Coulomb Blockade of Tunnel-Coupled Quantum Dots

A thesis presented

by

John Michael Golden

to

The Department of Physics

in partial fulfilment of the requirements

for the degree of

Doctor of Philosophy

in the subject of

Physics

Harvard University

Cambridge, Massachusetts

April 1997

REPORT DOCUMENTATION PAGE

Form Approved

AFRL-SR-BL-TR-98-

0368

1. AGENCY USE ONLY (Leave Blank)	2. REPORT DATE	3. REPORT TYPE
	April 1997	Final

4. TITLE AND SUBTITLE
Coulomb Blockade of Tunnel-Coupled Quantum Dots

5. FUNDING NUMBERS

6. AUTHORS
John Michael Golden

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Harvard University

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
AFOSR/NI
110 Duncan Avenue, Room B-115
Bolling Air Force Base, DC 20332-8080

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION AVAILABILITY STATEMENT
Approved for Public Release

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*
See attached.

14. SUBJECT TERMS

15. NUMBER OF PAGES

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	UL

Abstract

Quantum dots are small conducting islands that can often be usefully modeled as tiny capacitors. Though classical charging models can explain the Coulomb blockade of an isolated dot, they must be modified to explain the Coulomb blockade of dots coupled through the quantum mechanical tunneling of electrons. This thesis presents quantum mechanical models for pairs of tunnel-coupled dots and uses these models to follow the coupled-dot blockade as it evolves from that characteristic of two isolated dots to that characteristic of a single composite dot. The primary aim is to find the relation between two quantities: the fractional peak splitting f and the dimensionless interdot channel conductance g, both of which go from 0 to 1 as the isolated-dot blockade changes into the composite-dot blockade.

The thesis begins with Chapter 1, which introduces coupled quantum dots and the Coulomb blockade and highlights the contents of the succeeding chapters.

Chapters 2 and 3 present a transfer-Hamiltonian model for weakly coupled dots and a one-dimensional backscattering model for strongly coupled dots. The leading and subleading terms in the weak- and strong-coupling expressions for f as a function of g are derived. The weak-coupling calculation is performed via Rayleigh-Schrödinger perturbation theory about the endpoint $(g, f) = (0, 0)$. The strong-coupling calculation employs the bosonization approach about the endpoint $(g, f) = (1, 1)$. The results show substantial dependence on the number of interdot tunneling channels N_{ch}.

Chapter 4 goes beyond the work of Chapters 2 and 3, which rely upon the assumption that tunneling and backscattering amplitudes can be treated as energy-independent. Chapter 4 shows that, for realistic interdot barriers, the energy dependence of these amplitudes results in corrections of order U_ρ/W, where U_ρ is the energy cost associated with moving electrons between the dots and W is the energy scale over which transmission through the barrier varies significantly.

Acknowledgments

I would like to thank my thesis adviser Professor Bertrand I. Halperin for his help during my graduate studies at Harvard. I am also grateful to Professors Eric J. Heller and Robert M. Westervelt for serving on my thesis committee, to Professors Daniel S. Fisher and Michael Tinkham for being on the committee for my qualifying examination, to Professor Andrzej Lesniewski for advising me during my first two years of graduate study, and to Professor Gerald Holton for being my undergraduate advisor in physics. Special thanks should be extended to Professor Westervelt and members of his group—Aram Adourian, Catherine Crouch, Carol Livermore, and Fred Waugh—for explaining the experimental work that has inspired much of this thesis.

I thank my peers within the department, my friends, and my family for their gifts of advice, encouragement, and assistance. Particularly noteworthy have been the condensed matter theory "kids," a group which has included Alex Barnett, Martin Bazant, Seth Berger, Noam Bernstein, Tom de Swiet, Mark Goldman, Shilpa Jain, Raphael Lehrer, David Lubensky, Luca Marinelli, Aris Moustakas, Sharad Ramanathan, Jennifer Schwarz, Steven Simon, Igor Smolyarenko, and Joseph Watson. Among my compatriots and office-mates have been William Bies, Ismardo Bonalde and Irvith Carvajal, Hael Collins, Elizabeth Haley, Lev Kaplan, Matthew McIrvin, David Morin, Ron Rubin, Nathan Salwen, Savdeep Sethi, Jon Tyson, and Vyacheslav Zhukov. Finally, there is the select bunch who have borne the brunt of my existence outside of Jefferson and Lyman: Amy Douma, Paul Janzen, Ian Watson, and Brian K., Brian M. and Mary P. Golden.

Citations to Published Work

This thesis consists of an introduction followed by chapters based on individual research articles. Versions of Chapter 2 and Chapter 3, with the accompanying appendices, have been published in *Physical Review B*, and a version of Chapter 4 has been submitted to the same journal for publication. The reference information for the published articles is as follows: J. M. Golden and B. I. Halperin, Phys. Rev. B **53**, 3893 (1996); **54**, 16 757 (1996).

"Ah, you're moralizing. Compared to physics, that's called boring."

—the Marquise in Fontenelle's *Conversations on the Plurality of Worlds* [1]

Contents

List of Figures

Chapter 1

Introduction

The continued push for smaller and smaller electronic devices has led to the fabrication of submicron structures that are large on the scale of individual atoms but small enough that their quantum mechanical nature cannot be ignored. Modern lithographic techniques have permitted the creation of controllable *artificial atoms* or *quantum dots*, small conducting islands with discrete quantum eigenstates onto which experimentalists can put one electron at a time. By straddling the limits of quantum mechanical *microscopics* and classical *macroscopics*, such *mesoscopic* systems present physicists with a new array of challenges and opportunities. The capacity to construct a series of systems that interpolates between the macroscopic and microscopic limits forces researchers to come up with models and techniques that can effectively describe the gray areas in which classical and quantum mechanical approaches meet—in which, loosely speaking, $\hbar$ is neither zero nor one. The intermediate regime that mesoscopics occupies is almost inherently uncomfortable for physicists, who tend to prefer working via perturbation theory about the most extreme limits imaginable. Nevertheless, with care and patience, useful models can be constructed, and this thesis attempts to illustrate the potential for progress by presenting quantum mechanical models for the Coulomb blockade of systems of tunnel-coupled quantum dots. These models are intended to elucidate a key

issue in the world of quantum confinement—the question of how two isolated systems, with occupation numbers n_1 and n_2 that are both good quantum numbers, can evolve into a single well-connected supersystem, for which only the total number of electrons, $N_{\rm tot} = (n_1 + n_2)$ is a good quantum number.

In order to expedite the reader's assimilation of the information in this thesis, this chapter seeks both to introduce and to summarize the key points and concepts that lie around and within the detailed calculations that follow. As there are now a number of serviceable reviews of the general field of mesoscopic physics [2, 3, 4, 5] as well as the more specific subfields of single-electron tunneling [6, 7, 8, 9, 10, 11, 12, 14, 13, 15, 16] and quantum dots [17, 18, 19, 20, 21, 22], this chapter only covers the portions of these fields that are necessary to appreciate the contents of Chapters 2 through 4.

The design of the chapter is as follows. In Sec. 1.1, quantum dots are described, and the length and energy scales that characterize their physics are discussed. In Sec. 1.2, the phenomenon of the Coulomb blockade is presented, along with its orthodox explanation through a capacitive charging model. In Sec. 1.3, experimental observations of the Coulomb blockade of coupled quantum dots are reported, and the need to go beyond a narrowly capacitive model is observed. Sections 1.4 and 1.5 outline the structure and implications of quantum mechanical models for dot systems in which electron tunneling and backscattering are crucial components. Section 1.6 improves upon the results of the preceding sections by introducing a model that takes into account the finite nature of the potential barrier between the dots. In Sec. 1.7, the chapter is reviewed and the remainder of the thesis is outlined.

1.1 Defining Quantum Dots

1.1.1 Dot Apologia

To this point, the reader has been led to believe that this paper is concerned with *quantum dots* (QDs). In fact, this is less than half (or perhaps more than double) the story. The research presented in Chapters 2 through 4 is more properly described as concerned with *large planar quantum dots* (LPQDs), a somewhat awkward combination of words that this section seeks to render meaningful. The strategy for characterizing LPQDs consists of two parts. First, we present a theoretical conception of the large planar quantum dot—the quantum dot *Idea* [23]—which is based upon a particular mix of length and energy scales. We then discuss the empirical quantum dot, describing the primary sort of experimental system that affords concrete realizations of the LPQD idea.

1.1.2 The Quantum Dot Idea

A large planar quantum dot is a conducting island formed by laterally confining a region of a two-dimensional electron gas (2DEG). The characteristic vertical width w of the dot is that of the 2DEG itself and is less than one Fermi wavelength λ_F, where the Fermi energy E_F accounts for only the energy of the lateral degrees of freedom. Thus, the degrees of freedom corresponding to the vertical dimension can be ignored because the electrons within both the dot and the surrounding 2DEG are restricted to the lowest vertical quantum level. The LPQD is *planar*.

The LPQD is a *dot*, however, because it represents only a discrete portion of the 2DEG, and it is a *quantum* dot because its characteristic lateral dimension $L_{\rm dot}$ is much less than the phase decoherence length ℓ_ϕ over which the individual electronic wavefunctions decohere. The decoherence length diverges as the temperature of the system goes to zero, and, therefore, the assumption $L_{\rm dot} \ll \ell_\phi$ is at least always theoretically

achievable in the zero-temperature limit ($k_{\rm B}T \to 0$).

What remains to be explained is what makes the LPQD *large*. Quite simply, it is large because the dot length $L_{\rm dot}$ is much greater than the Fermi wavelength $\lambda_{\rm F}$. The average level spacing δ of the dot's single-particle eigenstates is therefore much less than the Fermi energy $E_{\rm F}$, and, hence, the number N of electrons on the dot is much greater than 1. We are particularly concerned with dots for which N equals a few hundred.

The explanation of the phrase *large planar quantum dot* is summarized by the following hierarchy of length scales:

$$w < \lambda_{\rm F} \ll L_{\rm dot} \ll \ell_\phi \,. \tag{1.1}$$

LPQDs can also be characterized by a ladder of energy scales. We have already encountered two such scales, the average level spacing δ and the much larger Fermi energy $E_{\rm F}$. Corresponding to the vertical height w of the 2DEG, we have a third energy scale E_w, the *confinement energy* of the 2DEG, which is the characteristic scale of the level spacing of the vertical modes and is, by assumption, larger than $E_{\rm F}$. There are two other important scales for the isolated quantum dot that we currently have in mind: the charging energy U and the thermal energy $k_{\rm B}T$. In addition to being low enough that $L_{\rm dot} \ll \ell_\phi$, the thermal energy $k_{\rm B}T$ must be at least about four times smaller than the average level spacing for individual dot eigenstates to be resolvable [18]. Since we wish only to worry about ground state wavefunctions, our dots should be in the limit $4k_{\rm B}T < \delta$.

The final energy scale, the charging energy U, results from the electrostatic repulsion between electrons on the dot. It equals the change in the chemical potential of the electrons when one electron is added to the dot. Consequently, at least for simple geometries in which the dot is essentially an isolated disk of charge [24], U is expected to scale roughly as $e^2/L_{\rm dot}$, whereas the level spacing δ scales as $2\pi\hbar^2/mL_{\rm dot}^2$, where m is the effective mass of an electron in the 2DEG. Thus, for large dots, U is an intermediate energy scale, much larger than δ but much less than the Fermi energy $E_{\rm F}$. The hierarchy

of energy scales is

$$4k_{\mathrm{B}}T \lesssim \delta \ll U \ll E_{\mathrm{F}} < E_w\,. \tag{1.2}$$

The development of the idea of the *large planar quantum dot* (*LPQD*) is now complete. It is time to descend to a more sublunary sphere to consider the *empirical* LPQD as realized in GaAs/AlGaAs heterostructures. Since the only quantum dots explicitly considered in this thesis are LPQDs, we will henceforth use the shorthand *quantum dots* (QDs) to refer to large planar quantum dots (LPQDs), unless explicitly noted otherwise.

1.1.3 The Empirical Quantum Dot

The sorts of Coulomb blockade and tunneling effects that are examined in this thesis have been observed in a number of different experimental systems, ranging from metallic films to silicon-inverson layers [13]. The primary systems for realization of large planar quantum dots (LPQDs) are, however, *split-gate semiconductor heterostructures* that consist of alternating layers of gallium arsenide (GaAs) and aluminum gallium arsenide ($\mathrm{Al}_x\mathrm{Ga}_{1-x}\mathrm{As}$) sitting beneath a layer of AlGaAs doped with silicon donors and an array of surface metallic gates [see Fig. 1.1(a)].

In such systems, electrostatic forces confine carrier electrons to a region only a few nanometers wide ($w \sim 4$ nm) at the interface between a pair of AlGaAs and GaAs layers [2]. The carrier electrons in the heterostructures are generally cooled to temperatures on the order of a tenth of a Kelvin. The result is a high-mobility two dimensional electron gas (2DEG) with a typical carrier concentration n_s of $4 \times 10^{11}/\mathrm{cm}^2$, an electron effective mass of about 0.067 m_{e} (where the free electron mass $m_{\mathrm{e}} = 9.1 \times 10^{-28}$ g), a Fermi wavelength of about 50 nm, and elastic and inelastic scattering lengths ℓ_{el} and ℓ_{inel} on the order of 10 μm [2].

Quantum dots are carved out of the 2DEG by applying negative voltages to the surface metallic gates [7, 15, 18, 19, 21, 22]. The negative potential repels electrons from the regions of the 2DEG lying directly under the gates. Hence, given an appropriate

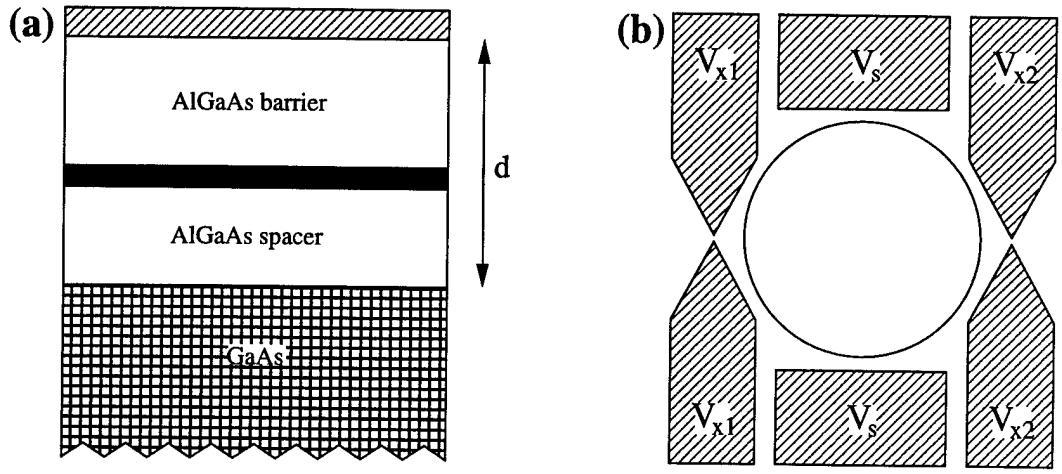

Figure 1.1: (a) Representative cross-section of a split-gate semiconductor heterostructure. The silicon donor layer (in black) lies between the AlGaAs spacer and barrier layers. The two-dimensional electron gas (2DEG) sits at the interface between the AlGaAs and GaAs layers, located a distance d beneath the surface metallic gates (striped). (b) A surface gate pattern that can be used to create the barriers delimiting a quantum dot. The region to which electrons on the dot are confined is suggested by the circle.

cookie-cutter pattern for the gates [see Fig. 1.1(b)], application of negative potentials results in the formation of *dots*, laterally confined islands of electrons. If the characteristic size of such an island, $L_{\rm dot}$, is several tenths of a μm, we have the proper relation of length scales for a large planar quantum dot (recall the inequalities of Eq. 1.1).

It is worth considering what the dot looks like to its electronic inhabitants. We can thereby gain a sense of the self-consistent potential in which the electrons move. Figure 1.2 gives a graphical representation of a possible potential landscape for a single quantum dot in a two-dimensional electron gas. The potential well within the interior of the dot is pictured as essentially flat. In reality, there are bumps in this plain, primarily as a result of the localized fields from silicon impurities in the donor layer. However, such roughness can generally be neglected compared to the smoothly modulated potential induced by the surface metallic gates, which delineates the dots and therefore determines the average intradot level spacings.

The most prominent portions of this smoothly modulated potential are the high-rise

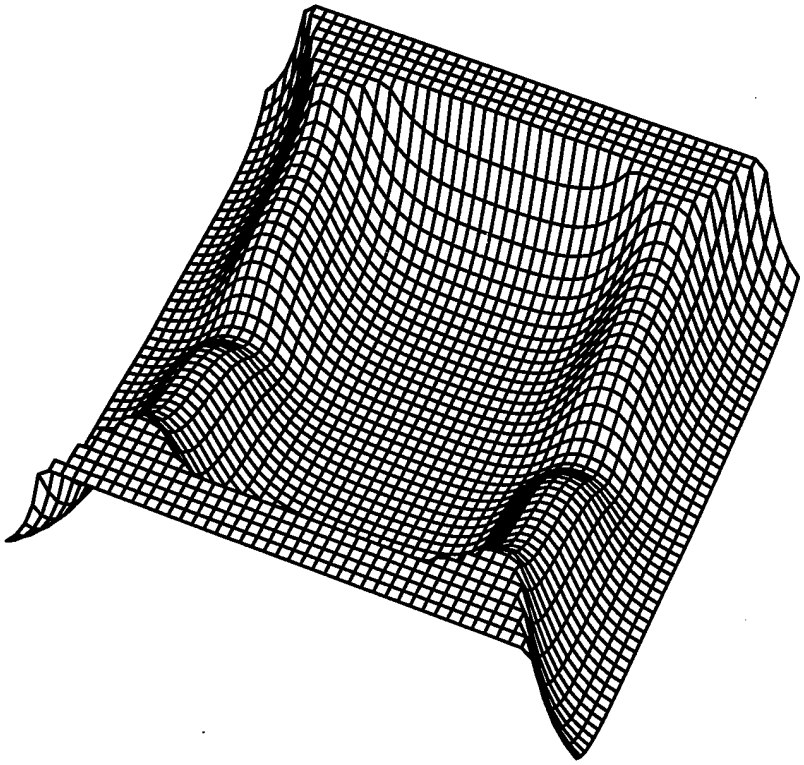

Figure 1.2: Representative ansatz for the potential landscape of a large planar quantum dot. Disorder due to the donor impurities is assumed small, so the interior of the dot can be pictured as essentially a plain. The "flat-tops" of the near and far edges are an artifact of truncating the potential at an arbitrary positive energy.

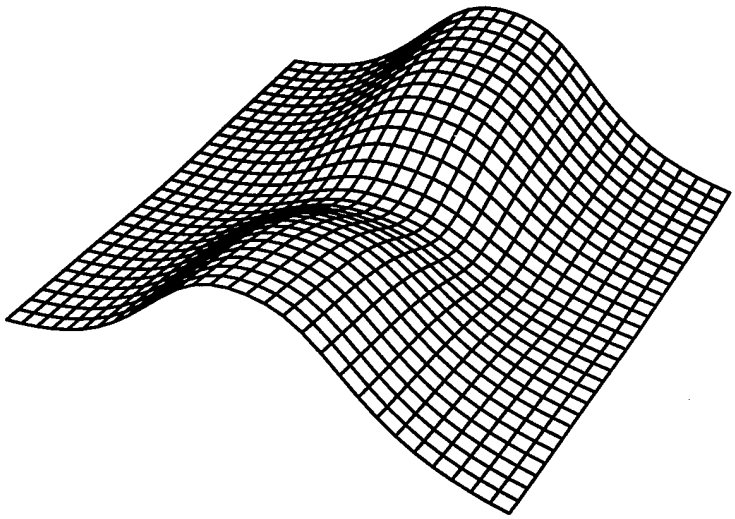

Figure 1.3: A saddle-shaped tunneling channel.

zones of the dot walls. In order for the dot to be well-defined, the maximum heights of these barriers must be at least of the order of the Fermi energy E_F. If there is a space between two of the surface gates used to define the dot [see Fig. 1.1(b)], the walls that define the dot can be punctured by a pass, a *tunneling channel* between the dot and the 2DEG that takes a characteristic *saddle shape* (see Fig. 1.3).

From the above discussion of the potential landscape, it is clear that there is an additional length scale ξ that was not covered in our contemplation of the quantum dot idea. This length scale is the characteristic distance over which the potential changes from its maximum values directly underneath the metallic gates to its sea-level values in the plains of the dot and the 2DEG. Often called the *device resolution*, ξ is expected—in the split-gate geometry of Fig. 1.1(a)—to be approximately equal to the distance d between the surface gates and the 2DEG [19, 25, 26]. Since the surface gates are generally between 50 nm and 200 nm from the 2DEG, ξ is usually larger than the Fermi wavelength λ_F, though of the same order of magnitude. The relative magnitude of ξ is vital in determining whether transmission through the saddle-shaped tunneling channels can be treated as independent of the energy of the incident electrons.

Having completed our discussion of the length scales of the empirical LPQD, we now consider the relevant energy scales. For a typical GaAs/AlGaAs LPQD with $L_{dot} \simeq 0.5$ μm and an electron temperature of 0.1 K, we have the following approximate equivalences: $k_B T \simeq 10$ μeV, $\delta \simeq 50$ μeV, $U \simeq 500$ μeV, and $E_F \simeq 14$ meV [2, 19, 22].

A final energy that we should consider is the energy scale W over which transmission through the saddle-shaped channels changes significantly. The size of this energy is related inversely to the device resolution ξ. If W is much greater than both the level spacing δ and the charging energy U, the transmission probability can probably be approximated as independent of the energy of the incident electrons. However, as noted in Sec. 1.6 and in more detail in Chapter 4, in typical LPQDs, the scale W is probably a bit closer to U than one might like, and therefore at least the leading effects due to

$U/W \neq 0$ should be considered.

1.1.4 Coupled Dots

With the addition of the transmission energy scale W, we have completed the list of length and energy scales that characterize a single *closed* quantum dot at low temperatures—i.e., a quantum dot that, to first approximation, does not exchange electrons with its surroundings and has a resolvable spectrum of eigenstates. To zeroth approximation, the closed quantum dots can be thought of as surrounded by infinite potential walls through which electrons cannot tunnel.

In fact, quantum dots are never completely closed off from their surroundings, and the tunneling rate between the dot and its environment must only fall below a certain threshold for the dot to be considered closed. The relevant criterion can be determined by considering the case of a single dot coupled via electron tunneling to a bulk lead [see 1.4(a)]. Via a number of arguments [21, 35, 36, 33, 38, 39], ranging from those employing relatively sophisticated quantum mechanical models to those based upon the RC time of a circuit model or upon the Thouless criterion [40], it has been shown that it is necessary that the tunneling rate Γ_j between the bulk lead and the dot's jth lowest-energy single-particle state satisfies the inequality [9, 14]

$$h\Gamma_j \ll \delta\,, \tag{1.3}$$

where the jth lowest state is assumed to be below or not far above the Fermi surface. Since $h\Gamma_j$ is the level width due to decay into the leads, satisfaction of this inequality is necessary for the individual single-particle eigenstates to be resolvable. It is also equivalent to requiring that the conductance between the dot and lead is much less than e^2/h.

Of course, what most of this thesis is about is not a single closed dot but, rather, a system of two coupled dots. Such a system can be created by applying negative voltages to the gate pattern shown in Fig. 1.4(b). If the external gate voltages V_{x1} and

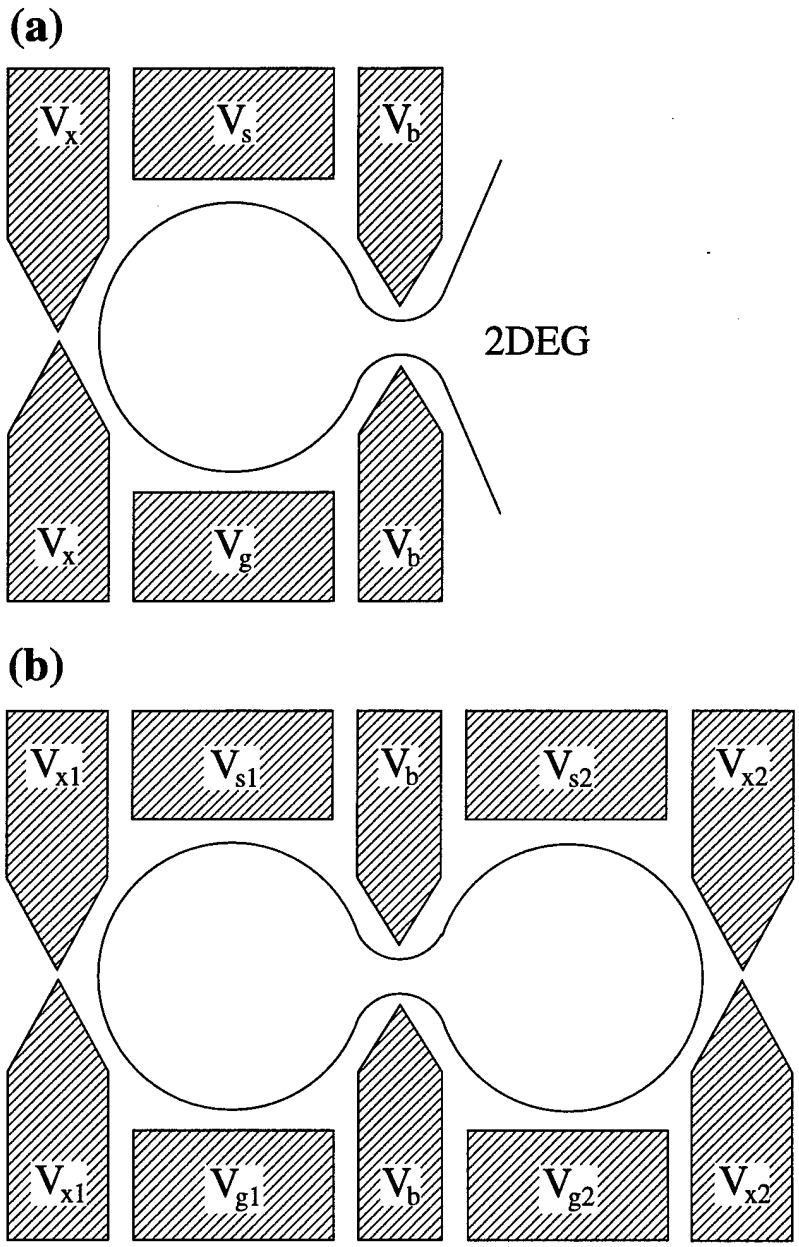

Figure 1.4: (a) Surface gate pattern for a single dot coupled to a bulk lead by a tunneling channel. Regions of significant electron occupation are suggested by the partially enclosed curve. The voltage V_b controls the rate of tunneling between the dot and the lead. V_g determines the average number of electrons on the dot. V_s and V_x are held constant, and tunneling through the barrier controlled by V_x is negligible compared to that through the barrier controlled by V_b. (b) Two adjacent tunnel-coupled dots. V_{g1} and V_{g2} control the average number of electrons on the individual dots. V_b controls tunneling between the two dots. The other voltages are fixed, and tunneling to the leads is assumed negligible compared to that between the dots.

V_{x2} are sufficiently negative, the two-dot system can be considered closed with respect to the surrounding leads. There is then a well-defined integer number of electrons on the *double dot.* On the other hand, if the central gate potential V_b is variable, there is not necessarily a well-defined number of electrons in either half of the double dot; the individual dots are not necessarily closed with respect to each other. When V_b is strongly negative, the conductance through the interdot barrier G_b is much less than e^2/h, and the double dot consists of two largely closed individual dots that are only *weakly coupled* through the tunneling of electrons between them. When the negativity of V_b is relaxed, the conductance G_b increases, and the dots become *strongly coupled*, with charge fluctuations rendering their individual occupation numbers ill-defined. The total number of electrons on the double dot remains a good quantum number, however, and, if the barrier V_b is lowered far enough (without, of course, being lowered so far that electrons leak out of the center of the double dot), one might expect that the system begins to behave like a single composite dot formed from the fusion of the two originally isolated dots. This thesis is largely about how such a fusion can occur.

