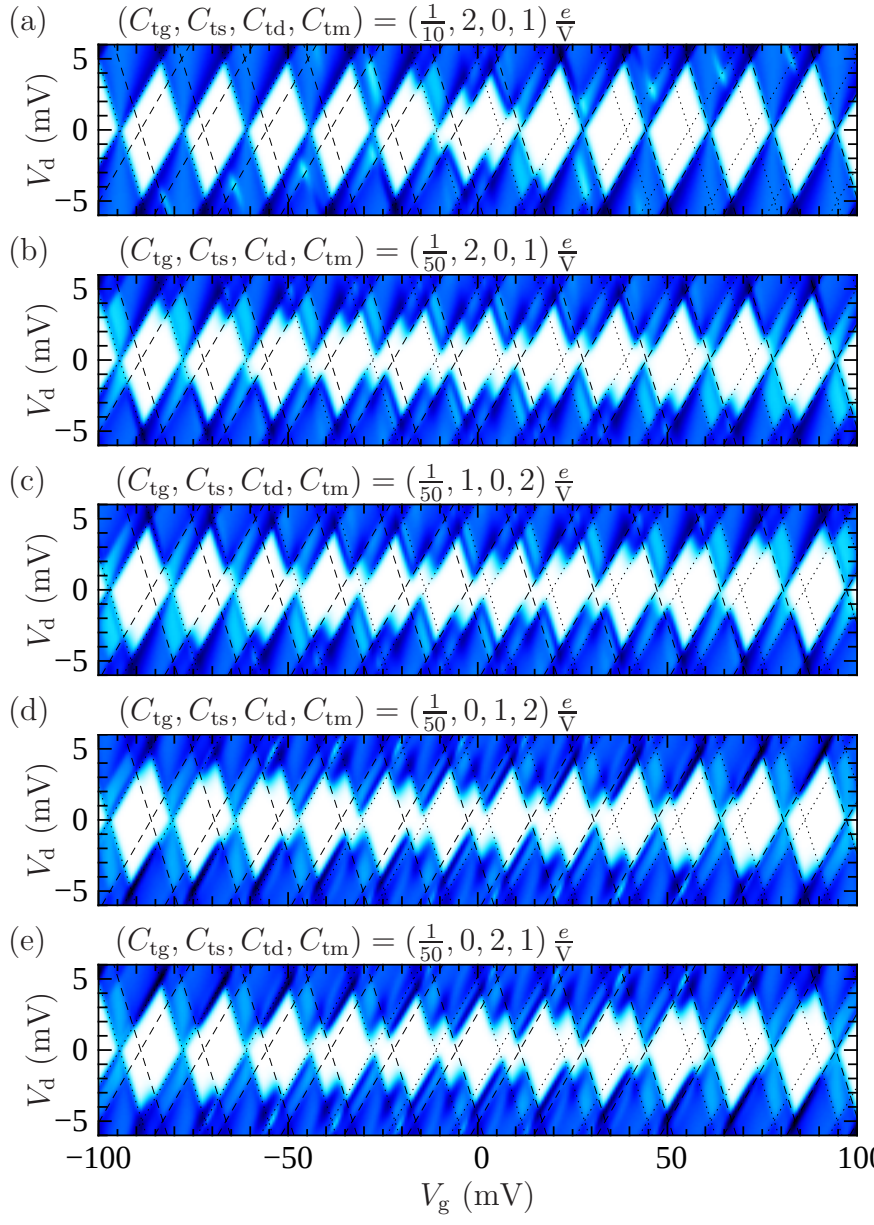

Figure 5.5: Calculated trap signatures for different sets of parameters. *(a) The trap is close to the source. (b) The coupling to the gate electrode is reduced by a factor of 5. The signature becomes wider. (c) The coupling to the source is reduced, the coupling to the dot increased. (d) The trap is placed on the drain side of the dot instead of the source side. (e) The coupling to the dot is reduced, the coupling to the drain increased.*

parameters for the trap instead of 3 (C_{tg}, C_t^s, C_{tm}). All 3 parameters of the trap are only significant if the trap can accommodate 2 or more electrons. In this case the spacing between the anomalies gives access to the absolute values of the trap capacitances.

So far we have considered traps on the source side of the dot. We will now describe how the signature depends on the position of the trap with respect to the dot. If as in the example of figure 5.4 the trap is on the source side, the zones of empty and occupied trap in the (V_g, V_d) plane are aligned with the negative slope of the diamonds (see middle panel of figure 5.4). This is due to the fact that along a line parallel to the negative slope of the diamonds the highest occupied level in the dot stays at constant energy with respect to the source Fermi level. As the trap is controlled mainly by source and dot, its occupation stays constant in this direction. The weak influence of the gate electrode on the trap causes small deviations from this direction. This can be seen in the middle panel of figure 5.4: the limits separating white (empty) and black (occupied) zones are not exactly parallel to the diamond slopes. Due to this almost perfect alignment, changes in the occupation number of the trap, and in consequence the teeth of width δV_g, occur only at the positive slopes of the diamonds. In the same way, the teeth appear on the negative slope if the trap is on the drain side of the dot.

As a summary, figure 5.5 illustrates the relation between the trap's capacitance matrix and its signature. If the teeth of constant width are visible at the positive slope of the Coulomb blockade diamonds, the trap is on the source side of the dot. If they are visible at the negative slope, the trap is on the drain side. The width of the teeth depends on β_t, the width of the anomalous region essentially on α_t (for β_t close to $\frac{1}{2}$, where the anomalies are most visible).

5.3.2 Position and nature of the traps

As an illustration, from the lower panel of figure 5.2 we infer $\alpha_t \approx 0.015$ and $\beta_t \approx 0.3$. These are the actual parameters that have been chosen for the simulation in figure 5.4, and the lower panels of figure 5.2 and figure 5.4 are indeed very similar. As for all impurities we observed, α_t is small. This is what we expect for a trap inside the silicon wire. The coupling to the gate electrode is much weaker than the coupling to the main dot or the source electrode because the dielectric constant of the oxide barrier ($\epsilon_{SiO_2} = 4$) is much smaller than that of bare silicon ($\epsilon_{Si} = 12$), which in addition is enhanced near the insulator–metal transition (see section 5.4). Indeed we do not expect traps outside the wire to cause such signatures:

- Traps located deep inside the oxide can be excluded because their transmissions would be too weak to observe statistical mixing of occupied and unoccupied trap states during our acquisition time below 1 s. Similar devices including intentional silicon nanocrystals at the interface between thermal oxide and deposited oxide have been studied in views of memory applications [87, 86]. The measured lifetime of charges in the nanocrystals exceeds 1 s by orders of magnitude already at room temperature and at low temperature gate voltages of about 5 V have to be applied in order to toggle the charge in the nanocrystals.

- Traps at the Si–SiO_2 interface are unlikely. The technology used attains less than $10^{11}\,\mathrm{cm}^{-2}$, corresponding to a few units per sample. As they are distributed throughout the entire band gap it is very unlikely to observe several of them in the small energy window $\alpha_\mathrm{t}({V_\mathrm{g}}^{\max} - {V_\mathrm{g}}^{\min}) \approx 30\,\mathrm{m}e\mathrm{V}$ that we scan in our measurement.

Therefore the most likely traps are defects in the silicon wire or As donor states. Given the volume of the access regions under the spacers and the doping level N_D, there are approximately 30 donor states under the spacers in devices of batch 9857. Typically we record between 3 and 10 anomalies but then we expect not all traps to be visible.

- The volume of the wire under the spacers is approximately the same as that of the quantum dot. As the capacitance is dominated by the capacitances towards source and drain (see section 5.4) we assume that the regions under the spacers have the same capacitance as the dot. We observe lever arm factors α approximately 10 times smaller for the traps than for the main dot. In the swept gate-voltage range we observe typically 200 resonances of the main dot. Thus we would expect to add approximately 20 electrons to the barriers, less than the estimated total number of traps.

- The charge on well-connected dopant sites is not quantified and should not cause clear trap signatures.

- Dopants very close to the dot ($\beta_\mathrm{t} \approx 1$) or to the reservoir ($\beta_\mathrm{t} \approx 0$) cause very small anomalies that are difficult to detect.

In batch B146 the doping level in the access regions is more than 10 times higher than in batch 9857. The whole Coulomb blockade spectrum should therefore be anomalous. Indeed, the spectrum in batch B146-1 at very low temperature is much less regular (see figure 5.1) than for batch 9857 but we cannot distinguish signatures as clear as in batch 9857. This is consistent because in batch B146 the mean distance between impurities is less than 3 nm and they are too well connected for the charge on them to be well quantized. In other words, the wire is very close to the insulator–metal transition. In bulk Si:As a critical concentration $N_\mathrm{c} = 8.6{\cdot}10^{18}\,\mathrm{cm}^{-3}$ was found [104]; the doping in batch B146 is $n_\mathrm{D} \approx 10^{19}\,\mathrm{cm}^{-3}$. In split 1 most of the samples are rather metallic, in split 2, where the doping is two times lower, most samples are rather insulating. The proximity of the metal–insulator transition explains the strong temperature dependence down to very low temperatures. In batch 9857 the access regions are much further on the insulating side so that activated processes are quenched at higher temperature. Nevertheless the conductance is comparable because the low doped regions are much shorter.

We have deduced that the observed traps lie inside the wire. The position of the trap along the wire can also be determined. First we can distinguish on which side of the dot the trap is: teeth on positive slope of the diamonds indicate a trap on the source side, teeth on the negative slope a trap on the drain side. Then the

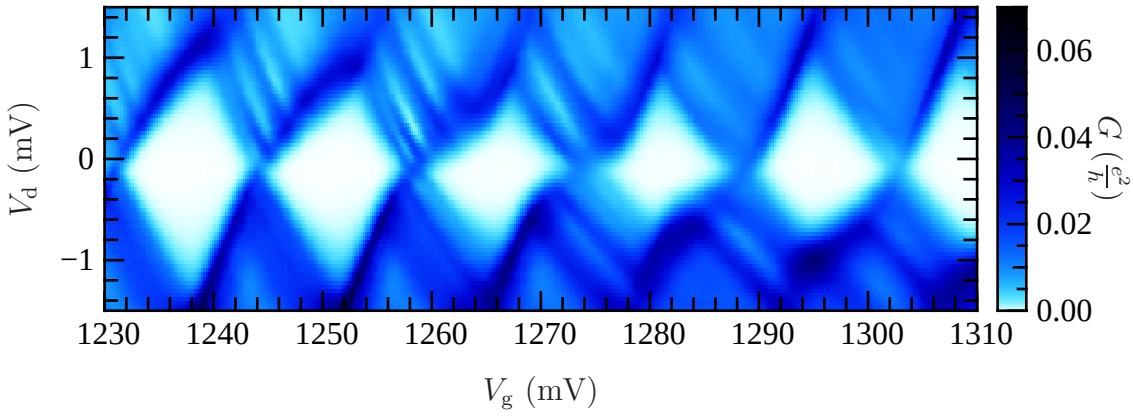

Figure 5.6: Trap signature with many-body effects. *This trap signature could be decomposed into trap signatures with different β-factors occurring at the same time (see figure 5.7). We interpret it as a charging event with different charge rearrangements in the barrier (see text). As the lines inside the conducting regions are parallel to the negative diamond slope we conclude that we deal with an effect in the source barrier (see argument on page 73). Sample 3.*

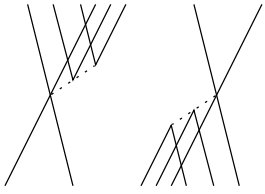

Figure 5.7: Interpretation of figure 5.6. *The kinks in the central diamond slopes of figure 5.6 are explained by several impurity states with different β_t-factors.*

parameter β_t gives the ratio between the capacitances towards the main dot and the source (or drain) electrode. As the dielectric constant of the wire is much higher than the surrounding silicon oxide, this ratio can be directly translated to a position in direction of the wire, in first approximation linearly. In the example of figure 5.2 with $\beta_t \approx 0.3$ we would expect the impurity to be located $\frac{2}{3}$ on the way from the dot (edge of the gate electrode) to the source reservoir (source side edge of the spacer).

5.3.3 Charging events with rearrangement

In the picture of traps we have drawn so far we considered only one trap at a time. In a more accurate picture one has to consider that, when one electron is added to a trap site, it can be energetically favorable to rearrange the electrons in the other sites in order to minimize the electron-electron interactions. In fact, this is expected to be the case for most charging events [106, 66, 67]. In addition, we drive a current through the barriers which also implies different charge configurations.

Some trap signatures do not have sharp teeth as described above but are more like the example in figure 5.6. Such smooth signatures seem more frequent in large samples and at high gate voltage. They could be decomposed in several teeth as

described above but with different β_t factors, each corresponding to a different (non-ground-state) distribution of the electrons over the traps (see figure 5.7)). We made numerical simulations where we placed the dopants arbitrarily in the barrier volume and set their mutual capacitance inversely proportional to their distance. According with [66] we found that configurations with the same total electron number are very close in energy so that, if one electron is added, there are several configurations available. Which configuration has the lowest energy depends on very small gate or bias voltage changes.

Such kinks have been observed in literature [115] but have been interpreted as excited states.

We think these almost degenerate excited electron configurations also to be responsible for the lines we observe inside the conducting regions near most trap signatures (upper panel of figure 5.2 and figure 5.6). Sometimes we observe such lines without clear trap signature (see upper panel of figure 5.2). The spacing of these lines is too high to be the excitation spectrum of the dots (see chapter 4.1.2). More importantly, the spacing between the lines is the same over several Coulomb blockade periods while, on the contrary, the excitation spectrum should be different for different electron numbers.

These lines can naturally be explained when we admit that there are multiple almost degenerate configurations of the electrons in the barriers. Each line would correspond to a new configuration entering the bias window. It is not yet clear why the lines are more marked near the trap signatures.

Trap signatures showing simple teeth as discussed before do not necessarily imply that no charge rearrangements occur. They only indicate that for n and $n+1$ electrons in the barrier there is respectively a clear ground state configuration with all others having much higher energy; yet the actual electron configuration for n and $n+1$ could be completely different. Consequently, the properties we attributed to a single trap have rather to be seen as the properties of the polaron that consists in adding one electron to the barrier and in rearranging the others in order to go to the new ground state. We will continue to speak of individual traps being charged but have to keep in mind that these events can be more complex.

5.3.4 Dynamics

Slow traps: random telegraph signal

So far we have used the changing occupation number of the traps to explain the anomalies in the Coulomb blockade conductance; yet the measurements of the mean current have not allowed us to measure the occupation number directly. But the currents through the main dot differ between empty and occupied trap because the positions of the Coulomb blockade resonances are shifted. Thus, at the anomalies where the mean occupation number of the trap is different from 0 and 1, the fluctuations of the occupation number should create a random telegraph signal (RTS) [64, 128, 129, 16] in the current through the main dot.

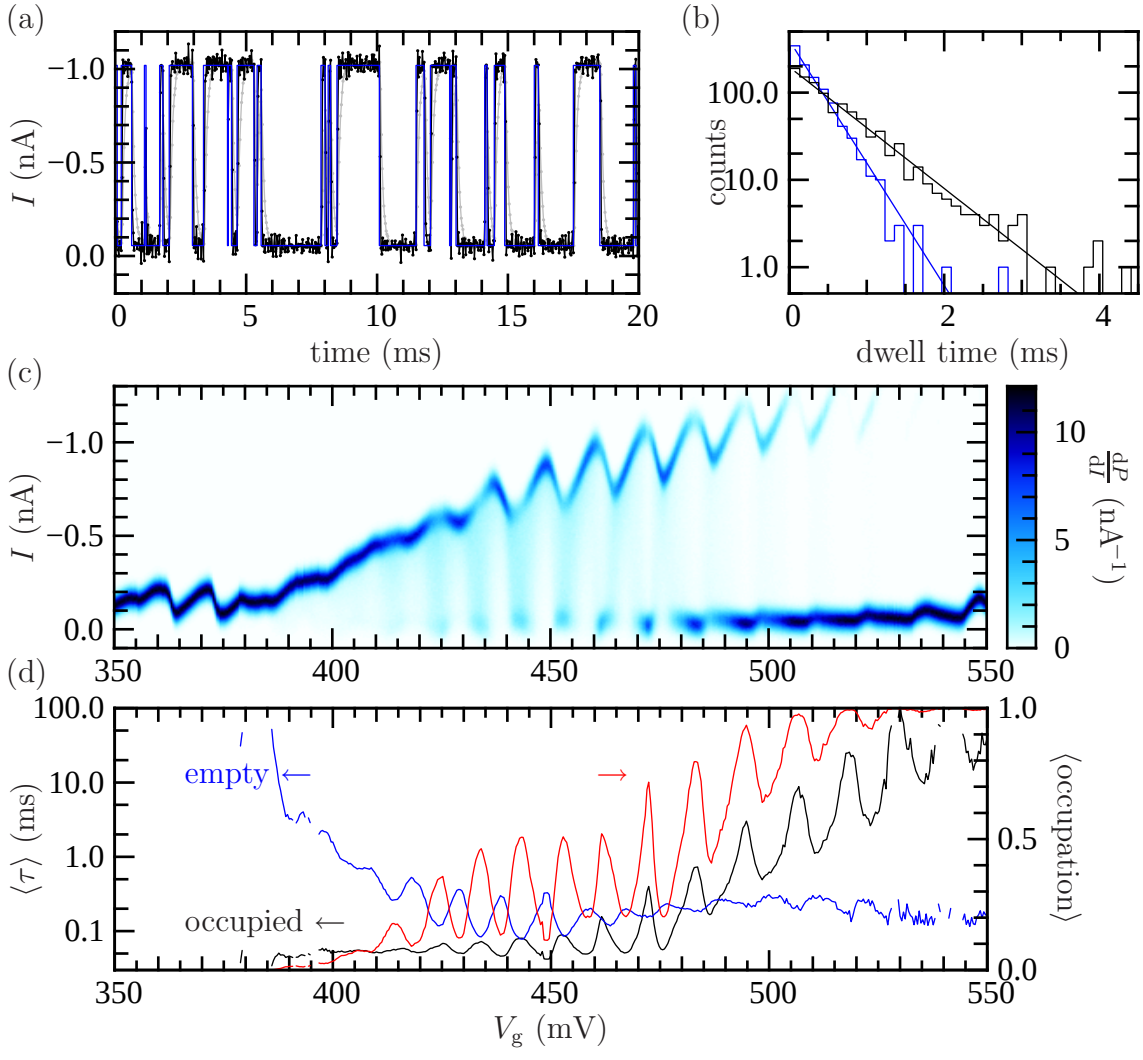

Figure 5.8: Analysis of a trap signature with switching. *(a) A RTS trace taken at* $V_g = 500\,\mathrm{mV}$, $V_d = -6\,\mathrm{mV}$. *Light gray trace: raw data. Black trace: data after compensation of the time constant* τ_{ampli} *of the current amplifier. Blue line: fitted signal (see appendix B.4). The detection time is approximately* $30\,\mu\mathrm{s}$. *(b) Histograms of the times passed in the weak current state (occupied trap, black) and the* $-1\,\mathrm{nA}$ *state (empty trap, blue). The time constants (averages of these times) are* $0.31\,\mathrm{ms}$ *and* $0.62\,\mathrm{ms}$. *The corresponding exponential distributions (straight lines) fit the histograms well. (c) Current histogram at* $V_d = -6\,\mathrm{mV}$. *The nonzero density between the two current levels is due to the finite rise time. The current for unoccupied trap is always higher than for occupied trap. (d) Time constants of the empty and occupied levels in function of gate voltage and occupation number of the trap. Sample 6.*

Near most of the trap signatures noise is not or only weakly increased. Only rarely are the trap signatures accompanied by a clear RTS signal. Indeed, the electrons become localized in a trap state as soon as their coupling is weaker than $\frac{e^2}{h}$. Even at a temperature of 150 mK this limit leads to tunneling rates in the GHz range. With our measurement in the kHz range we can only observe switching for very weakly coupled traps while the charge on traps with stronger coupling is still sufficiently quantized to produce clear signatures in the mean current.

An example where a clear RTS signal could be observed is given in figure 5.8(a). We detected the transitions in this signal with the algorithm described in appendix B.4. The distribution of the times spent in the two states follows the exponential distribution expected for a RTS (see figure 5.8(b)).

The color plot of the current distribution in figure 5.8(c) shows the evolution of the two current levels (the dark lines with high probability) with gate voltage. Above 380 mV the two levels are very different. This difference is not captured in the model and could be due to electrostatic interaction of the trap and the current path through the barrier: depending on the state of the trap, the dopants through which the main part of the current flows are well or poorly aligned in energy. Or, according to the previous subsection the configuration of the electrons in the barrier could have completely changed.

The fact that the current levels never cross greatly simplifies the assignment of the high and low current levels to the states of the trap. The high current trace being most likely at low gate voltage and the low current trace being most likely at high gate voltage allows to attribute the high current to the empty trap and the low current to the occupied trap.

The time constants of the empty and occupied state are plotted in figure 5.8(d). In accordance with panel (c), the time constant for the empty trap decreases with gate voltage while the time constant for the occupied trap increases. Superimposed with this slow change there are oscillations with a period of 12 mV, the peak spacing of the main dot. The observed time constants are bounded below by the detection time of 30 µs (see appendix B.4) and time constants close to the detection time are overestimated [89]. However, our finding of slow changes and oscillations of the time constants are not affected by this limitation. The mean occupation number is given by $\frac{\tau_{\text{occupied}}}{\tau_{\text{occupied}}+\tau_{\text{empty}}}$. This ratio is unbiased even if the time constants are overestimated [89]. It goes from 0 at low gate voltage to 1 at high gate voltage and oscillates strongly near $V_{\text{g}} = 450\,\text{mV}$ (see figure 5.8(d)). As explained in section 5.3.1 for the case of low bias, this oscillation is due to the discrete charge on the main dot which cycles the trap several times between empty and occupied state. It is not observed in RTS in larger devices without Coulomb blockade [64]. In figure 5.9 the occupation probability for different bias voltages is compared with simulation. As in figure 5.4, the oscillations are aligned parallel to the negative slopes of the Coulomb blockade diamonds indicating that the trap is on the source side of the dot.

RTS (i.e. current through the trap) only occurs when the trap is in the bias window. For large gate and bias voltage excursions where the charging energy of the main dot is negligible, the main dot can be considered as part of the drain

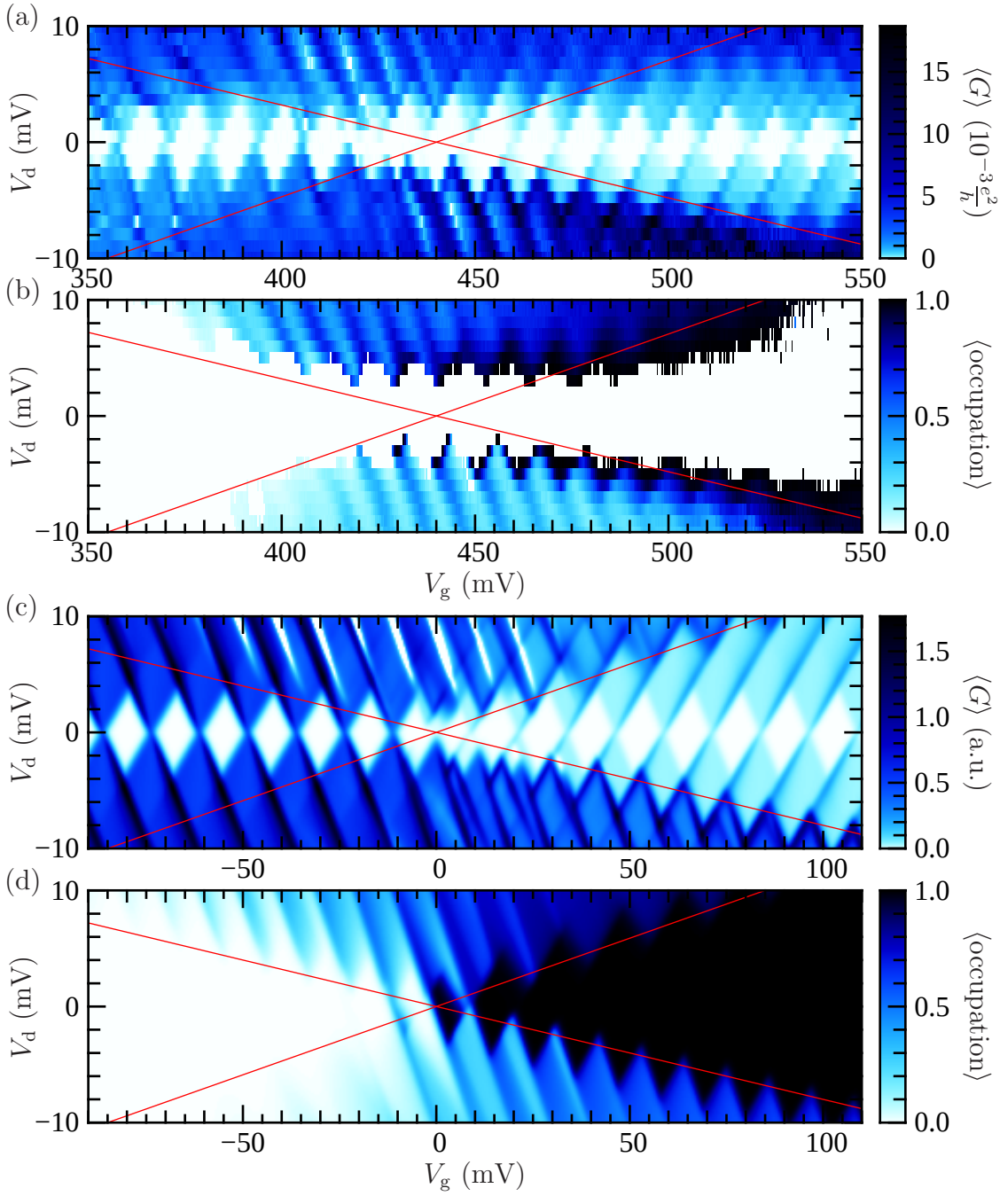

Figure 5.9: Comparison of measured occupation number and simulation. *Same trap as in figure 5.8 (sample 6). (a) Mean differential conductance obtained by numerical derivation of the mean current. (b) Occupation of the trap obtained from the duty cycle of the RTS signal. Regions where no clear RTS could be detected are left white. (c) and (d) Simulation with the following parameters: main dot:* $C_{\mathrm{mg}} = 13\,\mathrm{aF}$, $C_{\mathrm{ms}} = 10\,\mathrm{aF}$, $C_{\mathrm{md}} = 16\,\mathrm{aF}$*; trap:* $C_{\mathrm{tg}} = 0.013\,\mathrm{aF}$, $C_{\mathrm{ts}} = 0.10\,\mathrm{aF}$, $C_{\mathrm{td}} = 0$, $C_{\mathrm{tm}} = 0.16\,\mathrm{aF}$. *In units of the drain–dot barrier transmission, the source–dot barrier transmission is 10 for empty trap and* $\frac{1}{10}$ *for occupied trap, the source–trap barrier transmission is* $\frac{\Delta_1}{1000}$ *and the trap–dot barrier transmission is* $\frac{\Delta_1}{3000}$.

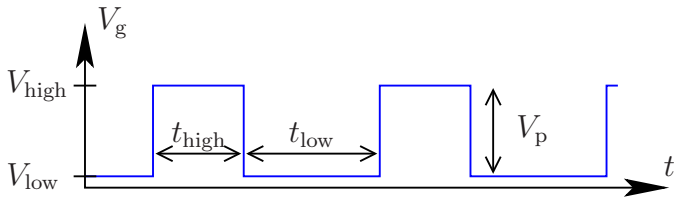

Figure 5.10: Pulse signal. *A pulse signal is applied to the gate electrode in order to probe the dynamics of fast traps.*

reservoir. The zone where the trap is in the bias window is then delimited by slopes $\frac{C_{tg}}{C_t}$ and $-\frac{C_{tg}}{C_{tm}}$ (indicated by straight lines in figure 5.9), just as for a single quantum dot. These slopes give a more straightforward access to the parameters α_t and β_t as the calculations in section 5.3.1. In accordance with the shape of the teeth in panel (a) we find $\alpha_t = 0.05$ and $\beta_t = 0.6$.

The mean occupation of the trap is higher for positive drain voltage than for negative drain voltage indicating a higher transmission rate of the trap towards the source than towards the main dot.

In figure 5.9(c) and (d) we try to reproduce panels (a) and (b). For this simulation we reduce by a factor of 100 the transmission of the source barrier of the main dot when the trap is occupied. This reproduces the lines of reduced differential conductance at positive drain voltage (compare figure 5.9(a) and (c)). In the simulation the oscillations of the trap occupation decay more rapidly with bias voltage than in the measurement. This could be related to our approximation of a thermal distribution of kinetic energies in the main dot, which is certainly not accurate at high bias voltage.

Fast traps: measurements with pulse excitation

As we said, only very few traps are sufficiently slow for their random telegraph signal to be measured. But dynamics faster than the current measurement can also be probed by applying pulse excitations (see figure 5.10) and measuring the average current [28, 26, 42, 99].

Figure 5.11 shows a current measurement under pulse excitation. The vertical and diagonal lines in the central panel correspond to the same Coulomb blockade resonances. The diagonal lines appear when a Coulomb blockade level is in the bias window for the high gate voltage level, the vertical lines when it is in the bias window for the low gate voltage level. In the central part of the figure the lines split at high pulse voltage (the right panel shows a zoom). Such a splitting can be explained by a trap too slow to follow the pulse excitation. Suppose the trap is empty at a static gate voltage $V_g = V_{low}$ and occupied at a static gate voltage $V_g = V_{high}$. If a pulsed gate voltage oscillating between V_{low} and V_{high} as in figure 5.10 is applied and if the period of the pulse signal is small compared to the time constant of the trap, then the trap cannot follow the pulse excitation and will be occupied with a certain probability — the same for $V_g = V_{high}$ as for $V_g = V_{low}$. This is why the lines split. Now the two states of the trap are possible now for one and the same gate voltage. So we observe the peaks of the main dot twice, once for empty and once (shifted by $\beta_t V_+$) for occupied trap.

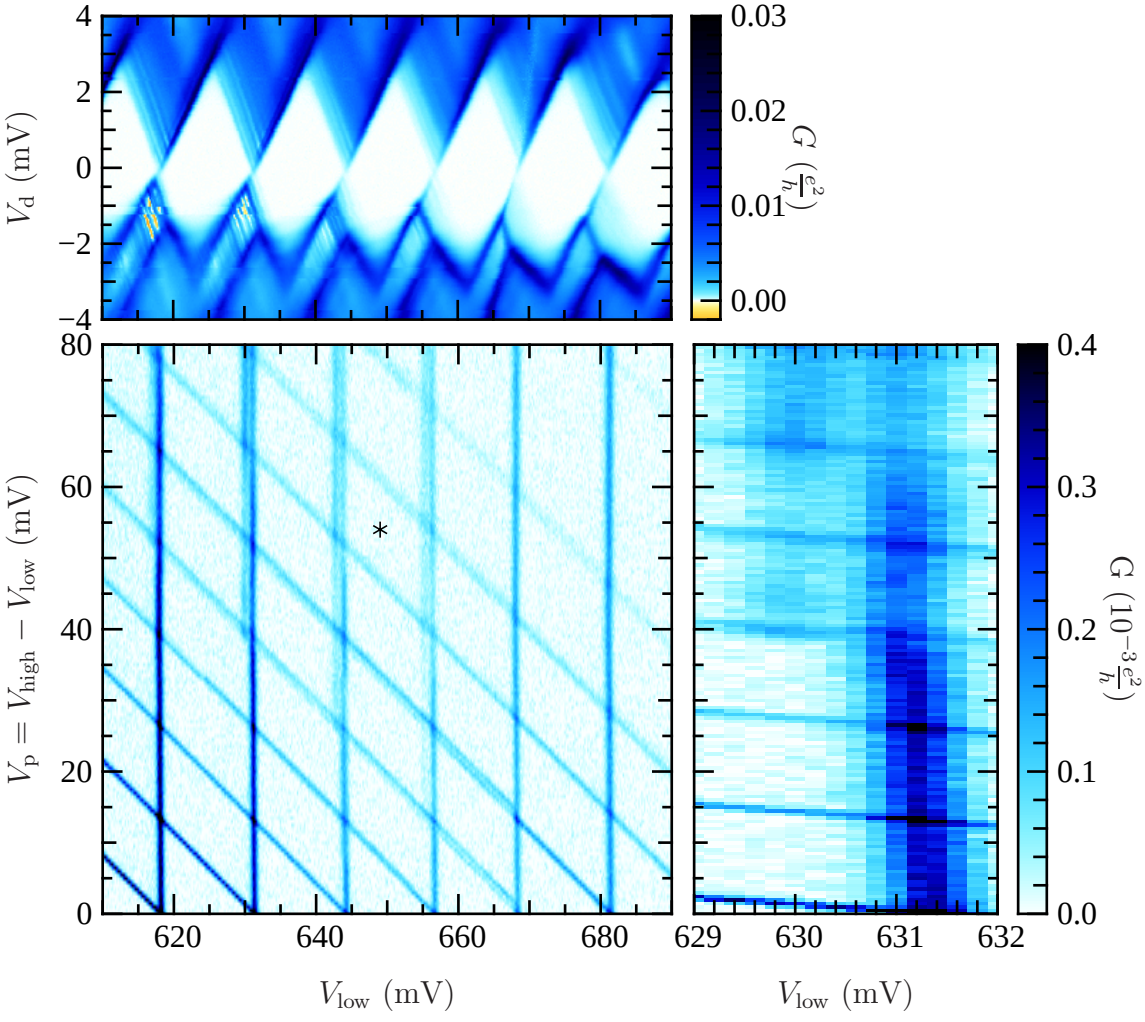

Figure 5.11: Trap under pulse excitation. *The pulse times (see figure 5.10) are $t_{high} = t_{low} = 1\,\mu s$. The vertical lines in the central panel are the Coulomb blockade resonances, measured when the gate voltage is low. The inclined lines are the Coulomb blockade resonances during the high gate voltage. The splitting of the lines around $V_{low} = 650\,mV$ and a pulse height of $V_p = 50\,mV$ reveals a charge trap (see text). The right panel shows a zoom of such a splitting. $\beta_t \approx 0.9$ is very close to 1 so that the trap signature is not well visible in the nonlinear conductance plot (top panel). Sample 3.*

In the right panel of figure 5.11 the left line of the pair appears only for large V_{high}. This signifies that it corresponds to the occupied state. This means the shift due to the trap is close to one Coulomb blockade period (the traps always shift the resonances towards positive gate voltage). With eq. (5.15) we get $\beta_t \approx 0.9$, i.e. a trap very close to the main dot.

In order to measure the lifetime of the trap state we apply a 3-step pulse to the gate electrode. First V_{low}, then V_{high}, and finally a probe step at a voltage between them. We choose V_{low} and V_{high} such that they do not fall on a Coulomb blockade

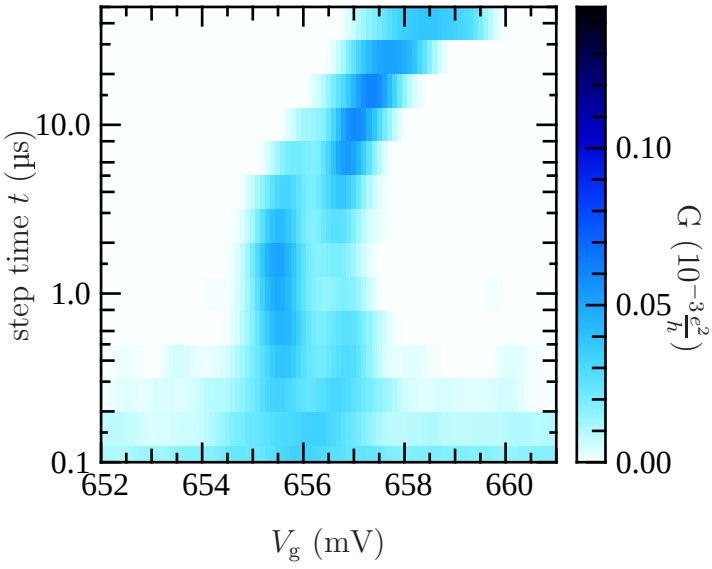

Figure 5.12: Relaxation time of a trap. *A 3-step pulse is applied to the gate electrode. First $V_{low} = 649\,\mathrm{mV}$ then $V_{high} = 702\,\mathrm{mV}$ is applied (asterisk in figure 5.11). There is no current at these two voltages. Finally a probe step with variable gate voltage (the horizontal axis) is applied. The duration t (the vertical axis) is the same for each of the three steps. Sample 3.*

resonance and the measured current can be fully attributed to the probe step. The result of such a measurement is shown in figure 5.12. The splitting of the line is clearly visible around $t = 1\,\mu s$. As before, the left line corresponds to the occupied state and the right one to the empty state. For very fast pulses the lines for empty and occupied trap have the same intensity. For $t \approx 1\,\mu s$ the line for occupied trap becomes more intense. This signifies that the time constant of the trap is of the same order as t, i.e. during the high gate voltage step it just has the time to get occupied but it takes almost the full time of the probe step for the trap to become empty. If the pulse time is increased further, the trap has time to relax to the empty state during the probe step. We conclude that the time constant of the trap is approximately 1 µs at the chosen gate voltages. The shift towards higher gate voltage at $t > 30\,\mu s$ is due to the setup (the pulse signal is fed through a capacitor causing a low frequency cut-off, see appendix A.2.4). The background for $t < 0.2\,\mu s$ is due to the finite raise time of the pulse signal (generated by a HP 33120A).

Very fast traps: measurements with sine excitation

The dynamics of most traps are still too fast to be captured by this measurement under pulse excitation. This is the case even when the HP 33120A is replaced with a high frequency pulse generator (Anritsu MP1763C). In this case the pulse length is limited to $\gtrsim 10\,ns$ by our setup, because due to the bonding process there are approximately 10 mm of unmatched line between the microwave line and the gate electrode which causes strong ringing at high frequency. Therefore, at very high frequency we prefer applying a sinus signal where no ringing can occur.

For the measurements with pulse excitation we used the fact that a trap shifts the Coulomb blockade resonances when it is occupied. But in section 5.3.1 we saw a second effect: a trap does also suppress the Coulomb blockade resonances. This is the effect we will probe with a sine excitation. As figure 5.13 explains, the Coulomb blockade resonances are restored when the gate voltage is swept so rapidly that

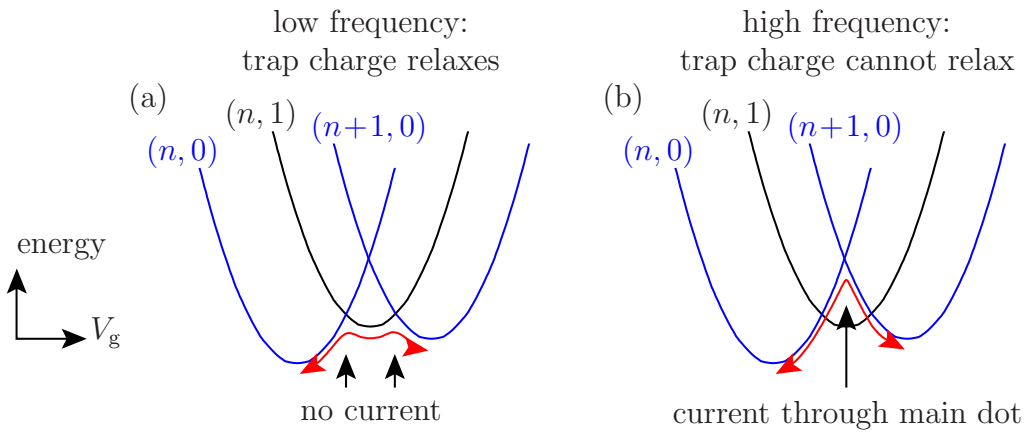

Figure 5.13: Energy diagram near a trap signature. *The numbers on top of the parabolas are the numbers of electrons on the dot and in the trap. (a) If the gate voltage is swept slowly, the degeneracy point between states* $(n, 0)$ *and* $(n + 1, 0)$ *is bypassed via state* $(n, 1)$*. The transitions* $(n, 0) \leftrightarrow (n, 1)$ *and* $(n, 1) \leftrightarrow (n, 0)$ *do not cause first order current through the dot so there is no Coulomb peak. (b) If the gate voltage is swept rapidly, the state* $(n, 1)$ *cannot be reached because the trap charge cannot relax. Thus the degeneracy point* $(n, 0) \leftrightarrow (n + 1, 0)$ *is reached and the corresponding Coulomb peak is restored. Analogue effects suppress and restore the resonance between states* $(n, 1)$ *and* $(n + 1, 1)$ *(not shown).*

the trap charge cannot relax. We do this by adding a RF sine wave to the DC gate voltage. The Coulomb blockade resonances should reappear as soon as the frequency of the RF signal is higher than the inverse time constant of the trap.

Figure 5.14 shows the corresponding measurement: the current through the dot is plotted as a function of the amplitude of the RF signal and the DC gate voltage for different frequencies. The behavior at 10 MHz is the same as at very low frequency. We make out 3 traps, one at $V_g = 770\,\text{mV}$, one at $V_g = 840\,\text{mV}$ and a very large signature centered at $V_g = 870\,\text{mV}$ manifested by suppressed current. At 100 MHz resonances appear in the signature at $V_g = 870\,\text{mV}$ and close it for sufficiently high pulse amplitude. We conclude that the lifetime of the trap is between 10 ns and 100 ns. The same is true for the trap at 840 mV. However, the signature at 770 mV is almost unchanged between 10 MHz and 100 MHz. Only at 1 GHz we observe this trap out of equilibrium.

Between 100 MHz and 10 GHz there is a second change in the trap signature near 870 mV. While at 100 MHz a peak-to-peak amplitude of $V_\sim = 10\,\text{mV}$ ($20\,\text{dB}_{\text{mVpp}}$) is necessary to produce the resonance at $V_g = 875\,\text{mV}$, it appears at much lower amplitude at 10 GHz. We see two possible explanations for this behavior.

- The trap is filled by photon-assisted tunneling. In this case it is not necessary to sweep the whole gate voltage range indicated in figure 5.13 but an electron can get the energy to enter or leave the trap by absorbing a photon. However, the energy of photons at 10 GHz is only 40 µeV. This is much smaller than

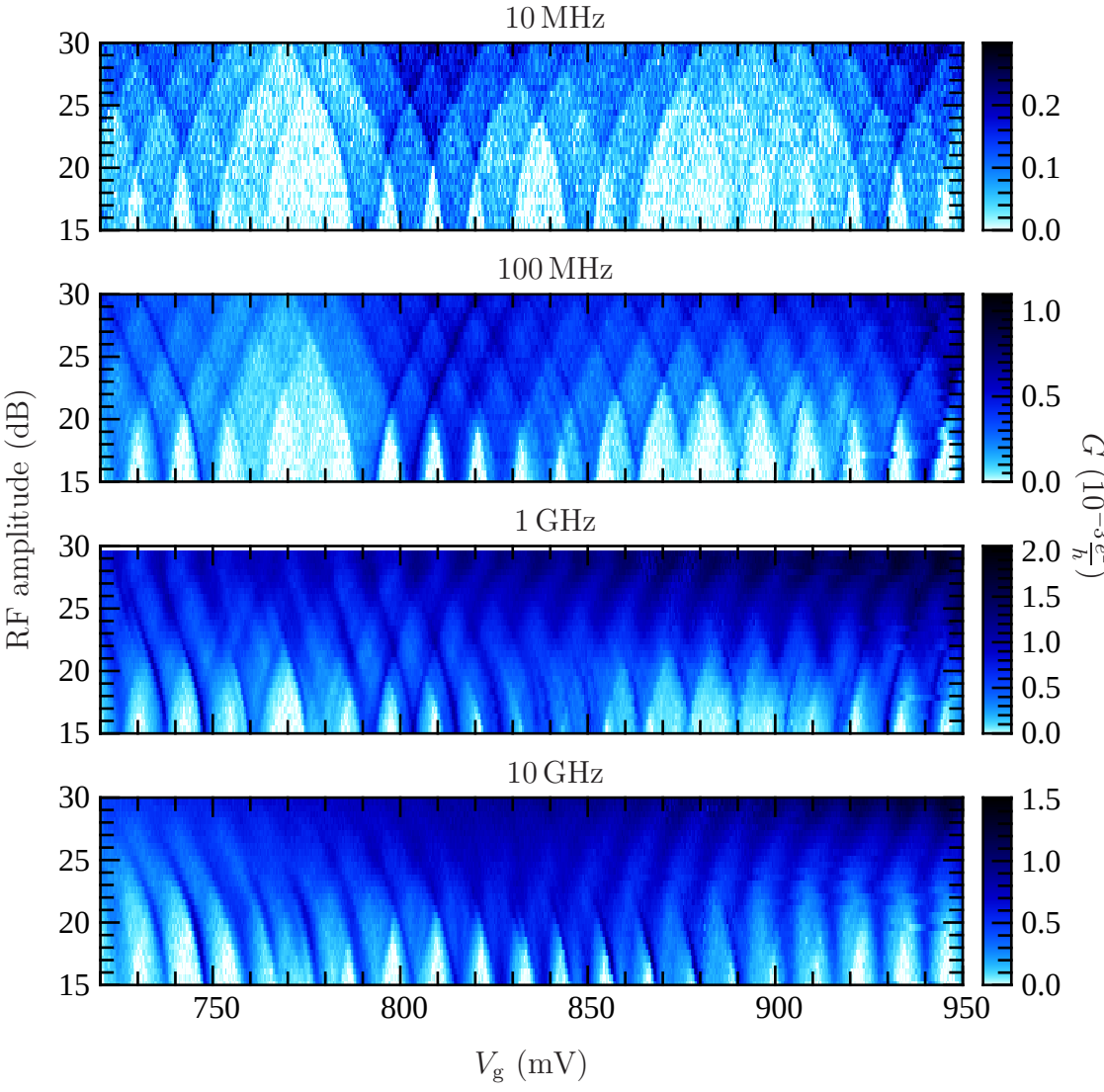

Figure 5.14: Traps under sinus excitation. *We apply a sinusoidal signal to the gate electrode. The frequency-dependent loss has been accounted for so that* 0 dB *corresponds to a peak-to-peak amplitude of* 1 mV*. These measurements indicate trap time constants in the order of* 1 ns *(see text). Sample 7.*

the energy of $e\alpha_m V_\sim = 2\,\mathrm{meV}$ corresponding to the pulse amplitude at a lever arm factor $\alpha_m = 0.2$. Therefore, a large number of photons would be needed to furnish sufficient energy. This makes the process extremely improbable.

- More probably, the pulse excitation is so fast that even the charge of the main dot and the charge rearrangements in the barriers (see section 5.3.3) cannot follow the signal any more. Then the gate voltage has a stronger effect on the trap because the charge on the main dot and the rearrangement cannot compensate the gate potential any more. Indeed, in the definition of the trap's lever arm factor α_t (cf. page 71) we have considered the charge of the main

dot as relaxed. Now that the charge of the main dot cannot follow the pulse signal any more and compensate it, the lever-arm factor of the trap increases considerably.

In order to see whether the pulse signal could be sufficiently fast to drive the charge on the dot out of equilibrium, let us suppose that the source and drain barriers have a conductance $G_\mathrm{T} = 2 \cdot 10^{-3}\,\frac{e^2}{h}$. This equals a peak conductance of $5 \cdot 10^{-4}\,\frac{e^2}{h}$ (see expr. (2.35)) which is what we observe. The time needed to charge one electron onto the main dot at a voltage difference $\alpha_\mathrm{m} V_\sim \approx 2\,\mathrm{mV}$ is then $\tau \approx \frac{e}{2 G_\mathrm{T} \alpha_\mathrm{m} V_\sim} \approx 0.4\,\mathrm{ns}$. Thus, 10 GHz should be a sufficiently high frequency to observe non-equilibrium charge states of the quantum dot itself.

For the 3 traps in figure 5.14 we found characteristic times of the order of magnitude of 10 ns. In the same sample at higher gate voltage we see even faster traps with characteristic times well below 1 ns. This is of the same order of magnitude as the time constant of the quantum dot itself. The fact that the traps can be as fast as the transport through the barriers is a further indication that the traps we observe are part of this transport process. This means they are inside the wire and not in the oxide or at the wire-oxide interface.