A point worth stressing is that the two-dot systems we study are always closed with respect to the surrounding leads. The tunneling rates Γ_j are such that $h\Gamma_j$ is not only much less than the average intradot level spacing δ but also on the order of a fifth of the thermal energy $k_{\mathrm{B}}T$. The conductance through the external barriers energized by V_{x1} and V_{x2} [recall Fig. 1.4(b)] is therefore not much more than a handful of hundredths of e^2/h and can be neglected in looking for the ground state of the two-dot system. In contrast, as the voltage V_b is varied, the conductance G_b goes from nearly zero to values on the order of or greater than e^2/h and is therefore not negligible.

With these comments on quantum dot closure, our tour of the definition of quantum dots has come to an end. Along the way, we have encountered a menagerie of energy scales centered upon the charging energy U. We now consider how these scales can be used to make sense of the Coulomb blockade.

1.2 The Orthodox Coulomb Blockade

1.2.1 Interactions in Artificial Atoms

The quantum dots described in Sec. 1.1 are sometimes referred to as *artificial atoms.* They differ significantly from natural atoms, however, both in terms of the nature of their confining potential, which lacks the high symmetry of a central potential [41], and in terms of their characteristic size $L_{\rm dot}$, which is on the order of thousands of angstroms and therefore much larger than an atomic radius. Since, at least in the simplest geometries, the average single-particle level spacing goes roughly as $1/L_{\rm dot}^2$ and the typical electronic energy of repulsion goes roughly as $1/L_{\rm dot}$, the size discrepancy between natural and artificial atoms means that the effects of Coulomb repulsion are much more important in the behavior of quantum dots than in natural atoms [20]. Perhaps the most dramatic example of such size-driven electron-electron effects is the phenomenon known as the *Coulomb blockade.*

1.2.2 Phenomenological Description of the Blockade

To describe the Coulomb blockade we refer to Fig. 1.5, which gives a caricature of a quantum dot coupled capacitively to a gate at the potential V_g and coupled both capacitively and via tunneling channels to bulk leads at the potentials V_{L1} and V_{L2}, respectively. In the absence of the Coulomb blockade, one might expect that, for an arbitrary value of V_g and a small value for the bias $\Delta V = V_{L2} - V_{L1}$, the current through the dot would be proportional to ΔV. However, when both $k_{\rm B}T$ and ΔV are tens of microelectronvolts or less, the experimentally observed current is essentially zero for most values of V_g, the exceptions being roughly evenly spaced values of V_g at which nonzero conductance peaks occur. The fact that the conductance through the dot is usually zero is the effect known as the Coulomb blockade.

Why do we suspect the blockade is Coulombic? If it were to dissolve when $k_{\rm B}T$ or

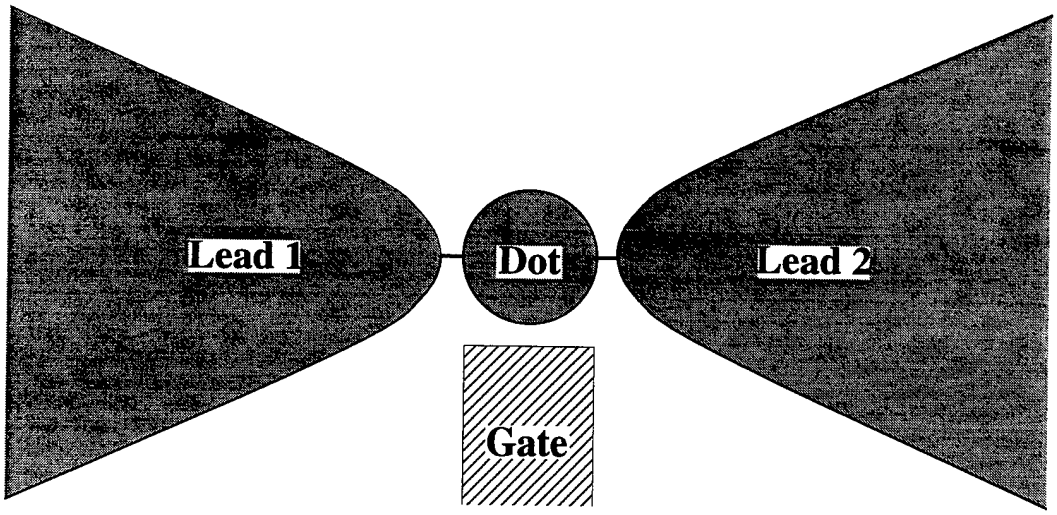

Figure 1.5: Caricature of a system of one dot tunnel-coupled to two bulk leads and capacitively coupled to a surface gate.

ΔV rose to values on the order of the average level spacing δ, it might be attributed to the discreteness of δ. However, it persists to values of $k_{\mathrm{B}}T$ and ΔV equal to hundreds of microelectronvolts—i.e., until $k_{\mathrm{B}}T$ and ΔV are of the order of the charging energy U.

1.2.3 Orthodox Model for the Single-Dot Blockade

To understand more precisely whence the Coulomb blockade arises, it is helpful to view the quantum dot as a capacitor characterized by the potential energy $Q^2/2C_\Sigma$, where Q is the net charge on the dot and C_Σ is its total capacitance. Having adopted this *orthodox model*, one can then construct a circuit diagram showing the couplings between the dot and the surrounding gates and leads. Calculation of the energy E of this circuit, including the work done in charging the dot-capacitor, yields the result [6, 7, 8, 10, 11, 12, 13, 15, 16]

$$E = \frac{U}{2}(n - \phi)^2 , \tag{1.4}$$

where $U = e^2/C_\Sigma$, N is the number of *excess* electrons on the dot, and ϕ is a dimensionless parameter that depends linearly on the gate voltage V_{gate}:

$$\phi = \frac{C_g V_g}{e} + \phi_0 \,, \tag{1.5}$$

where ϕ_0 is a constant and C_g is the capacitance between the dot and the gate.

For a dot that is isolated from its surroundings and cold ($h\Gamma_j$ and $k_B T$ are small), the number n assumes only well-defined integer values, and the integral nature of n provides a ready explanation for the Coulomb blockade. When ϕ does not equal $m + 1/2$, where m is an integer, the ground state of the circuit is unambiguously given by the state in which n is the integer n_ϕ that is nearest to ϕ. The energies of states in which $n \neq n_\phi$ exceed the ground-state energy by at least

$$\Delta E = \frac{U}{2}(1 - 2|n_\phi - \phi|) \,. \tag{1.6}$$

When both the thermal energy $k_B T$ and the voltage bias ΔV are much less than ΔE, processes that change the number of electrons on the dot are effectively blocked, and virtually no current can pass. The result is the Coulomb blockade.

1.2.4 Degeneracies and Conductance Peaks

What happens when ϕ equals $m + 1/2$? Then, there are two degenerate ground states for the circuit—one in which there are $n_- = (2\phi - 1)/2$ electrons on the dot and another in which there are $n_+ = n_- + 1 = (2\phi + 1)/2$ electrons on the dot. Thus, depending upon which of the $n_\pm$ states the circuit is in, an electron can be added or removed with energetic impunity. Current can flow. Consequently, as ϕ is varied, one expects the regions where there is essentially no current through the dot to be punctuated by evenly spaced conductance peaks centered at values of V_g that yield $\phi = m + 1/2$.

The existence and periodicity of the conductance peaks can be explained graphically by plotting the parabolic energy curves $E_n(\phi)$ [see Fig. 1.6(a)]. These curves give the energies of states in which the number of "excess" electrons on the dots is given by the

particular integer values n. For ϕ between $n-1/2$ and $n+1/2$, the curve $E_n(\phi)$ is the lowest-energy parabola, and there is a finite excitation energy to the next lowest-energy curves, $E_{n\pm1}(\phi)$. At $\phi = n \pm 1/2$, on the other hand, the $E_n(\phi)$ curve is degenerate with the $E_{n\pm1}(\phi)$ curve. At such points, current flows, and conductance peaks result [see Fig. 1.6(b)].

The fact that the conductance peaks have a finite width is a consequence of both thermal broadening ($k_{\rm B}T \neq 0$) and lifetime broadening ($h\Gamma_j \neq 0$). In our energy hierarchy (recall the discussion in Sec. 1.1.3), the lifetime broadening is negligible compared to the thermal broadening, and, since $4k_{\rm B}T \lesssim \delta$, the shape of conductance peaks near $\phi = m + 1/2$ is given by [9, 42]

$$G_{\rm dot} = \frac{e^2}{4k_{\rm B}T}\left(\frac{\Gamma_m^{L1}\Gamma_m^{L2}}{\Gamma_m^{L1}+\Gamma_m^{L2}}\right)\cosh^{-2}\left[\tilde{\epsilon}(m,\phi)/2k_{\rm B}T\right], \tag{1.7}$$

where $\tilde{\epsilon}(m,\phi) = \epsilon_m - \epsilon_{\rm F} + U\left(1/2 - |m-\phi|\right)$, ϵ_m is the kinetic energy of the mth lowest-energy single-particle state on the dot, and Γ_n^{Li} is the tunneling rate between this state and lead i.

1.2.5 Orthodox Model for the Double-Dot Blockade

Our analysis of the single-dot Coulomb blockade can be straightforwardly extended to that of the Coulomb blockade of two dots in series. The schema for the system is shown in Fig. 1.7. Once again there are two leads with voltages V_{L1} and V_{L2} providing a bias $\Delta V = V_{L2} - V_{L1}$. There are now, however, two side gate voltages, V_{g1} and V_{g2}, that can, in theory at least, be varied separately to control the number of electrons on each of the dots. If the two dots are well isolated from each other—i.e., if interdot electrostatic interactions and tunneling are negligible—the energy of the system equals the sum of the energies of the individual dots:

$$E = \sum_{i=1}^{2}\frac{U_i}{2}(n_i - \phi_i)^2, \tag{1.8}$$

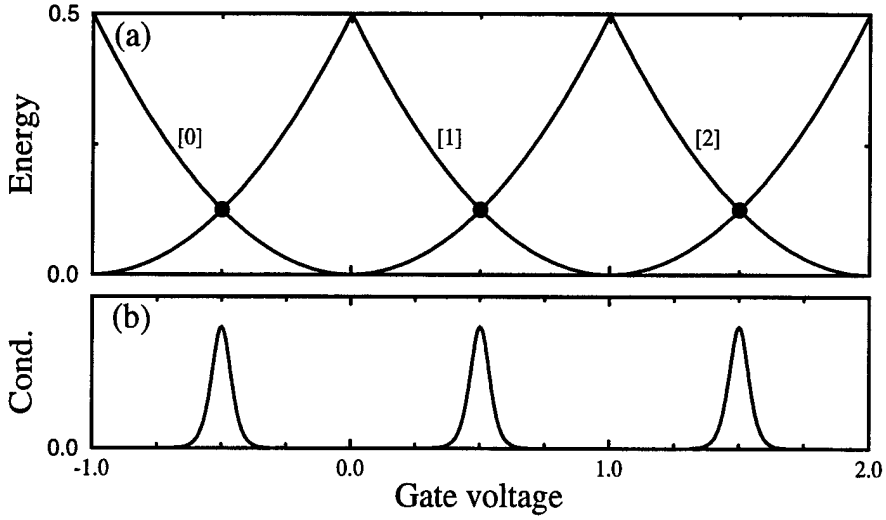

Figure 1.6: (a) Energy curves in the "orthodox model" for a single dot coupled capacitively to a surface gate. Energies are given in units of the dot charging energy U; the gate voltage is in units of e/C_g. For each integer number n of "excess" electrons on the dot, there is a corresponding energy curve, labeled $[n]$, that is a parabolic function of the gate voltage. (b) Conductance through the dot as a function of the gate voltage. The conductance peaks are schematic; for simplicity they are thermally broadened and symmetric in shape. The peaks occur at the points of degeneracy of the energy parabolas.

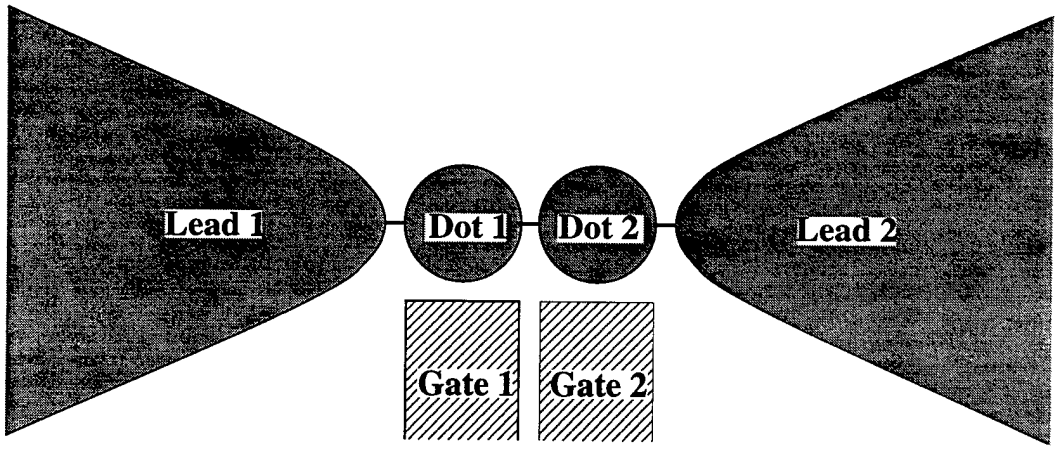

Figure 1.7: Caricature of a system of two tunnel-coupled dots with adjoining bulk leads and occupation-controlling surface gates.

where n_i is the number of electrons on dot i, $\phi_i = C_{gi}V_{gi}/e$, $U_i = e^2/C_{\Sigma i}$, $C_{\Sigma i}$ is the total capacitance of dot i, and C_{gi} is the capacitance between dot i and the ith side gate.

Studies of the implications of Eq. 1.8 for the Coulomb-blockade behavior of the double-dot system have led to an understanding of a rich variety of effects that result from asymmetry between the two dots—i.e., from $C_{\Sigma 1} \neq C_{\Sigma 2}$, etc. [47, 48]. However, since this thesis studies the change in the double-dot Coulomb blockade as a result of interdot tunneling, it is best to separate the issue of asymmetry from our concerns by considering a system of symmetric dots in which the effect of interdot coupling is most discernible and dramatic.

1.2.6 Uncoupled Symmetric Dots

First we should clarify what is meant by a system of *symmetric dots*. In essence, we mean that the dots are electrostatically symmetric—i.e., that electrostatic quantities that are essentially invariant under change of the gate voltages V_{gi}, such as the total capacitances $C_{\Sigma i}$ and the gate-to-dot capacitances C_{gi}, are the same on both dots.

Under these conditions, Eq. 1.8 simplifies to

$$E = \frac{U}{2}\sum_{i=1}^{2}(n_i - \phi_i)^2 . \tag{1.9}$$

If we assume an extra symmetry for the dots, $\phi_1 = \phi_2$, the energy becomes a function of a single voltage parameter $\phi = \phi_i$, and each pair of integer occupation numbers $\{n_1, n_2\}$ is associated with an energy function $E_{\{n_1,n_2\}}(\phi)$ that is parabolic in ϕ. As can be surmised from Fig. 1.8(a), these functions yield parabolas that are identical in shape, their only distinguishing feature being the locations of their minima.

1.2.7 Peak Splitting and Even-Odd Energy Shifts

Since the decisive issue for the Coulomb blockade is whether the ground state is degenerate, we are primarily interested in the lowest-energy parabolas $E_{N_{\rm tot}}(\phi)$ that correspond to specific values of the double-dot occupation number $N_{\rm tot} = (n_1 + n_2)$. The values of n_1 and n_2 for such parabolas are given by the formula $n_1 = n_2 = N_{\rm tot}/2$ when $N_{\rm tot}$ is even and by the formula $n_1 = n_2 \pm 1 = (N_{\rm tot} \pm 1)/2$ when $N_{\rm tot}$ is odd. For *even* $N_{\rm tot}$, the minima of the $E_{N_{\rm tot}}(\phi)$ parabolas all lie on the line $E = 0$. For *odd* $N_{\rm tot}$, the minima of the parabolas lie along a different line of higher energy, $E = U/4$. For both even and odd parabolas, the ϕ-coordinate of the minimum is $N_{\rm tot}/2$ (see Fig. 1.8).

For small bias, the conductance through the double dot is close to zero except near $\phi = \phi_m$, where $\phi_m = m + 1/2$. Around these values of ϕ, conductance peaks are observed. Unlike the single-dot case, however, these peaks are essentially double or degenerate because for $\phi = \phi_m$ it is true both that $E_{2m} = E_{(2m+1)}$ and that $E_{(2m+1)} = E_{(2m+2)}$. Thus, the ϕ_m-values mark the intersection of four parabolas, rather than two.

Now suppose that the lowest-energy odd parabolas shift down relative to the lowest-energy even parabolas by a distance Δ. Then, the single intersection at $\phi = \phi_m$ and $E = U/4$ is replaced by two separate intersections at $\phi = \phi_m \pm \Delta/U$. Thus, as indicated in Fig. 1.8, the single peak at ϕ_m splits into two distinct subpeaks equally spaced

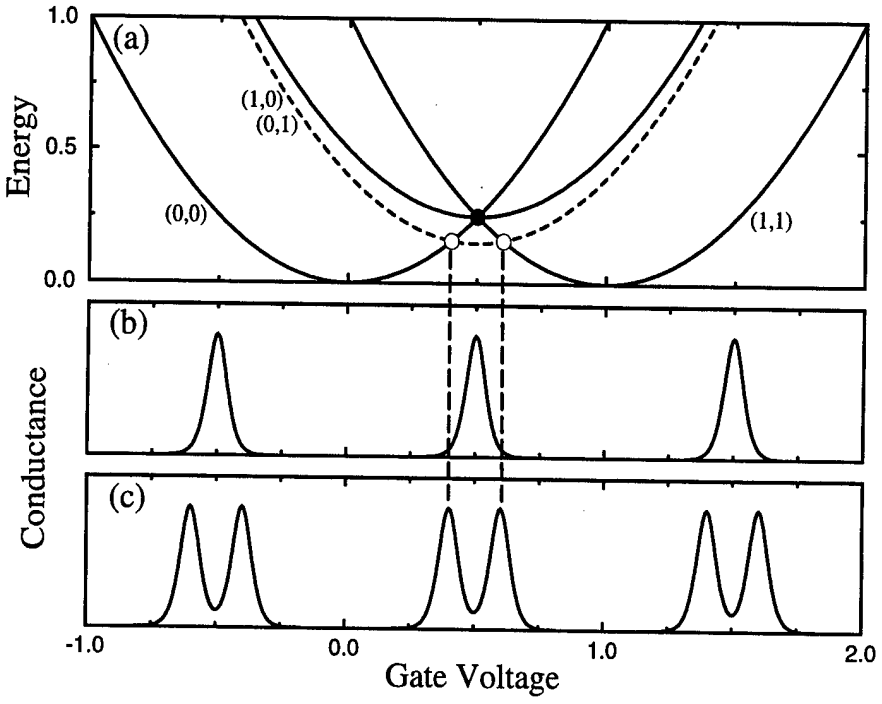

Figure 1.8: (a) Energy curves in the "orthodox model" of a symmetric two-dot system coupled capacitively to a gate potential. Energies are given in units of the charging energy U; the gate voltage is in units of e/C_g. In the absence of interdot coupling, each state with occupation n_i of the ith dot has a corresponding energy curve, labeled (n_1, n_2), which is a parabolic function of the gate voltage. The zero of energy coincides with the lowest possible energy for states in which the total two-dot occupation $N_{\rm tot} = (n_1 + n_2)$ is even. The solid odd-$N_{\rm tot}$ parabola gives the lowest-energy curve for $N_{\rm tot} = 1$ when there is no interdot coupling. The dotted parabola is the shifted-down energy curve that results when there is nonzero coupling. The relevant degeneracy points are indicated by a black dot for zero coupling and white dots for nonzero coupling. (b) Zero-coupling conductance through the double dot as a function of the gate voltage. The conductance peak shapes are schematic. The peaks occur at the points of degeneracy of the lowest-energy zero-coupling parabolas. (c) Nonzero-coupling conductance through the double dot as a function of the gate voltage. The peaks are aligned with the perturbed degeneracy points. Each zero-coupling peak has split into two separate peaks equally distant from the zero-coupling peak position. Increasing the interdot coupling increases the separation between the paired peaks until the full set of peaks is again regularly distributed, with half the original period.

about ϕ_m, with the magnitude of the *peak splitting*—the distance between the paired subpeaks—being linear in Δ.

If the odd parabolas shift down by exactly $\Delta = U/4$, their minima then lie along the line $E = 0$, and the subpeaks are distributed evenly at the positions $\phi_m^s = (2m + 1)/4$, where m is an integer. The conductance peak pattern is identical to the pattern characteristic of a single dot with total capacitance $2C_{\Sigma i}$ and dot-to-gate capacitance $2C_{gi}$. Accordingly, in terms of the phenomenological parameter Δ, the orthodox model provides a framework for understanding the evolution of the Coulomb blockade of two isolated dots into that of a single composite dot that is essentially the sum of the original individual dots.

1.2.8 Center-of-Mass Coordinates

Another way of using the orthodox model to understand the transformation of the two-dot blockade is to express the energy (recall Eq. 1.8) in the charging analog of *center-of-mass* coordinates—i.e., to express E in terms of the total double-dot occupation $N_{\rm tot} = (n_1 + n_2)$ and half the difference between the individual dot occupations $n = (n_2 - n_1)/2$. One then has [51, 52]

$$E = \frac{U_\Phi}{2}(N_{\rm tot} - \Phi_{\rm tot})^2 + U_\rho(n - \rho/2)^2\,, \tag{1.10}$$

where $\Phi_{\rm tot} = \phi_1 + \phi_2$ and $\rho = \phi_2 - \phi_1$. For the circuit of Eq. 1.9, in which the interdot capacitance is assumed to be zero, $2U_\Phi = U_\rho = U$, where U is the charging energy of the individual dots. In general, however, the interdot capacitance is not zero, and one then finds that $U_\Phi \neq U_\rho$. In particular, it turns out that $U_\Phi = e^2/2C_\Sigma$ and $U_\rho = e^2/(C_\Sigma + 2C_{\rm int})$, where $C_{\rm int}$ is the interdot capacitance and C_Σ is now defined to be the total capacitance of one of the dots minus the interdot capacitance [49].

From comparison with Eq. 1.4, it is clear that the first term on the right side of Eq. 1.10 is the energy for a single dot with a capacitance equal to the sum of the capacitances of the two individual dots. Consequently, transformation of the two-dot

system to a composite-dot system must come through cancellation or elimination of the term proportional to U_ρ. In situations where $\rho = 0$, this corresponds to shifting the lowest-energy odd parabolas down by $\Delta = U_\rho/4$ relative to the lowest-energy even parabolas.

1.2.9 Splitting through Capacitive Coupling

We have yet to suggest a physical mechanism for inducing such a shift. Perhaps the most obvious means of doing this is to send U_ρ itself to zero—or at least to some value on the order of the level spacing δ. This can be done by making the interdot capacitance C_{int} very large compared to C_Σ. Since C_{int} tends to be less than $0.1C_\Sigma$ when the dots are well isolated, sending U_ρ to zero requires that C_{int} grow by a factor on the order of 100. Such growth in C_{int} is difficult to achieve in a two-dimensional geometry. When coplanar regions of charge are moved closer together, the growth in their intercapacitance is at most logarithmic and therefore generally cut off before it can multiply by 100 [50]. Thus, though interdot capacitive coupling is expected to make a small and relatively constant contribution to the peak splitting, we must look for another means of ridding ourselves of the U_ρ-term. A series of recent coupled-dot experiments has done much to point us in the right direction.

1.3 Experimental Results for the Symmetric Double Dot

1.3.1 Measurements of Peak Splitting and Conductance

The double-dot experiments performed by F. Waugh, C. Crouch, C. Livermore, R. Westervelt, and their collaborators [43, 44, 45, 46] indicate that a primary mechanism for the emergence of composite-dot behavior is the fluctuation of charge between the dots when interdot tunneling channels are opened. Waugh *et al.* use the surface-gate pattern shown in Fig. 1.4(b) to create a system of symmetric dots that fits the characteristics of the two-dot system of Sec. 1.2. They then measure the conductance through the dot as a function of the side gate voltage V_g—i.e., as a function of $\phi = C_g V_g/e$—and repeat this measurement for different values of the voltage V_b, which controls the strength of the interdot barrier. They find that as the voltage V_b is made less negative—i.e., as the interdot barrier is lowered—the conductance peak distribution changes from that characteristic of two isolated dots to that characteristic of a single composite dot.

In addition to observing peak splitting and convergence to single-dot behavior, Waugh *et al.* measure the conductance G_b through the interdot barrier as a function of V_b. They do this by measuring the conductance through the double dot after exterior gates such as V_{x1} and V_{x2} have been de-energized. Decreasing the magnitudes of these negative gate potentials lowers the barriers that separate the dots from the bulk leads. After sufficient reduction of the exterior barriers, the charged regions formerly described as dots are essentially just peninsular extensions of the leads. There are no longer any capacitive charging energies exclusively associated with these regions. The capacitances associated with the bulk leads are, needless to say, very large, and the corresponding charging energies are therefore much less than the thermal energy $k_B T$. The Coulomb blockade has been eliminated.

Once the external gates have been turned off, the dominant resistance in going from lead 1 to lead 2 is the resistance due to the interdot barrier. Thus, leaving room

for correction due to the fact that lowering the external barriers results in a partial lowering of the interdot barrier [19, 22, 43, 44, 45, 46], one can find the conductance G_b by measuring the conductance between leads 1 and 2.

The experiments find that G_b rises in a series of *conductance steps*, each of height $2e^2/h$. This phenomenon of conductance quantization, familiar from studies of quantum point contacts and quantum wires [2, 3, 4, 27, 28, 29, 30, 31, 32], is explained by the fact that, as V_b is made less negative, the saddle-shaped potential between the dots becomes broader and flatter. The alteration in shape and height is accompanied by a decrease in the minimum energy $E_m^{\rm tr}$ of an electron in the mth lowest-energy transverse mode of the saddle region—where the *transverse* direction is the direction perpendicular to the direction of propagation through the constriction (the y-direction in Fig. 1.9). As $E_m^{\rm tr}$ descends through the Fermi energy $E_{\rm F}$, the conductance through the mth mode rises rapidly from nearly 0 to its saturation value of $2e^2/h$, giving one *conductance quantum* for each of the two spin-degenerate channels that have been opened. Thus, the $2e^2/h$ steps in the conductance G_b mark the successive opening of transverse orbital modes through the saddle.

The key finding of Waugh *et al.* is that, within the range of values of V_b in which G_b goes from nearly 0 to approximately $2e^2/h$, the Coulomb-blockade peak distribution goes from that characteristic of two isolated dots to that characteristic of a single composite dot. Based on our understanding of conductance quantization, it appears that the convergence to composite-dot behavior coincides with the opening of a single orbital mode through the saddle between the dots.