5.3.5 Spin

The spin of the trap state leads, via the Zeeman energy under magnetic field, to a gate-voltage shift of the trap signature of:

$$e\alpha_\mathrm{t} \frac{\partial V_\mathrm{g}}{\partial B} = g \mu_\mathrm{B} \Delta S_\mathrm{z} \tag{5.17}$$

where μ_B is the Bohr magneton and ΔS_z the change in spin quantum number of the trap state in direction of the magnetic field when an electron is added to the trap. It can take the values $\pm\frac{1}{2}$. If there are already electrons in the trap, higher changes are also possible, but they imply spin flips and such processes are expected to be very slow [126]. The Landé factor g for impurities in Si and SiO_2 has been measured by electron spin resonance [73]. The observed renormalizations are beyond the precision of our measurements, therefore we take $g = 2$. The gate-voltage lever-arm of the trap states α_t is very weak as we have shown above. The Zeeman shifts should therefore be strong.

Indeed, the magnetic field clearly shifts the trap signature in figure 5.15 to lower gate voltage. In order to identify the shift as the Zeeman effect, we compare it quantitatively with the prediction of our model. The shift of the resonances due to the trap is half the peak spacing (for example, the Coulomb blockade resonance at $V_\mathrm{g} = 350\,\mathrm{mV}$ shifts to 355 mV when at $B = 5 \ldots 8\,\mathrm{T}$), so $\beta_\mathrm{t} = \frac{1}{2}$ (see eq. (5.15)). The lever arm for the main dot for this gate voltage is $\alpha_\mathrm{m} = 0.26$ and the width of the trap signature varies from 2.5 periods without magnetic field to 1.5 periods at 16 T. This implies a gate-voltage lever arm for the trap of $\alpha_\mathrm{t} = 0.026 \ldots 0.043$ (see eq. (5.16)) which we interpolate as a linear function of magnetic field. Finally

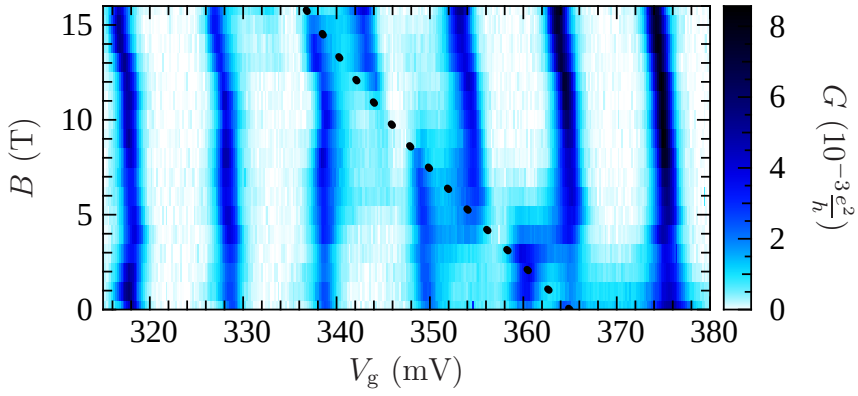

Figure 5.15: Shift of a trap signature with magnetic field. *The dotted line indicates the Zeeman shift expected for a trap state being occupied by a first electron. It depends on the gate-voltage lever-arm which in turn is determined by the width of the signature. This prediction of the Zeeman shift follows exactly the observed shift. Sample 2.*

the dotted line in figure 5.15 is obtained by putting this lever-arm and $S_z = -\frac{1}{2}$ in eq. (5.17). It is in very good agreement with the measured shift and confirms our model. The increase of the lever arm with magnetic field could be explained as follows. In the access regions the nanowire is close to the metal-insulator transition and the dopant states strongly increase the dielectric constant [19]. Under magnetic field they shrink [106], reducing the localization length and the dielectric constant in the wire. Therefore the coupling towards the main dot and the reservoir decreases while the gate capacitance dominated by the oxide capacitance remains unaffected.

We observe such Zeeman shifts in the majority of our samples. In most cases the trap signature shifts to lower gate voltage as in figure 5.15. This indicates that the traps are occupied with only one electron. If, on the contrary, they were occupied with a second electron it would have to occupy the energetically less favorable state whose energy is increased by the Zeeman effect. This would lead to a shift towards higher gate voltage under magnetic field, which we observe only very rarely. This single occupation is a further indication that the traps are actually the As-donor sites which can only be occupied by one electron due to Coulomb repulsion[2]. However, clusters of two donors could contain two or more electrons [11, 106]. For not too high doping levels, clusters should be rare though. Accordingly we observe much less shifts to higher than to lower gate voltage.

In devices based on similar technology Xiao *et al.* observed that all shifts occurred in direction of higher gate voltage [128, 129] indicating doubly occupied traps. With precise measurements of the Landé factor they located the traps inside the ox-

[2]Arsenic donors in silicon can be populated with 2 electrons but due to the electron-electron interaction the second electron is so weakly bound that in the scale of our devices it can be considered as delocalized. See for example ref. [106]

ide. This difference is further evidence that the traps we observe in our devices are not located in the oxide but inside the silicon wire.

5.4 Capacitance of the source and drain barriers

Coulomb blockade is a powerful tool for measuring small capacitances. The analysis in chapters 3.4 and 4 is based on precise measurements of the gate capacitance. We will now investigate the capacitance of the source and drain barriers which can be measured through the slopes in the Coulomb blockade diamonds (see eqs. (2.33) and (2.34)).

5.4.1 Increase of source and drain capacitance

Figure 5.16 shows the Coulomb blockade diamonds over a large gate-voltage range. The drain voltage height of the diamonds decreases from approximately 3 mV at $V_g = 200\,\text{mV}$ to 0.2 mV at $V_g = 2500\,\text{mV}$. This height is $\frac{e}{C}$ where C is the total capacitance of the quantum dot. It is the direct image of the charging energy $\frac{e^2}{C}$. As the width of the diamonds $V_+ = \frac{e}{C_g} = 13.2\,\text{mV} \pm 0.9\,\text{mV}$ is almost constant, this implies a strong increase of the source and drain capacitances. These capacitances are extracted from the Coulomb blockade diamonds with the algorithm described in appendix B.3. The result for figure 5.16 is shown in figure 5.17. The gate capacitance does not depend on gate voltage. The source and drain capacitances, however, increase by a factor 30.

How can this increase be explained? The simplest explanation would be that the length of the source and drain barriers decreases as gate voltage is increased. But this explanation cannot hold. The limit of the barriers on the reservoir side is well controlled by the abrupt increase of the doping level (see chapter 3.1, the precise control of sharp doping profiles is indeed one of the key features of CMOS technology). The gate voltage has only a small influence on this side of the barrier so that it cannot modify the potential significantly. But even the dot side limit of the barrier is not displaced by gate voltage because if it were, the size of the dot would change and therefore also its gate capacitance. This, however, is not the case as the constant peak spacing testifies.

As it is not the geometry of the barriers that changes, it must be their properties, i.e. the dielectric constant. The right hand axis of figure 5.17 is the dielectric constant of the source and drain barriers estimated from their capacitance through the planar capacitor approximation[3]. At low gate voltages we find dielectric constants close to the bare dielectric constant of silicon. In fact, the first points give $\epsilon < \epsilon_{Si}$. This signifies that we under-estimate the dielectric constant despite the fact that the planar capacitor approximation tends to over-estimate the dielectric constant

[3]Due to the high dielectric constant ($\epsilon \geq \epsilon_{Si} = 11.9$) compared to the dielectric constant of the surrounding silicon oxide ($\epsilon_{SiO_2} = 3.9$), the planar capacitor approximation is justified even though we estimate the barrier to be 35 nm long and only 17 nm thick.

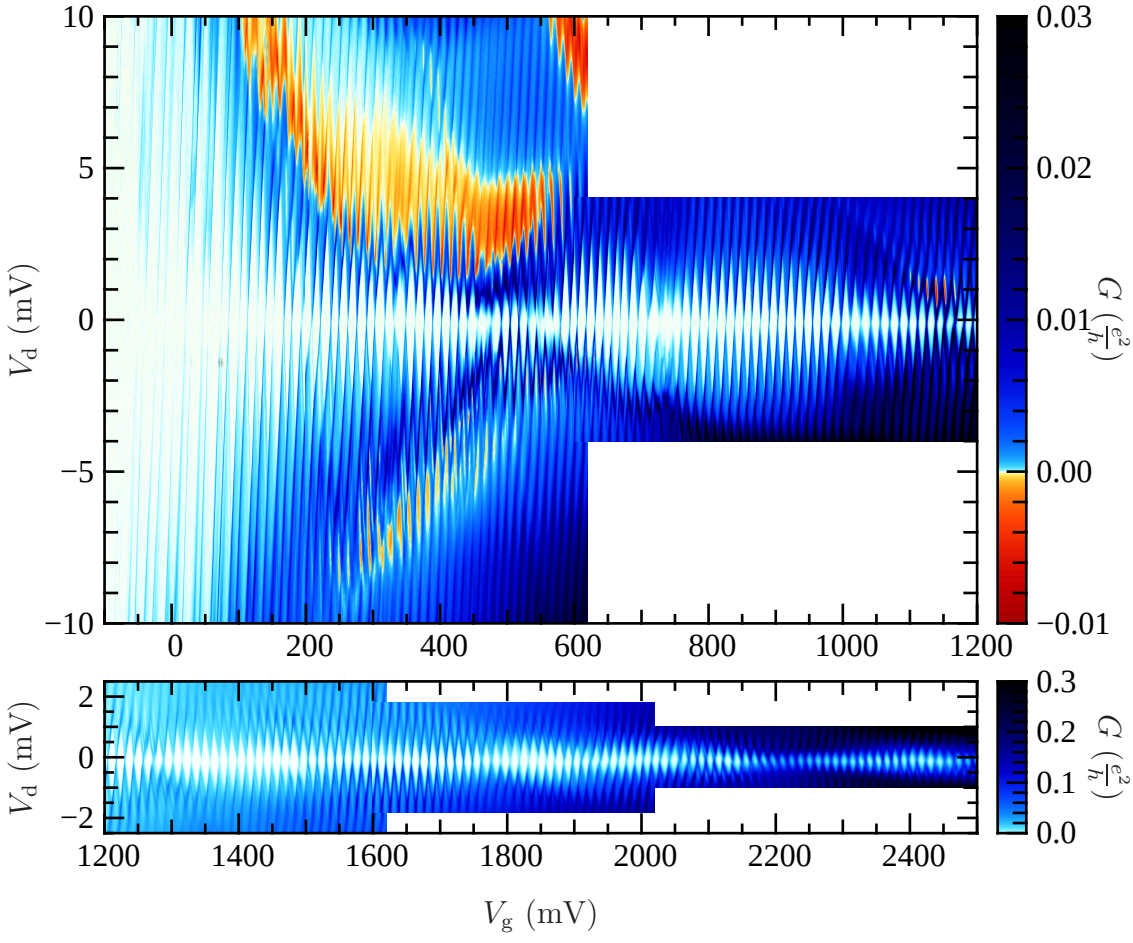

Figure 5.16: Global view of a stability diagram. *The height of the Coulomb blockade diamonds decreases dramatically with gate voltage. This implies that the total capacitance increases, and in consequence, as the gate capacitance is constant, that the source and drain capacitances increase. The V_g- and V_d-scales are the same for the two parts of the diagram, the color scales are not. Sample 3.*

because it does not consider edge effects. The reason for the under-estimation could be the fact that the quantum dot does not fill the whole wire but is only an accumulation layer close to the gate (see chapter 4.1.2). This makes that the effective length of the capacitor is somewhat longer than the distance between dot and barrier (see figure 5.18). However, this is a weak effect because the wire is thin compared to the length of the barriers. More likely, at very low gate voltage the quantum dot itself is close to the metal–insulator transition and screening inside it is not perfect, especially near the borders. So the dot itself will contribute to the dot-source and dot-drain capacitors and thus decrease the source and drain capacitances.

Nevertheless, we observe that, as the gate voltage is increased, the source and drain capacitances increase by a factor of 30. We will now try to explain this increase of the dielectric constant with our model of impurity states being charged as gate voltage is increased.

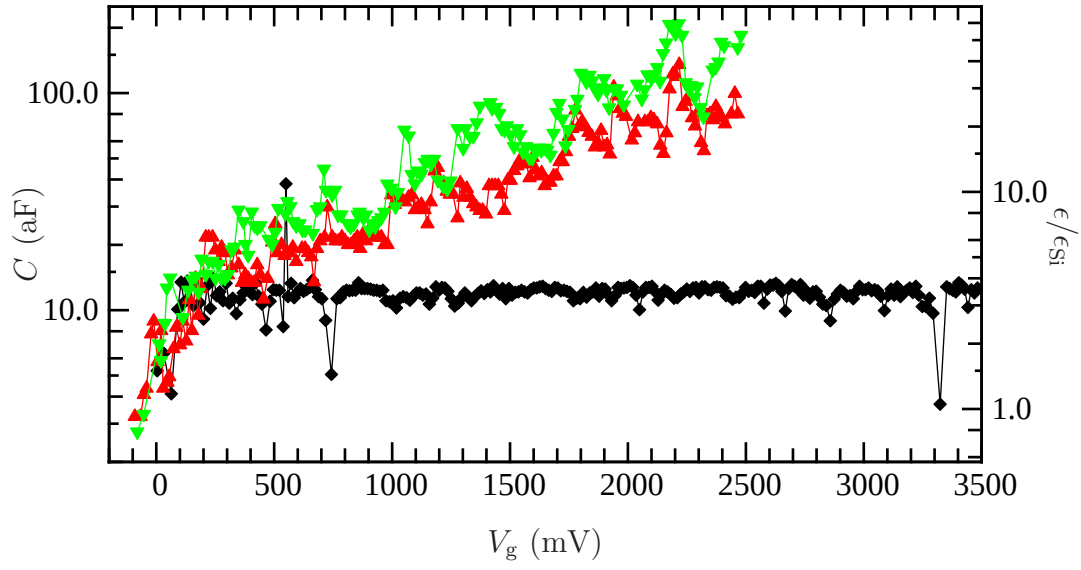

Figure 5.17: Typical evolution of the capacitances of a quantum dot with gate voltage. *(♦) gate capacitance, (▲) source capacitance, (▼) drain capacitance. These capacitances are extracted from the data of figure 5.16 (sample 3).*

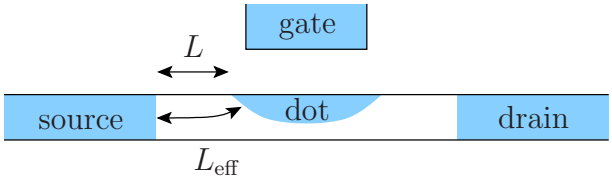

Figure 5.18: Effective barrier length. *Due to the fact that the dot is only a 2D layer the effective barrier length can be longer than* $L = 35\,\mathrm{nm}$.

5.4.2 Capacitances near charge trap anomalies

In 5.3 we have considered the effect of charging events in the barriers on the Coulomb blockade signature. In our model we supposed the trap capacitances to be small compared to those of the dot and in consequence neglected the indirect source or drain capacitance, i.e. the series capacitance of the dot–trap capacitor with the trap–source or trap–drain capacitor. To be precise, we did consider this indirect capacitance but in our model of constant capacitances, this capacitance also exists if the trap is empty. Therefore, the increase of the capacitance when the trap is populated is not taken into account by our model.

The increase ΔC of the barrier capacitor due to one dopant in its volume V can be estimated as

$$\frac{\Delta C}{C} \approx \frac{\alpha}{\epsilon_0 \epsilon_{\mathrm{Si}} V} \tag{5.18}$$

where α is the polarizability of the dopant. With the experimental value $\alpha = 10^{-35}\,\mathrm{Fm}^2$ (see page 65) and a typical barrier volume of $35\,\mathrm{nm} \times 50\,\mathrm{nm} \times 20\,\mathrm{nm} = 35000\,\mathrm{nm}^3$ one obtains $\frac{\Delta C}{C} = 0.003$. This is below the precision of our measurement. Yet, as figure 5.19 shows, the increase of capacitance when a trap is populated can

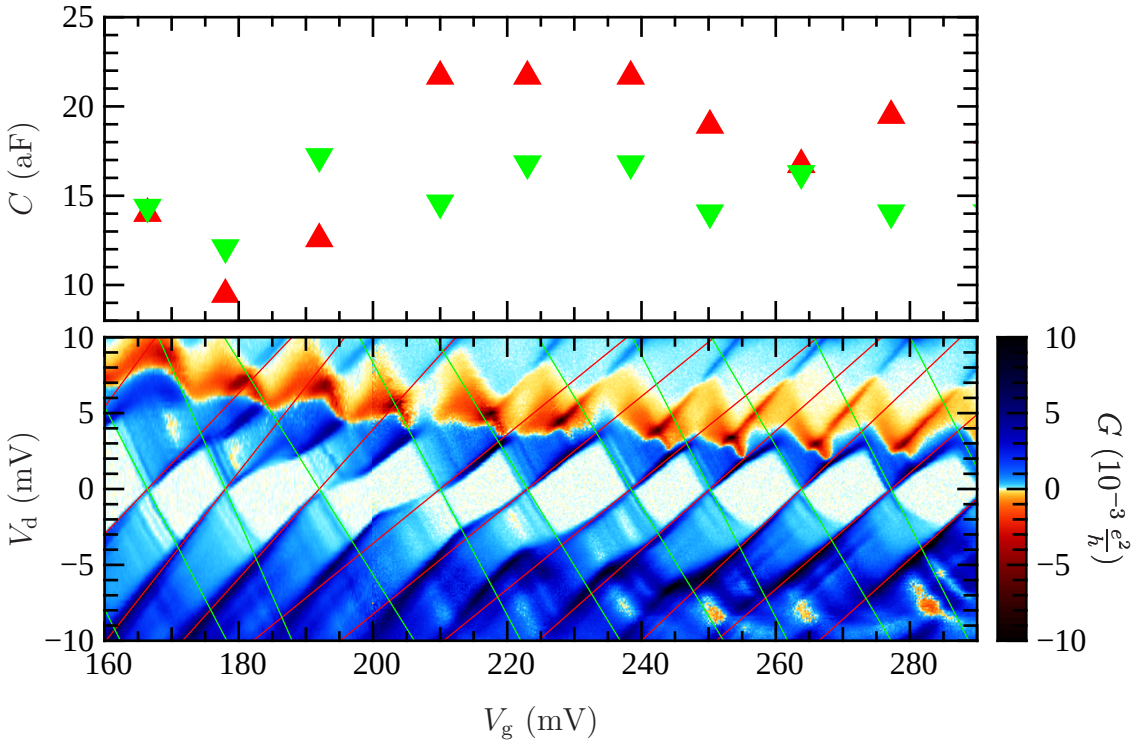

Figure 5.19: Increase of the capacitance at a trap signature. *Following the arguments of section 5.3 we conclude that the trap causing the signature around* $V_g = 200\,\mathrm{mV}$ *is on the source side of the dot. Accordingly we observe an increase of the source capacitance (▲) as the trap is populated while the drain capacitance (▼) remains unchanged. Another signature increasing the drain capacitance occurs at the right border of the figure. Sample 3.*

be quite strong (more than 100 % in this case). In accordance with figure 5.6 we attribute the charge trap signature in figure 5.19 to a charging event with reorganization in the source barrier (the lines inside the conducting regions are parallel to the negative diamond slopes). Accordingly, the source capacitance increases while the drain capacitance stays almost constant. Such a high increase of the capacitance when the number of electrons in the barrier goes from n to $n+1$ could be explained by the following scenario. For n electrons there is a clear energy separation between the ground state and excited states with different distributions of the electrons over the sites. The electrons can then screen a small external field only by deformation of the wave functions inside each site. This causes only a very small increase of the dielectric constant. As the $n+1$st electron is added, there could be several near-degenerate states with different distributions of the electrons over the sites. In this case the electrons can be displaced by a small external electric field and screen it efficiently. If the spacing of the excitation spectrum is smaller than the coupling width $\hbar\Gamma$ between sites, the charge density can be displaced continuously by an electrical field and the dielectric constant is significantly increased.

Such sudden increases of the capacitance near trap signatures seem quite fre-

quent in figure 5.17. After a jump, when the gate voltage is further increased, the capacitance seems to decrease smoothly until the next trap signature.

5.4.3 Average behavior of the capacitances

In addition to the strong mesoscopic fluctuations of the capacitances near trap signatures there is a strong increase of the capacitances in figure 5.17 on a long gate-voltage range. We will now try to describe this global behavior.

The arguments employed above to explain the increase of the capacitance with the addition of an electron to the barrier are similar in principle to the Mott-Anderson metal–insulator transition. We will try to explain the capacitance of the barriers as well as their conductance by a gate voltage driven metal–insulator transition in the barriers. First, we will give a short overview of the usual models for the dielectric constant in bulk metals and insulators and of the few models that exist for the dielectric constant near the metal-insulator transition.

Dielectric constant of metals and insulators and the transition between them

The dielectric constant $\epsilon(\boldsymbol{q},\omega)$ of a material with translational symmetry describes the response to a small potential

$$\delta U_{\mathrm{e}} = U_{\mathrm{e}} \mathrm{e}^{\imath \boldsymbol{q}\boldsymbol{r}} \mathrm{e}^{\imath\omega t} \mathrm{e}^{\alpha t} \tag{5.19}$$

with small positive α. The electrons will react to this potential by changing their charge distribution by $\delta\rho$. This charge redistribution creates itself a potential δU_{i} given by

$$\nabla^2 \delta U_{\mathrm{i}} = -\frac{\delta\rho}{\epsilon_0} \tag{5.20}$$

The dielectric constant is now defined as the ratio of the external potential δU_{e} over the total potential $\delta U_{\mathrm{t}} = \delta U_{\mathrm{e}} + \delta U_{\mathrm{i}}$.

$$\delta U_{\mathrm{t}} = \frac{\delta U_{\mathrm{e}}}{\epsilon(\boldsymbol{q},\omega)} \tag{5.21}$$

The charge redistribution $\delta\rho$ is driven by the total potential and can be calculated in first order perturbation theory and one finally obtains

$$\epsilon(\boldsymbol{q},\omega) = 1 + \frac{e^2}{\epsilon_0 q^2} \sum_{i,j} |\langle j|\mathrm{e}^{\imath\boldsymbol{q}\boldsymbol{r}}|i\rangle|^2 \frac{f(E_i) - f(E_j)}{E_j - E_i - \hbar\omega + \imath\hbar\alpha} \tag{5.22}$$

where f is the electron distribution function. Expression (5.22) is known as Lindhard's expression. It is the starting point for most calculations of the dielectric constant. For a more complete derivation of expression (5.22) see for example [132].

From now on we consider only the static dielectric constant, i.e. $\omega = 0$. In the case of a metal and for small wave vectors $\boldsymbol{q}$, the most significant contributions to

(5.22) come from states near the Fermi level, where the real part of the denominator is small. In a metal near the Fermi level the quasi particle states i and j can be approximated by plane waves. The matrix element becomes then $\delta(\boldsymbol{k}_j - \boldsymbol{k}_i - \boldsymbol{q})$. The fraction in (5.22) can be approximated as $-f'(E)$ and we finally get the static dielectric constant of a metal

$$\epsilon(\boldsymbol{q}, 0) = 1 + \frac{\lambda^2}{q^2} \tag{5.23}$$

with

$$\lambda^2 = -\frac{e^2}{\epsilon_0} \int_{-\infty}^{\infty} \mathrm{d}E f'(E) N(E) \approx \frac{e^2}{\epsilon_0} N(E_\mathrm{F}) \tag{5.24}$$

This approximation is valid if the density of states $N(E)$ is constant in the energy range where the distribution function goes from 1 to 0. This range is of order $k_\mathrm{B}T$ in equilibrium where f is the Fermi function.

The dielectric constant diverges for $\boldsymbol{q} \to 0$. This means that an external perturbation is completely screened out over long distances. For example, the screened potential U_t caused by a point charge falls off exponentially with $e^{-\lambda r}$. λ^{-1} is called the Thomas-Fermi screening length.

The divergence of the dielectric constant and the resulting screening are due to the high density of *delocalized* states at the Fermi level. They allow the electrons in them to be easily attracted by a positive electrical potential or repelled by a negative one. Therefore they cause a negative or positive charge density which compensates the electrical potential completely.