1.3.2 Arguments for Conductance-Controlled Peak Splitting

This correlation between the opening of one orbital mode and the emergence of composite-dot behavior suggests that the correct physical mechanism for the transformation of the Coulomb blockade is the back-and-forth movement of charge that occurs when an inter-

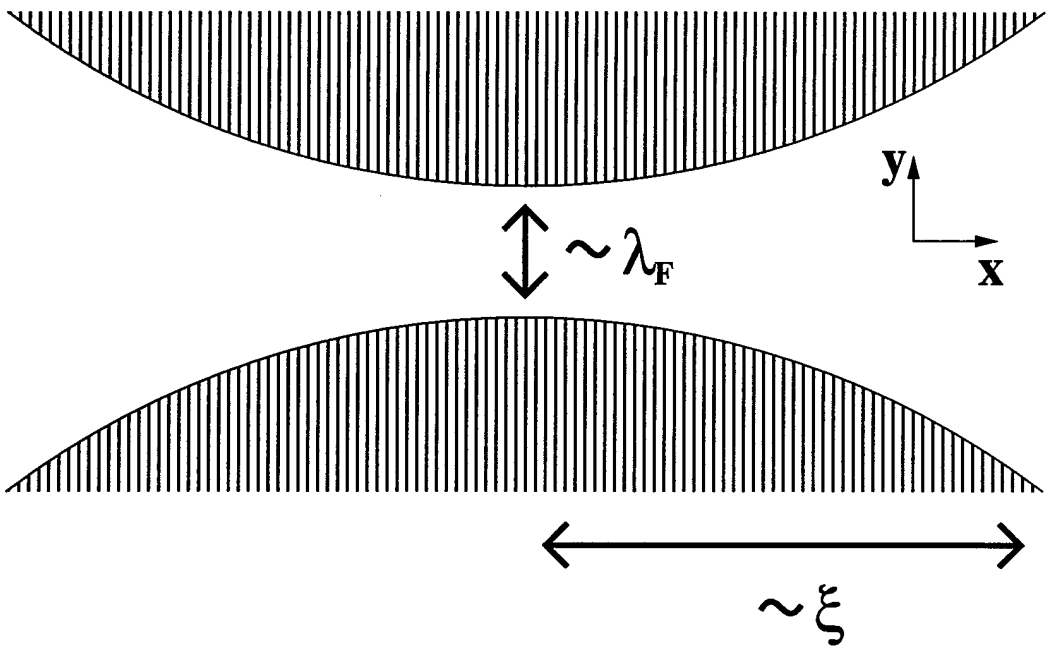

Figure 1.9: Top-down view of the constricted interdot connecting region. The width of the constriction is on the order of the Fermi wavelength λ_F. Its length is on the order of the device resolution ξ.

dot tunneling channel is opened. As the interdot conductance becomes of order e^2/h, fluctuations in the charge on the dots make the assumption of quantized integer occupation numbers n_1 and n_2 increasingly less valid. There is less and less reason to suppose that odd parabolas should be shifted upwards relative to even ones because the expectation value of $\hat{n}$, which makes the determinative distinction between the odd-$N_{\rm tot}$ and even-$N_{\rm tot}$ cases, is no longer a good quantum number.

The correlation between the opening of a single orbital mode and the transformation of the Coulomb blockade renders even more implausible explanations for the development of the composite-dot blockade via growth in the interdot capacitance $C_{\rm int}$. There is indeed some increase in $C_{\rm int}$ as the barrier between the dots is lowered and electrons on the dots become less separated. However, as the electrons generally only become closer in the vicinity of the narrow saddle, the interdot capacitance is not expected to grow significantly. Semiclassical simulations of two-dot geometries by M. Stopa confirm

that opening a single orbital mode results in only negligible growth in the corresponding intercapacitance [41].

Thus, it is reasonable to suspect that the principal physical cause of the observed peak splitting is the increase in the interdot conductance as the barrier between the dots is lowered. Recalling Eq. 1.10, we see that, for this to be true, the existence of a nonzero interdot conductance G_b must lead to a ρ-dependent energy shift that cancels the ρ-dependent term of Eq. 1.10.

1.3.3 Dimensionless Formulation of the Problem

It is worth taking a little time to formulate this last statement in a more mathematical manner. Let us define the *dimensionless interdot channel conductance* g by the formula $g = G_b/(N_{\rm ch}e^2/h)$, where, in the experiments under consideration, the number of tunneling channels $N_{\rm ch}$ equals two. Let us similarly define a dimensionless parameter f, the *fractional peak splitting*, which equals the ratio between the magnitude of the peak splitting at a given value of g and its saturation magnitude for $g = 1$. From Sec. 1.2, it follows that $f = \Delta/(U/4)$, where Δ is the coupling-induced shift of the odd-$N_{\rm tot}$ ground states relative to the even-$N_{\rm tot}$ ground states. It then follows that our goal is to develop physically meaningful models that allow determination of f as a function of both g and the number of tunneling channels $N_{\rm ch}$. We can then compare the result

$$f(N_{\rm ch}, g)$$

with the experimental data (see Fig. 1.10) and, if the comparison is favorable, gain some confidence in the credibility of explaining the peak splitting in terms of the interdot conductance.

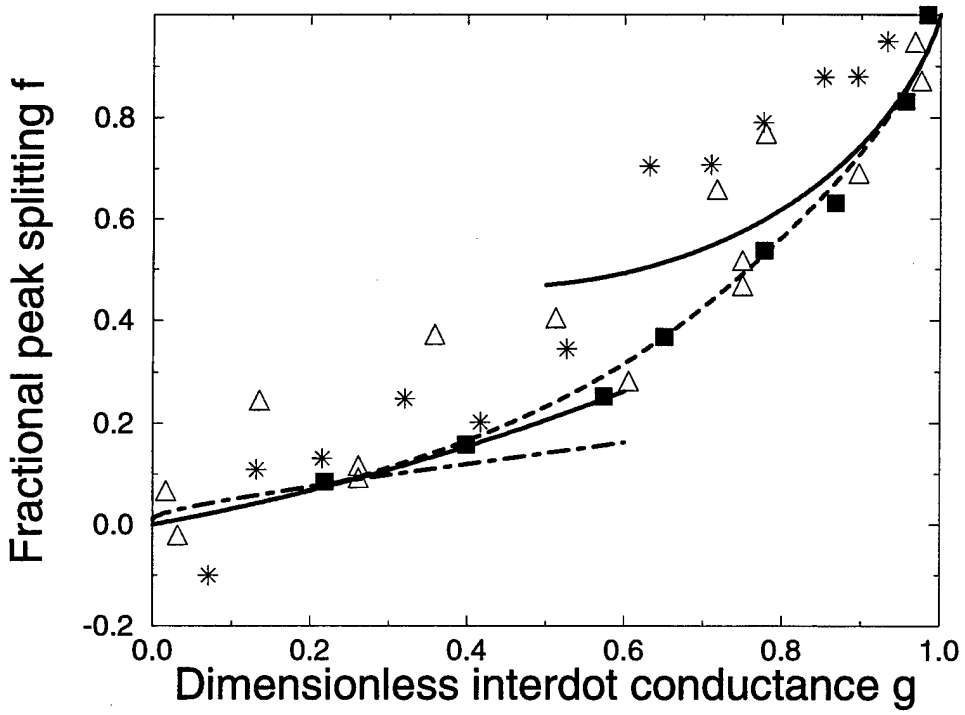

Figure 1.10: Experimental and theoretical results for the fractional peak splitting f as a function of the dimensionless interdot conductance g when $N_{\rm ch} = 2$. The solid lines are the results in the weak and strong-coupling limits when the tunneling amplitudes are assumed to be energy-independent. The small-dashed curve that extends from $(g, f) = (0, 0)$ to $(g, f) = (1, 1)$ is an interpolating curve that conforms with these results. The dot-dashed line is the $2\pi U/\hbar\omega = 1$ curve derived when the energy dependence of the tunneling amplitudes is taken into account. The stars, triangles, and squares symbolize different sets of experimental data from Waugh *et al.* [43, 44] and Livermore *et al.* [46], the squares being the most recent.

1.4 Theoretical Results for Weakly Coupled Dots

1.4.1 Transfer-Hamiltonian Model

In the limit of weakly coupled dots—i.e., for $g \ll 1$—we can use a transfer-Hamiltonian approach for an arbitrary number $N_{\rm ch}$ of interdot tunneling channels to find the relation between the dimensionless interdot conductance g and the fractional peak splitting f [51, 52, 53, 54]. In this approach, the Hamiltonian consists of three parts: the single-particle kinetic energies K, the capacitive charging energies $V_{\rm tot}$, and the hopping term H_T. For simplicity, the tunneling amplitudes between the dots are assumed independent of the particular tunneling channel, as is trivially the case for the two $SU(2)$-symmetric channels involved in the experiments described in Sec. 1.3. The three parts of the Hamiltonian then take the form

$$\begin{aligned} K &= \sum_{i=1}^{2} \sum_{k,\sigma} \epsilon_{ik\sigma} \hat{n}_{ik\sigma} \,, \\ V_{\rm tot} &= \frac{U_\Phi}{2} (\hat{N}_{\rm tot} - \Phi_{\rm tot})^2 + U_\rho (\hat{n} - \rho/2)^2 \,, \\ H_T &= \sum_{k_1,k_2,\sigma} t_{k_1 k_2} \left(c^\dagger_{2k_2\sigma} c_{1k_1\sigma} + \text{H.c.} \right) \,, \end{aligned} \tag{1.11}$$

where i is the dot index, σ is the channel index (which is equivalent to a spin index in recent $N_{\rm ch} = 2$ experiments), and k is a kinetic index which accounts for all electron degrees of freedom not included in the dot or channel indices. The number operators $\hat{N}_{\rm tot} = (\hat{n}_1 + \hat{n}_2)$ and $\hat{n} = (\hat{n}_2 - \hat{n}_1)/2$ are quantized versions of the occupation numbers $N_{\rm tot}$ and n. The single-particle energies $\epsilon_{ik\sigma}$ are assumed to be distributed relatively evenly with an average level spacing δ in an energy band $[\epsilon_0, \epsilon_0 + D]$. For zero temperature and $H_T = 0$, the single-particle states on either dot are presumed to be occupied up to the energy $\epsilon_0 + FD$, where F is the *filling fraction* of the single-particle bands. The bandwidth D is assumed to be at least of the order of the Fermi energy $E_{\rm F}$; consequently, $D \gg U_\Phi, U_\rho \gg \delta$.

Calculation of the interdot conductance g and the fractional peak splitting f is done

via perturbation in H_T and in the continuum limit $\delta \ll U_\Phi, U_\rho$. Since coupling to the external leads is neglected, the term in $V_{\rm tot}$ proportional to U_Φ is simply constant and can be dropped, leaving us with the differential charging energy [51]:

$$V = U_\rho(\hat{n} - \rho/2)^2 . \tag{1.12}$$

To make calculation of g and f manageable, we will assume that $t_{k_1 k_2}$ can be treated as independent of k_1 and k_2. The resulting model with $t_{k_1 k_2} = t$ is roughly equivalent to a system of two planar lattices linked by a hopping element T between their origins (see Fig. 1.11)—i.e., to two lattices connected by the hopping term

$$H_T = \sum_\sigma T(c^\dagger_{20_2\sigma} c_{10_1\sigma} + \text{H.c.}) , \tag{1.13}$$

where 0_i is the origin of lattice i [51] and $T = N_{\rm lt} t$, with $N_{\rm lt} \simeq D/\delta$ being the number of sites on the lattice [51].

The assumption $t_{k_1 k_2} = t$ is valid under two conditions. The first is that the charging energy U_ρ is much greater than the level spacing δ. The fact that $U_\rho \gg \delta$ does more than permit us to calculate summations over single-particle states via a continuum approximation. It implies that the positions of the Coulomb-blockade conductance peaks are not sensitive to the fluctuations in the tunneling probabilities proportional to $|t_{k_1 k_2}|^2$. Instead, the Coulomb blockade is determined by the average behavior of a nonzero number of tunneling probabilities—i.e., by averages of $|t_{k_1 k_2}|^2$ over a finite number (presumably some fraction of U/δ) of neighboring combinations of k_1 and k_2.

The second condition for taking $t_{k_1 k_2} = t$ is that the device resolution ξ is small enough to ensure that the transmission energy scale W exceeds U_ρ. If this condition fails to hold, the averages of the tunneling probabilities will themselves change significantly over the energy range U_ρ.

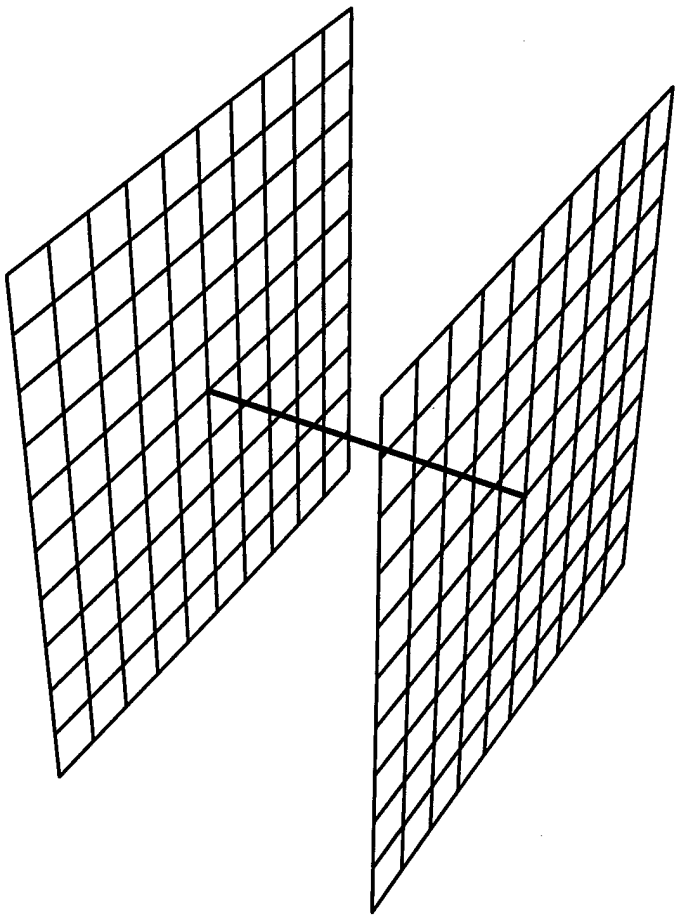

Figure 1.11: Two lattices coupled via a hopping channel between their origins.

1.4.2 Calculation of the Conductance and Peak Splitting

Having emphasized the stipulation that $\delta \ll U_\rho \lesssim W$, we can proceed to the results for the dimensionless channel conductance g and the fractional peak splitting f. As discussed in Sec. 1.3.1, the experimental measurement of the channel conductance g corresponds to setting $U_\rho = 0$ and finding the net flow of electrons from dot 1 to dot 2 when the dots are held at different chemical potentials. One might worry about the fact that de-energizing the gates also alters the level spacing δ. However, this does not affect the result, as can be seen by referring to the coupled-lattice interpretation of our model (recall Eq. 1.13 and Fig. 1.11). In this picture, de-energizing the gates corresponds to increasing the number of sites in the lattices ($N_{\rm lt} \to \tilde{N}_{\rm lt}$) while keeping T constant. It follows that this lattice expansion results in both the level spacing δ and the tunneling amplitude t being multiplied by the factor $N_{\rm lt}/\tilde{N}_{\rm lt}$. Since g is a function only of the ratio t/δ, the separate changes in t and δ cancel and can be ignored.

The result for g, which is calculated exactly to all orders in perturbation theory, is

$$g = \frac{4\chi}{|1 + (1 + i\eta)^2\chi|^2}\,, \tag{1.14}$$

where $\chi = |\pi t/\delta|^2$ and $\eta = (1/\pi)\ln[F/(1-F)]$ [51]. It is remarkable that, for all values of F, the maximal dimensionless conductance is 1. Still, the equation for g indicates that the weak-coupling model is not useful when t is on the order of or much greater than δ since, in this regime, the interdot conductance in the hopping model is squeezed off by the formation of a bound state at the points where the lattices are linked.

The fractional peak splitting f is computed by solving for the quantity $\tilde{f}_\rho$, which has the definition

$$\tilde{f}_\rho = \frac{\left[E^0_{\rm grd}(\rho) - E^0_{\rm grd}(0)\right] - \left[E^g_{\rm grd}(\rho) - E^g_{\rm grd}(0)\right]}{U/4}\,, \tag{1.15}$$

where $E^g_{\rm grd}(\rho)$ is the ground-state energy of the double-dot system for the given values of g and ρ and for an even total double-dot occupation $N_{\rm tot}$. In terms of $\Delta_T(\rho)$, the

ground-state energy shift induced by H_T,

$$\tilde{f}_\rho = \frac{\Delta_T(0) - \Delta_T(\rho)}{U/4}. \tag{1.16}$$

After determining that $\Delta_T(1)$ is the same, except for corrections of order δ/D, as the ground-state shift of the system for $N_{\rm tot}$ odd and $\rho = 0$ (see Chapter 2), one arrives at an equation for the fractional peak splitting:

$$f = \lim_{\rho \to 1} \tilde{f}_\rho \,. \tag{1.17}$$

1.4.3 Peak Splitting Expanded in the Conductance

Calculating via Rayleigh-Schrödinger perturbation theory to leading order in δ/D, one finds that $\tilde{f}_\rho$ can be expressed as a power series in $\chi = (\pi t/\delta)^2$. In Chapters 2 and 3, the terms in this power series are computed up to order χ^2. After inverting Eq. 1.14, one obtains an expansion of $\tilde{f}_\rho$ in powers of g [52]:

$$\tilde{f}_\rho = \sum_{n=1}^{\infty} \sum_{m=1}^{n} a_{mn}(\rho)\, (N_{\rm ch})^m\, g^n \,. \tag{1.18}$$

From Chapters 2 and 3, we know the coefficients $a_{mn}(\rho)$ for $n \leq 2$. The resulting equation for the fractional peak splitting is

$$f \simeq 0.1404 N_{\rm ch}\, g + 0.1491 N_{\rm ch}\, g^2 - 0.009798 (N_{\rm ch})^2\, g^2 + \ldots \tag{1.19}$$

The weak-coupling result for $N_{\rm ch} = 2$ is plotted in Fig. 1.10 and is seen to give good agreement with the experimental data. The key qualitative aspects of the f-versus-g curve for such small values of $N_{\rm ch}$ are that, for $g \simeq 0$, it has a relatively small slope and an upward curvature. For $N_{\rm ch} \gtrsim 10$, the f-versus-g curve looks significantly different, possessing a large upward slope and a downward curvature.

1.5 Theoretical Results for Strongly Coupled Dots

1.5.1 Effective One-Dimensionality of the Planar Dots

Though the weak-coupling results give a good qualitative sense of the f-versus-g curve, they are only expected to be quantitatively accurate for $g \ll 1$. To understand the behavior at and about the *strong-coupling* endpoint $(g, f) = (1, 1)$, we will need a different model that is independent of ρ for $g = 1$ and perturbative in $(1 - g)$ instead of g.

If the saddle region between the two dots is approximately adiabatic—i.e., if there is only negligible scattering between the transverse modes within the saddle—such a strong-coupling model can be obtained through the observation that the two-dot problem is effectively one-dimensional [33, 34]. If we presume that the entire double-dot system were adiabatic, with a well-defined separation of the transverse and longitudinal motions throughout, it follows immediately that the two-dot system consists simply of a number N_{ch} of separate one-dimensional tunneling channels that interact only through the capacitive charging energies of the dots. Electrons in transverse modes that do not penetrate the barrier region can be ignored as they simply sit inertly on one dot or the other.

What might be a bit surprising is that more realistic systems, in which adiabaticity does not hold outside the saddle, are still effectively one-dimensional. In such systems, the average energy spacing in the one-dimensional channels is the same as that of the two-dimensional dots, but the probability of an electron actually being in the saddle is weighted by the proportion of overlap between the electronic wavefunction and the relevant transverse mode. The combination of this *wavefunction weighting* with the two-dimensional level spacing yields an average overall normalization that is indistinguishable from a fully adiabatic system with a one-dimensional level spacing in the channels [26]. Of course, the wavefunction weights vary randomly for individual dot

eigenstates, and it is important to have $U_\rho \gg \delta$ so that the weights are averaged over a large number of single-particle eigenstates.

1.5.2 1D Fermionic Model and the Interdot Conductance

We can now proceed with a one-dimensional model for the coupled-dot system. Since we are currently concerned with the strong-coupling ($g \to 1$) limit of such a model, the particles in our one-dimensional system propagate nearly freely between the two dots. As a result, in the connecting region itself, we are primarily confronted with particles that can be characterized as mostly *right movers* or mostly *left movers*—i.e., particles that can be characterized as essentially going from dot 1 to dot 2, with a small amount of backscattering, or from dot 2 to dot 1, again with only little backscattering. Instead of indexing particles by the dots with which they are associated, we therefore index them by the directionality index j, where $j = 1$ for right movers and $j = 2$ for left movers. The channel index σ and the supplementary index k that were used in the weak-coupling limit remain serviceable.

The one-dimensional Hamiltonian is the following:

$$\begin{aligned} K &= \sum_{j=1}^{2} \sum_{k,\sigma} \xi_{k\sigma} \hat{n}_{jk\sigma} , \\ V &= U_\rho (\hat{n} - \rho/2)^2 , \\ H_B &= \sum_{k_1,k_2,\sigma} v_{k_1 k_2} (c^\dagger_{2k_2\sigma} c_{1k_1\sigma} + \text{H.c.}) , \end{aligned} \tag{1.20}$$

where $\xi_{k\sigma}$ is the kinetic energy of a particle in the σth channel with kinetic index k. The Hamiltonian once again consists of three parts, but the perturbative term represents backscattering from the interdot barrier rather than tunneling through it.

Just as we assumed that $t_{k_1 k_2}$ could be considered constant in the weak-coupling limit, we now assume that we can replace $v_{k_1 k_2}$ by the constant v, the assumption again being that $W \ll U_\rho$. The backscattering term H_B then represents the backscattering from a delta-function potential $V_0 \delta(x)$, and the relation between the magnitude of this

potential and the conductance through the barrier is given by [51, 52]

$$(1-g) = \tilde{V}^2 + O[\tilde{V}^4], \tag{1.21}$$

where $\tilde{V} = V_0/\hbar v_F$, v_F being the Fermi velocity.

1.5.3 Bosonized Euclidean Action and Strong-Coupling Endpoint

In order to find the ρ-dependence of the ground-state energy, we will use the *bosonization* formalism, in which the low-energy degrees of freedom of a one-dimensional system of fermions are expressed in terms of bosonic density and phase operators [55, 56, 57, 58, 59, 60, 61, 62, 63]. The bosonized Euclidean action consists of the following parts [33, 34, 51, 52]:

$$\begin{aligned} S_0 &= \frac{1}{\beta}\sum_{\omega_m,\sigma} |\omega_m||\tilde{\theta}_\sigma(\omega_m)|^2, \\ S_{int} &= U_\rho \int_0^\beta d\tau \left(\frac{1}{\sqrt{\pi}}\left[\sum_\sigma \theta_\sigma(\tau)\right] - \rho/2\right)^2, \\ S_B &= \frac{\tilde{V}D}{2\pi}\sum_\sigma \int_0^\beta d\tau \cos\left[2\sqrt{\pi}\theta_\sigma(\tau)\right], \end{aligned} \tag{1.22}$$

where D is the bandwidth, β is the inverse temperature ($\beta = 1/k_B T$), and the energies ω_m are $\hbar$ times the Matsubara frequencies ($\omega_m = 2\pi m/\beta$ with m an integer). The bosonic field $\theta_\sigma(\tau)$ measures the number of electrons in the σ channel that lie to one side of the barrier. $\tilde{\theta}(\omega_m)$ is its Fourier transform:

$$\theta_\sigma(\tau) = \frac{1}{\beta}\sum_{\omega_m} e^{-i\omega_m \tau}\,\tilde{\theta}_\sigma(\omega_m). \tag{1.23}$$

The first step in handling the action of Eq. 1.22 is to shift all the $\theta_\sigma(\tau)$ fields by a constant: $\theta_\sigma(\tau) \to \theta_\sigma(\tau) + \rho/2N_{ch}$. This moves all the ρ-dependence of the action from the capacitive charging term V to the backscattering term H_B. Hence, it makes clear that the action is independent of ρ when the backscattering parameter $\tilde{V}$ is zero—i.e., when $g = 1$. The bosonized action therefore yields the desired endpoint for the f-versus-g curve: $(g, f) = (1, 1)$.

1.5.4 Peak Splitting near the Strong-Coupling Endpoint

What about the behavior in the vicinity of the $g = 1$ endpoint? To find this we solve for the ground-state energy via perturbation in S_B. Once again, we look to find the quantity $\tilde{f}_\rho$, which is now formulated in terms of the ground-state energy shifts $\Delta_B(\rho)$ induced by backscattering between the dots:

$$\tilde{f}_\rho = \rho^2 - \frac{\Delta_B(\rho) - \Delta_B(0)}{U/4}. \tag{1.24}$$

It is clear from Eq. 1.24 that the fractional peak splitting f equals 1 minus nontrivial terms that go to zero when $(1 - g)$ goes to zero. However, in contrast to the weak-coupling limit, in which we are able to solve for the leading terms in g for any number of tunneling channels, in the strong-coupling limit we can—at the present time—only use the bosonized Euclidean action to compute the leading nontrivial terms for $N_{\rm ch} = 1$ and $N_{\rm ch} = 2$. For $N_{\rm ch} > 2$, there are qualitative arguments that the leading nontrivial dependence goes as $(1 - g)^{N_{\rm ch}/2}$ [33, 52], but, as we shall see shortly, in the case of $N_{\rm ch} = 2$ the leading nontrivial term goes as $(1-g)\ln(1-g)$, and therefore the $(1-g)^{N_{\rm ch}/2}$ proposition is at the very least susceptible to logarithmic corrections.

In any case, the leading nontrivial terms for $N_{\rm ch} = 1$ and $N_{\rm ch} = 2$ are derived in Chapters 2 and 3 of this thesis. For $N_{\rm ch} = 1$, the result is [51, 53, 54]

$$f \simeq 1 - \frac{8e^\gamma}{\pi^2}\sqrt{1-g}, \tag{1.25}$$

where $\gamma \simeq 0.577$ is the Euler-Mascheroni constant. For the experimentally relevant case of $N_{\rm ch} = 2$, the result is [52]

$$f \simeq 1 + \frac{16e^\gamma}{\pi^3}(1-g)\ln(1-g) - 0.425(1-g), \tag{1.26}$$

which is plotted in Fig. 1.10. The $N_{\rm ch} = 1$ and $N_{\rm ch} = 2$ results show an even more dramatic dependence on $N_{\rm ch}$ than the results for weak coupling. Also notable is that, whereas the weak-coupling calculation leads to an analytic expansion in powers of g, the strong-coupling calculation leads to terms that are singular at $(1 - g) = 0$, an outcome

that is not entirely shocking given that the infinite $g = 1$ one-dimensional system is translationally invariant but the infinite $g < 1$ system is not.

1.5.5 Concerning Coefficients

Readers possessing some familiarity with the perils of bosonization might worry about whether the coefficients in Eqs. 1.25 and 1.26 are well-determined [51]. Bosonization is most often used to derive scaling relations for which such coefficients are irrelevant. Moreover, closer inspection of the coefficients reveals that each one consists of at least two factors that separately depend upon the manner in which the high-energy bosonic degrees of freedom are cut off. Fortunately, as shown in Chapter 3, if corrections of order U_ρ/D are neglected, the coefficients appear to be independent of the nature of the ultraviolet cutoff [52]. Hence, it is reasonable to suppose that they are indeed well-defined.

1.5.6 Connection to the Weak-Coupling Results

With the leading terms of f-versus-g known for both the weak and strong-coupling limits, the curves that interpolate between these limits are relatively tightly constrained. A plausible interpolating curve that analytically approaches the limiting behaviors for $N_{\rm ch} = 2$ is displayed as a dashed line in Fig. 1.10. The curve shows remarkably good agreement with the most recent experimental data [46].

1.6 Finite-Barrier Model for Weakly Coupled Dots

1.6.1 Why Bother?

In the preceding sections, it has been shown that credible results are obtained by working in what might be characterized as the *universal* limit $U_\rho/W = 0$ in which f can be expressed in terms of $N_{\rm ch}$, g, and various universal coefficients. Having achieved so much by assuming constant tunneling and backscattering amplitudes, one might wonder what could motivate us to construct a model in which the amplitudes are variable. The first reason for considering a realistic $U_\rho/W \neq 0$ system is that we want more than a hand-waving justification for the approximation $U_\rho/W = 0$. Second, we want a more precise idea of when this approximation is valid. Third, we would like to possess a better understanding of the behavior of systems in which ξ is considerably larger than $\lambda_{\rm F}$ and for which it is presumably not reasonable to take $U_\rho/W = 0$. Finally, we hope that, with improved experimental precision, the non-universal corrections that result from $U_\rho/W \neq 0$ will become detectable even in systems where ξ is nominally small.

1.6.2 Semi-Localized Basis and Parabolic Barrier

In looking for these corrections, we consider the weak-coupling limit only. As in our earlier strong-coupling analysis, a crucial trick comes from the fact that the coupled-dot system is effectively one-dimensional. A second trick consists of choosing a particular basis of states that is formed from simple linear combinations of the exact single-particle eigenstates of the full double dot. The states in this basis are *semi-localized*: they have some tendency to be concentrated on one dot or the other, particularly at low energies, but they extend throughout the double dot. As a result, the operator $\hat{n}$ is not diagonal in this basis, and the perturbative quantity is $\hat{n} - \hat{n}_0$, where $\hat{n}_0$ is the number operator that counts the difference between the occupation of semi-localized states preferentially associated with dot 2 and that of semi-localized states preferentially associated with

dot 1.

To obtain a quantitative result with such a semi-localized basis, one must assume some concrete form for the self-consistent barrier between the dots. Given the relatively smooth nature of the electrostatic potentials, it is reasonable to model the barrier as parabolic, being given by the equation $V(x) = V_{\max}(1 - x^2/2\xi^2)$ for $|x| \leq \sqrt{2}\xi$, where ξ is the device resolution. The barrier is then an inverted parabolic well with an associated harmonic oscillator frequency $\omega = \sqrt{V_{\max}/m\xi^2}$. The differential conductance through such a barrier has been known since at least the mid-1930s:

$$g = \frac{1}{1 + e^{-2\pi(E_{\mathrm{F}} - V_{\max})/\hbar\omega}}. \tag{1.27}$$

It follows that the characteristic transmission energy scale is given by

$$W = \frac{\hbar\omega}{2\pi}. \tag{1.28}$$

For nontrivial values of g (i.e., for g a reasonably sized, finite distance from both 0 and 1), we must have $E_{\mathrm{F}} \simeq V_{\max}$. We then conclude that $W \simeq \hbar v_{\mathrm{F}}/2\pi\sqrt{2}\xi$ [26]. As expected, W depends inversely on the characteristic length ξ.

1.6.3 Peak Splitting as a Non-Universal Function of Conductance

Calculation of the fractional peak splitting f is, as usual, a bit more difficult than calculation of g. In Chapter 4, we see that, for $U_\rho/W \neq 0$ and $g \ll 1$, f is raised above its $U_\rho/W = 0$ value by an amount that is, at leading order, proportional to $U_\rho/W = 2\pi U_\rho/\hbar\omega$ [26]. In the extreme $g \to 0$ limit, the dependence of f on g is expected to be quite different from that for $U_\rho/W = 0$: the curve is concave and goes as $(2\pi U_\rho/\hbar\omega)(1/|\ln g|)$. The explanation for such behavior is that, for very small g, the well-localized occupied states that lie at and below the Fermi energy couple only weakly to similarly well-localized states but couple comparatively strongly to the mostly transmitting states that lie above the barrier. The penalty for coupling to these states

is the relatively large energy difference between them, which, from Eq. 1.27, is of order $(\hbar\omega/2\pi)|\ln g|$.

Though we do not calculate the correction to the $2\pi U_\rho/\hbar\omega = 0$ behavior in the strong-coupling ($g \simeq 1$) limit, we can surmise that the drop in the transmission probability as one goes to lower and lower energies below the Fermi surface probably means that the $2\pi U_\rho/\hbar\omega \neq 0$ corrections will bring about a decrease in the fractional peak splitting relative to that for $2\pi U_\rho/\hbar\omega = 0$.

Figure 1.10 displays the small-g part of the $N_{\rm ch} = 2$ curve for $2\pi U_\rho/\hbar\omega = 1$, which is the value for $2\pi U_\rho/\hbar\omega$ that appears to correspond to recent experiments (see Chapter 4 for details). In this case, the $2\pi U_\rho/\hbar\omega \neq 0$ corrections are relatively small in magnitude. Thus, the interpolating curve does not have to be significantly modified to account for them, and the experiments can indeed be characterized as being, for the most part, in a large W or small ξ limit. For larger values of ξ, this need not be the case.

1.7 Summary and Preview

This chapter has introduced both the general physics of quantum dots and the specific physics of the Coulomb blockade. It has also provided an overview of the theoretical results and calculations of the chapters that follow. Section 1.1 delved into the particulars of the definition of quantum dots, describing in detail the *large planar quantum dots* with which we are primarily concerned. Section 1.2 reviewed the Coulomb blockade and the *orthodox model* that does much to explain it. Section 1.3 described recent experiments with symmetric coupled dots in which peak splitting was seen to be correlated with the conductance G_b through the interdot barrier. It also introduced the essential program of this thesis, which is to develop realistic quantum mechanical models that give the relation between the *fractional peak splitting* f and the *dimensionless channel conductance* g for various values of the number of tunneling channels, $N_{\rm ch}$. Sections 1.4-1.6 presented three such models for the regimes of weak coupling ($g \simeq 0$), strong coupling ($g \simeq 1$), and weak coupling in the presence of a finite barrier ($g \simeq 0$ with $U_\rho/W \neq 0$), respectively.

The punchline of this chapter is that, in order to describe the transformation of the Coulomb blockade of two isolated dots into the blockade of one large composite dot, one must move beyond the classical physics of the orthodox model and adopt an approach that takes account of quantum mechanical tunneling. In practice, one has to come up with at least two such approaches—one for weak coupling that is perturbative in g about $(g, f) = (0, 0)$ and one for strong coupling that is perturbative in $(1 - g)$ about $(g, f) = (1, 1)$. Sections 1.4 and 1.5 presented the transfer-Hamiltonian model for weak coupling and the one-dimensional backscattering model for strong coupling that form the backbone of Chapters 2 and 3 of this thesis. Chapter 2 gives the leading results derived from these models; subleading terms are calculated in Chapter 3.

In Chapter 4, the third model for the double dot is introduced. It goes beyond the search for the curve $f(N_{\rm ch}, g)$ by taking into consideration the *non-universal* quantity

U_ρ/W, which is the ratio of the charging energy U_ρ and the characteristic transmission energy scale W. It concludes that the corrections due to nonzero U_ρ/W are potentially measurable and significant but that, within the range of current experimental precision, the $N_{\rm ch} = 2$ interpolating curve determined by the universal $U_\rho/W = 0$ results is fundamentally serviceable over the bulk of the domain of g.

The reader should be warned that there exist small but possibly nettlesome differences between the notational schemes of the various chapters, each of which is essentially a self-contained essay. For example, Chapters 2 and 3 use U_1 and U_2 where this chapter uses $2U_\Phi$ and U_ρ. Chapter 4 uses U for U_ρ.

Chapter 2

Leading Results for the Coupled-Dot Blockade

2.1 Introduction

Turning on a tunnel junction between a bulk lead and a quantum dot leads to progressive destruction of the single-dot Coulomb blockade [8, 13, 14, 15]. Experiments by Waugh *et al.* [19, 43, 44] and Molenkamp, Flensberg, and Kemerink [48] chronicle this eradication for two tunnel-coupled dots of equal and widely disparate charging energies, respectively. Inspired by the experimental results of Waugh *et al.*, the present chapter seeks to develop a simple model for the coherent tunneling of electrons between a pair of electrostatically identical quantum dots [see Fig. 2.1(a)]. The goal is to describe the evolution of the Coulomb blockade from that of two isolated dots to that of one composite dot in terms of parameters that determine the states of the isolated dots and the nature of the connection between them. In the limits relevant to the experimental situation, we find that the most important dimensionless parameters are the number N_{ch} of conducting channels between the two dots and the dimensionless *interdot barrier conductance* g of each channel, which is measured when the Coulomb blockade has been removed.

(Waugh *et al.* [19, 43, 44] measure the conductance through the interdot barrier by de-energizing the external potentials V_{xi} that separate the dots from the leads. This conductance is to be distinguished from the conductance measured in the double-dot Coulomb blockade measurements, which might be referred to as the *Coulomb blockade conductance* or *double-dot conductance.*)

The problem of coupled quantum dots and more generally, of the effect of tunnel-couplings upon the Coulomb blockade has received much attention. Ruzin *et al.* [47] examined the Coulomb blockade structure of two non-identical dots in series via a standard activation-energy approach. Stafford and Das Sarma [64, 65] as well as Klimeck, Chen, and Datta [66] have applied Hubbard-like models with and without interdot capacitances to determine the many-body wavefunctions for tunnel-coupling between a small array of single-dot eigenstates. Many investigators have studied the effect of tunneling upon the Coulomb blockade for metallic junctions, in which there are a large number of conducting channels [67, 68, 69, 70, 71, 72]. Relatively few have considered junctions with only one or two channels [33, 73, 74]. Furthermore, the work on one or two-channel junctions has been restricted to consideration of a single dot coupled to bulk leads rather than systems of coupled dots. A significant finding of this chapter is that by introducing a "fictional" difference between the gate voltages on the individual dots, one can map the two-dot problem onto the one-dot problem and adapt previously obtained results for strong interdot coupling between a single dot and a bulk lead.

In Sec. 2.2, we present a brief review of the experimental results that have motivated our investigation. In Sec. 2.3.1, we define a tunneling model which is useful for calculations in the limit of weak coupling between the two dots. In Sec. 2.3.2, we show how a "center-of-mass transformation" allows one to map the two-dot problem onto the one-dot problem. Section 2.4 presents the weak-coupling results for our theory, and Section 2.5 gives the strong-coupling results and offers plots of the data and theory for one and two-channel junctions. A summary of our findings is provided in Sec. 2.6.

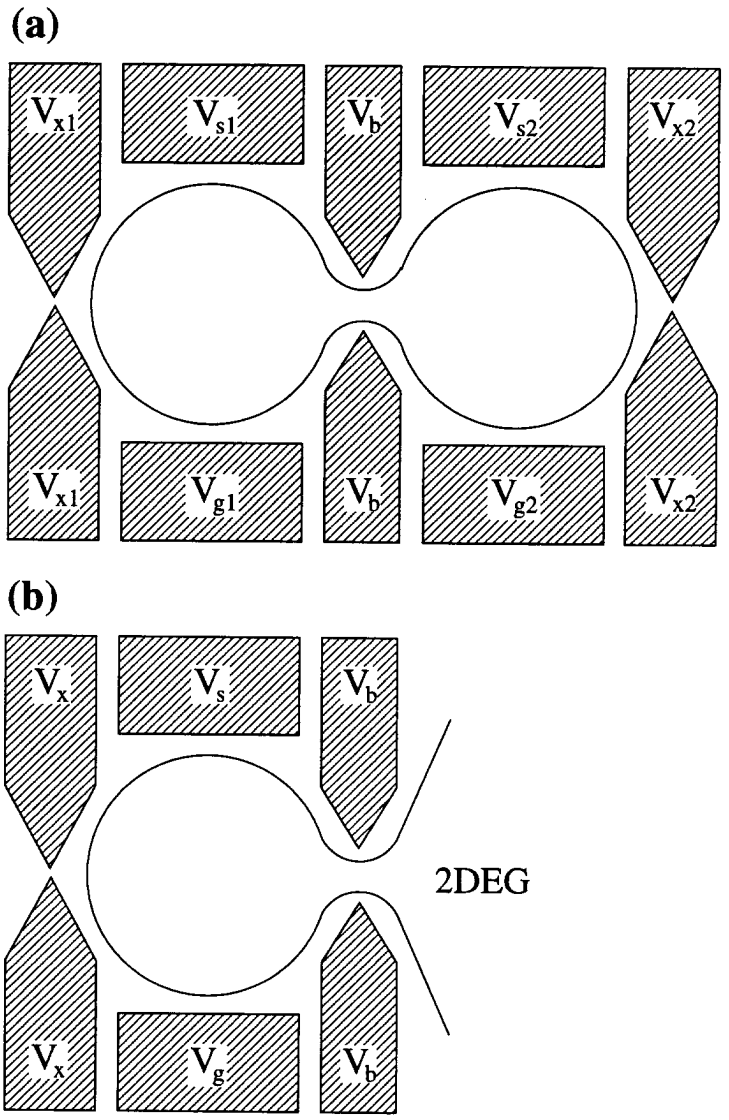

Figure 2.1: (a) Schematic diagram for the double dot. Negative potentials are applied to each of the gates to form the double-dot structure. The gate potentials V_{g1} and V_{g2} control the average numbers of electrons on the dots. These are the potentials that are varied to see the Coulomb blockade. V_b controls the rate of tunneling between the dots. V_{x1} and V_{x2} control the rate of tunneling to the adjacent bulk 2D electron gas (2DEG) leads. For calculations of the double-dot energy shifts, tunneling to the leads is assumed negligible compared to tunneling between the two dots. In measuring the barrier conductance G_b, however, the potentials V_{xi} are turned off so that each dot is strongly connected to its lead. The side-wall potentials V_{s1} and V_{s2} are fixed. (b) Schematic diagram for the single dot. V_b now controls tunneling between the dot and the bulk 2DEG. V_g determines the average number of electrons on the dot. For our purposes, V_s and V_x are constant, and tunneling to the bulk 2DEG through the barrier defined by V_x is negligible compared to tunneling through the barrier defined by V_b.

2.2 Motivation

The experiment of F. R. Waugh *et al.* [19, 43, 44] provides the primary motivation for this chapter. These authors study the effect that varying the interdot potential barriers has upon the Coulomb blockade conductance peak structure for arrays of n dots, where n equals 2 or 3. For their Coulomb blockade measurements, they energize the confining gates [V_{xi} in Fig. 2.1(a)] so that the conductance between the dots and the external leads is much less than $2e^2/h$. Having tuned the dots to be electrostatically identical—i.e., to have common gate and total dot capacitances C_g and C_Σ—they find that lowering the interdot barriers results in interpolation between the peak structure characteristic of the isolated individual dots and that characteristic of a single composite dot having capacitance nC_Σ: the initial isolated-dot peaks split into bunches of n sub-peaks, and the splitting within the bunched sub-peaks increases until they are essentially equally distributed with n-times the periodicity of the original peaks [see Figs. 2.2(b) and 2.2(c)]. For the double dot ($n = 2$), Waugh *et al.* also measure the conductance G_b of the barrier between the two dots after the exterior walls of the double dot have been removed and remark that plots of the sub-peak splitting and barrier conductance as functions of the barrier gate voltage appear substantially similar.

Waugh *et al.* use a $T = 0$ "capacitive charging model" to interpret their data. In this model, electrons on the dots are treated as charged particles with no kinetic energy that occupy each dot in integer amounts. In the absence of coupling, the energy is given by the sum of the potential energies of the individual dots. For two dots with common capacitances C_Σ and C_g, the expression for the energy has an especially simple form:

$$E = \frac{U}{2}\sum_{i=1}^{2}(n_i - \phi_i)^2, \tag{2.1}$$

where U is the charging energy for each individual dot, $U = e^2/C_\Sigma$; n_i is the number of electrons on the ith dot; and ϕ_i is the gate voltage parameter that determines the energy-minimizing value of n_i. For common gate voltages and gate-to-dot capacitances,

we have the relations $\phi_i = C_{gi}V_{gi}/e = C_gV_g/e = \phi$. Figure 2.1(a) should help put these parameters in context.

For each set of integer occupation numbers (n_1, n_2), the capacitive charging model with $\phi_i = \phi$ gives an energy $E_{(n_1,n_2)}$ that is a parabolic function of the common gate voltage parameter ϕ (see Fig. 2.2). All the parabolas are identical in shape, their only distinguishing features being the locations of their minima. The lowest-energy parabola $E_{N_{tot}}(\phi)$ for a given value of $N_{tot} = \sum_{i=1}^{2} n_i$ has $n_1 = n_2 = N_{tot}/2$ for N_{tot} even and $n_1 = n_2 \pm 1 = (N_{tot} \pm 1)/2$ for N_{tot} odd. In the former *even* case, the minima all lie on the line $E = 0$. In the latter *odd* case, the minima are displaced upward, sitting along $E = U/4$. For all parabolas, the ϕ-coordinate of the minimum is $N_{tot}/2$.

A prominent peak in the double-dot conductance occurs at values of ϕ such that the lowest-energy parabolas corresponding to consecutive values of N_{tot} cross—in other words, at values of ϕ for which $E_{N_{tot}}(\phi) = E_{N_{tot}+1}(\phi)$ for some integer N_{tot}. For the model of Eq. 2.1, this occurs whenever $\phi = m + 1/2$, where m is an integer. (One such crossing point is marked by the black dot in Fig. 2.2.)

In a model in which coupling between the dots is included, the lowest-energy parabolas for odd N_{tot} are shifted downward relative to the lowest-energy even-N_{tot} parabolas by an "interaction energy" E_{int}. This downward shift splits each of the initial crossing points into a pair of crossing points symmetric about the position of the initial degeneracy, from which they are separated by a distance proportional to E_{int}. As a result, each of the initial conductance peaks is similarly split into two sub-peaks with separation proportional to E_{int}. The sub-peak splitting reaches its saturation value when $E_{int} = U/4$—i.e., when the lowest-energy *even* and *odd* parabolas have the same minimum energy. At this point, the relevant crossing points occur for $\phi = \frac{1}{2}(m + \frac{1}{2})$. The corresponding conductance peaks are once again equally spaced, but their period is now that characteristic of a single dot with capacitance $2C_\Sigma$.

Thus, in the capacitive charging model, the problem of explaining the peak splitting

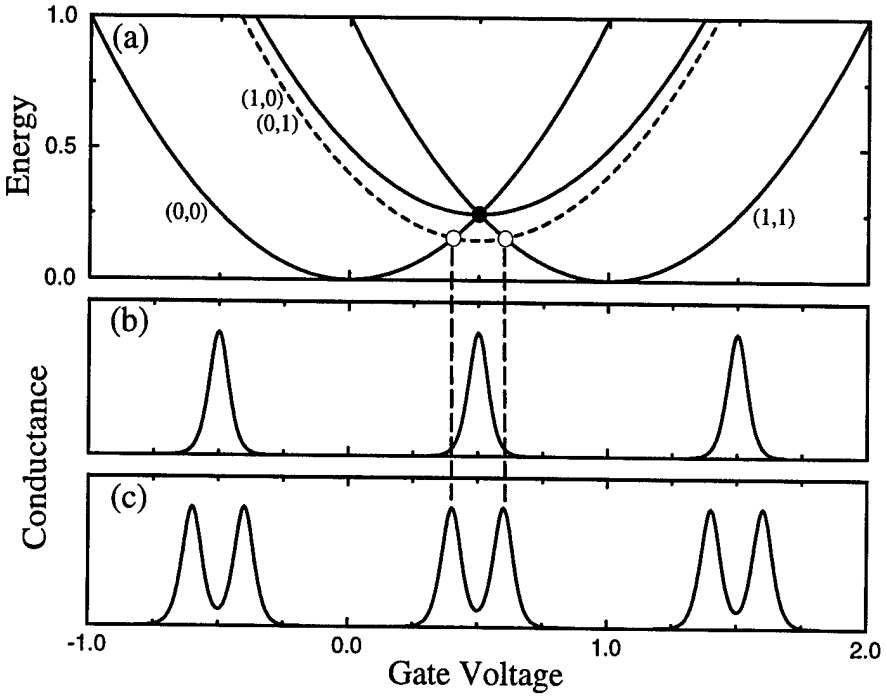

Figure 2.2: (a) Energy curves in the capacitive charging model for electrostatically identical dots with $V_{g1} = V_{g2}$ and zero interdot capacitance. Energies are given in units of the charging energy U; the gate voltage is given in units of e/C_g. Each zero-coupling eigenstate with definite particle number n_i on the ith dot gives rise to a parabola, labeled (n_1, n_2), which shows the state's energy as a function of the gate voltage. The zero of energy is chosen to coincide with the lowest energy possible for states with an even value for the total number of particles N_{tot}. The solid odd-N_{tot} parabola gives the lowest-energy curve only when there is no interdot coupling. The dotted parabola is the shifted-down energy curve for odd N_{tot} that results from finite coupling between the dots. The relevant degeneracy points are indicated by a black dot for zero coupling and white dots for finite coupling. (b) "Zero-coupling" conductance through the double dot as a function of the gate voltage. (For ease of viewing, peaks are depicted as symmetric with uniform finite widths and heights.) Conductance peaks are aligned with the zero-coupling degeneracy points such as the one shown in (a) and occur regularly with unit period. (c) Conductance through the dot for finite interdot coupling. Conductance peaks are aligned with the perturbed degeneracy points. Each zero-coupling peak has split into two separate peaks, equally distant from the zero-coupling peak position. Increasing the interdot coupling increases the separation between the paired peaks until the full set of peaks is again regularly distributed, with half the original period. (This figure for the capacitive charging model follows that of Waugh *et al.* [19, 43, 44].)

reduces to the problem of describing the shift in the ground state energy of a double dot containing a fixed total number of particles. Waugh [19, 43, 44] has shown that introduction of a capacitive coupling C_{int} between the two dots would allow one to obtain a picture in qualitative agreement with the experimental results: as the interdot capacitance goes to infinity, E_{int} converges to $U/4$. However, the magnitude of the interdot coupling necessary to fit the experimental data is much larger than what one would expect from an electrostatic interaction between two adjacent dots having a narrow tunneling channel between them. Waugh found that in order to bring about the saturation peak splitting, the interdot capacitance would have to grow from its zero-tunneling value by approximately a factor of 250, to a magnitude ten times larger than the single-dot capacitances at zero tunneling [19, 43, 44]. If one were to model the two dots as coplanar sheets of charge being moved closer together as the interdot channel conductance g is increased—essentially a best case scenario for those wishing to induce large interdot capacitances—the intercapacitance would depend only logarithmically on the separation, and the distance between the dots would have to be decreased to much less than an interatomic length to effect the required growth.

Consequently, the use of a large interdot capacitance C_{int} must be regarded as simply a reparametrization of the problem which replaces one unknown, E_{int}, with another unknown, C_{int}. What we really want is a theory which produces agreement with experiment and expresses E_{int} in terms of simple measurable quantities. Waugh *et al.* provide one candidate: the conductance G_b of the barrier between the two dots. The remainder of this chapter is devoted to developing a theory of the relation between E_{int} and the dimensionless conductance per tunneling channel $g = G_b/N_{ch}G_0$, where N_{ch} is the number of independent interdot tunneling channels (assumed to have identical conductances) and G_0 is the conductance quantum e^2/h. In the experiments of Waugh *et al.* [19, 43, 44], there is no applied magnetic field and the dots are connected by a narrow constriction allowing only a single transverse orbital mode with double spin

degeneracy. As a result, in this experimental case, N_{ch} equals 2.

2.3 Tunneling Model for the Double-Dot Coupling

2.3.1 Definition of the Model

Our goal can be stated a bit more precisely. For a general tunnel-coupling between two dots involving any number N_{ch} of identical, independent channels and dimensionless channel conductance g, our aim is to express the fractional energy shift $f = 4E_{int}/U$ as a function of g and N_{ch} plus any other parameters that might be found to be important. In order to derive an equation for f, we first choose a double-dot Hamiltonian. We will ignore electrostatic coupling of the dots for the moment: it will be noted at the beginning of Part B of this section that the presence of an interdot capacitance makes no substantive difference for our analysis. Interaction between the dots will occur solely via tunneling through the barrier between them. Such tunnel Hamiltonians have been found useful from the beginnings of Coulomb blockade theory [75], and the model we will use is a double-dot version of the Hamiltonian used, for example, by Averin and Likharev to investigate the conductance oscillations of small metal-to-metal tunnel junctions [76]. In particular, we have the Hamiltonian $H = H_0 + H_T$, where

$$\begin{aligned} H_0 &= K + V, \\ K &= \sum_{i=1}^{2} \sum_{\sigma} \sum_{\mathbf{k}} \epsilon_{i\mathbf{k}\sigma} \hat{n}_{i\mathbf{k}\sigma}, \\ V &= \frac{U}{2} \sum_{i=1}^{2} (\hat{n}_i - \phi_i)^2, \\ H_T &= \sum_{\sigma} \sum_{\mathbf{k}_1 \mathbf{k}_2} (t_{\mathbf{k}_1 \mathbf{k}_2} c^{\dagger}_{2\mathbf{k}_2\sigma} c_{1\mathbf{k}_1\sigma} + \text{H.c.}). \end{aligned} \tag{2.2}$$

In these equations, i is the dot index, σ is the channel index (which could signify different spin channels), and $\mathbf{k}$ is the index for all internal degrees of freedom not included in the channel index. In addition, $\hat{n}_i = \sum_{\mathbf{k}\sigma} \hat{n}_{i\mathbf{k}\sigma}$ is the number operator for the ith dot,

and $t_{\mathbf{k_1 k_2}}$ is the tunneling matrix element between a dot 1 wavefunction indexed by $\mathbf{k_1}$ and the dot 2 wavefunction lying in the same channel and indexed by $\mathbf{k_2}$. The gate voltage parameter ϕ_i has the same meaning as in Eq. 2.1. $\epsilon_{i\mathbf{k}\sigma}$ is the kinetic energy of the single-particle eigenstate of the ith dot having the indicated degrees of freedom. For simplicity, we will take these energies to be independent of dot and channel: $\epsilon_{i\mathbf{k}\sigma} = \epsilon_{\mathbf{k}}$.

The next step in focusing upon a model Hamiltonian is to choose a form for $t_{\mathbf{k_1 k_2}}$. Quite generally, $t_{\mathbf{k_1 k_2}}$ will be nonzero only when both k_1 and k_2 lie within some wavevector shell that maximally spans the space between the theory's low and high momentum cut-offs. The size of the wavevector shell depends on details of the barrier—in particular, the characteristic lengths of the channel both parallel and perpendicular to the voltage wall between the dots. If the barrier has an abrupt delta-function shape, the tunneling wavevector shell will span all of reciprocal space. If, on the other hand, the channel evolves adiabatically from the dots, the shell width will be small on the scale of a Fermi wavevector. Important questions are how many states lie within this shell—i.e., how large is the width W of the corresponding energy shell compared to the average level spacing δ between different states in the same channel (hereafter referred to as "the average level spacing" or just "the level spacing")—and for a given $\mathbf{k_1}$, for how many $\mathbf{k_2}$ is $t_{\mathbf{k_1 k_2}}$ nonzero. Thin-shell models with "one-to-one" hopping elements (i.e., for which $t_{\mathbf{k_1 k_2}} = 0$ unless $\mathbf{k_1} = \mathbf{k_2}$) have been applied to the coupled dot problem with some success,[5–7] especially for level spacings δ which are on the order of the charging energy U. For the nearly micron-sized dots used by Waugh *et al.*, however, U is approximately 400 μeV and δ is on the order of 30 μeV [19, 43, 44], so we expect that a tunnel-coupling sufficient to destroy the isolated-dot Coulomb blockade will involve a large number of single-dot eigenstates. Indeed, as it does appear that the characteristic size of the channel approximates a Fermi wavelength (40 nm) [77], it is reasonable to suppose that the wavevector shell width is on the order of a Fermi wavevector and, therefore, that the energy shell width W is comparable to the Fermi energy (13 meV),

which is much larger than U.

Consequently, assuming an abrupt tunnel barrier, we consider a thick-shell model that is the antithesis of the injective thin-shell model. Working in a regime where $W \gg U \gg \delta$, we use a tunneling matrix element t that is independent of k_1 and k_2 within the shell:

$$t_{\mathbf{k_1 k_2}} = t \ \forall \mathbf{k_1}, \mathbf{k_2} \text{ such that } \epsilon_0 < \epsilon_{\mathbf{k}_1}, \epsilon_{\mathbf{k}_2} < \epsilon_0 + W.$$

As the quantities we calculate are independent of the phase of t, we guiltlessly choose t to be real. This model is roughly equivalent to one in which each dot is represented by a tight-binding lattice with intersite hopping elements of order W/δ and where inter-dot tunneling occurs via a tunneling Hamiltonian with a single site-to-site connection. Choosing these tunneling sites to be at the origins $\mathbf{0_1}$ and $\mathbf{0_2}$ of the respective lattices, we may write

$$H_T = \sum_{\sigma} (T c^{\dagger}_{20_2\sigma} c_{10_1\sigma} + \text{H.c.}),$$

where $T = N_W t$ and $N_W = W/\delta$ is the number of orbital states per channel in each dot within the bandwidth W. (The equivalent lattice model should include second and further neighbor hopping so that the density of states is approximately constant between ϵ_0 and $\epsilon_0 + W$. The lattice constant is chosen by requiring that the product of N_W and the area of a unit cell equals the area of a single dot.)