In contrast, in the case of a band insulator the density of states is 0 at the Fermi level. Then Lindhard's expression (5.22) can be evaluated [132] as

$$\epsilon(\boldsymbol{q}, 0) \approx 1 + \frac{n_\mathrm{V} e^2 \hbar^2}{\epsilon_0 m E_\mathrm{g}^2} \tag{5.25}$$

where n_V is the number of electrons in the valence band per volume. E_g is the effective band gap; it is larger than the optical gap but smaller than the energy difference between the centers of the valence and conduction bands because the weighting due to the denominator in (5.22) favors small energy differences. In the case of silicon, given its dielectric constant $\epsilon = 11.9$ and $N = 1.25 \cdot 10^{28}\,\mathrm{m}^{-3}$, expression (5.25) gives an effective band gap of $1.26\,e\mathrm{V}$, somewhat larger than the optical band gap $E_\mathrm{g} = 1.17\,e\mathrm{V}$

An equivalent expression to (5.25) can be obtained in the more phenomenological model of atomic polarizability α, which is defined via

$$\boldsymbol{p} = \alpha \boldsymbol{E}_\mathrm{loc} \tag{5.26}$$

where $\boldsymbol{p}$ is the polarization of the atom and $\boldsymbol{E}_\mathrm{loc}$ the local electric field.

In this semi-classical model one obtains with the Lorentz-Lorenz local field correction the Clausius-Mossotti relation

$$\epsilon = 1 + 3\left(\frac{1}{1 - \frac{N\alpha}{3\epsilon_0}} - 1\right) \tag{5.27}$$

Expressions (5.25) and (5.27) of the dielectric constant are physically equivalent. In the case of small dielectric constants, i.e. $\frac{N\alpha}{\epsilon_0} \ll 1$, and if the atoms are considered as almost isolated harmonic oscillators the two expressions give exactly the same result because the band gap becomes the excitation energy of the harmonic oscillator which in turn is related to the polarizability via the spring constant.

Yet another way of obtaining the dielectric constant is by using one of the Kramers-Kronig relations:

$$\epsilon(0,\omega) = 1 + \frac{2}{\pi\epsilon_0}\mathcal{P}\int_0^\infty \frac{\sigma(\omega')}{\omega'^2 - \omega^2}\mathrm{d}\omega' \tag{5.28}$$

This equation is far more general than (5.25) or (5.27) and contains no approximations. It follows directly from Cauchy's integral formula and the fact that the generalized electrical susceptibility $\tilde{\chi} = \chi + \frac{\imath\sigma}{\omega}$ is the Fourier transform of a causal response function (i.e. a function which is 0 for negative time).

In the case of a metal with $\sigma(0) > 0$, the principal value integral (5.28) diverges for $\omega \to 0$.

In the case of an insulator, conduction can only begin when electrons from the valence band can be excited to the conduction band. Simply setting $\sigma(w) = \sigma_0\Theta(\hbar\omega - E_\mathrm{g})$, equation (5.28) gives $\epsilon(0,0) = 1 + \frac{2\hbar\sigma_0}{\pi\epsilon_0 E_\mathrm{g}}$, a finite static dielectric constant. A more precise model for $\sigma(w)$ is needed to produce the E_g^{-2} dependence of equation (5.25).

Now, what happens when an insulator is continuously converted into a metal? We will consider the example of the metal–insulator transition of n-silicon as the concentration of occupied dopant states is increased. As we want to consider the increase of the dielectric constant due to the donors and take the host dielectric constant as given, we have to replace ϵ_0 in the formulae above by the host dielectric constant $\epsilon_0\epsilon_\mathrm{Si}$.

Castner *et al.* [19] found the static dielectric constant of silicon to increase with increasing donor concentration and finally to diverge. In the case Si:As they found a critical concentration $(6.4 \pm 0.5) \cdot 10^{24}\,\mathrm{m}^{-3}$.

Already the Clausius-Mossotti relation predicts such a divergence of the dielectric constant as $\frac{N\alpha}{3\epsilon_0\epsilon_\mathrm{Si}} \to 1$. The resulting critical concentration is $N_\mathrm{c} = 3 \cdot 10^{25}\,\mathrm{m}^{-3}$ (we use the experimental polarizability (5.8)). This is of the same order as Castner's experimental result. Already in 1927, Herzfeld [43] showed that this critical concentration allows to discern metals and insulators. In his description the Lorentz-Lorenz field compensates the "spring constant" as the critical concentration is approached and finally sets the electrons free.

The problem of the metal-insulator transition in doped semiconductors is different inasmuch as the dopants do not form a periodic crystal. In fact, disorder is the main ingredient for this transition. The derivation of the Clausius-Mossotti relation, however, relies upon a highly symmetric (for example simple cubic) crystal structure, where the induced dipole moments of neighboring sites cancel out. Castner observed a strong upward derivation from the Clausius-Mossotti expression as the critical concentration was approached. Later, he attempted to describe this deviation by a rather cumbersome extension of the Clausius-Mossotti expression including disorder [18].

Imry *et al.* parameterized the evolution of the dielectric constant at the metal–insulator transition with the localization length ξ, which has been found to describe accurately the scaling of conductance near the metal–insulator transition in many different systems (see for example the reviews of Beenakker [9] and Mirlin [84]). The localization length can be seen as the average length over which a wave function spreads. It is finite in an insulator and infinite in a metal. Imry's main argument is simple. One starts from the metallic dielectric constant (5.23). Electrons can be considered as free on a range ξ. Thus, they can screen a potential with a wave length smaller than ξ just like a metal. If the external potential has a wave length larger than ξ the electrons cannot follow the potential completely because they are confined within a range of ξ: the singularity in (5.23) is cut off at $q \approx \xi^{-1}$. Therefore the static macroscopic dielectric constant is

$$\epsilon(0,0) = 1 + b\lambda^2\xi^2 \tag{5.29}$$

where b is a constant of order 1. For $q \gg \xi^{-1}$ the metallic dielectric constant is recovered.

For an infinite quasi 1D wire without interaction Efetov [23] calculated the prefactor b in the framework of the 1D σ-model. He found[4] $b = 8\,\zeta(3) \approx 9.62$, ζ being the Riemann ζ-function.

Matching conductance and capacitance

As we mentioned above, the localization length ξ used to parameterize the capacitance has also been used to describe the conductance near the metal-insulator transition. In the framework of the random matrix theory [9] one finds the conductance in a long quasi-1D wire to be log-normally distributed.

$$\mathrm{var}(\ln g) = -2\langle \ln g \rangle \tag{5.30}$$

Here g is the conductance in quantum units , i.e. $G = \frac{e^2}{h} g$. The average logarithmic conductance is —in the orthogonal ensemble, i.e. in the absence of magnetic field, magnetic impurities and spin–orbit coupling— given by

$$\langle \ln g \rangle \approx \frac{-2L}{\xi} + \ln g_0 \tag{5.31}$$

[4] He writes $\epsilon = 32\zeta(3)e^2 N_{\mathrm{F}} \xi^2$. (formula (11.73) in [23]). We corrected it because to our understanding this expression is in cgs units, and a factor π and the additional 1 have been omitted.

The constant g_0 is of order unity. In the framework of the 1D σ-model (like Efetov's expression of the dielectric constant), Mirlin [84] calculates the average conductance including prefactors.

$$\langle g \rangle = 2^{-5/2}\pi^{7/2}(\xi/l)^{3/2}\mathrm{e}^{-L/2\xi} \tag{5.32}$$

Orders higher than 2 of the conductance distribution could not yet be calculated in the σ-model, so that for the average of the logarithmic conductance only the RMT result exists.

These results suppose that the coherence length exceeds the sample size. In the barriers of our samples where the density of states is low, this condition should be satisfied at low temperature and low bias voltage. Yet we cannot access the source and drain conductance separately[5] and transport is not coherent through the entire sample because the spacing of the excitation spectrum in the dot is $\lessapprox k_\mathrm{B}T$. Therefore we have to consider the barriers separately and calculate the total conductance with expr. (2.35). The conductance is then

$$g_\mathrm{peak} = \frac{1}{2}\frac{g_\mathrm{s}g_\mathrm{d}}{g_\mathrm{s}+g_\mathrm{d}} \tag{5.33}$$

on top of the Coulomb blockade peaks. In consequence we cannot compare expressions (5.29) and (5.31) directly. Instead we proceed as follows. First we calculate the localization length based on capacitance via expression (5.29). We keep the prefactor $b\lambda^2$ as a fitting parameter, the same for all peaks in all samples. (We will discuss the fluctuations later on.) With this localization length we calculate the conductance of each barrier according to expression (5.31). The constant g_0 in (5.31) is taken as a second fitting parameter, again the same for all peaks and all samples. With expression (5.33) we finally get an estimation of the peak conductance based on the source and drain capacitance. In figure 5.20 we compare this estimated conductance to the measured peak conductance.

At high conductance we obtain a good agreement between this capacitance-based conductance estimation and the measured peak conductance. The the two fitting parameters are $b\lambda^2 = 0.05\,\mathrm{nm}^{-2}$ and $g_0 = 5$. These are quite reasonable values. g_0 is close to unity as predicted by theory and, with Efetov's prefactor, $\lambda^{-1} = 14\,\mathrm{nm}$ is close to the average nearest neighbor distance between dopants ($n_\mathrm{D}^{-1/3} = 10\,\mathrm{nm}$) as one would expect in a Coulomb glass (see [66] and chapter 4.1.2). The corresponding density of states is $2.1 \cdot 10^{43}\,\mathrm{m}^{-3}\mathrm{J}^{-1}$.

The plot contains 989 points[6] from 10 samples with various wire widths. 50 of these points have an estimated conductance too small to appear on the plot. The statistical properties of the data in figure 5.20 are plotted in figure 5.21. They

[5] Actually, the nonlinear conductance allows to separate source and drain conductance. But the expression is quite cumbersome in the metallic regime at finite temperature. The trap signatures (see above) and the fluctuating density of states in the barriers modify the nonlinear conductance significantly so that we consider an individual determination of source and drain conductance too daring.

[6] Approximately 200 peaks are detected per sample, but source and drain capacitance cannot be extracted reliably for all of them.

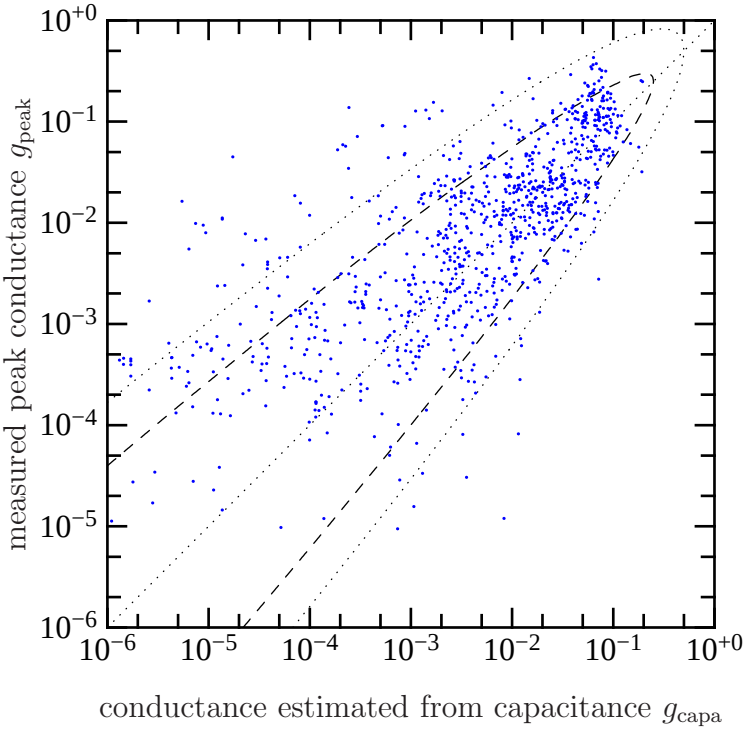

Figure 5.20: Correlation between capacitance and conductance. *The horizontal axis is the estimation of the drain–source conductance obtained by translating capacitance to conductance (see text). The only 2 fitting parameters are a prefactor 5 in equation (5.31) and a screening length of 14 nm. The dashed and dotted curves are estimations of the fluctuation based on RMT (see text).*

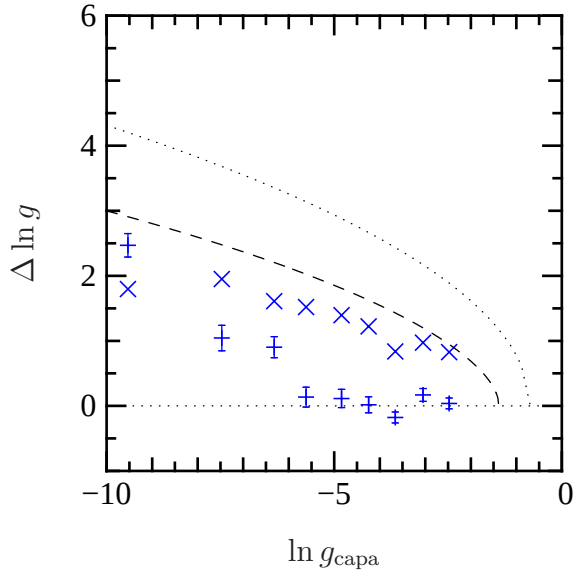

Figure 5.21: Statistical properties of data from figure 5.20. *The statistics are evaluated for groups of 100 points. The difference of vertical and horizontal axis* $\Delta \ln g = \ln g_{\mathrm{peak}} - \ln g_{\mathrm{capa}}$ *is plotted versus the capacitance based conductance estimation* $\ln g_{\mathrm{capa}}$. *(×) root mean square of* $\Delta \ln g$. *(+) average of* $\Delta \ln g$. *The error bars mark the standard deviation of this average. Deviations below* $\ln g_{\mathrm{capa}} = -6$ *are attributed to measurement errors (cf. figure 5.22).*

have been obtained in the following way. The points are first ordered by estimated conductance g_{capa}. Then we evaluate the average $\Delta \ln g = \ln g_{\mathrm{peak}} - \ln g_{\mathrm{capa}}$ for sets of approximately $n = 100$ points (marker + in figure 5.21). The error bars are the estimated standard deviation of the mean value

$$\sigma(\langle \Delta \ln g \rangle) = \sqrt{\frac{\langle (\Delta \ln g)^2 \rangle - \langle \Delta \ln g \rangle^2}{n}} \tag{5.34}$$

The observed deviations of $\langle \Delta \ln g \rangle$ from 0 are of the same order as these error bars. This means that the predictions (5.29) and (5.31) are consistent with our observations. Only for low conductance there is a significant deviation towards

positive $\langle \Delta \ln g \rangle$. The reason for this deviation can be seen in figure 5.22 where the influence of measurement errors has been simulated. For this plot we have chosen independent localization lengths for source and drain, equally distributed between 0 and the barrier length. Then we calculated the conductance with expressions (5.31) and (5.33) and the capacitance with expression (5.29) using Efetov's prefactor and a screening length $\lambda^{-1} = 14\,\mathrm{nm}$. After that, rather pessimistic "measurement" errors are added: $10^{-4}\,\frac{e^2}{h}$ noise floor, and 10 % relative error for the conductance and $5\,\epsilon_0$ absolute error and 50 % relative error for the dielectric constant. This is an overestimation of the RMS error but there are sporadic misinterpretations of the capacitance with completely random results (see appendix B.3). As figure 5.22 shows, the noise floor of the current measurement is the only measurement error that causes a significant deviation from perfect correlation. It is the cause for the upward deviation of $\langle \Delta \ln g \rangle$ at low conductance. Fortunately, of all measurement errors this noise floor is the easiest to characterize. One simply has to evaluate the current fluctuations when there is no measurable current, for example inside the Coulomb diamonds or at gate voltages below the first Coulomb blockade resonance. In our case its main contribution is the thermal current noise of the current amplifier's feedback resistance (see appendix A). The noise floor varies from sample to sample because the integration times for the lock-in measurements and the voltage excitations are different. In consequence, in the experimental data we cannot observe a sharp horizontal border as in the simulation.

The dispersions in figure 5.22 caused by the other simulated "measurement" errors are much smaller than the actual dispersion in the measured data. Thus, we can conclude that the dispersion in figure 5.20 — at least at higher conductance — is not due to measurement errors but is significant. We will discuss this dispersion now.

Correlation of conductance and capacitance fluctuations

We took the inverse screening length λ or, equivalently, the density of states at the Fermi level as a constant fitting parameter — the same for all samples. In other words we did not consider gate voltage dependent fluctuations or sample to sample fluctuations of this parameter. As only the capacitance and not the conductance depends on this parameter, its fluctuations cause a dispersion of the points in figure 5.20.

An individual value for λ for the source and drain barrier and for each peak cannot be calculated because we do not have access to the source and drain barrier conductance separately. And even if it could be calculated, such an analysis would not be more meaningful than our analysis because (5.29) and (5.31) describe the average capacitance and conductance and not their individually observed values. An individual λ for each barrier and each resonance would mean to project all possible fluctuations onto λ.

The fluctuations of conductance around its mean value defined by λ are given by eq. (5.30) ($\mathrm{var}(\ln g) \approx 4L/\xi, \xi \ll L$). For the fluctuations of capacitance there are, to

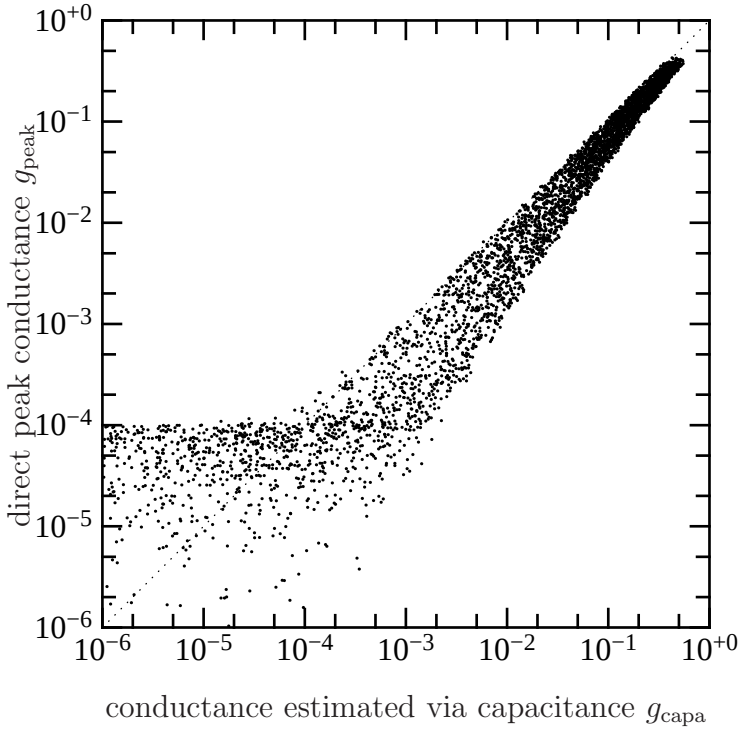

Figure 5.22: Simulated impact of measurement errors. *Conductance and capacitance are supposed to be fully correlated through expressions* (5.29), (5.31) *and* (5.33). *Dispersion is introduced by rather pessimistic Gaussian "measurement" errors in the capacitance and the peak conductance (see text). As the dispersion is much smaller than in the real data of figure 5.20, we conclude that the dispersion in figure 5.20 is significant.*

our knowledge, no theoretical predictions. However, conductance and capacitance fluctuations only cause dispersion in our plot as long as they are not correlated. Otherwise they would be interpreted as fluctuations of the localization length itself. Obviously the fluctuations are not fully correlated because we observe indeed a strong dispersion.

Now, on the contrary, what would be the dispersion if the fluctuations of capacitance and conductance were completely uncorrelated around an average behavior described by expressions (5.31) and (5.29)? We can find a lower boundary of the dispersion if we neglect the capacitance fluctuations and consider only the conductance fluctuations, for which the theoretical prediction (5.30) exists. To do this, we have to evaluate how fluctuations of the barrier conductances g_{s} and g_{d} are transferred to the peak conductance g_{peak}. This depends on the relative mean values of the barrier conductances. We evaluate the two extreme cases of very different or exactly equal mean values. In both cases we consider the fluctuations of g_{s} and g_{d} as independent.

In the case of very different mean values, i.e.

$$(\langle \ln g_{\mathrm{s}} \rangle - \langle \ln g_{\mathrm{d}} \rangle)^2 \gg \mathrm{var}(\ln g_{\mathrm{s}}) + \mathrm{var}(\ln g_{\mathrm{s}}) \tag{5.35}$$

the peak conductance can be approximated as

$$g_{\mathrm{peak}} \approx \frac{1}{2} \min\{g_{\mathrm{s}}, g_{\mathrm{d}}\} \tag{5.36}$$

Then the distribution of the logarithmic peak conductance follows the distribution of the smaller conductance. Thus

$$\mathrm{var}(\ln g_{\mathrm{peak}}) = -2(\langle \ln g_{\mathrm{peak}} \rangle + \ln 2) \tag{5.37}$$

The dotted lines in figures 5.20 and 5.21 are respectively the corresponding 15.9 %, 50 % (median) and 84.1 % quantiles of this distribution. As the distribution is a Gaussian, the 50 % quantile is the average and the 15.9 % and the 84.1 % quantiles are one standard deviation away from it.

In the other extreme $\langle \ln g_s \rangle = \langle \ln g_d \rangle$ the error transfer function is more complicated and we evaluate it via a Monte-Carlo simulation. We suppose the logarithmic source and drain conductance to have the same mean value and the same Gaussian distribution with variance given by (5.30) but to fluctuate independently. The dashed lines in figures 5.20 and 5.21 correspond respectively to the 15.9 % and 84.1 % quantiles of the resulting distribution which is close to a Gaussian but slightly asymmetric.

We expect the real distribution to be closer to the latter case because condition (5.35) requires huge differences of the average source and drain capacitances. On the other hand we observe quite different gate-voltage dependencies of the peak height in different samples with the same geometry. As there is no reason to believe that there are less differences between the source and drain barriers of one sample than there are between the barriers of samples with the same geometry, this leads to suppose that the source and drain barrier conductances can evolve very differently with gate voltage. Indeed, nowhere in our analysis we supposed a specific gate voltage dependence of the localization length.

In figure 5.20 the dispersion of the data points seems to be well described by our prediction for similar source and drain conductances (dashed lines). The comparison of the square-root of the variances in figure 5.21 reveals that the observed fluctuations are actually somewhat smaller than the prediction. This seems to indicate that conductance and capacitance fluctuations are correlated. But one has to be careful here because the RMT results (5.31) and (5.30) suppose $\xi \ll L$ in order to neglect all but the strongest conduction channel. In our samples however, ξ and L have the same order of magnitude and our need for a prefactor 5 in (5.31) could be an indication that more than one conductance channel participate significantly to transport.

We will now try to learn more about the fluctuations by studying the evolution of conductance and capacitance with gate voltage. Figure 5.23 shows the localization lengths $\xi = \sqrt{\frac{\epsilon/\epsilon_{\mathrm{Si}}-1}{b\lambda^2}}$ (cf. eq. (5.29)) related to source (▲) and drain (▼) capacitance. For the conductance we plotted the localization length corresponding to symmetric barriers, i.e. $g_{\mathrm{barrier}} = 4g_{\mathrm{peak}}$. We did so not only for the measured peak conductance (●) but also for the conductance estimated from the capacitances (♦).

The figure shows that the large scale evolution of estimated conductance and measured peak conductance are perfectly correlated. On a small gate-voltage scale however, there is no clear correlation.

This observation is confirmed by figure 5.24(a) where the correlation coefficient

$$c(\Delta V_g) = \frac{\langle \Delta_{\mathrm{peak}} \Delta_{\mathrm{capa}} \rangle}{\sqrt{\langle \Delta^2_{\mathrm{peak}} \rangle \langle \Delta^2_{\mathrm{capa}} \rangle}} \tag{5.38}$$

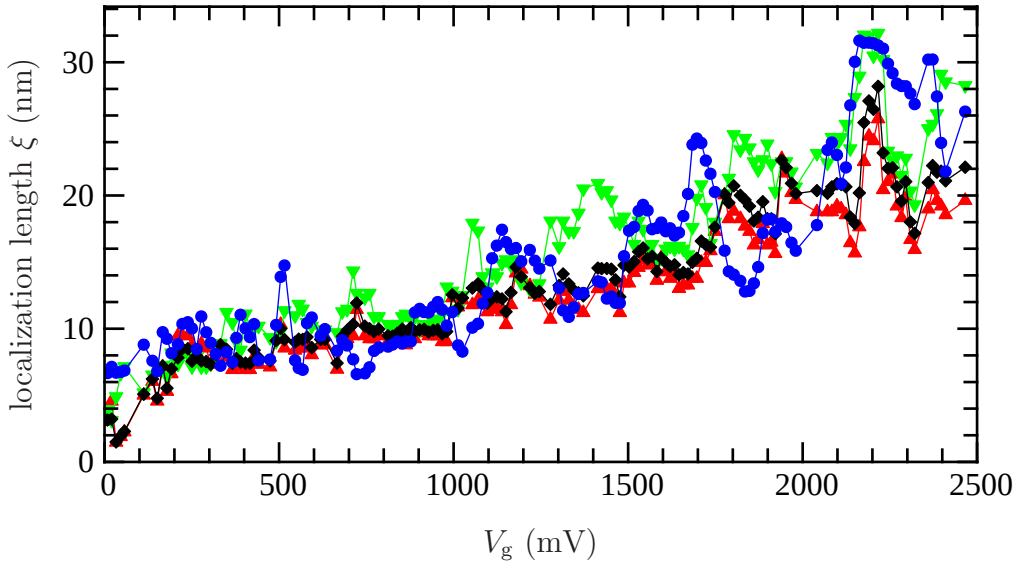

Figure 5.23: Evolution of the localization length with gate voltage. *The different localization lengths correspond to the following quantities: (▼) drain capacitance, (▲) source capacitance, (♦) conductance estimated from capacitance, (●) measured conductance. We use the global fitting parameters of figure 5.20 ($b\lambda^2 = 0.05\,\mathrm{nm}^{-2}$, $g_0 = 5$). Sample 3.*

with

$$\Delta_{\mathrm{peak}} = \ln g_{\mathrm{peak}}(V_g + \Delta V_g) - \ln g_{\mathrm{peak}}(V_g) \tag{5.39}$$

$$\Delta_{\mathrm{capa}} = \ln g_{\mathrm{capa}}(V_g + \Delta V_g) - \ln g_{\mathrm{capa}}(V_g) \tag{5.40}$$

is plotted. The average is taken over gate voltage V_g and the samples.

For large gate-voltage ranges ΔV_g, the correlation coefficient comes close to 1. This supports what we already know, namely that capacitance and conductance can be linked via the localization length. For small gate-voltage ranges $\Delta V_g \to 0$ however, the correlation coefficient tends to a very small value, approximately 0.1.

Figure 5.24(b) shows the distribution of Δ_{peak} versus Δ_{capa} with $\Delta V_g \approx 200\,\mathrm{mV}$. The distribution of Δ_{capa} is much less symmetric than that of Δ_{peak}: the capacitance seems to fluctuate only weakly most of the time. Only rarely are there strong changes which tend to go towards positive capacitance. In section 5.4.2 we have associated such abrupt changes with electrons being added to the barriers.

The small correlation coefficient between the changes of capacitance and conductance does not necessarily mean that changes in conductance and capacitance are unrelated. Interpreting figure 5.24(b) daringly one could see a "C"-like structure: for a positive change of capacitance the conductance either increases or decreases but only rarely stays the same. In terms of the description of electrons being added to the barriers this would imply that the capacitance most likely increases when an electron is added whereas the conductance can increase or decrease. In other words,

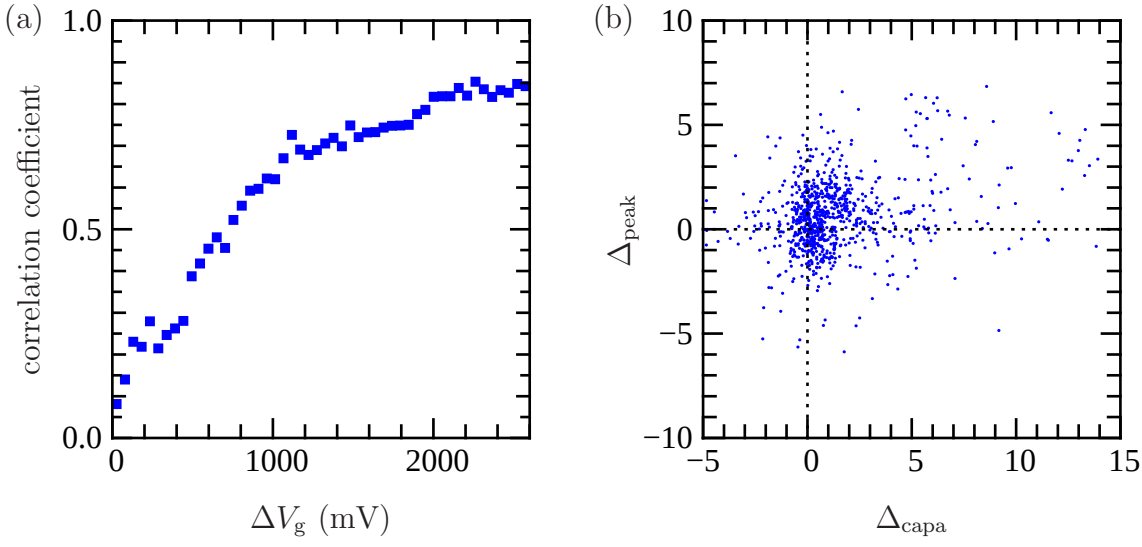

Figure 5.24: Correlations of capacitance and conductance fluctuations. *(a) Correlation coefficient* (5.38) *between increase over a gate voltage distance* ΔV_g *of measured and capacitance-estimated conductance. The correlation coefficient is averaged over gate voltage offset and 10 samples (b) Scatter plot from which the correlation coefficient at* $V_g = 200\,\mathrm{mV}$ *has been calculated.*

according to section 5.4.2, strong changes of capacitance and conductance occur in a correlated manner near trap signatures but as the conductance can change in both directions, the correlation coefficient is weak.

Such a behavior is similar to the frequency dependence of capacitance and conductance linked by the Kramers-Kronig relations. In fact, the gate voltage dependence of conductance and capacitance and their frequency dependence are related. In first order time-dependent perturbation theory the tunneling rate between reservoirs and the dot is proportional to $\delta(\Delta E \pm \hbar\omega)$ where ΔE is the energy difference between final and initial state. It depends linearly on gate voltage: $\Delta E = e\alpha V_g + E_0$. Thus, without interactions, the conductance at frequency ω and gate voltage V_g can be expressed by the conductances at zero frequency and gate voltages $V_g \pm \frac{\hbar\omega}{e\alpha}$. In presence of interactions the situation is more involved because the excitation spectrum changes with gate voltage and the number of electrons in the dot and the barriers (scrambling). Nevertheless, gate voltage dependence and frequency dependence are related.

5.4.4 Interactions

The theories leading to equations (5.29), (5.31) and (5.30) are one-particle theories and therefore do not take into account electron-electron interactions beyond mean field approximation. The most prominent effect of electron-electron interactions inside the barriers is to cause a gap in the single-particle density of states around the Fermi level [106]. This also reduces the localization length and thus affects

capacitance and conductance. However, the influence of the gap is not necessarily the same for conductance and capacitance because the conductance depends on the density of states in an energy range given by $\hbar\Gamma$ where Γ is the coupling between sites. According to equation 5.25 the capacitance should decrease in E_g^{-2} with the band gap E_g and has therefore no characteristic energy scale.

In our system we also have to consider the electrostatic interactions of the electrons in the barrier with the main dot. As we have seen in section 5.3 they suppress the low bias Coulomb blockade resonances when an additional electron is added to the barrier but do not lead to a decrease of the capacitance.

Yet, over a large gate voltage range we found the one-particle model to describe our data quite accurately. The interactions seem to screen themselves out just as in a metal. The explanation could be the following. Electrons in a metal near the Fermi level can, in first approximation, be considered as non-interacting particles (with renormalized properities) because the delocalized electron states make that electron-electron interactions are screened out. In our system the electron-electron interactions are not completely screened out at small energy because electrons are localized (thus the finite dielectric constant). At sufficiently high energies, however, electrons can be displaced from one localized state to another. This causes complete screening on a large energy (and spatial) scale. But this effect occurs in finite steps. Thus on a small energy scale screening is incomplete and interactions play an important role, while, averaged over a large energy scale, screening is complete and one-particle descriptions are quite accurate.

Chapter 6

Conclusion

6.1 Overview

In this work Coulomb blockade phenomena in etched silicon nanowires have been studied. These devices are fabricated on a CMOS platform (CEA/Leti) and benefit from the high maturity of silicon technology, in particular the high quality of the Si/SiO_2 interface and the good control of the doping level. At low temperature the devices are very stable and do not show slow relaxations or occasional shifts in their transport properties.

The wires contain no oxide barriers or constrictions. Nevertheless doping and gate voltage define threee distinct parts of the wire which, from the center to the ends, are:

the quantum dot: This is the part of the wire below the gate electrode which is tightly controlled by the gate voltage.

the barriers: These parts of the wire are only weakly controlled by the gate electrode. They provide (or not) the necessary confinement for Coulomb blockade.

the reservoirs: They are the parts of the device where the wire widens or the doping level is very high. Their resistance is negligible.

Depending on the doping level, the barriers and the quantum dot undergo a metal–insulator transition at different gate voltages, i.e. they pass from a mesoscopic Coulomb glass regime where electrons are localized in minima of the disorder potential and behave almost classically to a diffusive regime where electron states spread over the whole dot or barrier region.

When the barrier regions are heavily doped their conductance is high, above $\frac{e^2}{h}$. Then they are not controlled by the gate electrode any more and cannot provide confinement: they merge with the reservoirs. In this case we observe the metal–insulator transition in the quantum dot: when gate voltage is increased, the insulating mesoscopic Coulomb-glass regime, where the fluctuations of the Coulomb blockade peak

spacing are as strong as its average (see chapter 4.2), goes over into the diffusive regime with weak Coulomb blockade or universal conductance fluctuations.

When doping is low in the barrier regions, the Coulomb glass regime in the dot cannot be observed because the barriers become too opaque at low gate voltage. Now it is the barriers that are in the Coulomb glass regime and the quantum dot can be observed in the diffusive regime. In this configuration the fluctuations of the Coulomb blockade peak spacing are weak, of the order of the single-particle level spacing (see chapter 4.1). Our design where the gate electrode covers only a small part of the device allows then for controlled Coulomb blockade without the needs of oxide barriers or constrictions. The size and gate capacitance of the quantum dot is controlled by the surface area of the gate–wire overlap and the conductance by the doping levels (see chapter 3). Ironically it is the disorder in the barriers introduced by doping that allows the devices to be so well controlled. First, the strong disorder potential makes that the transition between strong and weak barriers occurs slowly with the potential offset. Therefore, the barriers are insensitive to potential fluctuations and no further gate electrodes are needed to control them. Second, the presence of resonant states allows the barriers to be thicker than direct tunnel barriers without becoming too opaque. The thicker barriers make the devices more robust. Nevertheless the barriers must not be too long, for otherwise hopping and Coulomb blockade become important in the barriers and they are once more very sensitive to small potential fluctuations with respect to a critical value and to temperature.

While at low barrier doping the Coulomb glass regime of the quantum dot is inaccessible, the Coulomb glass regime of the barriers can be observed in detail because the quantum dot forms a SET which allows not only to measure the transport properties of the barrier regions but also to gather information on the charge and the capacitance of the barrier. In chapter 5 we find a large increase of the barrier capacitance with gate voltage. We interpret it as an increase of the dielectric constant as the metal–insulator transition is approached in the barriers. We compare it with the increase of conductance with gate voltage and find a good agreement with scaling laws for the metal–insulator transition.

With the SET individual charging events of localized states in the barrier —most likely in dopant sites— have been studied in detail: their capacitance matrix, their ground state spin and their tunneling rates have been measured. The dynamics of very slow traps has been accessed by measuring directly the RTS signal they cause in the Coulomb blockade conductance. For most of the traps, however, this is not possible; they are too fast: we found tunneling rates in the high MHz to GHz range by studying the relaxation of the trap charges manipulated by applying a pulse signal to the gate voltage. The charge traps can be as fast as the SET itself.

The periodicity of the SET together with the fact that high gate voltage changes are needed to modify significantly the potential in the barriers, make it possible to control the SET and to drive the metal–insulator transition in the barriers or to manipulate individual charge traps with a single gate electrode: gate voltage

adjustments smaller than the peak spacing are used to control the SET, larger sweeps to control the barrier potential.

6.2 Outlook

We have shown that silicon nanowires can be good single electron transitors and we have focussed on the low doped parts of the wire that form the barriers. But this work is by no means complete. First of all, in order to verify the attribution of disorder and traps to dopants, similar devices without doping should be measured. The trap states in the barriers should then completely disappear. Actually, we have designed such devices where the reservoirs are silicides (metals) and the wire itself is undoped. These devices will soon be available. It will be particularly interesting to study the few-electron regime in such devices in order to see if disorder is still important, i.e. if the dot is still in the Coulomb glass regime. If this is the case, at CEA/Grenoble silicon nanowires of comparable diameters are also grown epitaxially. Comparing transport and Coulomb blockade in epitaxial nanowires with the etched nanowires could bring insight into the disorder created by the surface of the wire.

The spin of electrons in dopant sites in silicon is very interesting as a quantum bit because its relaxation time is expected to be long. We have measured the time-resolved occupation number of dopant states and their Zeeman shift. But so far we could not measure the relaxation of excited spin states because the trap has to be sufficiently slow and to be situated near the center of the barrier. Due to the relatively high doping level, such isolated traps are rare. In order to have slow traps in most samples, the source and drain regions and the part of the wire containing the traps could be separated. This way, while keeping the doping of the barriers sufficiently high, the doping for the trap region could be adjusted to have only one or two dopants on average. Combining our devices with the recently achieved controlled implantation of single dopants could bring ultimate control.

For these measurements of individual dopants we have used the fact that our quantum dot is very stable and that only one gate electrode is needed to control it. Both features make it also very interesting for applications such as the electron pump, a quantum current source.

Appendix A

Experimental setup

A.1 Cryogenics

The physics studied in this work has typical energy scales in the µV–mV range. The measurements have to be performed at temperatures with $k_B T$ below these energies, otherwise the effects would be masked. We use two different dilution refrigerators and a helium 3 cryostat to reach these temperatures.

The first dilution refrigerator (figure A.1(a)) is used for standard low frequency transport measurements. It is an Air Liquide dilution refrigerator without 1 K pot (pumped helium 4). Its base temperature is approximately 60 mK. The cryostat is equipped with a 2 T superconducting magnet. The corresponding Zeeman energy is of the order of the electronic temperature. Therefore, spin effects are difficult to measure on this cryostat.

The second cryostat (figure A.1(b)) has a much stronger superconducting magnet which can provide up to 16 T at 4 K or 18 T at 2 K. However, such high fields imply a small field bore — 42 mm in our case. The matching cryostat has only 37 mm external diameter over all its length. This small diameter implies very small cooling power (below 10 µW at 100 mK, the base temperature being approximately 70 mK). As the mixing chamber is inside the magnetic field it is made of plastic to avoid eddy currents when the field is swept. Thermalization at the mixing chamber temperature is achieved with a thin silver film.

The third cryostat (figure A.1(c)) is a helium 3 cryostat with a base temperature of approximately 300 mK. The larger sample space of this cryostat allows for good shielding and the use of radio-frequency equipment. Recently we have equipped this cryostat with a 8 T superconducting magnet.

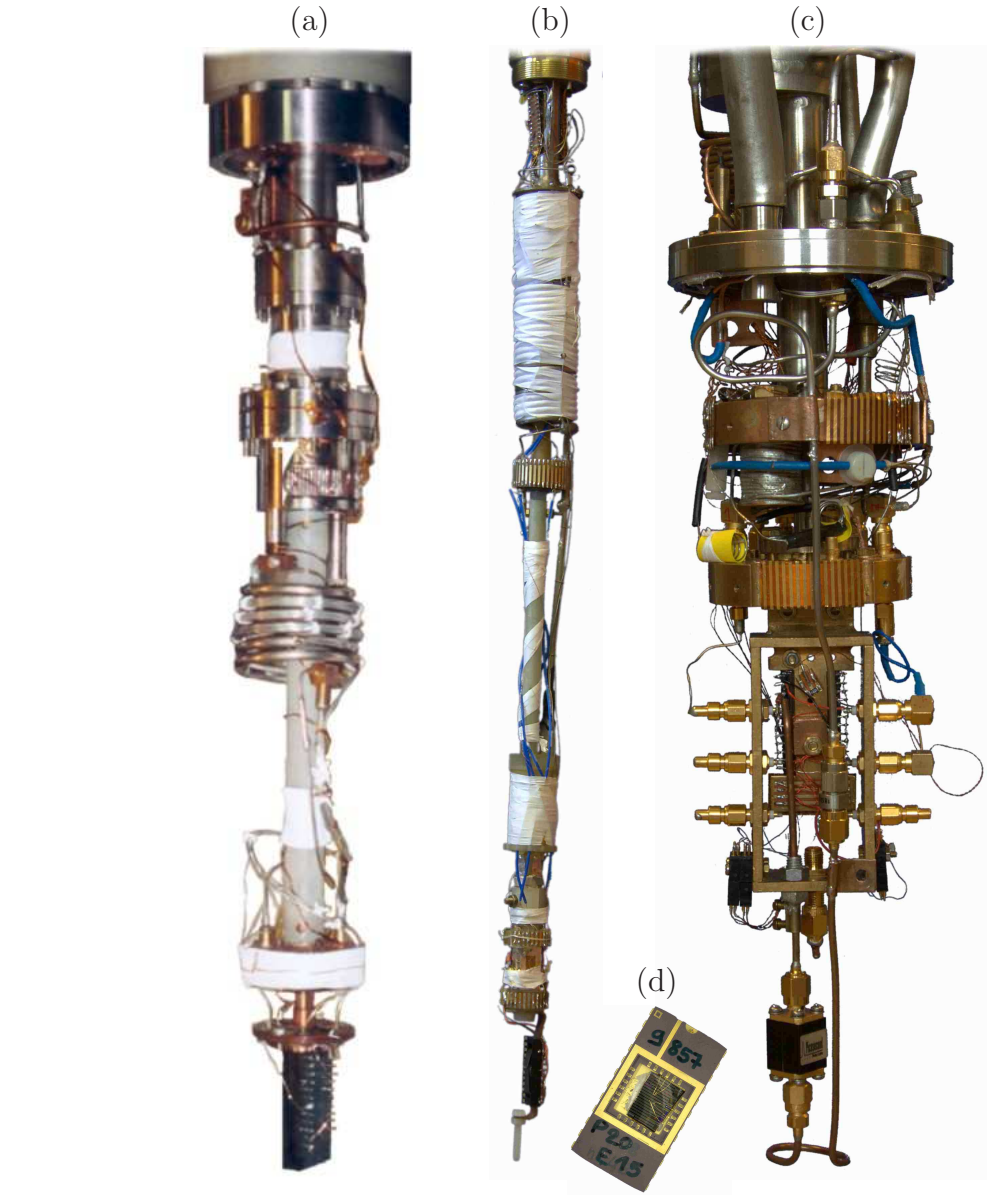

Figure A.1: Cryostats. *The sample shields are removed. (a) Air Liquide dilution refrigerator. (b) Dilution refrigerator for the* 18 T *magnet. (c) Helium 3 refrigerator used for high frequency measurements. (d) Sample carrier.*

A.2 Electrical measurements

A.2.1 Conductance measurement

All measurements but those of the random telegraph noise were performed in lock-in technique. Figure A.2 shows the wiring scheme. We use SR 830 lock-in amplifiers at frequencies between 17 Hz and 180 Hz. The AC signal of the SR 830 is first divided by 10 with a standard 50 Ω attenuator. It is then passed through a differential amplifier

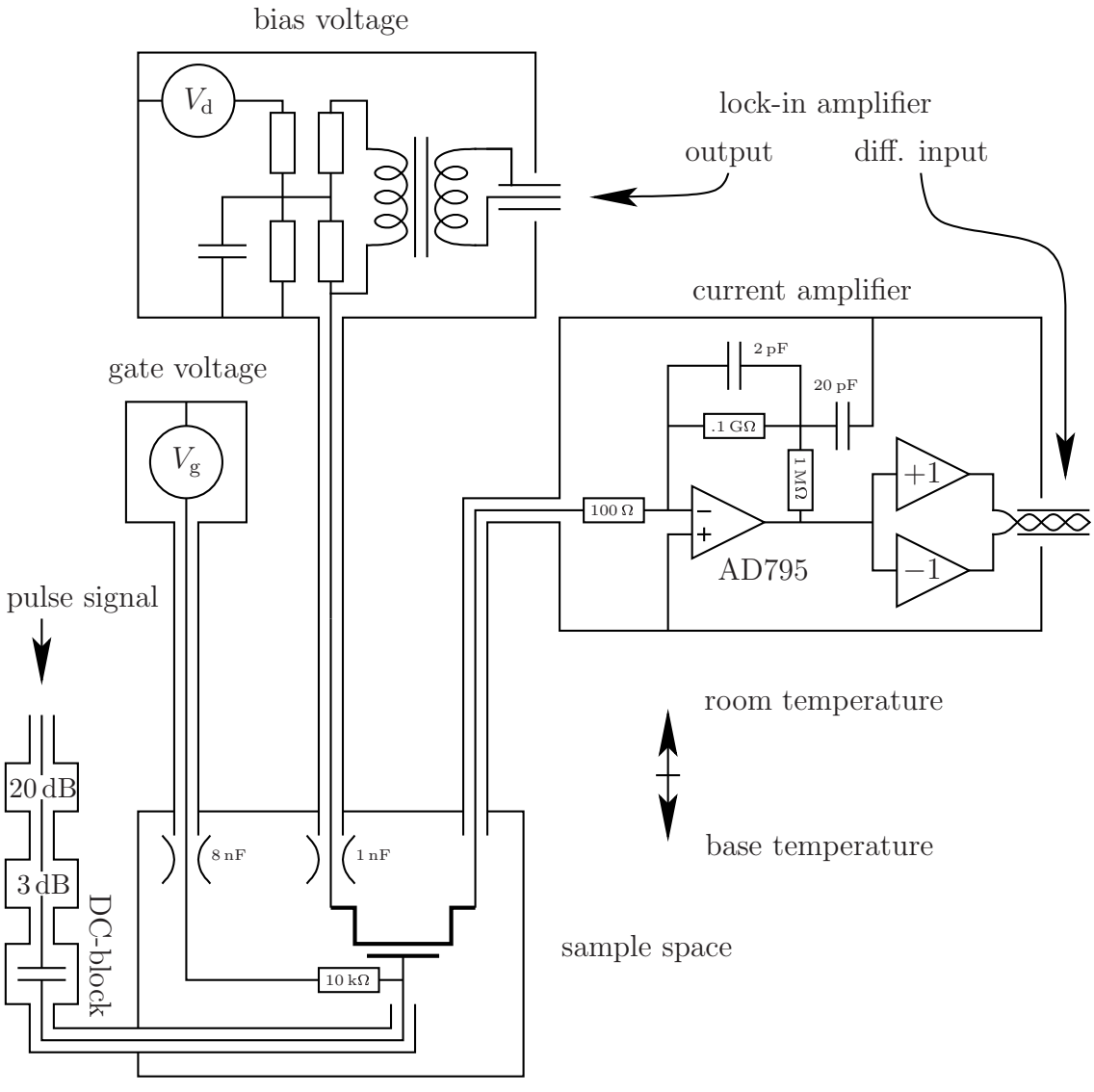

Figure A.2: Wiring scheme for the electrical measurements. *Only the home-made parts are detailed.*

supplied by batteries or an audio frequency transformer (this is the case depicted in A.2) to break ground loops (the source of the lock-in amplifier is not floating ground) and then divided once more by 1000 to obtain the needed excitations of the order of $V_{ac} = 10\,\mu V$ ($eV_{ac} < kT$). This signal is applied to the drain electrode of our samples. The current is measured at the source electrode. We use home made current amplifiers with a gain in the first stage of $50\,M\Omega$ or $100\,M\Omega$ and supplied by batteries. The $100\,M\Omega$ amplifier has an AD795 as first stage and has differential output (this amplifier is shown in figure A.2). This way the ground loop between

the lock-in amplifier and cryostat can be completely[1] broken. The bandwidth of the measurement is limited by two parameters:

- The capacitance of the measurement line, typically several 100 pF, has to be charged through the sample resistance R_{sample}. Thus the measurement line limits the bandwidth for high sample resistances. This capacitance also tends to make the amplifier unstable by increasing the gain at high frequency. This effect has to be compensated by a small capacitor in parallel with the feedback resistance. This capacitor causes a further bandwidth limitation. Therefore it has to be chosen in function of the capacitance of the measurement line. We use an additional RC low pass filter in the feedback loop (see figure A.2). Unlike the bandwidth limitations described above, this low pass causes a fixed cut-off frequency which makes the amplifier less sensitive to the capacitance of the measurement line and allows to use a smaller capacitor in parallel with the feedback resistance without making the amplifier unstable.