As the Fermi energy ϵ_F must be somewhere between ϵ_0 and $\epsilon_0 + W$, the meaning of ϵ_0 depends on the width of the band. For a maximally thick shell, ϵ_0 lies at the bottom of the conduction band, and W is an ultraviolet cut-off, which is chosen to be of order twice the Fermi energy. Alternatively, when the barrier between the dots has a broader spatial extent, the energy shell sits more narrowly about the Fermi energy, and the width W is on the order of the energy difference needed to produce a factor-of-two change in the magnitude of the transmission amplitude for an incident particle. We

define a dimensionless filling parameter

$$F = \frac{\epsilon_F - \epsilon_0}{W},$$

which gives the position of the Fermi level within the bandwidth W. Provided that $(1-F)W$ and FW are both large compared to U, our final results should be independent of the precise values of W or F.

2.3.2 Map between the Double- and Single-Dot Systems

The model we have constructed is basically the two-dot version of that used by L. I. Glazman and K. A. Matveev [73] and by H. Grabert [72] to study the charge fluctuations of a single conducting island connected via point-tunnel junctions to conducting leads [see Fig. 2.1(b)]. Indeed, by using an analog of the standard center-of-mass transformation of classical mechanics and fixing the total number of particles in the two-dot system, we can create an exact mapping between the two-dot and one-dot problems. Consider again the double-dot potential energy V. By transforming to the analog of center-of-mass coordinates, one generates the following form:

$$V = \frac{U_1}{4}(\hat{N}_{tot} - \Phi_{tot})^2 + U_2(\hat{n} - \rho/2)^2, \tag{2.3}$$

where $\hat{N}_{tot} = \sum_{i=1}^{2} \hat{n}_i$, $\Phi_{tot} = \sum_{i=1}^{2} \phi_i$, $\hat{n} = (\hat{n}_2 - \hat{n}_1)/2$, $\rho = \phi_2 - \phi_1$, and $U_1 = U_2 = U = e^2/C_\Sigma$ when the interdot capacitance is zero. The rationale for the normalizations for $\hat{n}$ and ρ will soon be made apparent. In the meantime, note that for our Hamiltonian, N_{tot} is a constant of motion. Thus, for given N_{tot}, Φ_{tot}, and U_1, we can drop the first term and insert in the Hamiltonian a reduced potential energy:

$$V_{red} = U_2(\hat{n} - \rho/2)^2. \tag{2.4}$$

The impact of a nonzero interdot capacitance can now be trivially included: its only effect is to decrease the value of U_2.

In particular, in unpublished work [49], C. H. Crouch and J. M. Golden have found that if C_Σ is defined to be the total capacitance for a single dot minus the interdot capacitance, introduction of a constant interdot capacitance C_{int} decreases U_2 from e^2/C_Σ to $e^2/(C_\Sigma + 2C_{int})$. The equality $U_1 = U = e^2/C_\Sigma$ is left unchanged. For a given value of the conductance parameter g, if f is the fractional peak splitting in the model with zero capacitive coupling between the dots, then the fractional splitting f' for a system with an interdot capacitance is simply related to f by the equation

$$(1 - f') = \frac{U_2}{U}(1 - f). \tag{2.5}$$

Capacitive coupling between the dots thus leads to a nonzero splitting ($f' \neq 0$) even when there is no tunneling between the dots ($f = g = 0$).

We can now return to Eqs. 2.3 and 2.3. Restrict N_{tot} to be even. Then, $\hat{n}$ has integer expectation values in all the unperturbed double-dot eigenstates. With the total number of particles in the two dots held constant and even, the Hamiltonian is exactly that of a single dot tunnel-coupled to an ideal lead. The dot has number operator $\hat{n}$, charging energy $2U_2$, and gate voltage parameter $\rho/2$. In the absence of tunneling and with the level spacing in both dots much less than U_2, the ground state is an eigenstate of $\hat{n}$ that minimizes the reduced potential energy, which in the future we consider equivalent to "the potential energy." For $\rho = 0$, the minimum potential energy is zero and is achieved when the eigenvalue n of $\hat{n}$ is zero—i.e., when there are equal numbers of particles in the two dots. All other values of n give higher potential energies. For $\rho = 1$, on the other hand, the minimum potential energy is $U_2/4$, and $n = 0$ and $n = 1$ give degenerate minima.

These no-tunneling distinctions between zero and $U_2/4$ and between nondegeneracy and double degeneracy are quite familiar: they characterized the *even* and *odd* double-dot ground states ($\rho = 0$ for both) discussed in Sec. 2.2. Indeed, what we called the "*even* double-dot ground state" is precisely the "N_{tot} even, $\rho = 0$ ground state." The "*odd* double-dot ground state" is not exactly the same as the "N_{tot} even, $\rho = 1$ ground

state"; there is no getting around the fact that one case has one more (or less) particle than the other. However, in terms of their ground state energies, the difference between the two will be down by a factor of FN_W or $(1-F)N_W$. For a wide shell somewhere in the vicinity of half-filling, both FN_W and $(1-F)N_W$ are much greater than one, and the above difference is negligible. Calculation of E_{int} with $\phi_1 = \phi_2$ for the double dot is therefore equivalent to calculating the relative shifts of the $\rho = 0$ and $\rho = 1$ ground states of a single dot tunnel-coupled to a bulk two-dimensional electron gas. Having arrived at this conclusion, we will find that we have made much easier the job of solving Waugh's two-dot problem in the strong coupling regime: we can now redirect earlier work on the one-dot problem to our purpose.

More generally, we observe that we have created a model that extends beyond Waugh's experiment to circumstances where the two dots have different gate voltage parameters. Such situations can also be mapped to the one-dot problem. As the minimum potential energy is periodic in ρ with period two and is also even in ρ, the general solution is given by that for ρ in the interval $[0,1]$. For ρ in this interval, the difference in the ground-state energies of the double-dot system for even N_{tot} and odd N_{tot} is related to the difference in the ground-state energies of the single-dot system for gate voltage parameters ρ and $1-\rho$. The theory developed in this chapter permits calculation of the relative downward shift of the $\rho \neq 0$ ground state to the $\rho = 0$ ground state. Dividing by the zero-tunneling energy difference of the two ground states, we find that our emended aim is to calculate

$$f_\rho = \left(\frac{E_{int}(\rho)}{U_2\rho^2/4}\right) = \Psi_\rho(g, N_{ch}, u, N_W, F), \tag{2.6}$$

where $0 < \rho \leq 1$, $u = U_2/W$, $N_W = W/\delta$, and $E_{int}(\rho)$ is the ground-state energy relative to the ground-state energy for $\rho = 0$.

2.4 Results in the Weak-Coupling Limit

2.4.1 Barrier Conductance in the Weak-Coupling Limit

Before we can derive our equation for f_ρ in terms of g, we must find a formula for the barrier conductance. Measurement of the barrier conductance G_b with the exterior gates turned off can be modeled by calculating the tunnel junction conductance for $U_1 = U_2 = 0$. As mentioned before, we assume the different conducting channels to be identical yet independent—their individual conductances are the same and they do not interfere with one another. These assumptions are certainly reasonable for the two spin channels in the experiment. Using the Lippmann-Schwinger equation with H_T inserted for the scattering potential [78], one can solve for the perturbed electron eigenfunctions. The Heisenberg equation of motion for $\hat{n}_1$ can then be used to solve for the particle flow from dot 1 to dot 2 for a given voltage bias. Solving the resulting expression for the linear conductance gives the following equation for the dimensionless conductance per channel:

$$g = \frac{G_b}{N_{ch}G_0} = \frac{4\alpha}{|1+\chi\alpha|^2}, \tag{2.7}$$

where $\alpha = (\pi T/W)^2 = (\pi t/\delta)^2$ and $\chi = [1 + \frac{i}{\pi}\ln(\frac{F}{1-F})]^2$. This equation generalizes H. O. Frota and K. Flensberg's result for half-filling ($F = 0.5, \chi = 1$), derived via a Green's function-Kubo formula approach [74]. It is reassuring to note that despite χ's imaginary part for $F \neq 0.5$, the maximal dimensionless conductance is one for all filling fractions.

The calculated conductance G_b exhibits rather curious behavior: it first rises to a maximum of $N_{ch}G_0$ corresponding to N_{ch} fully open channels and then falls asymptotically to zero as $(T/W = t/\delta) \to \infty$. As Frota and Flensberg note [74], the asymptotic damping of the conductance results from the fact that formation of bonding and anti-bonding states at the tunnel junction makes the cost of passing through prohibitively high. The limit of $(T/W = t/\delta) \to \infty$ is in some sense unphysical: we do not expect

a point-to-point hopping coefficient T to significantly exceed the tunneling shell width; nor do we expect the tunneling matrix element t to be much greater than the average level spacing. Nevertheless, the apparent absence of any good reason to truncate the theory at a particular value of t indicates that the model is at best unwieldy in the limit of strong coupling. To get the correct limiting behavior for strong coupling, it is more convenient to use a different approach, suitable for perturbation about the $g = 1$ limit. This will be described in Sec. 2.5.

2.4.2 Relative Energy Shift of Even and Odd States in the Weak-Coupling Limit

In the meantime, the site-to-site tunneling model is still useful in the weak coupling regime. So we plod ahead, calculating via standard Rayleigh-Schrödinger perturbation theory the second order shift in the ground state energy for $\rho \neq 0$ minus that for $\rho = 0$. The $\rho = 1$ shift will be taken to equal the limit of the general $0 \leq \rho < 1$ shift as $\rho \to 1$. It might be objected—correctly— that this limit fails properly to account for the degeneracy of the ground state at $\rho = 1$. Such a failing is pardonable, however, for the contributions that are left out are all smaller by a factor of FN_W or $(1-F)N_W$ from those which are retained. Since we assume that t/δ is finite, F is of order $\frac{1}{2}$, and N_W is large, the omitted terms are negligible.

For $N_W \gg 1$, the perturbation theory sums can be approximated as integrals. Observing that $u = U_2/W \ll 1$, we divide the difference between the second order shifts by $U_2\rho^2/4$ to get the leading approximation to f_ρ:

$$\begin{aligned} f_\rho^{(1)} &= 4N_{ch}\frac{t^2}{\delta^2}\left(\frac{1}{\rho^2}\right)[(1-\rho)\ln(1-\rho) \\ &\quad +(1+\rho)\ln(1+\rho)+O(u\rho^2)]. \end{aligned} \tag{2.8}$$

The second-order term indicates a significant feature of f_ρ: it is even in ρ. This property has also been noted by H. Grabert [72] and results from the fact that at any order of

perturbation theory, every tunneling process contributing to the energy shift has a twin with the roles of dots 1 and 2 interchanged. In any intermediate virtual state with eigenvalue n for $\hat{n}$, the potential energy is greater than that for the unperturbed ground state by $\delta V(\rho) = U_2 n(n-\rho)$. Therefore, when dots 1 and 2 are interchanged, $\delta V(\rho) \to \delta V(-\rho)$ for all the intermediate states. If we represent one of the twin terms by $\Delta(\rho)$, the other is $\Delta(-\rho)$, and we see that f_ρ is constructed of sums that are even in ρ.

Using the second-order (in t/δ) parts of g and f_ρ, we can now write a first-order equation for f_ρ in terms of g:

$$\begin{aligned} f_\rho^{(1)} &= \frac{N_{ch} g}{\pi^2}\left(\frac{1}{\rho^2}\right)[(1-\rho)\ln(1-\rho) \\ &\quad +(1+\rho)\ln(1+\rho)+O(u\rho^2)], \end{aligned} \tag{2.9}$$

a result consistent with the large-N_{ch} calculation of the effective capacitance of a single dot at $\rho = 0$ [67]. Setting $\rho = 1$ to calculate the relative shifts of the original *even* and *odd* states, we find

$$f^{(1)} = \frac{2\ln 2}{\pi^2} N_{ch} g + O(ug, g^2), \tag{2.10}$$

where we have used the fact that f as originally defined without the subscript is equivalent in our limits to $f_{\rho=1}$. The above equation indicates that the plot of f as a function of gate voltage is not just a replica of the plot for g as a function of gate voltage —as a prima facie look at Waugh *et al.*'s data might lead one to suppose [19, 43, 44]. In particular, for $g \ll 1$ and $N_{ch} = 2$, Eq. 2.10 gives a slope of approximately 0.28 for $f(g)$, rather than unity. Thus, in this regime, the fractional splitting f of the double-dot conductance peaks should lag g, the dimensionless barrier conductance per channel.

2.5 Connection to the Strong-Coupling Limit

If we blithely extended our perturbative equation for f to the limit $g \to 1$, the large-N_{ch} f would greatly overshoot its mark and the one or two-channel f would fall substantially

short. The real issue is not, however, how badly such a naive extrapolation fails, but whether we can connect these weak-tunneling results to those that can be calculated for the strong-tunneling limit. Having discussed the equivalence of the two-dot and one-dot problems at length in Sec. 2.3.2, we can turn to see what the current literature on the one-dot problem offers. For the large-N_{ch} limit, a reasonable interpolation between the solutions for weak and strong coupling has already been found.[8,9,11]

The situation is less clear for the case with which we are most concerned, in which N_{ch} equals one or two. Flensberg and Matveev [33, 34] have proposed a useful Luttinger liquid approach in which the nearly transparent link between a single dot and an electrode is modeled as a one-dimensional channel with a slightly reflective potential barrier. Convergence to the single composite-dot limit is achieved naturally and neatly, and E_{int} is calculated perturbatively in r, where r is the reflection amplitude, and $g = 1 - |r|^2$. Using the map between the two-dot and one-dot problem, we can translate Matveev's calculations of the leading term for $(1 - g) \ll 1$ to our language. We find that for $N_{ch} = 1$ (i.e., assuming spin polarization), the fractional peak splitting in the two-dot problem when $\rho = 0$ is given by the following:

$$f = 1 - C_1 \frac{8e^{\gamma}}{\pi^2} \sqrt{1 - g} + \ldots \tag{2.11}$$

where $\gamma \simeq 0.577$ is the Euler-Mascheroni constant and C_1 is an error factor on the order of 1 that we have inserted to guard against the possible imprecision of calculating in Luttinger liquid theory with a finite cut-off [79]. For the case relevant to the experiments of Waugh *et al.* [19, 43, 44], that of $N_{ch} = 2$, adaptation of Matveev's calculation gives

$$f = 1 + C_2 \frac{16e^{\gamma}}{\pi^3} (1 - g) \ln(1 - g) + \ldots \tag{2.12}$$

where C_2 is an error factor analogous to C_1. Except for the logarithmic factor in the second formula, these equations are of the form suggested by the scaling analysis of Flensberg [33], which predicts effective charging energies behaving as $(1 - g)^{N_{ch}/2}$. Matveev's initial two-channel solution is, in fact, linear in $(1 - g)$ but diverges logarith-

mically as $U_2/\delta \to \infty$. A higher-order analysis to eliminate the divergence [34] replaces the logarithm with argument U_2/δ by one with argument $(1-g)^{-1}$.

In Fig. 2.3, we show the f-versus-g plots given by the weak and strong coupling formulas (2.10), (2.11), and (2.12) for $N_{ch} = 1$ and $N_{ch} = 2$ with $C_1 = C_2 = 1$. In each case, a possible interpolation between the weak and strong coupling limits is given by a dashed curve. For $N_{ch} = 2$, the corresponding experimental data of Waugh *et al.* [19, 43, 44] is also plotted. Given the experimental error implicit in the dispersion of the data points themselves, the data is seen to be in reasonable agreement with theory.

It is clear, however, that, unlike the calculations for $N_{ch} \gg 1$, for $N_{ch} = 2$ the order of calculation completed so far does not really allow confident interpolation between the weak and strong coupling limits. On the strong-tunneling side, $C_2 \approx 1.5$ would effect greater agreement with our suggested interpolation: Luttinger liquid theory's prediction of $C_1 = C_2 = 1$ must certainly be checked. With respect to the weak-tunneling results, calculation of higher orders in perturbation theory should improve the matching, but such computations are made difficult by the fact that the correlations induced by the strong Coulomb interaction make normal Green's functions methods inapplicable [71]. Different time orders must be treated separately, and as appears to occur quite generally in Coulomb-blockade problems [80], the number of diagrams grows pathologically with the order in perturbation theory. Nevertheless, calculation of the g^2-term in the weak-tunneling limit is conceivable, and this term may permit a more reliable interpolation between the weak and strong coupling regimes.

Irrespective of the difficulty of connecting the strong and weak coupling limits, it should be emphasized that despite the uncertainty in the coefficients C_1 and C_2, the strong coupling results do give an important constraint on the form of the theoretical f-versus-g curves—viz., the value of f must reach 1 at the point where g equals 1. Thus, a model that treats the Coulomb blockade peak splitting as a function of the interdot channel conductance produces the experimentally observed saturation splitting for a

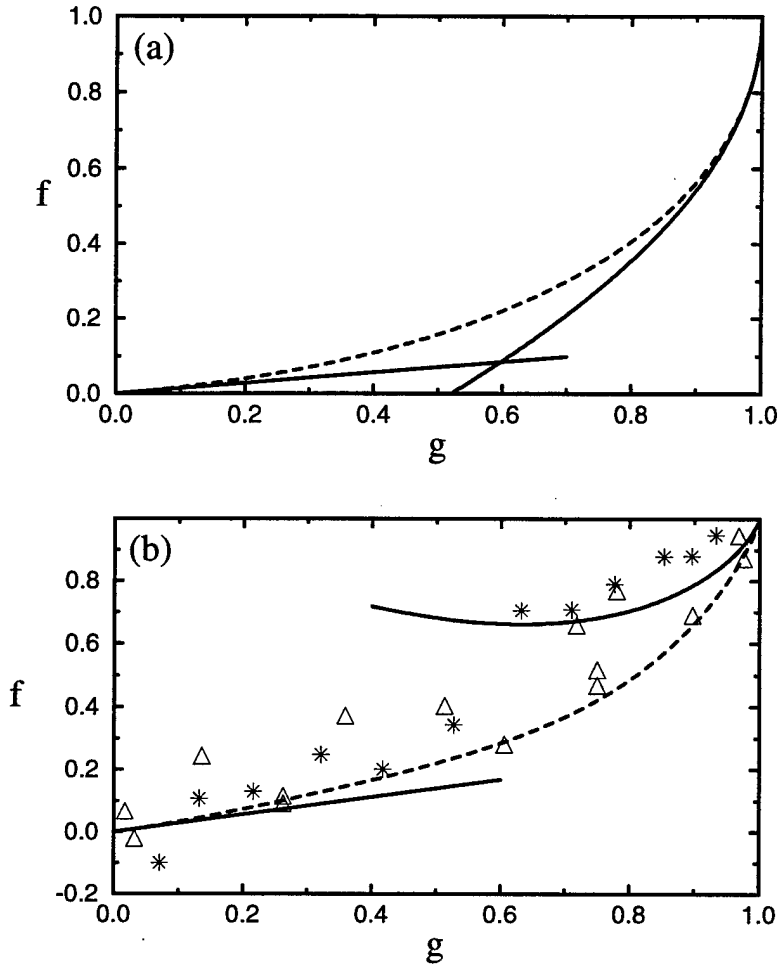

Figure 2.3: Plots of the fractional Coulomb blockade conductance peak splitting f as a function of the dimensionless conductance per channel g in the weak and strong tunneling limits for (a) $N_{ch} = 1$ and (b) $N_{ch} = 2$ with coefficients $C_1 = C_2 = 1$ in Eqs. 2.11 and 2.12. Possible interpolating functions are shown by dashed curves. Data points from Waugh *et al.* [19, 43, 44] are given as triangles or stars; the two different symbols correspond to different data sets. The value of f for the experimental data has been extracted from the measured splitting fraction f' as discussed in Sec. 2.3.2 with $U_2/U \approx 0.9$. This choice of U_2/U corresponds to the constant interdot capacitance of 20 aF and total single-dot capacitance of 0.4 fF estimated for the experiments of Waugh *et al.* [19, 43, 44].

reasonable physical value of the parameter g that marks the strength of the interdot coupling. This fact can be understood by noting that a non-zero interdot conductance results in charge fluctuations between the dots for which the natural energy scale is U_2, the energy scale that characterizes the difference between the $\rho = 0$ and $\rho \neq 0$ ground states. As g increases, larger and larger charge fluctuations, in which multiple electrons move from one dot to the other, become increasingly significant, and the initial $g = 0$ difference between the $\rho = 0$ and $\rho \neq 0$ ground states becomes less relevant to the energy of the $g \neq 0$ ground states, which after all are superpositions of a great number of $g = 0$ eigenstates with a wide variety of charge distributions.

The decrease in the ρ-dependence of the ground state energy for $g \neq 0$ could be described, at least approximately, by an "effective interdot capacitance." However, the introduction of such a fictive and, as noted in Sec. 2.2, unphysical mediator merely begs the question of how such a large effective interaction is produced. Tunneling provides an answer by allowing electrons to hop back and forth between the dots, interacting directly with their "neighbors" through the pre-existing $g = 0$ two-dot capacitances.

2.6 Conclusion

Following the work of Waugh *et al.* [19, 43, 44], we have investigated the relation between the barrier conductance and the Coulomb blockade for two electrostatically equivalent dots connected by one or more identical tunneling channels and have found an explanation for the evolution of the double-dot Coulomb blockade that does not rely upon unphysically large values for the interdot capacitance, the intradot level spacing, or the number of conducting channels. We propose to write the fractional peak splitting f of the Coulomb blockade conductance peaks as a function of the number of channels N_{ch} and the dimensionless barrier conductance per channel g, assuming that the energy level spacing δ is small compared to the interdot Coulomb blockade energy U_2 and that U_2 is small compared to the bandwidth W of states over which the amplitudes for trans-

mission through the barrier are roughly constant. Using a "uniform thick-shell model" for the tunneling term in the Hamiltonian, we solve for this function to leading order in the limit of weak interdot coupling. We find that in this limit, the peak splitting should evolve linearly with the total barrier conductance with a slope substantially less than one.

In order to solve for the strong-coupling limit, we have introduced a "fictional" difference between the gate voltages on the individual dots. Such a generalization of the two-dot problem makes it relatively straightforward to adapt our analysis to situations where there is a voltage bias between the two dots [45]. Its purpose here is to allow for a map between the previously unsolved two-dot problem and a better known one-dot problem. The strong-coupling results that we obtain via this mapping give an asymptotic form for the peak-splitting that behaves as $(1-g)\ln(1-g)$.

In the case of $N_{ch} = 2$, which is pertinent to the experimental results of Waugh *et al.* [19, 43, 44], the present limiting forms for strong and weak coupling do not match up well enough to allow a reliable quantitative interpolation between the two limits. Nevertheless, a plausible interpolating curve is in qualitative agreement with existing experimental data. More precise experimental results would allow for a test of the slope predicted for the weak tunneling limit. An extension of current theory is still necessary to permit a convincing connection between the two asymptotic limits.

2.7 Acknowledgments

The authors are grateful for very helpful discussions with F. R. Waugh, R. M. Westervelt, Gergely T. Zimanyi, C. A. Stafford, C. Crouch, C. Livermore, and Steven H. Simon. J. M. G. thanks the United States Air Force for financial support. The work for this chapter was also supported by the NSF through the Harvard Materials Research Science and Engineering Center, Grant No. DMR94-00396. After the separate manuscript corresponding to this chapter was essentially complete, the authors received a copy of a

paper, prior to publication, by K. A. Matveev, L. I. Glazman, and H. U. Baranger [53] in which similar ideas were independently developed.

Chapter 3

Subleading Results for the Coupled-Dot Blockade

3.1 Introduction

The opening of tunneling channels between two quantum dots leads to an erosion of the individual dots' Coulomb blockade [8, 13, 14, 15]. For a pair of electrostatically identical quantum dots (see Fig. 2.1(a) for a schematic view of the double-dot structure), the progress of this erosion can be chronicled by tracking the splitting of the Coulomb-blockade conductance peaks as they evolve from doubly degenerate single-dot conductance resonances to nondegenerate double-dot peaks with twice the original periodicity [19, 43, 44, 51, 53, 54]. For a system in which the tunneling channels can be treated as having the same individual conductances and in which the Coulomb charging energies are large compared to the single-particle level spacings but small compared to the tunneling channel bandwidths, the fractional peak splitting f can be expressed as a function of two parameters: N_{ch}, the number of tunneling channels between the two dots, and g, the dimensionless conductance per tunneling channel. (In this chapter, the *conductances* indicated are always *dimensionless conductances*, by which we mean the

actual conductance divided by the conductance quantum, e^2/h.)

In particular, for weakly coupled dots ($g \to 0$), the fractional peak splitting can be expressed perturbatively as a sum of terms of the form $a_{m,n}(N_{\rm ch})^m g^n$, where $1 \leq m \leq n$ and $a_{m,n}$ is independent of $N_{\rm ch}$ and g. Chapter 2 has produced the leading term in this expansion. However, as this term is simply linear in the total interdot tunneling conductance, $g_{\rm tot} = N_{\rm ch} g$, it does not distinguish between behavior in the large-$N_{\rm ch}$ and small-$N_{\rm ch}$ limits. To make such a distinction, one must calculate to second order in g, in which case one obtains two sets of terms, one proportional to $N_{\rm ch} g^2 = g_{\rm tot}^2/N_{\rm ch}$ and the other proportional to $(N_{\rm ch})^2 g^2 = g_{\rm tot}^2$.

As in Chapter 2, the fractional peak splitting is calculated by adding an additional dimensionless parameter ρ to the problem, where ρ represents a capacitively weighted voltage difference between the two dots. The fractional peak splitting f is then found to be given by the more general function of ρ, f_ρ, at $\rho = 1$. The introduction of the parameter ρ allows for a clear mapping between the problem of two tunnel-coupled dots and that of a single dot coupled to a bulk lead [51]. It also allows for consideration of experimental situations in which there is a voltage bias and $\rho \neq 0$ [45]. In addition, the introduction of this parameter allows for comparison of the results of our calculations with those of workers in the field of metallic junctions [67, 68, 70, 72, 81], who have been concerned primarily with calculating quantities such as "effective charging energies" $U_{\rm eff}$ (or, alternatively, "effective capacitances" $C_{\rm eff} = e^2/2U_{\rm eff}$), which are related to derivatives with respect to ρ of $\rho^2 f_\rho$. Thus, the $(N_{\rm ch})^2 g^2$ terms in the expansion of f_ρ that we derive in this chapter can be compared with weak-coupling calculations from the theory of metallic junctions [67, 72] in which quantities such as the effective charging energy are expanded perturbatively in powers of $N_{\rm ch} g$. (The reader should note that, for the purpose of computing such derivative quantities as $U_{\rm eff}$, the weak-coupling calculations performed in this chapter are only useful when ρ lies far from the singular point $\rho = 1$ [82].)

Some of the most interesting work on large-$N_{\rm ch}$ metallic junctions has been concerned not with this weak-coupling limit but, rather, with the strong-coupling regime in which such a simple perturbative expansion in $N_{\rm ch}g$ is inapplicable [68, 70, 81]. Study of the strong-coupling limit $g \to 1$ for small-$N_{\rm ch}$ junctions has also proven fruitful, revealing a dramatic dependence of the peak splitting on $N_{\rm ch}$. In the cases of $N_{\rm ch} = 1$ and $N_{\rm ch} = 2$, the leading corrections to a fractional peak splitting equal to one ($f = 1$) have been found to be proportional to $\sqrt{1-g}$ and $(1-g)\ln(1-g)$, respectively [34, 51, 53, 54], and it has been hypothesized by Molenkamp, Flensberg, and Kemerink [33, 48] that, for $N_{\rm ch} > 2$ but finite, the leading nontrivial term is proportional to $(1-g)^{N_{\rm ch}/2}$. This last suggestion appears to correspond to calculations of the "effective charging energy" $U_{\rm eff}$ for metallic junctions [68, 70, 81], where, once again, the effective charging energy is proportional to the second derivative of $\rho^2 f_\rho$ at $\rho = 0$. Consequently, $U_{\rm eff}$ can be expected to scale with $(1-g)$ in the same manner as the ρ-dependent corrections to the fractional peak splitting $f = f_{\rho=1}$, and it is reassuring that the metallic-junction limit gives an effective charging energy proportional to $e^{-g_{\rm tot}/2}$, which is equivalent to $(1 - g_{\rm tot}/N_{\rm ch})^{N_{\rm ch}/2}$ in the limit $N_{\rm ch} \to \infty$.