- The gain of the first-stage operational amplifier. The sample and the current amplifier together form a voltage amplifier with gain $\frac{R_{\text{feedback}}}{R_{\text{sample}}}$. The operational amplifier must have a higher gain than the gain defined by the resistances. Thus, the bandwidth is limited by the gain-bandwidth product of the amplifier. (approximately 2 MHz for the AD795). This limitation is important for low sample resistances.

These bandwidths can be as low as 100 Hz. Measurements have thus to be done at low frequency.

For the measurement of random telegraph signals we applied a DC voltage to the sample and connected the current amplifier with 50 MΩ gain to a SR 785 spectrum analyzer that we used in this case only as 16 bit analog–digital converter.

A.2.2 Voltage sources

As voltage sources we use either an Adret 103A, a high precision, low noise GPIB programmable DC voltage reference source or home made voltage sources. The latter are supplied by batteries and can be mounted directly on top of the cryostat to avoid noise created or captured by flexible cables. They are programmed by computer via optical fibers to avoid ground loops. To avoid high frequency noise created by the sources themselves, the clock of the logic circuit is only activated when commands are received. 16 bit DACs with a voltage step of 100 µV or smaller are used. The nonlinearities of the DACs are smaller than that.

[1]The input of the SR830 lock-in amplifier can be configured as "float" but even then the shield is grounded by a 10 kΩ resistor. The resulting ground loop currents can cause significant perturbations in the measurement.

A.2.3 Signal generators

For some of the measurements we applied time-dependent signals to the gate electrode. We use an HP 33120A function generator for signals in the kHz and low MHz range. Above we use an Anritsu MP1763C two level pulse pattern generator. The length of the low and high levels can be chosen as a multiple of the clock period which can go down to 80 ps. For sine wave signals we use a Hewlett Packard 8350B sweep oscillator with a 83592B RF plug-in ranging from 10 MHz to 20 GHz.

A.2.4 Wiring

Low frequency

At low temperatures electrons become more and more decoupled from the phonon bath. This is particularly true for nanostructures where the electrons of interest are confined into a very small volume and the single-particle level spacing is high. Thus the temperature of the electron distribution in the sample is almost entirely controlled by electrons and photons arriving from the leads. At room temperature the maximum of thermal radiation is at approximately 100 THz (Planck's law). Very few photons of such high energies are sufficient to perturb the electron distribution in the device dramatically [32]. Unfortunately such photons in the high GHz and THz range cannot by blocked by usual electronic filters (networks of capacitors, resistors, inductors) because their size exceeds the wave length. Therefore, special precautions have to be taken to measure mesoscopic devices at low temperature. The high THz and infrared radiation can be stopped by putting the sample inside a metal shield. More difficult to shield is the GHz and low THz radiation which can be transmitted by the measurement lines.

First, the lines have to be thermally anchored to the base temperature (and before that to the still and to helium temperature). The high GHz and low THz photons have then to be filtered. Typically copper powder filters [78], lithographed micrometer-sized meanders [123] or lossy coaxial lines [133] are used ([12] gives an overview). We use the latter, either Microcoax 0.5 mm coaxial cable with PTFE dielectric and inner and outer conductors of stainless steel or Philips Thermocoax 1 Ac Nc 05, a 0.5 mm thick stainless steel coaxial cable. Its inner conductor is CrNi and the dielectric is MgO powder. Its resistance is 50 Ω/m. This cable is designed as heating element. Yet, as shown in [133], it has excellent properties for our purposes. It attenuates efficiently above 100 MHz and its resistance varies only by a few percent between room temperature and 4 K. We use lines of approximately 2.5 m thus having a total resistance of approximately 125 Ω. With respect to Microcoax, Thermocoax is much cheaper and the dielectric has a better thermal conductivity, allowing to thermalize the inner conductor more efficiently. The Thermocoax can be soldered or even brazed without risking damages to the dielectric. However, this cable is very difficult to connect properly. First it is difficult to remove the outer conductor without breaking the central conductor because the dielectric is strongly compressed. Then, the center conductor seems to be passivated and even with strong acids it

is difficult to solder it reliably. In addition, the MgO dielectric is hygroscopic and once it has absorbed water, resistances down to 1 MΩ can shunt inner and outer conductor, making the cable useless as measurement line. Thus, the ends have to be sealed rapidly after cutting; we use Stycast 2850 FT epoxy.

At the low temperature end of the lines we put additional filter capacitors in the nF range to filter the low MHz range which is not well filtered by the coaxial lines. In the current measurement line, however, the capacitor would cause severe band-width limitations or instabilities of the current amplifier (see section A.2.1) and has to be omitted.

To absorb remaining GHz and MHz photons we cover the inside of the sample shield with copper-powder charged Stycast 2850 FT.

High frequency

We have equipped the helium 3 cryostat with a high frequency line. In order to apply high frequency signals the attenuation has to be almost constant over the whole frequency range of the transmitted signals up to the 10 GHz range. Therefore, the filtering scheme described above with its strongly frequency dependent attenuation cannot be used and strong frequency independent attenuation has to be used instead.

We use 50 Ω, 2.2 mm thick stainless steel semi-rigid coaxial lines to transmit the high frequency signals. The inner conductor is silver covered to decrease loss. The signal is attenuated by a 20 dB resistive attenuator in liquid helium and enters the calorimeter through a hermetic SMA connector. At base temperature the signal is once more attenuated by a 3 dB resistive attenuator. Obviously, the DC gate voltage of several volts cannot be applied through the attenuators. Therefore, before reaching the sample, the signal is passed through a DC-block. We use a Picosecond 5508-110 with a $C = 2.2\,\mu F$ capacitor (-3 dB frequency range into 50 Ω from 700 Hz to > 26 GHz). Both the high frequency line through the DC-block and the DC line through a $R = 10\,k\Omega$ resistor are connected to the gate electrode (see figures A.1(c) and A.2). The crossover frequency should be $f = (2\pi RC)^{-1} = 7$ Hz but the capacitance of the DC-block decreases to 19.4 nF at 4.2 K so that the crossover frequency is approximately 300 Hz. This is lower than the signal frequency but still high compared to our sampling frequency < 10 samples/s. The heat carried by the coaxial line increases the base temperature of the cryostat by approximately 60 mK.

A.2.5 Effective electronic temperature

The effective electronic temperature is measured by first determining the lever-arm factor individually for each peak (see appendix B.3) and then fitting the Coulomb blockade resonances with a thermally broadened Coulomb blockade resonance (equation (2.35)).

In the helium 3 cryostat the effective electronic temperature is approximately 400 mK for our nanowire samples. This is very close to the base temperature. However, the electronic temperature in the dilution refrigerators is always above 250 mK.

In the cryostat for the 18 T magnet where no decent shielding is possible it can even be as high as 700 mK.

In other silicon SETs however, we observe a much lower electronic temperature of only 100 mK [46]. Therefore, the high electronic temperature seems to be specific to the nanowire samples. We see two possible explanations.

- Cooling the electron distribution in the dot becomes difficult when $k_B T < \Delta_1$, the single-particle level spacing. I.e. an effective electronic temperature in the reservoirs much smaller than Δ_1 is needed in order to freeze out all electron excitations in the dot. Indeed, we know of no experiment where the effective electron temperature is significantly below the single-particle level spacing at high electron numbers.

- Transport through the barriers occurs by resonant tunneling through a percolation path of dopant states (see chapter 5). The coupling between the dot and the nearest dopant states as well as the coupling between neighboring dopant states can be high while the total conductance is low. Therefore, on a short time scale the dot is in a low impedance environment despite a low peak conductance so that lifetime broadening could increase the peak width. Using equation 2.35 to describe the peak width will then overestimate the electron temperature.

Appendix B

Data analysis

The conductance through mesoscopic systems, in particular quantum dots shows a huge amount of different features. Most of them cannot be modeled precisely. So the data often cannot be fitted directly to theoretical predictions but the features of interest have to be extracted first. In this appendix we present the most important algorithms we use.

All data analysis and simulations in this work have been coded in Yorick [136], an interpreted programming language with a syntax similar to C. An important difference is that, as in IDL or Matlab, mathematical operators and functions act directly on multidimensional arrays without the need of loops. This speeds up not only coding but also execution because the loops on large arrays are done in compiled code and not in interpreted code.

B.1 Color plots

Color plots, especially of the differential conductance, usually cannot be made by simply mapping the conductance linearly onto a color palette, because often there are features in the conductance at $10^{-4}\,\frac{e^2}{h}$ as well as at $\frac{e^2}{h}$. Such a dynamic range cannot be covered on a computer monitor or in print. A logarithmic scale is not a good choice either because the differential conductance can be negative. Another disadvantage of a logarithmic scale is that features emerging only very little from the noise floor cannot be visualized because noise has to be cut off. Therefore we often use a mapping that is linear for small conductances near the noise floor and becomes logarithmic for higher absolute conductances:

$$G \mapsto \operatorname{sgn} G \cdot \log\left(\frac{|G|}{G_0} + 1\right) \tag{B.1}$$

Typically we choose G_0 in the order of $10^{-3}\,\frac{e^2}{h}$. The result is then mapped linearly to the color palette.

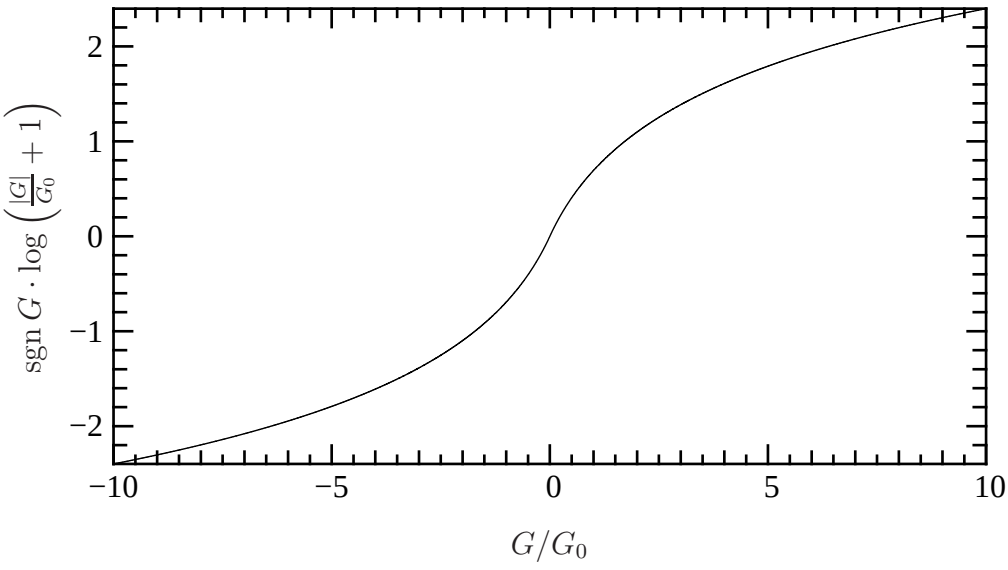

Figure B.1: Mapping used for color plots of conductance.

B.2 Peak detection

For the peak-spacing statistics (see chapter 4) the maxima in the $G(V_\mathrm{g})$ plot have to be detected precisely. The following algorithm has been used:

The conductance is at first sampled with a fixed gate-voltage step which is typically a few 10 µV at base temperature. A sample i is considered a peak if it fulfills the following criteria:

- $G_i > \max\{G_{i-l} \ldots G_{i-1}\}$
- $G_i \geq \max\{G_{i+1} \ldots G_{i+l}\}$
- $\min\{G_{i-w} \ldots G_{i+w}\} - \mathrm{avg}\{G_{i-l} \ldots G_{i-w-1}, G_{i+w+1} \ldots G_{i+l}\} \geq G_0$

We typically choose $l \approx 20$, $w \approx 3$, $G_0 \approx 10^{-5}\,\frac{e^2}{h}$. l times the gate voltage step is kept at least two times smaller than the average Coulomb blockade peak spacing, otherwise the peak-spacing distribution could be cut off at small peak spacings. These conditions allow to avoid spurious peaks due to noise. The last condition rejects spurious peaks consisting of only one data point (they can be due to vibrations when the setup is touched). If such peaks are not present we replace "min" by "avg" so that peaks much closer to the noise floor can be detected.

The precise peak position is still subject to noise and the finite gate voltage step. Therefore we fit the peak between the two half-height points with a thermally broadened resonance (see eq. (2.35)). If the conductance between peaks is much smaller than the peak height, the width of the fitted resonance allows to access the parameter $\frac{e\alpha}{kT}$ where α is the gate voltage lever arm and T the effective electron temperature. As we try to have a gate voltage step several times smaller than the

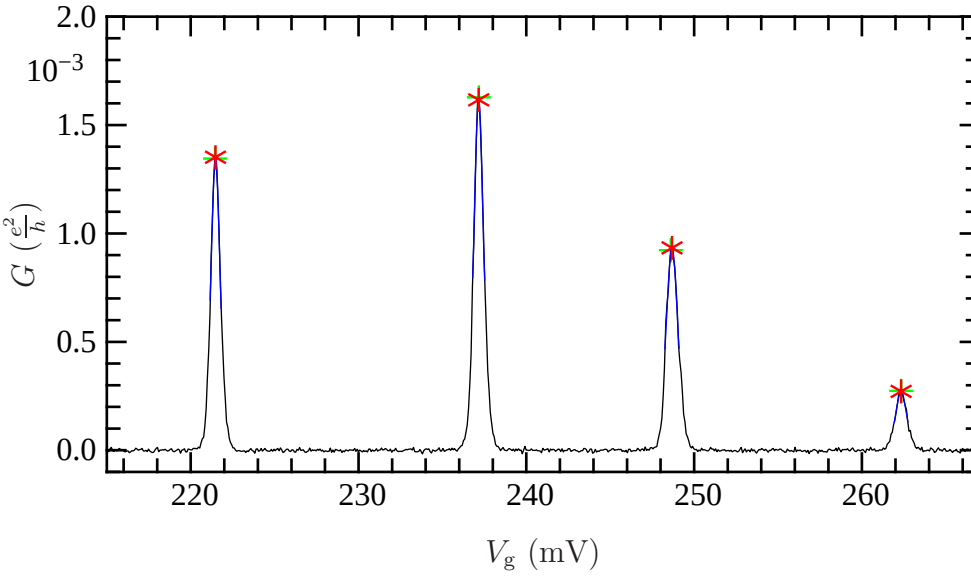

Figure B.2: Peak detection. *—: data, +: detected maximum, —: fit, ∗: maximum from fit. Sample 3.*

peak width, this fit reduces noise in the peak position and height. Figure B.2 shows an example.

Near trap signatures (see 5.3) the line shape tends to be asymmetric. In such cases the fit does not work well and fitting parameters give biased results. Yet, these errors are usually much smaller than the direct effect of the trap on peak height and position. Such peaks give tails in the distributions of peak spacing and peak height which we suppress in our analysis (see chapter 4).

B.3 Slopes of the Coulomb blockade diamonds

In section 5.4 we perform a statistical analysis of the source and drain capacitances. These capacitances are related to the slopes of the Coulomb blockade diamonds through equations (2.33) and (2.34). To have statistically meaningful results several hundred slopes on several samples have to be extracted. We tried several methods (by hand, detecting the Coulomb gap as a function of gate voltage, gradient direction) which turned out to be slow, unreproducible, biased, unreliable, or very sensitive to the chosen parameters. The following algorithm produced much better results and has been retained for the statistical analysis.

B.3.1 Algorithm

First characteristic features in the data are detected. For this purpose the following filter is used

$$F[G] = \frac{G - U * G}{\sqrt{(U * G)^2 + G_0^2}} \tag{B.2}$$

$*$ is the convolution operator, G is a 2-dimensional function (in our case $G(V_\mathrm{g}, V_\mathrm{d})$). G_0 is a scalar and U is a 2-dimensional unsharp filter. We use a Gaussian:

$$U(x, y) = \frac{1}{2\pi\sigma_\mathrm{x}\sigma_\mathrm{y}} \exp\left(-\frac{x^2}{2\sigma_\mathrm{x}^2} - \frac{y^2}{2\sigma_\mathrm{y}^2}\right) \tag{B.3}$$

The filter $G - U * G$ alone is known as unsharp masking in image editing[1]. It subtracts the unsharpened data from the original data. Thus all slowly evolving background is suppressed. This filter is a high pass filter and its effect is similar to the Laplace operator, i.e. a 2-dimensional second derivative.

The denominator in (B.2) normalizes the result with the unsharpened data. This is necessary because the conductance we measure has a very high dynamic range. Like in expression (B.1) the parameter G_0 serves to cut off the noise floor.

This filter has been used to transform the upper panel of figure B.3 into the middle panel. The filter detects accurately the borders of the Coulomb blockade diamonds.

As a second step a Hough transform [22] is applied to the filtered data. Hough transforms are used to detect features described by a set of parameters, usually in binary data. The general idea is that each non-zero point in the data votes for all sets of parameters that make the feature contain this point. Features well represented in the data will receive a higher vote than those that only cross some non-zero points by chance.

In our case the features to be detected are the edges of the different Coulomb blockade diamonds. We parameterize them by

$$v_\mathrm{g} = V_\mathrm{g} + s v_\mathrm{d} \tag{B.4}$$

Each data point $G(v_\mathrm{g}, v_\mathrm{d})$ votes with weight G for all offsets V_g and slopes s that verify (B.4). These are the points (V_g, s) on the straight line with slope $\frac{\mathrm{d}V_\mathrm{g}}{\mathrm{d}s} = -y$ and intersect the V_g axis at v_g.

In this simple case the Hough transform resumes to the integral

$$\tilde{G}(V_\mathrm{g}, s) = \int \mathrm{d}v_\mathrm{d} G(V_\mathrm{g} + s v_\mathrm{d}) \tag{B.5}$$

where $\tilde{G}$ is the Hough transform of G.

[1] Usually the output of this filter is added to the original image to form a sharpening filter

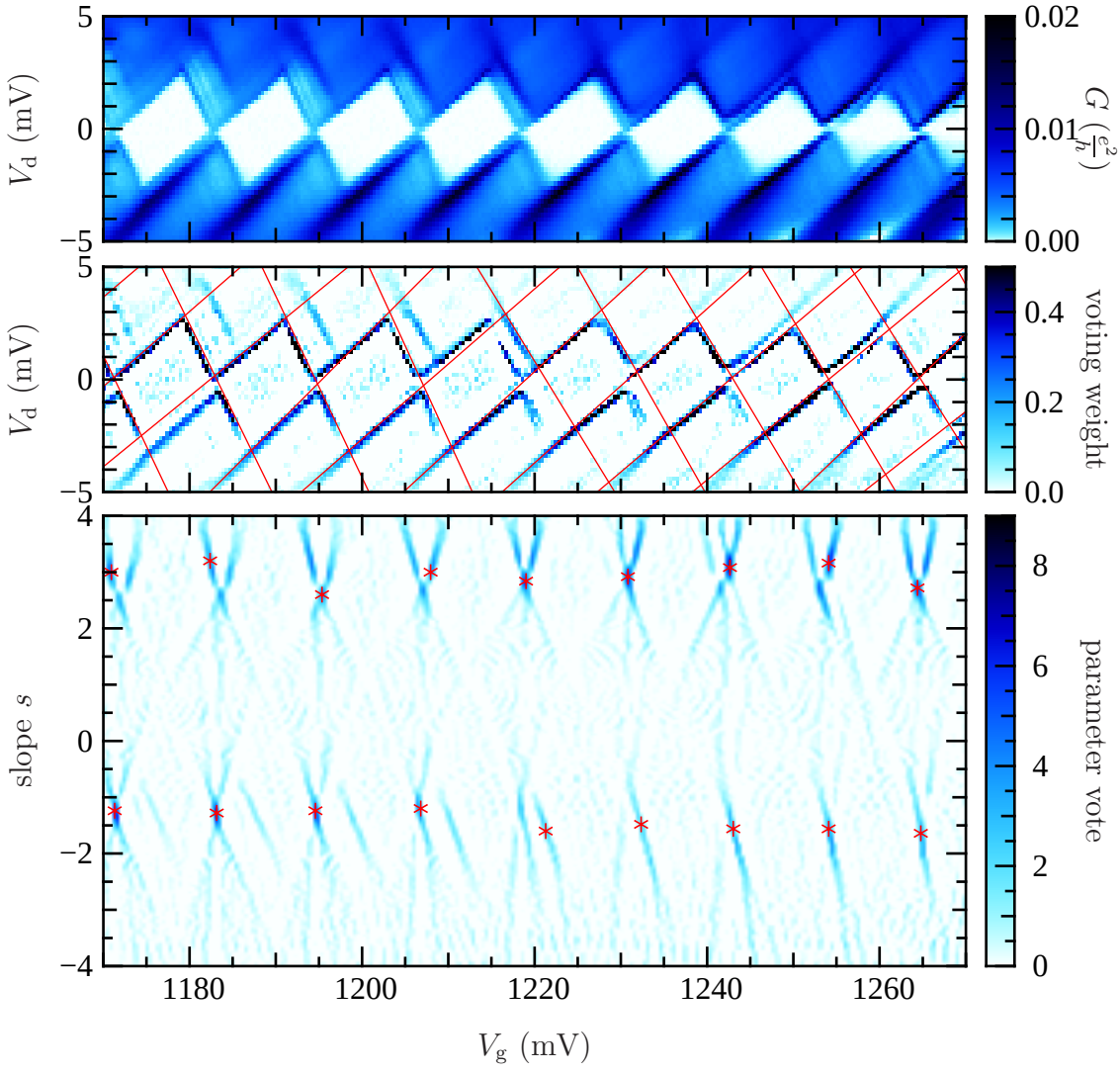

Figure B.3: Detection of diamond slopes via Hough transform. *Top panel: raw data. Middle panel: after filter* (B.2)*. Bottom panel: Hough transform of the middle panel. The red markers show the detected diamond slopes. They correspond to the red lines in the middle panel. Sample 7.*

Finally we apply the convolution kernels $(-\frac{1}{4}, -\frac{1}{4}, \frac{1}{2}, \frac{1}{2}, -\frac{1}{4}, -\frac{1}{4})$ in V_g direction and $(\frac{1}{2}, \frac{1}{2})$ in s direction to remove the continuous background and to smooth out oscillations due to binning. The lower panel of figure B.3 shows the result of such a Hough transform.

The last step is to detect the maxima in the Hough transform. We first search the maximum at positive and negative slope for each V_g and then use the peak detection algorithm described in B.2 to detect the peaks in V_g direction within these maxima. The detected maxima are marked in red in figure B.3.

The correct detection of the slopes is testified by the middle panel of B.3.

B.3.2 Error estimation

We will now try to estimate the errors of the detected slopes. As an example, the positive slope near 1195 mV in the middle panel of figure B.3 has an offset for $V_\mathrm{d} > 0$ with respect to $V_\mathrm{d} < 0$. In Hough space (lower panel) this is represented by a $\times$-shaped maximum near $V_\mathrm{g} = 1195\,\mathrm{mV}$, $s = 3$. The $\diagdown$-part of the maximum corresponds to positive drain voltage. It occurs at somewhat lower gate voltage than the $\diagup$-part corresponding to positive drain voltage. The individual maxima of both parts occur at the same slope $s = 3$. Yet the algorithm detects a smaller slope at the intersection of the two lines where the total vote is highest. The detected slope is therefore smaller (steeper in the middle panel) because it tries to pass through both lines, at positive and at negative drain voltage. However, such misinterpretations must fall within the uncertainty of the individual slopes. This uncertainty is given by the ratio of the width in V_g-direction δV_g to the length of the line in V_d-direction ΔV_d.

$$\delta s \approx \frac{\delta V_\mathrm{g}}{\Delta V_\mathrm{d}} \tag{B.6}$$

If we suppose the positive and negative slopes to have the same magnitude, the length of the lines in direction of V_d is $\Delta V_\mathrm{d} = \frac{V_+}{2s}$. We get a constant relative error of the slopes of

$$\frac{\delta s}{s} \approx 2\frac{\delta V_\mathrm{g}}{V_+} \tag{B.7}$$

The Coulomb blockade peak spacing V_+ is typically 10 mV in our samples. The width of the lines in the $G(V_\mathrm{g}, V_\mathrm{d})$ plot after filtering depends very much on the sample and the sampling rate in V_g direction. We estimate the relative error to be between 10 % and 50 % depending on the sample. In addition there are sporadic errors giving arbitrary slopes. They occur when background noise is detected as peak. In order to suppress such errors we eliminate slopes more than 50 % off the previous and the next ones.