Despite the recent progress in study of the strong-coupling limit, for the case of most direct experimental interest, $N_{\rm ch} = 2$ [19, 43, 44, 48], the leading-term calculation [34] that has previously been used fails to be completely satisfactory for at least three reasons. The first is that this calculation does not answer the question of whether the coefficient of $(1-g)\ln(1-g)$ is affected by the manner in which the ultraviolet cutoff is imposed in the low-energy bosonization approach [51]. The second is that the coefficient of the subleading term linear in $(1-g)$ is both unknown and naively infinite [34]. Finally, there is the worry—which also applies to the weak-coupling result—that, for $N_{\rm ch} = 2$, interpolation between the solutions for weak- and strong-coupling is difficult because the respective f-versus-g curves do not come especially close [51].

This chapter addresses these three concerns for the two-channel problem and also

extends earlier results for the general $N_{\rm ch}$-channel problem in the limit of weak coupling. In so doing, it illuminates the difference between large-$N_{\rm ch}$ and small-$N_{\rm ch}$ behavior for $g \approx 0$, creates a theory that can be more realistically compared to experimental results for $N_{\rm ch} = 2$, and argues for the universality of the results, which should be independent of the nature and magnitude of the ultraviolet cutoffs. Section 3.2 presents the g^2 extension of the weak-coupling theory and checks the result against calculations in the $N_{\rm ch} \to \infty$ limit. Section 3.3 gives the $(1-g)$ correction to the leading dependence in the strong-coupling limit for $N_{\rm ch} = 2$ and shows a plot of the experimental results and revised theoretical predictions for two-channel interdot junctions. Section 3.4 argues that the strong coupling results of Section 3.3 are independent of the nature of the way the ultraviolet cutoff is imposed and do not change when one allows the fermionic theory to stray from half filling. Section 3.5 summarizes the results, and Appendices A and B present technical details of calculations in Sections 3.2 and 3.3, respectively.

3.2 The Weak-Coupling Limit for Arbitrary $N_{\rm ch}$

For weakly coupled quantum dots, we use a model "site-to-site" hopping Hamiltonian [51] and calculate perturbatively in the tunneling term H_T:

$$\begin{aligned} H &= H_K + H_C\,, \\ H_K &= \sum_{i=1}^{2} \sum_{\sigma} \sum_{\mathbf{k}} \epsilon_{\mathbf{k}} \hat{n}_{i\mathbf{k}\sigma}\,, \\ H_C &= U_2(\hat{n} - \rho/2)^2\,, \\ H_T &= \sum_{\sigma} \sum_{\mathbf{k_1 k_2}} t(c^{\dagger}_{2\mathbf{k}_2\sigma} c_{1\mathbf{k}_1\sigma} + \text{H.c.})\,. \end{aligned} \tag{3.1}$$

As in Chapter 2, in these equations, i is the dot index; σ is the channel index; $\mathbf{k}$ is the index for all internal degrees of freedom not included in the channel index; H_C is the part of the electrostatic potential energy that is affected by interdot tunneling; $\hat{n}$ is half the difference in dot occupation numbers, $\hat{n} = (\hat{n}_2 - \hat{n}_1)/2$; ρ is a differential gate voltage

parameter and is restricted to values between 0 and 1 (as permitted by the system's unit periodicity); and U_2 is the differential charging energy, which, for electrostatically equivalent dots, is given by the formula $U_2 = e^2/(C_\Sigma + 2C_{\rm int})$, where $C_{\rm int}$ is the interdot capacitance and C_Σ is the total single-dot capacitance minus the interdot capacitance. If the dots are not electrostatically equivalent, the formula for U_2 and the definition of ρ are more complicated [53, 54]. However, the model is still applicable, and the results for f_ρ can still be used to obtain the peak splitting.

These calculations are made palatable by assuming that U_2 is much smaller than the tunneling-channel bandwidth W yet much greater than the average intrachannel level-spacing δ: $W \gg U_2 \gg \delta$. This assumption leaves us with a theory that we can consider to be in the continuum limit and that we can hope to be independent of ultraviolet cutoffs. As the bandwidth is presumably of the order of the Fermi energy ϵ_F, these assumptions are reasonable for the micrometer-sized dots of Waugh *et al.* [19, 43, 44], for which $\epsilon_F \approx 10$ meV, $U_2 \approx 400$ μeV, and $\delta \approx 30$ μeV.

As in Chapter 2, our primary goal is to calculate the fractional peak splitting f—i.e., the ratio of the distance between split Coulomb-blockade subpeaks for a given g and their maximal separation in the strong-coupling ($g \to 1$) limit. It was shown in Chapter 2 that, if the total number of electrons on the two dots is assumed even, the problem of solving for f is a corollary to the problem of solving for a more general quantity f_ρ, which characterizes the ground-state energy of the double dot when the difference between the external potentials applied to the dots is nontrivial and the total number of electrons on the two dots is fixed and even. Recall the equation for f_ρ:

$$f_\rho = \frac{\Delta_0 - \Delta_\rho}{U_2\rho^2/4}, \tag{3.2}$$

where Δ_ρ is the shift in the ground state energy induced by tunneling at a given value of the gate voltage parameter ρ and $U_2\rho^2/4$ is the difference between the unperturbed ground-state energies for the given ρ and $\rho = 0$. In Chapter 2, it was shown that, for

symmetric dots,

$$f = f_{\rho=1} \,. \tag{3.3}$$

It was also determined that f_ρ exhibits the following leading behavior as $g \to 0$:

$$\begin{aligned} f_\rho^{(1)} &= \frac{N_{\rm ch} g}{\pi^2 \rho^2}[(1-\rho)\ln(1-\rho) + (1+\rho)\ln(1+\rho) \\ &\qquad + O(\rho^2/\psi)]\,, \end{aligned} \tag{3.4}$$

where $\psi = W/U_2 \gg 1$. Thus, the corresponding leading behavior for f is

$$f^{(1)} = \frac{2\ln 2}{\pi^2} N_{\rm ch} g + O(N_{\rm ch} g/\psi)\,. \tag{3.5}$$

Extending perturbation theory beyond this result—i.e., beyond first-order in g—requires some laborious computation. The next-leading contributions come from two sources. The first, which we shall call $f_\rho^{(2A)}$, arises from a combination of the second-order energy shift that has already been calculated and the second term in the formula that relates the tunneling amplitude t to the channel conductance g. (The first term in this formula was used to derive Eq. 3.4.) The second source of g^2 terms, $f_\rho^{(2B)}$, is the shift in the ground-state energy provided by terms that are fourth-order in t.

The first contribution is relatively easy to calculate. The equation for g in terms of t has been derived for half filling by Frota and Flensberg [74] and for arbitrary filling in Chapter 2. In the latter calculation, the system is assumed to have a constant density of states between single-particle energies ϵ_0 and $(\epsilon_0 + W)$, the density of states being zero elsewhere. The system's level of "filling" is then characterized by the filling fraction $F = (\epsilon_F - \epsilon_0)/W$, where ϵ_F is the Fermi energy. In accordance with the half-filling result, one then finds the following:

$$g = \frac{4\chi}{|1 + (1+i\eta)^2\chi|^2}\,, \tag{3.6}$$

where $\chi = (\pi t/\delta)^2$ and $\eta = (1/\pi)\ln[F/(1-F)]$. Inverting this expression, one discovers that

$$\frac{t^2}{\delta^2} = \frac{g}{4\pi^2}\left[1 + \frac{1-\eta^2}{2} g + O(g^2)\right]\,. \tag{3.7}$$

Consequently, our first g^2 term is equal to the right side of Eq. 3.4 multiplied by $(1-\eta^2)g/2$:

$$f_\rho^{(2A)} = (1-\eta^2)\frac{N_{\rm ch}g^2}{2\pi^2\rho^2}[(1-\rho)\ln(1-\rho) + (1+\rho)\ln(1+\rho) + O(\rho^2/\psi)]. \quad (3.8)$$

This term is of the expected form $a_{1,2}^{(2A)}N_{\rm ch}g^2$, where $a_{1,2}^{(2A)}$ is a function of ρ.

On the other hand, $a_{1,2}^{(2A)}$ is dependent on the filling fraction F, a fact which appears to imperil our dreams of a theory that is universal in that it is insensitive to the details of the high-energy behavior (including whether, for example, certain high-energy states exist and therefore have a role in determining the filling fraction F). We shall see, however, that the F-dependence of $f_\rho^{(2A)}$ actually serves our end, for it exactly cancels the F-dependence of $f_\rho^{(2B)}$. As a result, we can further conclude that, through second order in the channel conductance g, expression of the fractional peak splitting in terms of the channel conductance is not only convenient for comparison with experiment but is also necessary and sufficient for constructing a result that can be hoped to be universal.

To support this claim, we must actually determine the value of $f_\rho^{(2B)}$. Sadly, it cannot be obtained as effortlessly as $f_\rho^{(2A)}$. There are 24 separate terms that contribute to the fourth-order energy shift. One 12-member subset consists of terms proportional to $(N_{\rm ch})^2$; the second consists of those simply linear in $N_{\rm ch}$. All but four of the twenty-four terms correspond to a specific series of four tunneling events that begin and end with the double-dot system's unperturbed ground state. The remaining four, which belong to the $(N_{\rm ch})^2$ subset, correspond to the fourth-order terms in Rayleigh-Schrödinger perturbation theory that are products of the second-order energy shift and a propagator squared. These four have been described by Grabert as *diagrams with insertions* [72].

In general, the nature of the twenty-four fourth-order terms is most digestibly summarized via a diagrammatic representation that looks essentially like one of time-ordered single-particle diagrams (see Fig. 3.1). Despite the superficial single-particle nature of this representation, it is important to remember that the propagators that enter into the

energy calculations are the propagators for the entire double-dot system, which depend upon both the tunneling particles' individual kinetic energies and the system's multiparticle potential energy. The presence of the multiparticle potential energy makes it impossible to reduce the calculation to the normal Feynman diagrams, for which one can write the problem entirely in terms of single-particle propagators. The presence of exchange terms, which do not appear among the diagrams proportional to $(N_{\mathrm{ch}})^2$, makes a pseudo-single-particle representation necessary.

Within this time-ordered perturbation theory scheme, the individual fourth-order terms are plagued by both ultraviolet and infrared divergences. Every term is divergent as the bandwidth goes to infinity and four of the $(N_{\mathrm{ch}})^2$ terms are divergent as $\rho \to 1$. (A different set of four is divergent as $\rho \to -1$.) From the result for $f_\rho^{(1)}$, we might hope to cancel the ultraviolet divergences and to obtain an answer for the ground-state energy that is infrared-singular but not infrared-divergent. Indeed, as Grabert has noted [72], the ultraviolet divergences of the $(N_{\mathrm{ch}})^2$ terms must drop out since, in the limit $U_2 \to 0$, these terms correspond to disconnected diagrams or insertion diagrams that exactly cancel one another and thus do not appear as Feynman diagrams. In contrast, the N_{ch} diagrams do have nontrivial Feynman-diagram analogs. As a whole, they correspond to a single totemic Feynman diagram—an individual ring marked by four tunneling events. The ultraviolet divergences of these diagrams are therefore expected to be persistent but irrelevant because we are interested only in the relative shift between the ground-state energies for arbitrary ρ and for $\rho = 0$ (recall Eq. 3.2). Accordingly, we expect that, when one subtracts the fourth-order shift for $\rho = 0$ from that for arbitrary ρ, the fourth-order terms produce a result that is neither ultraviolet- nor infrared-divergent but is infrared-singular as $|\rho| \to 1$. A brief summary of the actual calculation of these terms follows. Those interested in more detail are invited to peruse Appendix A, which offers a fuller description of the calculation of the $(N_{\mathrm{ch}})^2$ diagrams and a step-by-step computation of the contribution from one representative N_{ch} term.

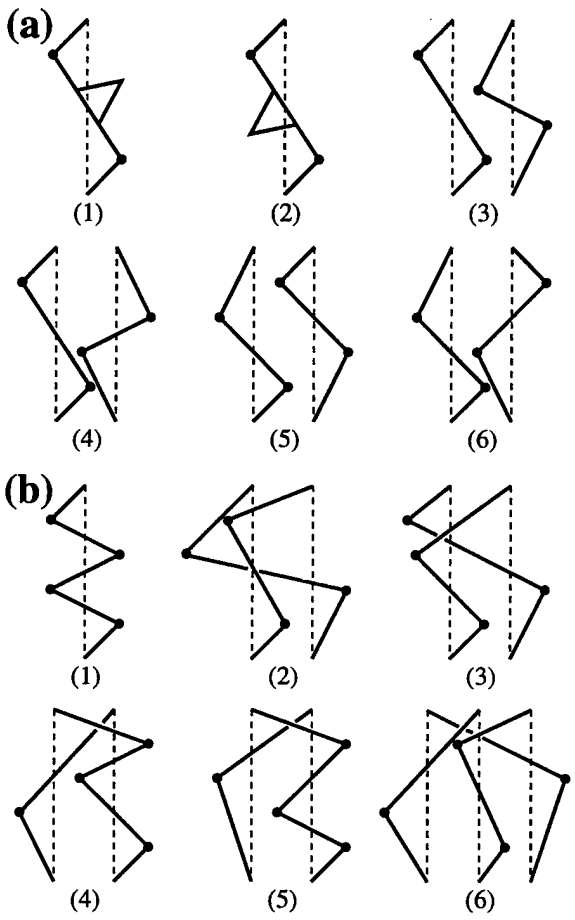

Figure 3.1: Diagrams for half of the (a) fourth-order, $(N_{ch})^2$ terms and (b) fourth-order, N_{ch} terms. The remaining terms are represented by diagrams that are mirror images of these. A vertical dashed line is drawn for each of the m particles that tunnels at least once from one dot to the other. This line stands for the corresponding particle's initial state, a state that must be filled at the end of the four tunneling events in order to recover the unperturbed ground state from which the system starts. A particle begins at the bottom of its vertical initial-state line. Particles in dot 1 propagate upward and rightward. Particles in dot 2 propagate upward and leftward. A tunneling event for a particle is signaled by a solid dot that coincides with a bend in the particle-propagation path. Each particle must end on one of the dashed vertical lines, meaning that it ends in the single-particle state that corresponds to that line. *Insertions* (see Sec. 3.2) are represented by triangles that project off a single-particle propagation line. If the projection points up, the insertion corresponds to the term in the second-order energy shift for which a particle tunnels off the dot occupied by the propagating particle. If the projection points down, the insertion corresponds to the second-order term for which a particle tunnels onto the dot occupied by the propagating particle. In the absence of exchange, all particles end on their own *initial-state lines.* A two-particle exchange carries a minus sign and results in each of two particles ending on the other's initial-state line. Three-particle exchange carries no sign (alternatively, one can view it as carrying two canceling minus signs) and results in each of three particles ending on one of the others' initial-state lines.

For the less scrupulous, there are still a few facts worthy of note. A prominent feature of the fourth-order calculation is that each term involves the integration over four energy variables (ϵ_i, where i ranges from 1 to 4) of a product of three propagators. In the $(N_{\rm ch})^2$ diagrams, the energy variables "pair off": ϵ_1 and ϵ_3 only appear as parts of the combination $\epsilon_{\rm I} = (\epsilon_3 - \epsilon_1)$, and ϵ_2 and ϵ_4 only appear as parts of the combination $\epsilon_{\rm II} = (\epsilon_4 - \epsilon_2)$. As a result, calculation of these terms reduces to the performance of double integrations over $\epsilon_{\rm I}$ and $\epsilon_{\rm II}$— albeit with a nontrivial density of states.

The $N_{\rm ch}$ diagrams cannot be handled in this way, for they involve particle exchanges that frustrate any desire to pair off the energy variables. The quadruple integration over the ϵ_i cannot be eluded. It can, however, be expedited by differentiating twice with respect to ρ while integrating out the energy variables and, then, integrating twice with respect to ρ in the end. One might worry about the fact that, by differentiating twice with respect to ρ, one has lost knowledge of terms constant and linear in ρ. However, these terms are unimportant. As noted in Chapter 2, the ground-state energy (perturbed or unperturbed) is symmetric in ρ. Therefore, terms linear in ρ must cancel out of the fourth-order energy shift when all the terms are summed. Constant terms are similarly negligible since, as usual, we are only interested in the relative energy shift $\Delta_\rho - \Delta_0$.

After the aforementioned tricks for calculating the $(N_{\rm ch})^2$ and $N_{\rm ch}$ diagrams have been used, the only real wrinkles that remain are integrals of the form

$$\mathcal{P}\int_0^{R\psi} dx\, \frac{\ln(x+B)}{(x+A)},$$

where $0 \leq |A| < B$, R is either F or $(1-F)$, and, as before, $\psi = W/U_2$. The symbol $\mathcal{P}$ indicates that, for $A < 0$, only the principal value of the integral is calculated. These integrals can be done by rewriting the argument of $\ln(x+B)$ as $[(x+A)+(B-A)]$ and Taylor-expanding about $(x+A)$ for $(B-A) < |x+A|$ and about $(B-A)$ for $(B-A) > |x+A|$. The result of such an integration may be sensitive to whether the system is below half filling $[F < (1-F)]$, at half filling $[F = (1-F)]$, or above half

filling $[F > (1-F)]$. However, the system as a whole has particle-hole symmetry, so one expects that the final result—once all the terms are summed—is symmetric under exchange of F and $(1-F)$. If there is no jump discontinuity when the system is precisely half full, the result for $F < (1-F)$ should determine the answer for all "finite" F, by which we mean all F such that $F\psi, (1-F)\psi \gg 1$. This thesis has been explicitly confirmed.

Indeed, the $(N_{\rm ch})^2$ part of the fourth-order relative energy shift is found to be independent of the filling fraction. Its contribution to f_ρ has a rather lengthy explicit form:

$$\begin{aligned} f^{(2B)}_{\rho,(N_{\rm ch})^2} &= \frac{(N_{\rm ch})^2 g^2}{4\pi^4\rho^2}\{-\frac{\pi^2}{6}\rho^2 + 4(1-\rho)\ln(1-\rho) \\ &+ \frac{1-\rho^2}{2}\ln^2(1-\rho) - 2(2-\rho)\ln[2(2-\rho)] \\ &+ \frac{1+\rho^2}{2}\ln(1+\rho)\ln(1-\rho) \\ &- 2\ln(3-\rho)\ln(1-\rho) + \ln^2(3-\rho) \\ &+ \frac{(3-\rho)(1-\rho)}{2}[\ln(1-\rho) - \ln(3-\rho)]^2 \\ &- (5-4\rho+\rho^2)\sum_{n=1}^{\infty}\frac{(-1)^{n+1}}{n^2}\left(\frac{1-\rho}{3-\rho}\right)^n \\ &- \frac{5\ln^2 3}{2} + 8\ln 2 + 5\kappa \\ &+ [\rho \to -\rho]\,\}, \end{aligned} \tag{3.9}$$

where the contents of the last pair of brackets indicate that one sums over all the terms in the curly braces again after replacing ρ with $-\rho$ and the quantity κ is given by

$$\kappa = \sum_{n=1}^{\infty}\frac{(-1)^{n+1}}{n^2}\left(\frac{1}{3}\right)^n. \tag{3.10}$$

As mentioned in Sec. 3.1, this result can be compared with a calculation by Grabert in the $N_{\rm ch} \to \infty$ limit [72]. Grabert calculates an average value for $\hat{n}$ in the ground state that is given by the formula

$$\langle\hat{n}\rangle = \frac{1}{4}\frac{\partial(\rho^2 f_\rho)}{\partial\rho} \tag{3.11}$$

for ρ between 0 and 1. It has been confirmed numerically that one-fourth the derivative of $\rho^2 f^{(2B)}_{\rho,(N_{\rm ch})^2}$ with respect to ρ agrees with the $(N_{\rm ch}g)^2$ term in Grabert's perturbative expansion of $\langle \hat{n} \rangle$.

A further source of comparison with $f^{(2B)}_{\rho,(N_{\rm ch})^2}$ comes from Golubev and Zaikin's weak-coupling calculation of the "effective capacitance" in the $N_{\rm ch} \to \infty$ limit [67]. Taking their effective capacitance $C_{\rm eff}$ to be related to an effective charging energy $U_{\rm eff}$ by $e^2/2C_{\rm eff} = U_{\rm eff}$, one finds that

$$U_{\rm eff}/U_2 = 1 - \frac{1}{2} \left. \frac{\partial^2(\rho^2 f_\rho)}{\partial \rho^2} \right|_{\rho=0} . \tag{3.12}$$

In Chapter 2, it was confirmed that the three weak-coupling calculations—ours, Grabert's, and Golubev and Zaikin's—give the same value for the effective charging energy through first order in $N_{\rm ch}g$. The $(N_{\rm ch}g)^2$ terms, however, do not agree. Our second-order result, as determined from Eq. 3.9, equals that derived from Grabert's calculation but is approximately seventeen times larger than that found by Golubev and Zaikin. The present computation therefore provides an important check on the large-$N_{\rm ch}$ calculations in the limit of weak coupling, resolving an apparent contradiction in the literature.

There are no comparable calculations for the terms that are linear in $N_{\rm ch}$ as these are negligible in the large-$N_{\rm ch}$ limit. However, knowing that $f = 1$ when $g = 1$ and that $f^{(1)} \approx 0.14 N_{\rm ch}g$, one might conjecture that the sign of the g^2 term changes from positive to negative when $N_{\rm ch}$ is of order 10. With respect to the expansion of f, such a crossover would imply that the coefficient of $N_{\rm ch}g^2$ is positive and approximately 10 times the size of the negative coefficient of $(N_{\rm ch})^2 g^2$.

To check this conjecture, we need to know the value of the fourth-order, linear-in-$N_{\rm ch}$

contribution to the fractional peak splitting. Our results for this quantity are

$$
\begin{aligned}
f^{(2B)}_{\rho,N_{\rm ch}} &= \frac{N_{\rm ch}g^2}{4\pi^4\rho^2}\{2\pi^2\eta^2(1-\rho)\ln(1-\rho) \\
&\quad + \frac{4}{3}(1-\rho)\ln^3(1-\rho) \\
&\quad - 2(2+\ln 2)(1-\rho)\ln^2(1-\rho) \\
&\quad + 4\left(2+\ln 2-\ln 2\ln 3-\kappa+\frac{\pi^2}{6}\right) \\
&\qquad \times (1-\rho)\ln(1-\rho) \\
&\quad - 2\left(\frac{\pi^2}{3}-2\kappa-2\ln 2\ln 3\right) \\
&\qquad \times [(3-\rho)\ln(3-\rho)-3\ln 3] \\
&\quad - \frac{2}{3}[(3-\rho)\ln^3(3-\rho)-3(3-\rho)\ln^2(3-\rho) \\
&\qquad + 6(3-\rho)\ln(3-\rho)-3\ln^3 3+9\ln^2 3-18\ln 3] \\
&\quad - 2\ln 2[(3-\rho)\ln^2(3-\rho)-2(3-\rho)\ln(3-\rho) \\
&\qquad - 3\ln^2 3+6\ln 3] \\
&\quad - 2\sum_{i=1}^{11} A_i(\rho) \\
&\quad + [\rho\to-\rho]\,\}\,,
\end{aligned}
\tag{3.13}
$$

where κ is given by Eq. 3.10, $\eta = (1/\pi)\ln[F/(1-F)]$, and the $A_i(\rho)$ are defined below:

$$
\begin{aligned}
A_1(\rho) &= -\int_0^\rho dx\,\frac{(\rho-x)\ln^2(3-x)}{1-x} \\
A_2(\rho) &= 2\int_0^\rho dx\,\frac{(\rho-x)\ln(3-x)\ln(1-x)}{1-x} \\
A_3(\rho) &= -2\ln 2\int_0^\rho dx\,\frac{(\rho-x)\ln(3-x)}{1-x}
\end{aligned}
$$

$$
\begin{aligned}
A_4(\rho) &= 2(1-\rho)\int_0^\rho dx\,\ln\left(\frac{2-x}{3-x}\right)\ln\left(\frac{1-\rho}{1-x}\right) \\
&\qquad \times\left(\frac{1}{3-x}-\frac{1}{1-x}\right) \\
A_5(\rho) &= 2\int_0^\rho dx\,(\rho-x)\ln\left(\frac{2-x}{3-x}\right)\left(\frac{1}{3-x}-\frac{1}{1-x}\right) \\
A_6(\rho) &= \int_0^\rho dx\,\frac{(\rho-x)\ln^2(1-x)}{3-x} \\
A_7(\rho) &= -2\int_0^\rho dx\,\frac{(\rho-x)\ln(3-x)\ln(1-x)}{3-x} \\
A_8(\rho) &= -2\ln 2\int_0^\rho dx\,\frac{(\rho-x)\ln(1-x)}{3-x} \\
A_9(\rho) &= -2(3-\rho)\int_0^\rho dx\,\ln\left(\frac{2-x}{3-x}\right)\ln\left(\frac{3-\rho}{3-x}\right) \\
&\qquad \times\left(\frac{1}{3-x}-\frac{1}{1-x}\right) \\
A_{10}(\rho) &= -2\int_0^\rho dx\,(\rho-x)\ln\left(\frac{2-x}{3-x}\right) \\
&\qquad \times\left(\frac{1}{3-x}-\frac{1}{1-x}\right) \\
A_{11}(\rho) &= -\int_0^\rho dx\,\ln^2(1-x)\ln(1+x)
\end{aligned}
\tag{3.14}
$$

The characterization of the fourth-order energy shift is now essentially complete. The result is more unwieldy than we would like. However, there are a few highlights that are easy to draw out. As expected, the fourth-order shift is neither ultraviolet- nor infrared-divergent but is singular as $|\rho| \to 1$, the leading singularities being in agreement with an earlier calculation by Glazman and Matveev [73]. In addition and quite gratifyingly, the solution is independent of the filling fraction F. As discussed earlier, the dependence of $f_\rho^{(2B)}$ on the filling fraction, which is concentrated in the η^2 term of the first line of Eq. 3.13, exactly cancels that of Eq. 3.8. Hence, there is some reason to believe that, when expressed in terms of the channel conductance g, the result is universal in the sense that it is independent of the details of the band structure for energies much greater than U_2, where U_2 is much less than the bandwidth W.

It is difficult to get a better handle on this algebraic smorgasbord by mere inspection.

One can add some precision to the picture of what has been accomplished by first assembling the g^2 terms of f_ρ and then plugging in $\rho = 1$ to obtain the contribution to the symmetric-dot fractional peak splitting f. Upon recalling that

$$f_\rho^{(2)} = f_\rho^{(2A)} + f_{\rho,(N_{\rm ch})^2}^{(2B)} + f_{\rho,N_{\rm ch}}^{(2B)} ,$$

one can evaluate the $A_{\rm i}$ integrals numerically for $\rho = 1$ to obtain

$$\begin{aligned} f^{(2)} \;\approx\; & [0.1491] N_{\rm ch} g^2 - [0.009798] (N_{\rm ch})^2 g^2 \\ & + O[(N_{\rm ch})^2 g^2/\psi] . \end{aligned} \tag{3.15}$$

We see that the conjecture about the $(N_{\rm ch})^2$ and $N_{\rm ch}$ contributions to $f^{(2)}$ is correct: the terms have opposite sign, and the ratio of their magnitudes is on the order of 10. For the case of $N_{\rm ch} = 2$, the g^2 term provides the desired upward correction to the f-versus-g curve.

Before specializing to the result for $N_{\rm ch} = 2$, we should explore the consequences of having a term proportional to $N_{\rm ch} g^2$. This term makes the result sensitive to the "fine structure" of the interdot conductance. As remarked in the introduction, terms of the form $(N_{\rm ch} g)^n$ can be rewritten as a simple power of the total conductance between the dots: $(N_{\rm ch} g)^n = (g_{\rm tot})^n$. Should the conductances in the various tunneling channels be allowed to differ, the form of these terms when written in terms of $g_{\rm tot}$ would remain unchanged. The only alteration would be in the equation for $g_{\rm tot}$ itself, which would revert to the more fundamental form

$$g_{\rm tot} = \sum_\sigma g_\sigma , \tag{3.16}$$

where g_σ denotes the dimensionless conductance of the σth channel.

For terms proportional to $(N_{\rm ch})^m g^n$ with $m \neq n$, the situation is quite different. Consider the $N_{\rm ch} g^2$ term in Eq. 3.15. If we had suspended the sum over channels until the end of our calculation, we would have seen that these terms are proportional to

$$[g^2]_{\rm tot} = \sum_\sigma g_\sigma^2 . \tag{3.17}$$

Only when symmetry considerations constrain all the individual channel conductances to be equal can we safely use $[g^2]_{\rm tot} = (g_{\rm tot})^2/N_{\rm ch}$.

Consequently, for the general situation in which the conductances in the separate channels are not necessarily equal, the appropriate equation for the fractional peak splitting is the following:

$$f \approx 0.1405(g_{\rm tot}) + 0.1491[g^2]_{\rm tot} - 0.009798(g_{\rm tot})^2 + \ldots \tag{3.18}$$

If we extended the expression to nth order in the dimensionless conductances, it would contain factors such as

$$[g^m]_{\rm tot} = \sum_\sigma g_\sigma^m ,$$

where $m \leq n$ and these factors appear both alone and in combination up to nth order in dimensionless conductance. The details of the "fine structure" are fully characterized by the set of $[g^m]_{\rm tot}$ for $1 \leq m \leq N_{\rm ch}$, and the fractional peak splitting can be expressed in terms of these. Further modifications might be thought necessary to account for the "hyperfine structure" that results from allowing the tunneling amplitude t in Eq. 3.1 to be a nontrivial function of $\mathbf{k_1}$ and $\mathbf{k_2}$. However, as long as the tunneling amplitude varies little over an energy range of order U_2, one would not expect Eq. 3.18 to be changed substantially.

3.3 The Strong-Coupling Limit for $N_{\rm ch} = 2$

The g^2 correction to the two-channel solution boosts confidence in the small-g end of our f-versus-g interpolation (see Fig. 3.2) but does little to improve the precision of theoretical predictions in the strong-coupling limit, a fact of particular concern for the experimentally relevant case of two interdot tunneling channels [19, 43, 44, 48]. The sections of the chapter that follow improve the strong-coupling theory for $N_{\rm ch} = 2$ in two substantial ways. The first contribution, presented here in Sec. 3.3, is the calculation of the second term in the $(1 - g)$ expansion about the $g = 1$ ground state. This term,

which is linear in $(1-g)$, is of interest both because it is significant in determining the shape of the f-versus-g curve and because, in the calculation that yields the primary $(1-g)\ln(1-g)$ term [34], the $(1-g)$ term is naively ultraviolet-divergent. The second important contribution, which comes in Sec. 3.4, is the provision of powerful evidence that the coefficients of the leading terms in the $(1-g)$ expansion are indeed independent of the high-energy structure of the theory.

To calculate in the limit of $g \to 1$, we model the tunneling link between the dots as a one-dimensional channel with a delta function scattering potential at its center. This model was originally developed for the problem of a single dot connected to a bulk lead [33, 34] but was shown in Chapter 2 and in a paper by Matveev, Glazman, and Baranger [53] to be easily adaptable to that of a pair of coupled dots. Within this ansatz, the value of the double-dot charging energy is a simple reflection of the total number of electrons that have been transferred through this channel from one side of the barrier (dot 1) to the other (dot 2). In addition, as the system is effectively one-dimensional, the fermionic degrees of freedom can be bosonized, and the Euclidean action assumes a characteristic Luttinger-liquid form [33, 34, 62]:

$$\begin{aligned} S &= S_0 + S_{\rm int} + S_{\rm b}\,, \\ S_0 &= \frac{1}{\beta}\sum_{\sigma}\sum_{\omega_m} |\omega_m||\tilde{\theta}_\sigma(\omega_m)|^2\,, \\ S_{\rm int} &= U_2 \int_0^\beta d\tau \left(\frac{1}{\sqrt{\pi}}\left[\sum_\sigma \theta_\sigma(\tau)\right] - \frac{\rho}{2}\right)^2\,, \\ S_{\rm b} &= \frac{\tilde{V}W}{2\pi}\sum_\sigma \int_0^\beta d\tau\, \cos[2\sqrt{\pi}\theta_\sigma(\tau)]\,. \end{aligned} \tag{3.19}$$

In these formulas, $\theta_\sigma(\tau)$ is a bosonic field that tracks the displacement of the one-dimensional electron gas at the barrier ($x = 0$), and $\tilde{\theta}_\sigma(\omega_m)$ is its Fourier transform:

$$\theta_\sigma(\tau) = \frac{1}{\beta}\sum_{\omega_m} e^{-i\omega_m\tau}\tilde{\theta}_\sigma(\omega_m)\,, \tag{3.20}$$

where τ is an imaginary time divided by $\hbar$, β is the inverse temperature ($\beta = 1/k_{\rm B}T$), and ω_m is $\hbar$ times a bosonic Matsubara frequency ($\omega_m = 2\pi m/\beta$). In addition, $\tilde{V}$ is a

measure of the barrier strength defined by $\tilde{V} = V_0/\hbar v_F$ for the delta-function potential $V_0\delta(x)$. As for the remaining parameters, v_F is the Fermi velocity and, as in the weak-coupling theory, W is the bandwidth—the difference between the lowest and highest single-particle energies in the channel. The inverse temperature β will be taken to infinity in calculating the energy of the ground state.

As in the weak-coupling theory, we ultimately want to parametrize the coupling between the dots by the dimensionless channel conductance g, rather than the barrier strength $\tilde{V}$. Accordingly, we need to find the relation between g and $\tilde{V}$. In our single-mode channel, g equals the single-particle transmission probability T, and $(1-g)$ equals the reflection probability R. The leading dependence of the channel conductance on $\tilde{V}$ equals what one would guess from the reflection probability of a single particle incident upon a one-dimensional delta-function potential [83]:

$$(1-g) = \tilde{V}^2 + O(\tilde{V}^4)\,. \tag{3.21}$$

Inverting this formula, we find that

$$\tilde{V}^2 = (1-g) + O[(1-g)^2]\,. \tag{3.22}$$

To lowest order, we have the approximation of Matveev [34], $\tilde{V} = \sqrt{1-g}$, which—it will be seen—is all that is required for the calculations in this chapter.

Having prepared ourselves to switch from a solution in terms of $\tilde{V}$ to one in terms of g, we proceed with the calculation of the ground-state energy. Our first move is to reorganize the action, expressing it in terms of bosonic fields that characterize the net charge and pseudospin degrees of freedom, where the pseudospin degrees of freedom correspond to "true spin" only if the two channels correspond to spin-up and spin-down, respectively. Defining the charge field by $\theta_c = \theta_1 + \theta_2 + \sqrt{\pi}\rho/2$ and the pseudospin field

by $\theta_{\rm s} = \theta_1 - \theta_2$, we find

$$\begin{aligned} S &= S_0^{(\rm s)} + S_0^{(\rm c)} + S_{\rm b}\,, \\ S_0^{(\rm s)} &= \frac{1}{2\beta}\sum_{\omega_m} |\omega_m| |\tilde{\theta}_{\rm s}(\omega_m)|^2\,, \\ S_0^{(\rm c)} &= \frac{1}{2\beta}\sum_{\omega_m} \left(|\omega_m| + \frac{2U_2}{\pi}\right) |\tilde{\theta}_{\rm c}(\omega_m)|^2\,, \\ S_{\rm b} &= \frac{\tilde{V}W}{\pi}\int_0^\beta d\tau\, \cos\left[\sqrt{\pi}\theta_{\rm c}(\tau) + \frac{\pi\rho}{2}\right] \cos\left[\sqrt{\pi}\theta_{\rm s}(\tau)\right]\,. \end{aligned} \tag{3.23}$$

The Euclidean action has now been written in terms of "high-energy" charge modes and "low-energy" pseudospin modes. We advance by integrating out the "high-energy" charge degrees of freedom. This integration is analogous to a renormalization in which one integrates out the higher-energy degrees of freedom within a particular channel [84]. One begins with the generating functional for the Euclidean action of Eq. 3.23:

$$Z = \int D[\theta_{\rm s}(\tau)] \int D[\theta_{\rm c}(\tau)]\, e^{-S[\theta_{\rm s}(\tau), \theta_{\rm c}(\tau)]}\,, \tag{3.24}$$

where, as usual, time-ordering is implicit within the functional integral approach. One then performs the integration over the fast modes to obtain the generating functional for an effective action depending only on the slow modes:

$$\begin{aligned} Z_{\rm s} &= \int D[\theta_{\rm s}(\tau)]\, e^{-S_{\rm eff}[\theta_{\rm s}(\tau)]}\,, \\ e^{-S_{\rm eff}[\theta_{\rm s}(\tau)]} &= \frac{e^{-S_0^{(\rm s)}} \int D[\theta_{\rm c}(\tau)]\, e^{-S_0^{(\rm c)}} e^{-S_{\rm b}}}{\int D[\theta_{\rm c}(\tau)]\, e^{-S_0^{(\rm c)}}}\,. \end{aligned} \tag{3.25}$$

Equation 3.25 determines the effective action $S_{\rm eff}$. To solve for it, one Taylor-expands the exponential factor $e^{-S_{\rm b}}$, performs the integral over charge degrees of freedom, and re-exponentiates the result. Before doing any of this, it is useful to make the following definition:

$$\langle \hat{A} \rangle_{\rm c} = \frac{\int D[\theta_{\rm c}(\tau)]\, \hat{A}\, e^{-S_0^{(\rm c)}}}{\int D[\theta_{\rm c}(\tau)]\, e^{-S_0^{(\rm c)}}}\,. \tag{3.26}$$

One can then rewrite Eq. 3.25 as follows:

$$\begin{aligned} e^{-S_{\rm eff}} &= e^{-S_0^{(s)}} \langle e^{-S_b} \rangle_c \\ &= e^{-S_0^{(s)}} \left[1 - \langle S_b \rangle_c + \frac{1}{2} \langle S_b^2 \rangle_c + O(\tilde{V}^3) \right] . \end{aligned} \tag{3.27}$$

Upon re-exponentiation, one obtains

$$S_{\rm eff} = S_0^{(s)} + \langle S_b \rangle_c - \frac{1}{2} \left\langle [S_b - \langle S_b \rangle_c]^2 \right\rangle_c + O(\tilde{V}^3) . \tag{3.28}$$

It is clear that to solve for the effective action to order $\tilde{V}^2 = (1-g)$, we must solve for both corrections to $S_0^{(s)}$ on the right side of Eq. 3.28.

Details of the calculation of these terms are presented in Appendix B. The result is that $S_{\rm eff} = S_0^{(s)} + S_b^{(1)} + S_b^{(2)}$, where

$$\begin{aligned} S_b^{(1)} &= \frac{\tilde{V}W}{\pi} e^{-\frac{\pi}{2}K_c(0)} \cos\left(\frac{\pi\rho}{2}\right) \int_0^\beta d\tau \, \cos\left[\sqrt{\pi}\theta_s(\tau)\right] , \\ S_b^{(2)} &= \left(\frac{\tilde{V}W}{\pi}\right)^2 e^{-\pi K_c(0)} \int_0^\beta d\tau_1 \int_0^{\tau_1} d\tau_2 \\ &\qquad \times \{\cos^2\left(\frac{\pi\rho}{2}\right) \left[1 - e^{-\pi K_c(\tau_1-\tau_2)}\right] \\ &\qquad\quad - \sinh[\pi K_c(\tau_1-\tau_2)]\} \\ &\qquad \times \cos\left[\sqrt{\pi}\theta_s(\tau_1)\right] \cos\left[\sqrt{\pi}\theta_s(\tau_2)\right] . \end{aligned} \tag{3.29}$$

The function $K_c(\tau)$ that appears in these formulas is the charge-channel correlation function, $K_c(\tau) = \langle \theta_c(\tau)\theta_c(0) \rangle_c$. Its numerical value can be found from the formula

$$K_c(\tau) = \frac{1}{\pi} {\rm Re} \int_0^\infty d\omega \, \frac{e^{-(\frac{2}{W}+i\tau)\omega}}{\omega + \frac{2U_2}{\pi}} . \tag{3.30}$$

To progress further, we define a new "unperturbed action" $S_{\rm New} = S_0^{(s)} + S_b^{(1)}$. We then write down the Hamiltonian that corresponds to this action:

$$\begin{aligned} H_{\rm New} &= H_0^{(s)} + H_b^{(1)} , \\ H_b^{(1)} &= \frac{\tilde{V}W}{\pi} e^{-\frac{\pi}{2}K_c(0)} \cos\left(\frac{\pi\rho}{2}\right) \cos\left[\sqrt{\pi}\theta_s(0)\right] . \end{aligned} \tag{3.31}$$

This is the Hamiltonian diagonalized by Matveev [34] through a process of "debosonization" (see Appendix B) in which the Hamiltonian is rewritten in terms of fermion operators d_k and d:

$$\begin{aligned} H_0^{(s)} &= \int_{-\Lambda}^{\Lambda} dk\, \xi_k d_k^\dagger d_k\,, \\ H_{\rm b}^{(1)} &= \lambda \int_{-\Lambda}^{\Lambda} dk \left[d_k^\dagger (d + d^\dagger) + (d + d^\dagger) d_k \right]. \end{aligned} \tag{3.32}$$

Here the single-particle energy ξ_k, the fermion interaction parameter λ, and the wavevector cutoff Λ have the formulas $\xi_k = \hbar v_F k$, $\lambda = \tilde{V} \cos(\pi\rho/2)\sqrt{2e^\gamma \hbar v_F U_2/\pi^3}$, and $\Lambda = W/2\hbar v_F$.

Since the Hamiltonian is now quadratic in fermion operators, a Bogoliubov transformation brings it to the desired diagonal form:

$$H_{\rm New} = E_{\rm New}^{(0)} + \int_0^{\Lambda} dk\, \xi_k \left(C_k^\dagger C_k + \tilde{C}_k^\dagger \tilde{C}_k \right), \tag{3.33}$$

where, if we write down only the terms of lowest order in $\tilde{V}$, replacing all others by an ellipsis,

$$\begin{aligned} \tilde{C}_k &= \frac{1}{\sqrt{2}}(d_k + d_{-k}^\dagger), \\ C_k &= \frac{1}{\sqrt{2}}(d_k - d_{-k}^\dagger) + \ldots \end{aligned} \tag{3.34}$$

The correction to the $\tilde{V} = 0$ ground-state energy is produced by the omitted terms in C_k (for details, see Appendix B). In particular, using $\Delta_{\rm str}^{(1)}$ to represent the difference between the ground-state energies of $H_{\rm New}$ for $\tilde{V} = 0$ and for arbitrary $\tilde{V}$, respectively, one finds that

$$\begin{aligned} \Delta_{\rm str}^{(1)}(\rho) &= \frac{4e^\gamma U_2}{\pi^3} \tilde{V}^2 \cos^2\!\left(\frac{\pi\rho}{2}\right) \\ &\quad \times \left(\ln\!\left[\tilde{V}^2 \cos^2\!\left(\frac{\pi\rho}{2}\right)\right] - \ln\!\left[\frac{\psi}{2}\right] \right. \\ &\quad \left. -1 + \ln\!\left[\frac{8e^\gamma}{\pi^2}\right] \right). \end{aligned} \tag{3.35}$$

As before, $\psi = W/U_2$, where the bandwidth $W \gg U_2$.

We see that the result for this first correction contains terms that are quadratic in $\tilde{V}$ and logarithmically divergent in ψ. This ultraviolet divergence is circumvented by the statement that one should replace W by U_2 because keeping only the first term from charge-channel integration is only a good approximation for energies less than the charging energy U_2 [34]. The terms in Eq. 3.35 that are merely quadratic in $\tilde{V}$ are thereby rendered finite and can be dropped in favor of the leading $\tilde{V}^2\cos^2(\pi\rho/2)\ln[\tilde{V}^2\cos^2(\pi\rho/2)]$ dependence.

To eliminate the logarithmic divergence more formally, one must calculate the shift in the ground-state energy that is induced by $S_{\rm b}^{(2)}$ (recall Eq. 3.29). As this term is itself quadratic in $\tilde{V}$ and as we are only interested in knowing the ground-state energy to order $\tilde{V}^2$, we can drop all but the leading part of the $S_{\rm b}^{(2)}$-induced shift. In expressing $S_{\rm b}^{(2)}$ in terms of the diagonalizing operators of $H_{\rm New}$, one may use the truncated formulas of Eq. 3.34. The relevant shift in the ground-state energy is then found by calculating the expectation value of $S_{\rm b}^{(2)}$ in the ground state of $H_{\rm New}$ (see Appendix B):

$$\begin{aligned}\Delta_{\rm str}^{(2)}(\rho) &= \frac{4e^{\gamma}U_2}{\pi^3}\tilde{V}^2\cos^2\left(\frac{\pi\rho}{2}\right) \\ &\quad\times\int_0^{\infty}dx\,\left[1-e^{-\pi K_{\rm c}\left(\frac{2x}{W}\right)}\right]\frac{1-e^{-x}}{x}, \end{aligned} \tag{3.36}$$

where units have been chosen such that $\hbar = 1$ and terms independent of ρ have been dropped since they are not relevant to evaluation of the fractional peak splitting f. It is not too hard to see that the factor $[1-e^{-\pi K_{\rm c}(2x/W)}]$ in the integrand makes for an ultraviolet cutoff of order $\psi = W/U_2$ (see Appendix B). It is even easier to see that $(1-e^{-x})$ provides an infrared cutoff of order 1. Thus, one can surmise that the leading term from the integral is $\ln(\psi/2)$, which is precisely what is needed to cancel the ultraviolet divergence in $\Delta_{\rm str}^{(1)}$.

What remains is to calculate the rest of the integral in Eq. 3.36, which we call Φ:

$$\begin{aligned}\Phi &= \lim_{\psi\to\infty}\left(\int_0^{\infty}dx\,\left[1-e^{-\pi K_{\rm c}\left(\frac{2x}{W}\right)}\right]\frac{1-e^{-x}}{x}\right. \\ &\quad\left.-\ln\left[\frac{\psi}{2}\right]\right). \end{aligned} \tag{3.37}$$

Numerical approximation of the integral in the limit $\psi \to \infty$ gives $\Phi = 0.1703 \pm 0.0002$.

One can now sum $\Delta_{\rm str}^{(1)}$ and $\Delta_{\rm str}^{(2)}$ to get the strong-coupling energy shift through order $\tilde{V}^2$. Having dropped terms that are independent of ρ, one has

$$\begin{aligned}\Delta_{\rm str}(\rho) &= \frac{4e^{\gamma}U_2}{\pi^3}\tilde{V}^2\cos^2\left(\frac{\pi\rho}{2}\right)\left(\ln\left[\tilde{V}^2\cos^2\left(\frac{\pi\rho}{2}\right)\right]\right. \\ &\qquad \left. -1+\Phi+\ln\left[\frac{8e^{\gamma}}{\pi^2}\right]\right)+\ldots \end{aligned} \tag{3.38}$$

We can now straightforwardly compute the fractional peak-splitting f in terms of the dimensionless conductance g. As mentioned before, if we are only interested in obtaining the ground-state energy to order $(1-g)$, only the leading term of Eq. 3.22 is relevant in converting Eq. 3.38 to an expression in terms of $(1-g)$. The value of $f = f_{\rho=1}$ follows from the fact that, in the strong-coupling limit,

$$f_{\rho} = 1 - \frac{\Delta_{\rm str}(\rho) - \Delta_{\rm str}(0)}{U_2\rho^2/4}. \tag{3.39}$$

In particular, Eqs. 3.22, 3.38, and 3.39 yield

$$\begin{aligned} f &= 1+\frac{16e^{\gamma}}{\pi^3}(1-g)\ln(1-g) \\ &\quad -\frac{16e^{\gamma}}{\pi^3}\left[1-\ln\left(\frac{8e^{\gamma}}{\pi^2}\right)-\Phi\right](1-g)+\ldots \end{aligned} \tag{3.40}$$

Since $\Phi \approx 0.1703$, we have

$$f \approx 1 + 0.919(1-g)\ln(1-g) - 0.425(1-g) + \ldots \tag{3.41}$$

Having determined the first corrections to the leading behaviors for both $g \to 0$ and $g \to 1$, we now have a more plausible picture for the connection between the $N_{\rm ch} = 2$ weak- and strong-coupling limits (see Fig. 3.2). The fit to the data could be improved if the interdot capacitance were larger than experimentally estimated [19, 44, 51] or if asymmetry between the dots were important [53, 54]. In any case, whether or not such further emendations should be made, the theory is within the range of present experimental error. The corrections introduced in this chapter have moved the

weak- and strong-coupling predictions by reasonable amounts in the desired directions, increasing both the ease and the precision of interpolation between the weak- and strong-coupling limits.

3.4 Insensitivity to the High-Energy Density of States

3.4.1 Insensitivity to the Functional Form of the Bosonic Cutoff

To have confidence that our coupled-dot calculations can be usefully compared to empirical data, we should make sure that the result, f expressed as a function of g, is independent of the details of the band structure far from the Fermi surface, where the assumption of a constant density of states becomes invalid. We have done much to confirm such robustness in the regime of weak coupling, for we have shown there that, through second order in g, $f(g)$ is independent of the bandwidth W and the filling fraction F as long as both FW and $(1-F)W$ are much larger than the charging energy U_2. Such dual invariance indicates that we can simply shear off a nontrivial number of high-energy states without affecting the result. We would expect then that we could make less Draconian modifications of the high-energy density of states with similarly perfect impunity.

With regard to the strong-coupling theory, matters have been left less assured. In Chapter 2, we introduced a factor of C_2 multiplying the first term in Eq. 3.40 to guard against the possibility that the coefficient of the energy shift calculated via bosonization was partly a product of the approach itself and, in particular, the manner in which the ultraviolet cutoff was imposed. Concern about such a possibility arises from the fact that the leading term in the $(1-g)$ expansion is proportional to the product of $e^{-\frac{\pi}{2}K_c(0)}$ and $\mathcal{A}$, where $\mathcal{A}$ is the generalization to non-exponential cutoffs of the normalization factor in Eq. B.10 that gives the proportionality between the fermionic position operators and

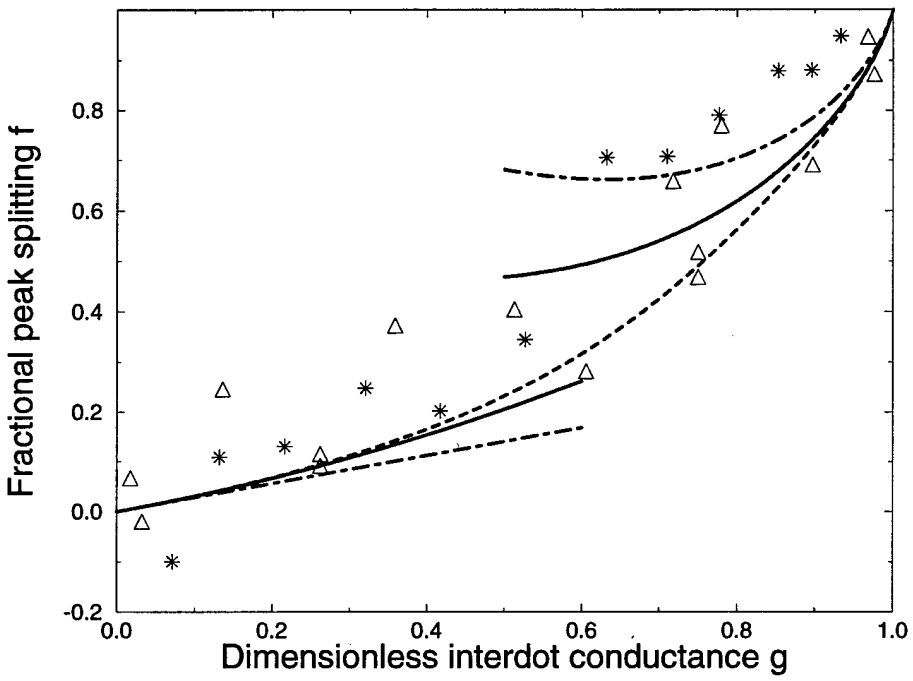

Figure 3.2: Graph of the fractional Coulomb-blockade peak splitting f as a function of the dimensionless conductance per channel g in the weak- and strong-tunneling limits for $N_{\rm ch} = 2$. The new theoretical curves are depicted as solid lines. The old theoretical curves from Chapter 2 are dot-dashed lines. The dashed curve shows a possible interpolating function. Data points from Waugh *et al.* [19, 43, 44] are given as triangles or stars; the two different symbols correspond to different data sets. The value of f for the experimental data has been extracted from the measured splitting fraction f' by using the method discussed in Chapter 2 with experimentally estimated values of 20 aF for the constant interdot capacitance and 0.4 fF for the total single-dot capacitance [19, 43, 44].

the exponentials of bosonic fields:

$$\psi_f^\dagger(0,\tau) = \mathcal{A} e^{i\sqrt{\pi}\phi_f(\tau)} . \tag{3.42}$$

Changing the nature of the bosonic cutoff [e.g., from the exponential $e^{-\alpha|\omega|}$ to the Gaussian $e^{-(\pi/4)\alpha^2\omega^2}$] causes the value of $e^{-\frac{\pi}{2}K_c(0)}$ to be multiplied by a constant factor. Although one would hope that a similar shift in the value of $\mathcal{A}$ compensates for the change in $e^{-\frac{\pi}{2}K_c(0)}$, to the authors knowledge, such a happy circumstance has not previously been checked to be true.

Similar questions could be asked about the prefactor for the term linear in $(1-g)$, with which we associate a factor C_3, where $C_3 = 1$ for the Luttinger-liquid approach with the standard exponential cutoff. This term is proportional both to $|\mathcal{A}|^2 e^{-\pi K_c(0)}$ and to an integral that depends upon $e^{-\pi K_c(\tau)}$ (see Eq. 3.36 in Sec. 3.3 and Eq. B.26 in Appendix B). Hence, in order to prove that the two leading strong-coupling terms do not vary with the choice of cutoff function, one must show that neither the product $\mathcal{D}_1 = |\mathcal{A}|^2 e^{-\pi K_c(0)}$ nor the integral

$$\begin{aligned} \mathcal{D}_2 \;=\; & \frac{1}{\beta W}\int_0^{\beta W/2} dx \left(\frac{\beta W}{2} - x\right) \\ & \times \left[1 - e^{-\pi K_c(2x/W)}\right] \frac{1-e^{-x}}{x} \end{aligned} \tag{3.43}$$

assumes different values when the shape of the cutoff is changed. Though we do not have a general proof that $\mathcal{D}_1$ and $\mathcal{D}_2$ are independent of the cutoff function, we can show that they remain the same for a whole class of functions that includes the exponential cutoff and that they are similarly unchanged when one replaces the exponential cutoff by a Gaussian. We believe that these facts are convincing evidence that the prefactors in Eq. 3.40 are insensitive to the nature of the high-energy cutoff.

First, we prove that $\mathcal{D}_1$ and $\mathcal{D}_2$ are the same for all cutoffs of the form

$$\nu(\omega,\alpha,\{b_m\}) = e^{-\alpha|\omega|}\left(1 + \sum_{m=1}^{M} b_m \alpha^m |\omega|^m\right), \tag{3.44}$$

where either M is finite or, for large m, b_m falls to zero faster than $m^{-\zeta}/m!$ for some real $\zeta > 0$. As usual, it is assumed that $\alpha U_2 \ll 1$, where $\alpha = 2/W$. We add the further assumption that $[(m-1)!\, b_m \alpha U_2] \ll 1$ for all m.