B.4 Random telegraph signal

In 5.3.4 we study the random telegraph signal produced by a charge trap. Here we describe how the signal is fitted.

As a first step we correct for the response function of the current amplifier. Usually it can be described accurately by the dominating time constant τ_amp (which depends on the sample resistance, see appendix A). For a continuous signal its effect would be corrected by $I' = I + \tau_\mathrm{amp}\frac{\mathrm{d}I}{\mathrm{d}t}$. For discrete sampling at intervals Δt we use

$$I'_n = \frac{1}{2}(I_{n+1} + I_n) + \frac{\tau_\mathrm{amp}}{\Delta t}(I_{n+1} - I_n) \tag{B.8}$$

After correction the rise time is of the order of Δt (see figure B.4, 30 µs in this example). Inevitably, this correction increases noise. In the present example the

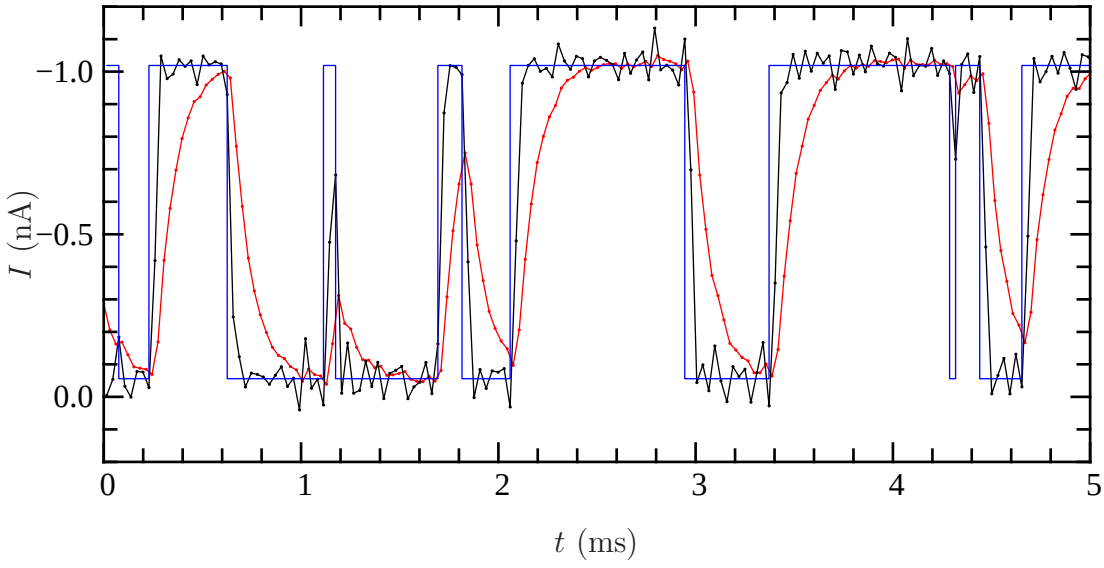

Figure B.4: Fitting an RTS signal. *Red: raw data, black: data after correction of the current amplifier's time constant, blue: fitted signal. This is a zoom of figure 5.8(a). Sample 6.*

signal to noise ratio is very good. So it is favorable loosing precision for bandwidth. With smaller signals it is more favorable not to apply the correction. We will see below that the signal to noise ratio should be at least 6 for a good detection of the RTS signal.

B.4.1 Histogram analysis

The next step is to detect the two current levels I_0 and I_1. For this purpose we make a histogram of the data points with a bin size $\delta I/3$. δI is the current noise (not including the random telegraph signal) which we determine beforehand. In the example presented in this work $\delta I = 34\,\mathrm{pA}$.

In the histogram relative maxima with more than 3 counts are detected. If there are more than 2 maxima, the bin width is multiplied by a factor $\gtrsim 1$ and a new histogram is made. This procedure is repeated until 2 or less maxima are found. If there are 2 maxima we take them as the current levels I_0 and ${I_1}^2$. If there are less than two maxima, no RTS signal is detected.

In this histogram we do not weight the data points equally but use a weight

$$w_n = \exp\left(-\frac{(\Delta I_{n-1})^2 + (\Delta I_n)^2 + (\Delta I_{n+1})^2}{6\,\delta I^2}\right) \tag{B.9}$$

[2]Actually, as the bins can be quite large at the end of the loop, we take the maximum of a parabola through the maximum and the 2 neighboring bin heights in order to increase the precision of the current levels

with $\Delta I_m = I_m - \frac{1}{3}(I_{n-1} + I_n + I_{n+1})$. This weighting suppresses data points on the flanks of the RTS signal and lets the stationary levels I_0 and I_1 emerge more clearly.

Having determined these two current levels we can estimate the duty cycle of the RTS. This is trivial as long as the difference of the levels $I_1 - I_0$ is much higher than the noise δI: we simply have to count the points near the levels I_0 and I_1. If, however, there is no clear separation between $I_1 - I_0$ and δI, then there will be measured currents I which can be explained by both, state 0 and state 1. Formally this is expressed by the probabilities $\mathrm{d}P(I|0)$ and $\mathrm{d}P(I|1)$ to measure the current I when the trap is respectively in state 0 and 1. They are both non-zero in this case. If we suppose Gaussian current noise (not including the RTS), the probability $\mathrm{d}P(I|0)$ is given by

$$\mathrm{d}P(I|0) = \frac{1}{\sqrt{2\pi}\delta I} \exp\left(-\frac{(I - I_0)^2}{2(\delta I)^2}\right) \mathrm{d}I \tag{B.10}$$

$P(I|1)$ is defined in the same way. But these are not the probabilities we are interested in. What we need to know is what state the measured point favors, i.e. the probabilities $P(0|I)$ and $P(1|I)$. According to Bayes' theorem

$$P(0|I) = \frac{\mathrm{d}P(I|0)P(0)}{\mathrm{d}P(I)} = \frac{\mathrm{d}P(I|0)P(0)}{\mathrm{d}P(I|0)P(0) + \mathrm{d}P(I|1)P(1)} \tag{B.11}$$

We first calculate the probabilities $P(I_i|0)$ and $P(I_i|1)$ for each data point $I_1, \ldots, I_N$ and then solve the system iteratively with

$$P_{n+1}(0) = \frac{1}{N}\sum_{i=1}^{N} P_{n+1}(0|I_i) = \sum_{i=1}^{N} \frac{\mathrm{d}P(I_i|0)P_n(0)}{\mathrm{d}P(I_i|0)P_n(0) + \mathrm{d}P(I_i|1)P_n(1)} \tag{B.12}$$

and the starting point $P_0(0) = P_0(1) = \frac{1}{2}$. The loop is terminated when the probabilities change by less then 10^{-5}.

If the two current levels are clearly separated ($I_1 - I_0 \gg \delta I$) and the time constant of the random telegraph signal is comparable to the rise time, we observe a finite and almost constant probability density between the two levels. This is due to points on the flanks of the RTS signal. To account for them we allow for a third state —let us call it T— if $I_1 - I_0 > 5\delta I$. We use the probability density

$$\mathrm{d}P(I|\mathrm{T}) = \frac{\mathrm{erf}\left(\frac{I-I_0}{\sqrt{2}\delta I}\right) - \mathrm{erf}\left(\frac{I-I_1}{\sqrt{2}\delta I}\right)}{2(I_1 - I_0)} \mathrm{d}I \tag{B.13}$$

to describe it. In this case we initiate the loop with $P_0(0) = P_0(1) = P_0(\mathrm{T}) = \frac{1}{3}$. Knowing that there are as much transitions $0 \to 1$ as there are transitions $1 \to 0$ we add 50 % of the final probability P(T) to $P(0)$ and 50 % to $P(1)$.

The accuracy of this algorithm is testified by figure B.5.

We have also tried to fit the laws described above to the histogram by using the standard Levenberg-Marquardt method. But it turned out to be less stable than the algorithm above.

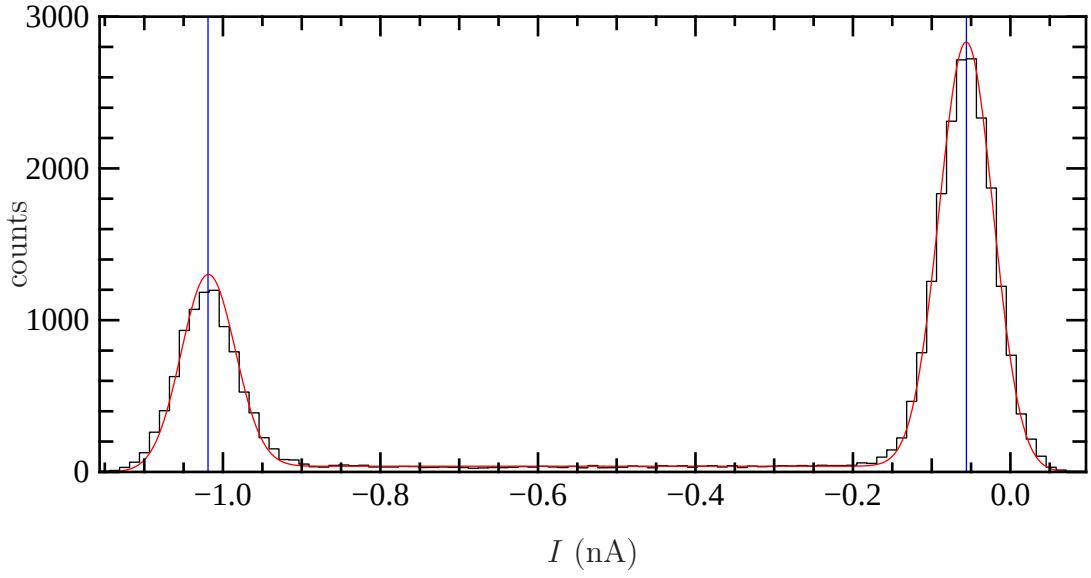

Figure B.5: Current histogram. *This plot serves as a control for the histogram analysis. Black: current histogram without weighting. The bin number is higher than for detecting the current levels. Blue: the two detected levels. Red:* $N\Delta I \mathrm{d}P(I)/\mathrm{d}I$ *with* $\mathrm{d}P(I) = \mathrm{d}P(I|0)P(0)+\mathrm{d}P(I|1)P(1)+\mathrm{d}P(I|\mathrm{T})P(\mathrm{T})$ *and where* ΔI *is the bin width,* $N = 30719$ *the total number of points, and* $P(0) = 0.28$, $P(1) = 0.62$, $P(\mathrm{T}) = 0.09$ *the results of the iterative process. The current noise* $\delta I = 34\,\mathrm{pA}$ *is determined beforehand. Sample 6.*

B.4.2 Transition detection

For the full counting statistics it is necessary to detect the transitions in the random telegraph signal. As a first step points above I_1 are set to I_1 and points below I_0 to I_0. Then our criterion for a transition $0 \to 1$ is

$$\text{trap in state 0 at point } n \text{ and } I_{n+1} - I_n > a\frac{I_1 - I_n}{I_1 - I_0}\delta I \tag{B.14}$$

and symmetrically for the transition $1 \to 0$

$$\text{trap in state 1 at point } n \text{ and } I_{n+1} - I_n < a\frac{I_0 - I_n}{I_1 - I_0}\delta I \tag{B.15}$$

The value of the threshold parameter depends on the rise time of the signal τ. In order to minimize spurious detections a has to be > 1. We got best results with $a = 4$. This choice is justified below. The blue line in figure B.4 has been obtained with the criteria. To test the accuracy we compare the result of this fit with the analysis of the histogram. We find good agreement, via

$$\frac{N_0}{N_0 + N_1} \approx P(0) \quad \text{and} \quad \frac{N_1}{N_0 + N_1} \approx P(1), \tag{B.16}$$

as long as $I_1 - I_0 > 5\,\delta I$. N_0 and N_1 are the number of data points for which the transition detection finds state 0 or 1 respectively; $P(0)$ and $P(1)$ are the results of the histogram analysis.

B.4.3 Accuracy

The conditions (B.14) and (B.15) involve only the currents I_n and I_{n+1}. Thus, the minimum detection time is the sampling interval Δt. What is now the probability to detect a transition within Δt? We consider the case of a transition $0 \to 1$; the probability of the transition $1 \to 0$ is calculated in the same way. If the rise time is τ then $I_{n+1} - I_n \approx \Delta t \frac{\mathrm{d}I}{\mathrm{d}t} = \frac{\Delta t}{\tau}(I_1 - I_n)$. If we suppose independent Gaussian errors with standard deviation δI [3] then the noise on $I_{n+1} - I_n$ has a standard deviation $\sqrt{2}\delta I$ and the transition is detected with a probability of

$$P_{n,n+1} = \frac{1}{2}\,\mathrm{erfc}\left(\frac{a}{2}\frac{I_1 - I_n}{I_1 - I_0} - \frac{\Delta t}{\tau}\frac{I_1 - I_0}{2\delta I}\right) \tag{B.17}$$

This probability is $> 92\,\%$ if $a < \frac{\Delta t}{\tau}\frac{I_1 - I_0}{\delta I} - 2$. For this estimation we used the worst case $I_n = I_0$ which occurs at the onset of the transition. But even if $P_{n,n+1} < 1$, the probability to detect a transition increases rapidly with the number of samples: $1 - P_{n,n+m} = (1 - P_{n,n+1}) \cdot \ldots \cdot (1 - P_{n+m-1,n+m}) > (1 - P_{n,n+1})^m$. The estimation "$>$" is due to the fact that the transition is more easily detected as $I_n \to I_1$ at the end of the transition. In our case the rise time τ and the sample interval Δt are almost equal (after correction for the current amplifier) so that with our choice $a = 4$ we expect a $> 92\,\%$ detection rate for dwell times $\geq \Delta t$ if $I_1 - I_0 > 6\,\delta I$.

The probability of spurious transitions is

$$P_{\text{trigger-happy}} = \frac{1}{2}\,\mathrm{erfc}\left(\frac{a}{2}\right) \tag{B.18}$$

It is smaller than $0.25\,\%$ if $a \geq 4$. Here we used the fact that the standard deviation of $I_{n+1} - I_n$ is $\sqrt{2}\delta I$. This is an overestimation because we clipped the fluctuations beyond $[I_0, I_1]$ (see above). So, if the local average current is I_0 or I_1, half of the fluctuations are suppressed. The standard deviation of $I_{n+1} - I_n$ is then approximately $0.83\,\delta I$ so that with $a = 4$ we get a spurious error rate smaller than 10^{-6}. The clipping is active if the system stays sufficiently long in states 0 and 1 to reach the levels I_0 and I_1. Therefore long dwell times are unlikely to be interrupted by spurious transitions.

To summarize, the spurious-transition error rate is negligible as long as the dwell times of the RTS exceed the rise time. But even if this condition is not fulfilled, spurious errors should not exceed $0.25\,\%$. With our choice of parameters, actual transitions should be detected with $> 92\,\%$ probability within a sampling interval of $30\,\mu\mathrm{s}$ if $I_1 - I_0 > 200\,\mathrm{pA}$.

[3] This is an overestimation when the current is close to I_0 or I_1 because we clipped the fluctuations beyond $[I_0, I_1]$. See below for a more detailed discussion.

Appendix C

Properties of silicon

The following table resumes the properties of undoped silicon. The main sources are [113] and [5].

Stable isotopes	${}^{28}_{14}\mathrm{Si}$ ($I = 0$): 92.2 % ${}^{29}_{14}\mathrm{Si}$ ($I = \frac{1}{2}$): 4.7 % ${}^{30}_{14}\mathrm{Si}$ ($I = 0$): 3.1 %
Crystal structure	diamond (lattice: fcc, base: atoms at $(0,0,0)$ and $\frac{a}{4}(1,1,1)$)
lattice constant (300 K)	$a = 0.543\,\mathrm{nm}$
Conduction band bottom	
minima	g_v = 6-fold degenerate at approximately $\frac{3\pi}{2a}\langle 0,0,1\rangle$
effective mass	$m_\mathrm{l} = 0.916 m_\mathrm{e}$ in the direction of the minimum and
	$m_\mathrm{t} = 0.190 m_\mathrm{e}$ perpendicular
effective DoS mass	$m^* = \sqrt[3]{g_v^2 m_\mathrm{l} m_\mathrm{t}^2}$
including degeneracy	$= 1.06 m_\mathrm{e}$
Bohr radius[1]	$a_\mathrm{B} = 4\pi\epsilon_0\epsilon_\mathrm{Si}\frac{\hbar^2}{m_\mathrm{t} e^2}$
	$= 3.3\,\mathrm{nm}$
Band gap	$E_\mathrm{g} = 1.17\,e\mathrm{V} - 4.73\cdot 10^{-4}\,\frac{e\mathrm{V}}{\mathrm{K}}\frac{T^2}{T+636\,\mathrm{K}}$
	$= 1.12\,e\mathrm{V}$ at 300 K
Electron affinity	$\chi = 4.05\,\mathrm{V}$
Dielectric constant	$\epsilon_\mathrm{Si} = 11.9$

[1] We use m_t as the effective mass in the Bohr radius because in each direction the conduction band minima with the lightest mass dominate the exponential tail of an impurity wave function [106].

Nomenclature

Vectors are printed in bold characters (e.g. $\boldsymbol{a}$). Operators and matrices have a hat (e.g. $\hat{H}$, $\hat{A}$). $\langle\cdot|\cdot\rangle$ (or $\langle\cdot|\cdot|\cdot\rangle$ if an operator/matrix is included) note the scalar product — both of classical vectors and wave functions. Variable indices are in italics; designative indices are printed upright (e.g. $\sum_i V_i$ vs. $V_\mathrm{g} + V_\mathrm{d} + V_\mathrm{s}$). The following symbols are used frequently.

α $= \frac{C_\mathrm{g}}{C_\mathrm{g}+C_\mathrm{s}+C_\mathrm{d}}$ gate voltage lever arm

a_B effective Bohr radius in silicon

C_g gate capacitance

C_s source capacitance

C_d drain capacitance

d_spin $= 2$ spin degeneracy

d_valley valley degeneracy

Δ_1 mean spacing between excitations at fixed particle number

Δ_+ addition energy, i.e. the energy needed to add one electron to the dot, usually much higher than Δ_1

e $\approx 1.602 \cdot 10^{-19}\,\mathrm{C}$ absolute value of the electron charge

e ≈ 2.718 Euler's number

E_F Fermi energy

E_g band gap ($E_\mathrm{g} \approx 1.17\,e\mathrm{V}$ in silicon at low temperature)

ϵ_0 $\approx 8.854 \cdot 10^{-12}\,\mathrm{F/m}$ vacuum dielectric constant

$\epsilon_{\mathrm{SiO_2}}$ ≈ 3.9 relative dielectric constant of silicon dioxide

ϵ_Si ≈ 11.9 relative dielectric constant of silicon

g_d drain barrier conductance in quantum units

g_s source barrier conductance in quantum units

g_peak Coulomb blockade peak conductance in quantum units

g_capa Coulomb blockade peak conductance in quantum units, estimated from capacitance

$g_\square$ dimensionless (Thouless) conductance in quantum units

G_T conductance of a tunnel barrier (between continous spectra)

$\imath$ imaginary unit $\imath^2 = -1$

I_T current of a tunnel barrier (between continous and discrete spectrum)

k_F Fermi wave number

k_B $\approx 1.381 \cdot 10^{-23}\,\mathrm{J/K}$ Boltzmann factor

ℓ mean free path

λ Thomas-Fermi screening wave number

m_e $\approx 9.109 \cdot 10^{-31}\,\mathrm{kg}$ electron mass

m^* effective mass

m_l $\approx 0.916\,m_\mathrm{e}$ longitudinal effective mass at the bottom of the silicon conduction band

m_t $\approx 0.190\,m_\mathrm{e}$ transversal effective mass at the bottom of the silicon conduction band

μ_n electron mobility

μ_B $\approx 57.88\,\mu\mathrm{eV/T}$ Bohr magneton

N density of states (per energy and per volume)

N_F density of states at the Fermi level

n_ch channel number

n_D donor concentration

n_S sheet carrier density

n_V volume carrier density

r_s gas parameter (normalized Wigner-Seitz radius)

$\mathcal{T}$ barrier transmission

t_{ox} gate oxide thickness

V_{+} $= \frac{e}{C_{\mathrm{g}}}$ Coulomb blockade peak spacing

V_{g} gate voltage

V_{s} source voltage

V_{d} drain voltage

V_{t} threshold voltage

V_{p} amplitude of pulse excitation on gate voltage

ξ localization length

Bibliography

[1] Y. Alhassid. *The statistical theory of quantum dots*. Rev. Mod. Phys. **72**, 895–968 (2000).

[2] Y. Alhassid and S. Malhotra. *Spin and interaction effects in quantum dots: A Hartree-Fock-Koopmans approach*. Phys. Rev. B **66**, 245313 (2002).

[3] Y. Alhassid and T. Rupp. *Effects of spin and exchange interaction on the Coulomb-blockade peak statistics in quantum dots*. Phys. Rev. Lett. **91**, 056801 (2003).

[4] S. Amasha, K. MacLean, I. Radu, D. M. Zumbuhl, M. A. Kastner, M. P. Hanson and A. C. Gossard. *Measurements of the spin relaxation rate at low magnetic fields in a quantum dot*. cond-mat (2006).

[5] T. Ando, A. B. Fowler and F. Stern. *Electronic properties of two dimensional systems*. Rev. Mod. Phys. **54**, 437–672 (1982).

[6] J. Bardeen and W. H. Brattain. *The transistor, a semi-conductor triode*. Phys. Rev. **74**, 230–231 (1948).

[7] J. Bardeen and W. Shockley. *Deformation potentials and mobilities in non-polar crystals*. Phys. Rev. **80**, 72–80 (1950).

[8] C. W. J. Beenakker. *Theory of Coulomb-blockade oscillations in the conductance of a quantum dot*. Phys. Rev. B **44**, 1646–1656 (1991).

[9] C. W. J. Beenakker. *Random-matrix theory of quantum transport*. Rev. Mod. Phys. **69**, 731–808 (1997).

[10] R. Berkovits, F. von Oppen and Y. Gefen. *Theory of charge sensing in quantum-dot structures*. Phys. Rev. Lett. **94**, 076802 (2005).

[11] R. N. Bhatt and T. M. Rice. *Single particle energy levels in doped semiconductors at densities below the metal-nonmetal transition*. Phys. Rev. B **23**, 1920–1935 (1981).

[12] K. Bladh, D. Gunnarsson, E. Hürfeld, S. Devi, C. Kristoffersson, B. Smålander, S. Pehrson, T. Claeson and P. Delsing. *Comparison of cryogenic filters for use in single electronics experiments*. Rev. Sci. Inst. **74**, 1323–1327 (2003).

[13] Y. M. Blanter, A. D. Mirlin and B. A. Muzykantskii. *Fluctuations of conductance peak spacings in the Coulomb blockade regime: Role of electron-electron interaction*. Phys. Rev. Lett. **78**, 2449–2452 (1997).

[14] M. Boehm, M. Hofheinz, X. Jehl, M. Sanquer, M. Vinet, B. Previtali, D. Fraboulet, D. Mariolle and S. Deleonibus. *Size scaling of the addition spectra in silicon quantum dots*. Phys. Rev. B **71**, 033305 (2005).

[15] W. H. Brattain and J. Bardeen. *Nature of the forward current in germanium point contacts*. Phys. Rev. **74**, 231–232 (1948).

[16] T. M. Buehler, D. J. Reilly, R. P. Starrett, V. C. Chan, A. R. Hamilton, A. S. Dzurak and R. G. Clark. *Observing sub-microsecond telegraph noise with the radio frequency single electron transistor*. J. Appl. Phys. **96**, 6827–6830 (2004).

[17] J. Bylander, T. Duty and P. Delsing. *Current measurement by real-time counting of single electrons*. Nature **434**, 361–364 (2005).

[18] T. G. Castner. *Donor-polarizability enhancement as the insulator-metal transition is approached from the insulating side*. Phys. Rev. B **21**, 3523–3538 (1980).

[19] T. G. Castner, N. K. Lee, G. S. Cieloszyk and G. L. Salinger. *Dielectric anomaly and the metal–insulator transition in n-type silicon*. Phys. Rev. Lett. **34**, 1627–1630 (1975).

[20] E. Conwell and V. F. Weisskopf. *Theory of impurity scattering in semiconductors*. Phys. Rev. **77**, 388–390 (1950).

[21] M. H. Devoret and R. J. Schoelkopf. *Amplifying quantum signals with the single-electron transistor*. Nature **406**, 1039–1046 (2000).

[22] R. O. Duda and P. E. Hart. *Use of the Hough transformation to detect lines and curves in pictures*. Comm. ACM **15**, 11–15 (1972).

[23] K. Efetov. *Supersymmetry in disorder and chaos*. Cambridge University Press (1997).

[24] G. Feher. *Electron spin resonance experiments on donors in silicon. I. Electronic structure of donors by the electron nuclear double resonance technique*. Phys. Rev. **114**, 1219–1244 (1959).

[25] G. Feher and E. A. Gere. *Electron spin resonance experiments on donors in silicon. II. Electron spin relaxation effects*. Phys. Rev. **114**, 1245–1256 (1959).

[26] T. Fujisawa, D. G. Austing, Y. Tokura, Y. Hirayama and S. Tarucha. *Nonequilibrium transport through a vertical quantum dot in the absence of spin-flip energy relaxation*. Phys. Rev. Lett. **88**, 236802 (2002).

[27] T. Fujisawa, T. H. Oosterkamp, W. G. van der Wiel, B. W. Broer, R. Aguado, S. Tarucha and L. P. Kouwenhoven. *Spontaneous emission spectrum in double quantum dot devices*. Science **282**, 932–935 (1998).

[28] T. Fujisawa, Y. Tokura and Y. Hirayama. *Transient current spectroscopy of a quantum dot in the Coulomb blockade regime*. Phys. Rev. B **63**, 081304 (2001).

[29] A. Fujiwara, H. Inokawa, K. Yamazaki, H. Namatsu, Y. Takahashi, N. W. Zimmerman and S. B. Martin. *Single electron tunneling transistor with tunable barriers using silicon nanowire metal-oxide-semiconductor field-effect transistor*. Appl. Phys. Lett. **88**, 053121 (2006).

[30] T. A. Fulton and G. J. Dolan. *Observation of single-electron charging effects in small tunnel junctions*. Phys. Rev. Lett. **59**, 109–112 (1987).

[31] A. Furusaki and K. A. Matveev. *Coulomb blockade oscillations of conductance in the regime of strong tunneling*. Phys. Rev. Lett. **75**, 709–712 (1995).

[32] D. C. Glattli, P. Jacques, A. Kumar, P. Pari and L. Saminadayar. *A noise detection scheme with 10 mK noise temperature resolution for semiconductor single electron tunneling devices*. J. Appl. Phys. **81**, 7350–7356 (1997).

[33] D. Goldhaber-Gordon, J. Göres, M. A. Kastner, H. Shtrikman, D. Mahalu and U. Meirav. *From the Kondo regime to the mixed-valence regime in a single-electron transistor*. Phys. Rev. Lett. **81**, 5225–5228 (1998).

[34] J. Gorman, D. G. Hasko and D. A. Williams. *Charge-qubit operation of an isolated double quantum dot*. Phys. Rev. Lett. **95**, 090502 (2005).

[35] H. Grabert and M. H. Devoret (editors). *Single charge tunneling: Coulomb blockade phenomena in nanostructures*, volume 294 of *NATO ASI series B: Physics*. Plenum Press (1992).

[36] D. E. Grupp, T. Zhang, G. J. Dolan and N. S. Wingreen. *Dynamic offset charges in single-electron transistors*. Phys. Rev. Lett. **87**, 186805 (2001).

[37] L. Guo, E. Leobandung and S. Y. Chou. *A silicon single-electron transistor memory operating at room temperature*. Science **275**, 649–651 (1997).

[38] S. Gustavsson, R. Leturcq, B. Simovic, R. Schleser, T. Ihn, P. Studerus, K. Ensslin, D. C. Driscoll and A. C. Gossard. *Counting statistics of single electron transport in a quantum dot*. Phys. Rev. Lett. **96**, 076605 (2006).

[39] F. Gutfleisch, H. Singer, K. Förger and J. A. Gomollon. *Calculation of high-voltage fields by means of the boundary element method*. IEEE Trans. Power Delivery **9**, 743–749 (1994).

[40] Y. Hada and M. Eto. *Electronic states in silicon quantum dots: Multivalley artificial atoms*. Phys. Rev. B **68**, 155322 (2003).

[41] R. Hanson, L. H. W. van Beveren, I. T. Vink, J. M. Elzerman, W. J. M. Naber, F. H. L. Koppens, L. P. Kouwenhoven and L. M. K. Vandersypen. *Single-shot readout of electron spin states in a quantum dot using spin-dependent tunnel rates*. Phys. Rev. Lett. **94**, 196802 (2005).

[42] R. Hanson, B. Witkamp, L. M. K. Vandersypen, L. H. W. van Beveren, J. M. Elzerman and L. P. Kouwenhoven. *Zeeman energy and spin relaxation in a one-electron quantum dot*. Phys. Rev. Lett. **91**, 196802 (2003).

[43] K. F. Herzfeld. *On atomic properties which make an element a metal*. Phys. Rev. **29**, 701–705 (1927).

[44] M. Hofheinz. *Confinement quantique et effets monoélectroniques dans les nanoMOSFETs*. Master's thesis, Université Grenoble I Joseph Fourier and Ecole Nationale Supérieure de Physique de Grenoble (2003).

[45] M. Hofheinz. *Coulomb blockade and spin in nanoscale silicon transistors*. Master's thesis, Universität Karlsruhe (2004).

[46] M. Hofheinz, X. Jehl, M. Sanquer, R. Cerutti, A. Cros, P. Coronel, H. Brut and T. Skotnicki. *Measurements of capacitances in multi-gate transistors by Coulomb blockade spectroscopy*. to be published in IEEE Trans. Nanotech. (2006).

[47] M. Hofheinz, X. Jehl, M. Sanquer, O. Cueto, G. Molas, M. Vinet and S. Deleonibus. *Capacitance measurements in nanometric silicon devices using Coulomb blockade*. to be published in Solid State Electronics (special edition ULIS 2006) .

[48] M. Hofheinz, X. Jehl, M. Sanquer, G. Molas, M. Vinet and S. Deleonibus. *Individual charge traps in silicon nanowires*. Eur. Phys. J. B **54**, 299–307 (2006).

[49] M. Hofheinz, X. Jehl, M. Sanquer, G. Molas, M. Vinet and S. Deleonibus. *A simple and controlled single electron transistor based on doping modulation in silicon nanowires*. Appl. Phys. Lett. **89**, 143504 (2006).

[50] Y. Imry, Y. Gefen and D. J. Bergman. *Dielectric anomalies near the Anderson metal-insulator transition*. Phys. Rev. B **26**, 3436–3439 (1982).

[51] G.-L. Ingold and Y. V. Nazarov. *Charge tunneling rates in ultrasmall junctions*. In Grabert and Devoret [35], pages 21–108.

[52] I. Ionica. *Effet de champ et Blocage de Coulomb dans des nanostructures de silicium élaborées par microscopie à force atomique*. Ph.D. thesis, Institut National Polytechnique de Grenoble (2006).

[53] H. Ishikuro and T. Hiramoto. *Quantum mechanical effects in the silicon quantum dot in a single-electron transistor*. Appl. Phys. Lett. **71**, 3691–3693 (1997).

[54] C. Jacoboni, C. Canali, G. Ottaviani and A. A. Quaranta. *A review of some charge transport properties of silicon*. Solid State Electron. **20**, 77 (1977).

[55] P. Jarillo-Herrero, S. Sapmaz, C. Dekker, L. P. Kouwenhoven and H. S. J. van der Zant. *Electron-hole symmetry in a semiconducting carbon nanotube quantum dot*. Nature **429**, 389–392 (2004).

[56] X. Jehl, M. Hofheinz, M. Boehm, M. Sanquer, G. Molas, M. Vinet and S. Deleonibus. *Peak spacing statistics in silicon single-electron transistors: Size and gate oxide thickness dependence*. Physica E **34**, 620–623 (2006).

[57] H. Jiang, H. U. Baranger and W. Yang. *Spin and conductance-peak-spacing distributions in large quantum dots: A density-functional theory study*. Phys. Rev. Lett. **90**, 026806 (2003).

[58] H. Jiang, D. Ullmo, W. Yang and H. U. Baranger. *Electron-electron interactions in isolated and realistic quantum dots: A density functional theory study*. Phys. Rev. B **69**, 235326 (2004).

[59] P. Joyez, V. Bouchiat, D. Estève, C. Urbina and M. H. Devoret. *Strong tunneling in the single-electron transistor*. Phys. Rev. Lett. **79**, 1349–1352 (1997).

[60] B. E. Kane. *A silicon-based nuclear spin quantum computer*. Nature **393**, 133–137 (1998).

[61] B. E. Kane, N. S. McAlpine, A. S. Dzurak, R. G. C. G. J. Milburn, H. B. Sun and H. Wiseman. *Single-spin measurement using single-electron transistors to probe two-electron systems*. Phys. Rev. B **61**, 2961–2972 (2000).

[62] M. W. Keller, A. L. Eichenberger, J. M. Martinis and N. M. Zimmerman. *A capacitance standard based on counting electrons*. Science **285**, 1706–1709 (1999).

[63] D. H. Kim, S.-K. Sung, J. S. Sim, K. R. Kim, J. D. Lee, B.-G. Park, B. H. Choi, S. W. Hwang and D. Ahn. *Single-electron transistor based on a silicon-on-insulator quantum wire fabricated by a side-wall patterning method*. Appl. Phys. Lett. **79**, 3812–3814 (2001).

[64] M. J. Kirton and M. J. Uren. *Noise in solid-state microstructures: A new perspective on individual defects, interface states and low-frequency (1/f) noise*. Adv. Phys. **38**, 367–468 (1989).

[65] F. H. L. Koppens, J. A. Folk, J. M. Elzerman, R. Hanson, L. H. W. van Beveren, I. T. Vink, H. P. Tranitz, W. Wegscheider, L. P. Kouwenhoven and L. M. K. Vandersypen. *Control and detection of singlet-triplet mixing in a random nuclear field*. Science **309**, 1346–1350 (2005).

[66] A. A. Koulakov, F. G. Pikus and B. I. Shklovskii. *Statistics of the charging spectrum of a two-dimensional Coulomb-glass island*. Phys. Rev. B **55**, 9223–9226 (1997).

[67] A. A. Koulakov and B. I. Shklovskii. *Charging spectrum and configurations of a Wigner crystal island*. Phys. Rev. B **57**, 2352–2367 (1998).

[68] I. O. Kulik and R. I. Shekhter. *Kinetic phenomena and charge discreteness effects in granulated media*. Sov. Phys.–JETP **41**, 308–316 (1975).

[69] V. V. Kuznetsov, A. K. Savchenko, D. R. Mace, E. H. Linfield and D. A. Ritchie. *Resonant tunneling spectroscopy of interacting localized states: Observation of the correlated current through two impurities*. Phys. Rev. B **56**, 15533–15536 (1997).

[70] V. V. Kuznetsov, A. K. Savchenko, M. E. Raikh, L. I. Glazman, D. R. Mace, E. H. Linfield and D. A. Ritchie. *Reentrant resonant tunneling*. Phys. Rev. B **54**, 1502–1505 (1996).

[71] P. Lafarge, P. Joyez, D. Estève, C. Urbina and M. H. Devoret. *Two-electron quantization of the charge on a superconductor*. Nature **365**, 422–424 (1993).

[72] L. Landau and E. Lifchitz. *Physique théorique, Tome 3, Mécanique quantique*. Mir Moscow (1975).

[73] P. M. Lenahan and J. F. Conley. *What can electron paramagnetic resonance tell us about the Si/SiO_2 system?*. J. Vac. Sci. & Tech., B **16**, 2134–2153 (1998).

[74] D. Loss and D. P. DiVincenzo. *Quantum computation with quantum dots*. Phys. Rev. A **57**, 120–126 (1998).

[75] W. Lu, Z. Ji, L. Pfeiffer, K. W. West and A. J. Rimber. *Real-time detection of electron tunnelling in a quantum dot*. Nature **423**, 422–425 (2003).

[76] S. Lüscher, T. Heinzel, K. Ensslin, W. Wegscheider and M. Bichler. *Signatures of spin pairing in chaotic quantum dots*. Phys. Rev. Lett. **86**, 2118–2121 (2001).

[77] Y. Léger, L. Besombes, J. Fernández-Rossier, L. Maingault and H. Mariette. *Electrical control of a single Mn atom in a quantum dot*. Phys. Rev. Lett. page 107401 (2006).

[78] J. M. Martinis, M. H. Devoret and J. Clarke. *Experimental tests for the quantum behavior of a macroscopic degree of freedom: The phase difference across a Josephson junction*. Phys. Rev. B **35**, 4682–4698 (1987).

[79] K. A. Matveev. *Coulomb blockade at almost perfect transmission*. Phys. Rev. B **51**, 1743–1751 (1995).

[80] K. A. Matveev. *Coulomb blockade oscillations of conductance in the regime of strong tunneling*. Phys. Rev. Lett. **75**, 709–712 (1995).

[81] S. M. Maurer, S. R. Patel, C. M. Marcus, C. I. Duruöz and J. S. Harris. *Coulomb blockade fluctuations in strongly coupled quantum dots*. Phys. Rev. B **83**, 1403–1406 (1999).

[82] U. Meirav, M. A. Kastner, M. Heiblum and S. J. Wind. *One-dimensional electron gas in GaAs: Periodic conductance oscillations as a function of density*. Phys. Rev. B **40**, 5871–5874 (1989).

[83] M. Miller, D. Ullmo and H. U. Baranger. *Statistics of wave functions in disordered systems with applications to Coulomb blockade peak spacing*. Phys. Rev. B **72**, 045305 (2005).

[84] A. D. Mirlin. *Statistics of energy levels and eigenfunctions in disordered systems*. Phys. Rep. **326**, 259–382 (2000).

[85] G. Molas. *Fabrication, caractérisation et modélisation de dispositifs memoire décananométriques à nanocristaux de Silicium sur substrat SOI*. Ph.D. thesis, Institut National Polytechnique de Grenoble (2004).

[86] G. Molas, X. Jehl, M. Sanquer, B. D. Salvo, D. Lafond and S. Deleonibus. *Manipulation of periodic Coulomb blockade oscillations in ultra-scaled memories by single electron charging of silicon nanocrystal floating gates*. IEEE Trans. Nanotech. **4**, 374–379 (2005).

[87] G. Molas, B. D. Salvo, G. Ghibaudo, D. Mariolle, A. Toffoli, N. Buffet, R. Puglisi, S. Lombardo and S. Deleonibus. *Single electron effects and structural effects in ultrascaled silicon nanocrystal floating-gate memories*. IEEE Trans. Nanotech. **3**, 42–48 (2004).

[88] G. E. Moore. *Cramming more components onto integrated circuits*. Electronics **38** (1965).
URL `http://download.intel.com/research/silicon/moorespaper.pdf`.

[89] O. Naaman and J. Aumentado. *Poisson transition rates from time-domain measurements with a finite bandwidth*. Phys. Rev. Lett. **96**, 100201 (2006).

[90] Y. V. Nazarov. *Coulomb blockade without tunnel junctions*. Phys. Rev. Lett. **82**, 1245–1248 (1999).

[91] Y. Ono, A. Fujiwara, K. Nishiguchi, H. Inokawa and Y. Takahashi. *Manipulation and detection of single electrons for future information processing*. J. Appl. Phys. **97**, 031101 (2005).

[92] C. Pasquier, U. Meirav, F. I. B. Williams, D. C. Glattli, Y. Jin and B. Etienne. *Quantum limitation on Coulomb blockade observed in a 2D electron system*. Phys. Rev. Lett. **70**, 69–72 (1993).

[93] S. R. Patel, S. M. Cronenwett, D. R. Stewart, A. G. Huibers, C. M. Marcus, C. I. Duruöz, J. S. Harris, K. Campman and A. C. Gossard. *Statistics of Coulomb blockade peak spacings*. Phys. Rev. Lett. **80**, 4522–4525 (1998).

[94] M. E. Raikh, L. I. Glazman and L. E. Zhukov. *Two-electron state in a disordered 2D island: Pairing caused by the Coulomb repulsion*. Phys. Rev. Lett. **77**, 1354–1357 (1996).

[95] L. P. Rokhinson, L. J. Guo, S. Y. Chou and D. C. Tsui. *Spin transitions in a small Si quantum dot*. Phys. Rev. B **63**, 035321 (2001).

[96] L. P. Rokhinson, L. J. Guo, S. Y. Chou, D. C. Tsui, E. Eisenberg, R. Berkovits and B. L. Altshuler. *Coherent electron transport in a Si quantum dot dimer*. Phys. Rev. Lett. **88**, 186801 (2002).

[97] I. M. Ruzin, V. Chandrasekhar, E. I. Levin and L. I. Glazman. *Stochastic Coulomb blockade in a double-dot system*. Phys. Rev. B **45**, 13469–13478 (1992).

[98] M. Sanquer, M. Specht, L. Ghenim, S. Deleonibus and G. Guegan. *Coulomb blockade in low-mobility nanometer size Si-MOSFETs*. Phys. Rev. B **61**, 7249–7252 (2000).

[99] S. Sasaki, T. Fujisawa, T. Hayashi and Y. Hirayama. *Electrical pump-and-probe study of spin singlet-triplet relaxation in a quantum dot*. Phys. Rev. Lett. **95**, 056803 (2005).

[100] A. K. Savchenko, V. V. Kuznetsov, A. Woolfe, D. R. Mace, M. Pepper, D. A. Ritchie and G. A. C. Jones. *Resonant tunneling through two impurities in disordered barriers*. Phys. Rev. B **52**, R17021–R17024 (1995).

[101] R. J. Schoelkopf, P. Wahlgren, A. A. Kozhevnikov, P. Delsing and D. E. Prober. *The radio-frequency single-electron transistor (RF-SET): A fast and ultrasensitive electrometer*. Science **280**, 1238–1242 (1998).

[102] S. R. Schofield, N. J. Curson, M. Y. Simmons, F. J. Rueß, T. Hallam, L. Oberbeck and R. G. Clark. *Atomically precise placement of single dopants in Si*. Phys. Rev. Lett. **91**, 136104 (2003).

[103] J. H. F. Scott-Thomas, S. B. Field, M. A. Kastner, H. I. Smith and D. A. Antoniadis. *Conductance oscillations periodic in the density of a one-dimensional electron gas*. Phys. Rev. Lett. **62**, 583–586 (1989).

[104] W. N. Shafarman, D. W. Koon and T. G. Castner. *dc conductivity of arsenic-doped silicon near the metal-insulator transition*. Phys. Rev. B **40**, 1216–1231 (1989).

[105] T. Shinada, S. Okamoto, T. Kobayashi and I. Ohdomari. *Enhancing semiconductor device performance using ordered dopant arrays*. Nature **437**, 1128–1131 (2005).

[106] B. I. Shklovskii and A. L. Efros. *Electronic properties of doped semiconductors*. Number 45 in Solid State Sciences. Springer (1984).

[107] W. Shockley. *The theory of p-n junctions in semiconductors and p-n junction transistors*. Bell Syst. Tech. J. **28**, 435 (1949).

[108] F. Simmel, D. Abusch-Magder, D. A. Wharam, M. A. Kastner and J. P. Kotthaus. *Statistics of the Coulomb-blockade peak spacings of a silicon quantum dot*. Phys. Rev. B **59**, R10441–R10444 (1999).

[109] F. Simmel, T. Heinzel and D. A. Wharam. *Statistics of conductance oscillations of a quantum dot in the Coulomb-blockade regime*. Europhys. Lett. **38**, 123–128 (1997).

[110] U. Sivan, R. Berkovits, Y. Aloni, O. Prus, A. Auerbach and G. Ben-Yoseph. *Mesoscopic fluctuations in the ground state energy of disordered quantum dots*. Phys. Rev. Lett. **77**, 1123–1126 (1996).

[111] A. Steane. *Quantum computing*. Rep. Prog. Phys. **61**, 117–173 (1998).

[112] M. Steffen, M. Ansmann, R. C. Bialczak, N. Katz, E. Lucero, R. McDermott, M. Neeley, E. M. Weig, A. N. Cleland and J. M. Martinis. *Measurement of the entanglement of two superconducting qubits via state tomography*. Science **313**, 1423–1425 (2006).

[113] S. M. Sze. *Physics of semiconductor devices*. John Wiley & Sons, New York, 2nd edition (1981).

[114] Y. Tan, T. Kamiya, Z. A. K. Durrani and H. Ahmed. *Room temperature nanocrystalline silicon single electron transistors*. J. Appl. Phys. **94**, 633–637 (2003).

[115] S. J. Tans, M. H. Devoret, R. J. A. Groeneveld and C. Dekker. *Electron-electron correlations in carbon nanotubes*. Nature **394**, 761–764 (1998).

[116] S. Tarucha, D. G. Austing, Y. Tokura, W. G. van der Wiel and L. P. Kouwenhoven. *Direct Coulomb and exchange interaction in artificial atoms*. Phys. Rev. Lett. **84**, 2485–2488 (2000).

[117] J. R. Tucker. *Complementary digital logic based on the "Coulomb blockade"*. Appl. Phys. Lett. **72**, 4399–4413 (1992).

[118] K. Uchida, J. Koga, R. Ohba, S. ichi Takagi and A. Toriumi. *Silicon single-electron tunneling device fabricated in an undulated ultrathin silicon-on-insulator film*. J. Appl. Phys. **90**, 3551–3557 (2001).

[119] D. Ullmo and H. U. Baranger. *Interactions in chaotic nanoparticles: Fluctuations in Coulomb blockade peak spacings*. Phys. Rev. B **64**, 245324 (2001).

[120] G. Usaj and H. U. Baranger. *Spin and e-e interactions in quantum dots: Leading order corrections to universality and temperature effects*. Phys. Rev. B **66**, 155333 (2002).

[121] G. Vasseur, D. Weinmann and R. Jalabert. *Coulomb blockade without potential barriers*. Eur. Phys. J. B **51**, 267–275 (2006).

[122] D. Vion, A. Aassime, A. Cottet, P. Joyez, H. Pothier, C. Urbina, D. Estève and M. H. Devoret. *Manipulating the Quantum State of an Electrical Circuit*. Science **296**, 886–889 (2002).

[123] D. Vion, P. F. Orfila, P. Joyez, D. Estève and M. H. Devoret. *Miniature electrical filters for single electron devices*. J. Appl. Phys. **77**, 2519–2524 (1995).

[124] F. R. Waugh, M. J. Berry, D. J. Mar, R. M. Westervelt, K. L. Campman and A. C. Gossard. *Single-electron charging in double and triple quantum dots with tunable coupling*. Phys. Rev. Lett. **75**, 705–708 (1995).

[125] D. Weinmann. *Quantum transport in nanostructures*. Ph.D. thesis, Universität Hamburg (1994).

[126] D. Weinmann, W. Häusler and B. Kramer. *Spin blockades in linear and nonlinear transport through quantum dots*. Phys. Rev. Lett. **74**, 984–987 (1995).

[127] D. K. Wilson and G. Feher. *Electron spin resonance experiments on donors in silicon. III. Investigation of excited states by the application of uniaxial stress and their importance in relaxation processes*. Phys. Rev. **124**, 1068–1083 (1961).

[128] M. Xiao, I. Martin and H. W. Jiang. *Probing the spin state of a single electron trap by random telegraph signal*. Phys. Rev. Lett. **91**, 078301 (2003).

[129] M. Xiao, L. Martin, E. Yablonovitch and H. W. Jiang. *Electrical detection of the spin resonance of a single electron in a silicon field-effect transistor*. Nature **430**, 435–439 (2004).

[130] M. J. Yoo, T. A. Fulton, H. F. Hess, R. L. Willett, L. N. Dunkleberger, R. J. Chichester, L. N. Pfeiffer and K. W. West. *Scanning single-electron transistor microscopy: Imaging individual charges*. Science **276**, 579–582 (1997).

[131] Z. Zhong, Y. Fang, W. Lu and C. Lieber. *Coherent single charge transport in molecular scale silicon nanowires*. Nano Lett. **5**, 1143–1146 (2005).

[132] J. M. Ziman. *Principles of the theory of solids*. Cambridge University Press (1964).

[133] A. B. Zorin. *The thermocoax cable as the microwave frequency filter for single electron circuits*. Rev. Sci. Inst. **66**, 4296–4300 (1995).

[134] *Intel: Technology and research page*.
URL `http://www.intel.com/technology/silicon`.

[135] *International Technology Roadmap for Semiconductors: Reports*.
URL `http://www.itrs.net/reports.html`.

[136] *Yorick Homepage*.
URL `http://yorick.sourceforge.net`.