The first step in our proof is to solve for the change in $e^{-\pi K_c(0)}$ when one goes from the standard exponential cutoff $\nu(\omega, \alpha, \{0\})$ to the more general form $\nu(\omega, \alpha, \{b_m\})$. The formula for $K_c(\tau)$ (recall Eq. 3.30) becomes

$$K_c(\tau) = \frac{1}{\pi} \mathrm{Re} \int_0^\infty d\omega\, \nu(\omega, \alpha, \{b_m\}) \frac{e^{-i\tau\omega}}{\omega + \frac{2U_2}{\pi}}. \tag{3.45}$$

We can write the change in $K_c(0)$ as

$$\delta K_c(0) = \frac{1}{\pi} \sum_{m=1}^{M} b_m \int_0^\infty d\omega \frac{\alpha^m \omega^m e^{-\alpha\omega}}{\omega + \frac{2U_2}{\pi}}. \tag{3.46}$$

Using $\omega = (\omega + 2U_2/\pi - 2U_2/\pi)$ and the binomial theorem, we can expand ω^m in powers of $(\omega + 2U_2/\pi)$. The integration is then straightforward and yields

$$\pi \delta K_c(0) = \sum_{m=1}^{M} b_m (m-1)! \,[1 + O(\alpha U_2)]. \tag{3.47}$$

Dropping the correction, we have the result

$$e^{-\pi K_c(0)} = \left[e^{-\sum b_m (m-1)!}\right] e^{-\pi K_{c,0}(0)}, \tag{3.48}$$

where $K_{c,0}(\tau)$ is the correlation function for the standard exponential cutoff.

Calculation of the change in the normalization constant $\mathcal{A}$ is more complicated. Following V. J. Emery [55], we find

$$|\mathcal{A}|^{-2} = \int_{-\infty}^{\infty} dx\, e^{\int_0^\infty d\omega\, \nu(\omega,\alpha,\{b_m\}) \frac{e^{i\omega x/\hbar v_F} - 1}{\omega}} + c.c. \tag{3.49}$$

A bit of calculation reveals that

$$\begin{aligned} |\mathcal{A}|^{-2} \;=\; & \left[e^{-\sum_{m=1}^{M} b_m (m-1)!}\right] \int_{-\infty}^{\infty} dx \left(\frac{\alpha}{\alpha - ix}\right) \\ & \times e^{\sum b_m (m-1)! \left(\frac{\alpha}{\alpha - ix}\right)^m} + c.c. \end{aligned} \tag{3.50}$$

It is apparent that the bracketed factor exactly cancels the factor that multiplies $e^{-\pi K_{c,0}(0)}$ in Eq. 3.48. Thus, in order for $\mathcal{D}_1 = |\mathcal{A}|^2 e^{-\pi K_c(0)}$ to be unaltered, the value of the integral in Eq. 3.50 cannot change as the b_m are varied. In short, the partial derivative of the integral with respect to each of these coefficients must be zero. The partial derivative with respect to b_m is given by the following formula:

$$\begin{aligned} P_m &= (m-1)! \int_{-\infty}^{\infty} dx \left(\frac{\alpha}{\alpha - ix}\right)^{m+1} \\ &\quad \times e^{\sum b_m (m-1)! \left(\frac{\alpha}{\alpha - ix}\right)^m} + c.c. \end{aligned} \tag{3.51}$$

Let $z = \alpha/(\alpha - ix)$. The resulting integral in the complex z plane follows a closed path, beginning and ending at $z = 0$:

$$P_m = -i\alpha(m-1)! \oint dz\, z^{m-1}\, e^{\sum b_m (m-1)!\, z^m} + c.c. \tag{3.52}$$

For M finite or b_m falling off faster than $m^{-\zeta}/m!$, the integrand is analytic throughout the region enclosed by the contour. By Cauchy's Theorem, $P_m = 0$.

Since the integral of Eq. 3.50 does not vary with b_m, we can make the statement

$$|\mathcal{A}|^2 = \left[e^{\sum_{m=1}^{M} b_m (m-1)!}\right] |\mathcal{A}_0|^2, \tag{3.53}$$

where $\mathcal{A}_0$ is the normalization factor for the standard exponential cutoff. Equations 3.48 and 3.53 yield

$$\mathcal{D}_1 = |\mathcal{A}|^2 e^{-\pi K_c(0)} = |\mathcal{A}_0|^2 e^{-\pi K_{c,0}(0)}. \tag{3.54}$$

We have now shown that, for the class of cutoffs $\nu(\omega, \alpha, \{b_m\})$, C_2 is constant.

What about C_3? To determine its fate, we must find the change in the quantity $\mathcal{D}_2$ (recall Eq. 3.43). After substituting α for $(2/W)$, we follow essentially the same path that we blazed in determining the change in $K_c(0)$ and find that the change in $K_c(\alpha x)$ is given by the formula

$$\delta K_c(\alpha x) = \frac{1}{2\pi} \sum_{m=1}^{M} b_m (m-1)! \left[\frac{1}{(1+ix)^m} + c.c.\right]. \tag{3.55}$$

Employing this, we can break the integral on the right side of Eq. 3.43 into two parts, the first of which is from 0 to $\psi^{1-\epsilon}$, where $0 < \epsilon < 1$ and $\psi = (W/U_2) \gg 1$. In this interval, the contribution from the entire term proportional to $e^{-\pi K_c(\alpha x)}$ can be shown to be zero in the limit $\psi \to \infty$. In the remainder, $\delta K_c(\alpha x)$ is on the order of $1/x^2$, which implies that the correction due to the generalization of $\nu(\omega, \alpha, \{b_m\})$ is proportional to

$$\int_{\psi^{1-\epsilon}}^{\beta W/2} dx \, \frac{e^{-\pi K_{c,0}(2x/W)}}{x^3} \leq \int_{\psi^{1-\epsilon}}^{\beta W/2} \frac{dx}{x^3},$$

which also equals zero in the limit $\psi \to \infty$. Therefore, $\mathcal{D}_2$ is constant, and we have proven that our strong-coupling results are insensitive to varying the cutoffs within the class $\nu(\omega, \alpha, \{b_m\})$.

The values of $\mathcal{D}_1$ and $\mathcal{D}_2$ can be shown to be similarly unaltered when we switch from the exponential cutoff to a Gaussian:

$$\nu_G(\omega, \alpha) = e^{-\frac{\pi}{4}\alpha^2\omega^2}. \tag{3.56}$$

Solving for $e^{-\pi K_c(0)}$ with this weight function, one discovers that

$$e^{-\pi K_{c,G}(0)} = \frac{e^{\gamma/2}}{\sqrt{\pi}} \alpha U_2, \tag{3.57}$$

where γ is once again the Euler-Mascheroni constant. The normalization coefficient $\mathcal{A}_G$ has not been solved for analytically. However, starting from Eq. 3.49, one finds that

$$\begin{aligned} |\mathcal{A}_G|^{-2} &= \pi\alpha \int_0^\infty dx \cos\left[\frac{\pi}{2}\mathrm{Erf}(x/2)\right] \\ &\quad \times e^{\frac{i\sqrt{\pi}}{2}\int_0^x dy\, e^{-y^2/4}\,\mathrm{Erf}(iy/2)}, \end{aligned} \tag{3.58}$$

where $\mathrm{Erf}(x) = (2/\sqrt{\pi})\int_0^x dt\, e^{-t^2}$ is the error function. It has been confirmed numerically that through at least 12 digits the product $\mathcal{D}_{1,G} = |\mathcal{A}_G|^2 e^{-\pi K_{c,G}(0)}$ agrees with the exponential cutoff. By arguments similar to those used for the class of cutoff functions studied above, it has also been shown that in the limit $\psi \to \infty$, the integral $\mathcal{D}_{2,G}$ is the same as for the exponential. The coefficients in Eq. 3.40 are again unaltered, and it seems reasonable to suppose that the invariance is general.

3.4.2 Insensitivity to the Fermionic Filling Fraction

Thus, it appears fairly certain that modifying the high-energy density of states in the bosonized theory does not affect the results of Sec. 3.3. Nevertheless, having solved the weak-coupling model for the general case of a fermionic system not necessarily at half filling and having seen that, when expressed as functions of the tunneling amplitude, both the conductance and the fractional peak splitting depend upon the filling fraction, one might wonder what happens to the strong-coupling results when one begins with a fermionic system that is not necessarily at half filling. Since Luttinger-style bosonization assumes symmetry between occupied and empty states, such a system can only be properly bosonized after the asymmetric fermion states have been integrated out. For example, if the system is below half filling $[F < (1-F)]$ and the zero of energy is at the Fermi surface ($\epsilon_F = 0$), the fermionic single-particle states with energies between FW and $(1-F)W$ must be integrated out, leaving a symmetric effective theory with single-particle energies ranging from $-FW$ to FW. Only after this symmetrization can the theory be bosonized without losing knowledge of the fermionic filling fraction F.

The task before us, therefore, is to "symmetrize" the fermionic theory that lies behind the bosonized action of Eq. 3.19. The archetypal fermionic Hamiltonian consists of the usual three parts: the single-particle kinetic energies, the multiparticle potential energy, and the backscattering barrier. The Hamiltonian therefore takes the following form:

$$\begin{aligned} H &= H_K + H_C + H_B\,, \\ H_K &= \sum_{j=1}^{2}\sum_{\sigma}\sum_{k} \xi_k c^{\dagger}_{jk\sigma} c_{jk\sigma}\,, \\ H_C &= U_2(\hat{n}-\rho/2)^2\,, \\ H_B &= \sum_{\sigma}\sum_{k_1 k_2} v(c^{\dagger}_{2\mathbf{k}_2\sigma} c_{1\mathbf{k}_1\sigma} + \text{H.c.})\,, \end{aligned} \tag{3.59}$$

where $\xi_k = \hbar v_F k$ and j is the index that distinguishes between right-movers ($j = 1$) and left-movers ($j = 2$).

The operator $\hat{n}$ is now somewhat more complicated than in the weak-coupling theory. In its simplest form, it can be written as

$$\hat{n} = \frac{1}{2}\sum_{j=1} \int dx \, [\Theta(x) - \Theta(-x)] \, \psi_j^\dagger(x)\psi_j(x) \,, \tag{3.60}$$

where $\Theta(x)$ is the Heaviside step function and ψ_j is the annihilation operator in position space for a right-moving ($j = 1$) or a left-moving ($j = 2$) fermion. After writing the components of the integrand in the momentum representation and integrating over x, one finds that, for a one-dimensional system of length L,

$$\hat{n} = \frac{-i}{L}\sum_j \sum_\sigma \sum_{k_1 k_2} \frac{c^\dagger_{jk_2\sigma} c_{jk_1\sigma}}{k_2 - k_1}(1 - \delta_{k_1,k_2}) \,, \tag{3.61}$$

which is equivalent to the integral version obtained by Matveev [34] from the observation that $d\hat{n}/dt$ equals the current operator at $x = 0$, the point of "division" between the two dots. (This point is, of course, not entirely well-defined in the limit $g \to 1$.)

The above equations for the Hamiltonian and number operator are presented as discrete sums. For future reference in implementing the symmetrization of the theory, we write the components of our Hamiltonian in integral form:

$$\begin{aligned} H_K &= \left(\frac{\hbar v_F}{\delta}\right) \sum_{j=1}^{2} \int dk \, \xi_k \, c^\dagger_{jk\sigma} c_{jk\sigma} \,, \\ H_C &= U_2(\hat{n} - \rho/2)^2 \,, \\ H_B &= v\left(\frac{\hbar v_F}{\delta}\right) \sum_\sigma \int dk_1 \int dk_2 \, (c^\dagger_{2\mathbf{k}_2\sigma} c_{1\mathbf{k}_1\sigma} + \text{H.c.}) \end{aligned} \tag{3.62}$$

where δ is the level-spacing for the one-dimensional system ($\delta = 2\pi\hbar v_F/L$) and

$$\hat{n} = \frac{-i}{2\pi}\sum_j \sum_\sigma \mathcal{P} \int dk_1 \int dk_2 \frac{c^\dagger_{jk_2\sigma} c_{jk_1\sigma}}{k_2 - k_1} \,. \tag{3.63}$$

For the fermionic strong-coupling model of Eq. 3.59, calculation of the channel conductance between the dots proceeds along the same lines as for weak coupling (see Chapter 2). In fact, since the density of states is constant in both theories, setting

$U_2 = 0$—the first step in the conductance calculation—renders them essentially identical, the only differences being in the last term, where t has been replaced by v and the index i for dot-1 or dot-2 fermions has been replaced by the index j for right-movers or left-movers. Accordingly, unlike the weak-tunneling term H_T, the perturbation H_B scatters fermions backward instead of transporting them forward and therefore causes a reduction in the conductance of the unperturbed system. Recalling the size of the conductance induced by H_T in the weak-coupling model, it is not hard to see that the channel conductance in the strong-coupling model is given by

$$g = 1 - \frac{4\chi}{|1 + (1 + i\eta)^2\chi|^2} , \tag{3.64}$$

where $\chi = (\pi v/\delta)^2$ and $\eta = (1/\pi)\ln[F/(1-F)]$. As in the weak-coupling theory, the result becomes troublesome as χ becomes large. However, we should be able to trust its testimony that the filling fraction does not affect the interdot conductance through second-order in $(\pi v/\delta)$.

This is all we need to know, for $(\pi v/\delta)$ can be straightforwardly written in terms of our previous strong-coupling parameter $\tilde{V}$. The relation is $\tilde{V} = 2(\pi v/\delta)$, and it follows that $\chi = (\tilde{V}/2)^2$. We recover the leading-order result of Eq. 3.22 and see that, to order $\tilde{V}^2$, the channel conductance g is independent of the fermionic filling fraction F. If we can likewise show that the relation between $\tilde{V}$ and the differential energy shift

$$\delta\Delta(\rho) = [\Delta_{\rm str}(\rho) - \Delta_{\rm str}(0)] \tag{3.65}$$

does not depend on the filling fraction F, we will know that the same is true for our final strong-coupling result, the expression for $f(g)$ in Eq. 3.40.

To prove $\delta\Delta$'s invariance with respect to F, we symmetrize the fermionic theory through a renormalization in which we integrate out all single-particle states at an energy distance of $W'/2$ or more from the Fermi surface, where $U_2 \ll W' \ll W$. The resulting symmetric theory with bandwidth W' can be bosonized without further qualm. However, as renormalization generates terms that are not present in the original

Hamiltonian, we must check to see what relevant effects these have upon the low-energy theory. We must also keep track of any contributions to $\delta\Delta$ that arise from the high-energy degrees of freedom alone.

Before we go about doing this, a comment on our approach is in order. One might view the proposed renormalization as occurring in two distinct stages: first, we integrate out the asymmetric particle-hole states; then, we integrate both particle and hole states down to energy W'. Since all that we will need to consider are the general scaling properties of the terms generated during the renormalization process, the distinction between the stages is of no importance and is henceforth ignored.

The argument resumes. Since our interest is in the Coulomb blockade, the renormalization scheme we use is designed to leave the Coulombic interaction term H_C unchanged. After wave-vectors between the original wave-vector cutoff Λ and the new wave-vector cutoff Λ/b (where $b > 1$) have been integrated out, the theory is re-scaled by writing it in terms of a new set of wave-vectors $k_b = bk$. Invariance of H_C is achieved by re-scaling the fermion creation and annihilation operators as well: $c^\dagger_{jk_b\sigma} = b^{-1/2} c^\dagger_{jk\sigma}$. [One might prefer to say that the coherent-state Grassman variables that correspond to the operators are re-scaled (see Ref. [97]).] The effect of renormalization upon the parameters $(\hbar v_F/\delta)$, U_2, and $v(\hbar v_F/\delta)$ of Eq. 3.62 is as follows:

$$\begin{aligned} \left[\frac{\hbar v_F}{\delta}\right]' &= b^{-1}\left[\frac{\hbar v_F}{\delta}\right], \\ [U_2]' &= U_2, \\ [v(\hbar v_F/\delta)]' &= b^{-1}\,[v(\hbar v_F/\delta)]\,. \end{aligned} \tag{3.66}$$

The backscattering H_B is revealed to be dangerously irrelevant. Though it scales like an irrelevant term, we cannot safely set it to zero as we know from Eq. 3.35 that the energy shift is singular as $v \to 0$.

In addition to rescaling the terms in the original Hamiltonian, renormalization generates terms of its own. It is not hard to see that all but the new backscattering terms

are irrelevant. The original Hamiltonian consists of the kinetic energy H_K, a two-body interaction H_{C2}, a one-body interaction H_{C1}, and a backscattering term H_B. H_{C2} and H_{C1} are normal-ordered operators given by the following formulas:

$$\begin{aligned} H_{C2} &= \frac{-U_2}{(2\pi)^2}\sum_{j_1,j_2}\sum_{\sigma_1,\sigma_2}\mathcal{P}\int dk_1\ldots dk_4 \\ &\qquad \times \frac{c^\dagger_{j_2k_4\sigma_2}c^\dagger_{j_1k_2\sigma_1}c_{j_1k_1\sigma_1}c_{j_2k_3\sigma_2}}{(k_4-k_3)(k_2-k_1)}\,, \\ H_{C1} &= -\rho U_2\hat{n} + \frac{-U_2}{(2\pi)^2}\sum_j\sum_\sigma\mathcal{P}\int dk_1\int dk_2\,\frac{c^\dagger_{jk_2\sigma}c_{jk_1\sigma}}{k_2-k_1} \\ &\qquad \times\left(\ln\left|\frac{\Lambda-k_1}{\Lambda+k_1}\right| - \ln\left|\frac{\Lambda-k_2}{\Lambda+k_2}\right|\right)\,, \end{aligned} \tag{3.67}$$

where $\Lambda = W/\hbar v_F$. H_{C2}, H_{C1}, and H_B can be represented by Feynman graphs [see Fig. 3.3(a)], which can then be connected to construct the terms that renormalization adds to the Hamiltonian. As usual, the internal lines of the second-generation graphs carry only high-energy momenta which lie within the shell of wave-vectors that are integrated out.

Given such rules for constructing the second-generation terms, one can deduce that, whenever one creates a new term by connecting lines emanating from the H_{C2} and H_{C1} graphs [see Fig. 3.3(b) for examples], one picks up a scaling factor of b^{-1}. For example, diagram 2 of Fig. 3.3(b) represents a two-body interaction produced by contracting one H_{C2} with one H_{C1}. This new interaction term is similar to H_{C2} except that the denominator contains only one power of (k_4-k_3) or (k_2-k_1) and, consequently, is less singular than H_{C2}, which is fixed under rescaling. Thus, the second-generation term must shrink under renormalization. Indeed, all such graphs formed from contracting the Coulombic interaction terms are similarly irrelevant and scale to zero under renormalization. They can be ignored in the effective theory. We should expect this result. Otherwise, our Coulomb-blockade model would probably never have been useful at all.

As for graphs that involve the backscattering term H_B [see Fig. 3.3(c)], we need only consider these to order v^2, for we go no further in calculating $f(g)$. Depending

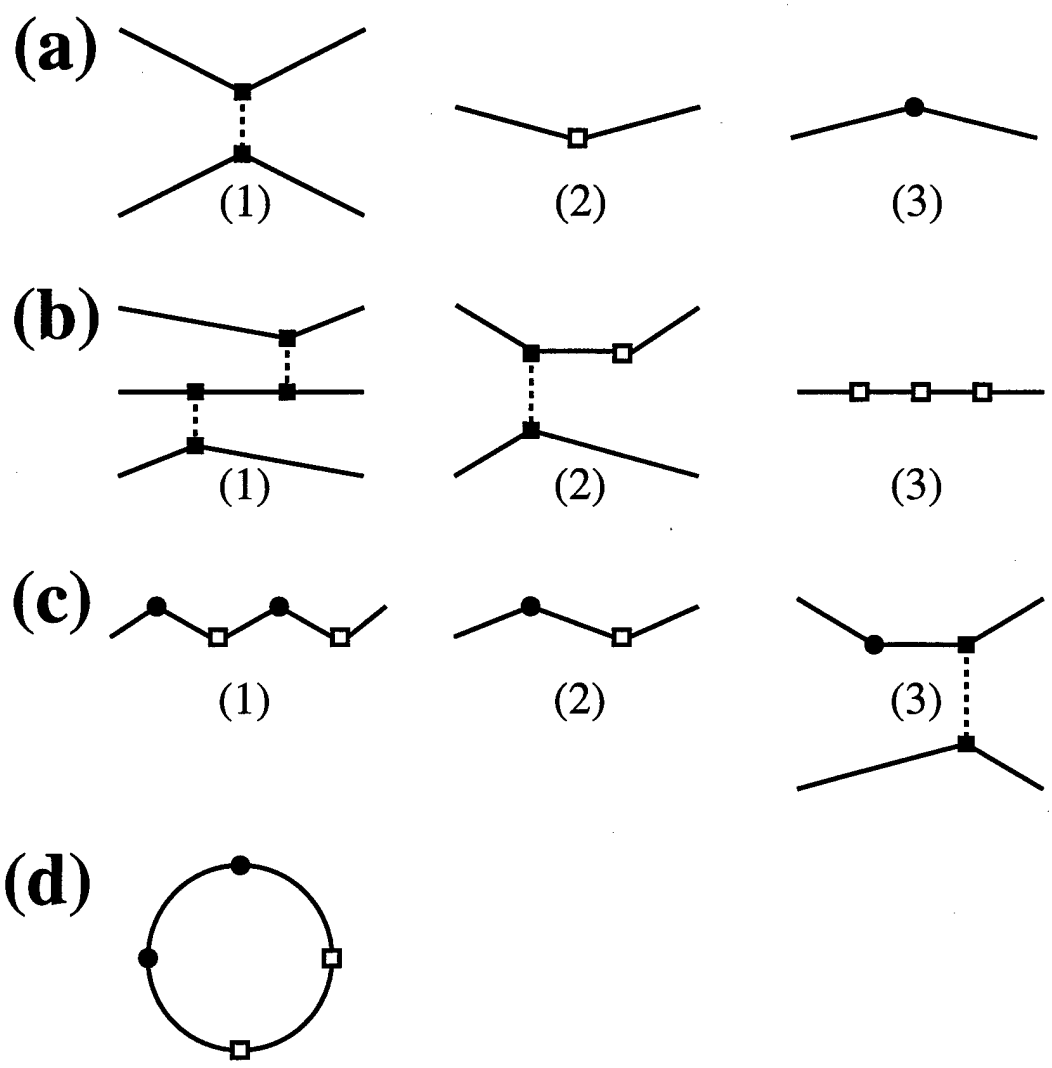

Figure 3.3: Feynman diagrams for integrating out single-particle energies a distance greater than W' from the Fermi surface in the fermionic version of the strong-coupling theory. (a) The three building-block Feynman graphs. Diagram 1 corresponds to the two-body Coulomb interaction H_{C2}. Diagram 2 corresponds to the one-body Coulomb term H_{C1}. Diagram 3 represents the backscattering H_B. (b) Second-generation m-body graphs constructed by contracting H_{C2}'s and H_{C1}'s. These terms are all irrelevant to the low-energy theory, scaling to zero under renormalization. (c) Second-generation graphs formed from combinations of H_B, H_{C2}, and H_{C1}. Terms such as diagram 1 that contain an even number of H_B's are irrelevant under renormalization. Diagrams 2 and 3 involve odd numbers of H_B's and are therefore dangerously irrelevant. Nevertheless, they are negligible in size compared to corresponding low-energy graphs and therefore can be safely discarded. (d) An example of a closed diagram used to calculate the contribution to the energy shift from the degrees of freedom that correspond to single-particle energies more than W' from the Fermi surface.

upon how many Coulombic interaction terms are introduced, the second-generation graphs that contain H_B all scale down by at least a factor of b^{-1}. Consequently, all but those which contribute to low-energy backscattering are irrelevant. Thus, we can drop graphs such as diagram 1 of Fig. 3.3(c) that contain an even number of H_B's. Graphs containing an odd number of H_B's are dangerously irrelevant but can ultimately be ignored because they are negligible compared to the corresponding graphs that can be constructed from the low-energy portions of the original H_{C2}, H_{C1}, and H_B. Diagrams 2 and 3 of Fig. 3.3(c), for example, are of order $v(U_2/W')$. If we had renormalized down to W'', where $U_2 \ll W'' \ll W'$, we would have found the corresponding graphs to be of order $v(U_2/W'')$. The contribution from energies above W' is therefore seen to be merely perturbative in relation to the contribution from energies between W'' and W'. The conclusion is that we can drop the parts of the graphs produced by integrating over energies greater than W'. Returning to our original renormalization down to W', we see that the graphs produced here have been shown to be negligible. The argument that the symmetrizing renormalization does not cause any significant changes in the low-energy Hamiltonian is complete.

Having disposed of the concern that the process of symmetrization might leave us with important new low-energy terms, we now show that any constant terms produced are similarly insignificant. Such constant terms correspond to closed diagrams constructed from the original Feynman graphs. Since all lines are internal, they all represent the propagation of high-energy excitations. There are obviously an infinite number of closed diagrams. Fortunately, we can limit our attention to a certain subset. We need not concern ourselves with diagrams involving less than two H_{C1} graphs: diagrams with only one H_{C1} graph must sum to zero as $\Delta(\rho)$ is even in ρ; diagrams with zero H_{C1} graphs cannot contribute to the differential energy shift $\delta\Delta$. Similarly, in any pertinent closed graph, H_B must appear a nonzero and even number of times. It cannot be absent as terms that do not include it shift all relevant ground-state ener-

gies equally and are therefore unimportant. Furthermore, in any closed graph, it must appear an even number of times because H_B is the only term that exchanges right- and left-movers. Thus, all the diagrams we need consider consist of a nonzero and even number of H_B's, at least two H_{C1}'s, and an arbitrary number of H_{C2}'s [see Fig. 3.3(d) for a canonical example].

Each such diagram corresponds to a number of time-ordered terms in Rayleigh-Schrödinger perturbation theory. For a Feynman diagram with r internal lines, the associated Rayleigh-Schrödinger terms have r integrations over momenta and $(r-1)$ propagators with denominators linear in the momenta. If the H_{C2} graph appears m_2 times in the Feynman diagram and the H_{C1} graph appears m_1 times, there are $m = (2m_2 + m_1) \geq 2$ additional denominators linear in the momenta, which have their origin in the wave-vector denominator of $\hat{n}$ (recall Eqs. 3.63 and 3.67). The propagator denominators are always on the order of W' or greater. The $\hat{n}$ denominators are of the form $(k - k')$, where k and k' are both in the high-energy wave-vector shell. Thus, these denominators can go to zero. However, the contribution from the regions where they become zero is negligible, the somewhat simplified explanation being that, when one of them goes to zero, the rest of the integrand can be treated as essentially constant, and we have

$$\mathcal{P}\int_{-\epsilon\Lambda'}^{\epsilon\Lambda'} \frac{dk}{k}\left[1 + O\left(\frac{\hbar v_F k}{W'}\right)\right] = O(\epsilon),$$

where $\Lambda' = W'/\hbar v_F$ and the constant $\epsilon \ll 1$. It follows that contributions to the overall result only come when the $\hat{n}$ denominators are themselves of order $W'/\hbar v_F$.

As a result, what remains is a nonsingular integration over r momenta of an integrand that is proportional to $[\bar{k}_1 \ldots \bar{k}_{(r-1+m)}]^{-1}$, where the $\bar{k}_i$ are linear in the momenta over which we integrate. Noting that the only other momentum dependence comes from the logarithmic term of H_{C1}, we see that, in energy units, the result of the integration is of order $(1/W')^{m-1}$. We now multiply the result of our integration by the various factors of U_2, v, and δ that stand aside the integral. For a closed diagram in which H_B, H_{C2},

and H_{C1} appear j, m_2, and m_1 times, respectively, the contribution to the energy shift is readily seen to be of order $U_2(v/\delta)^{2j}(U_2/W')^{m-1}$, where $m = (2m_2 + m_1) \geq 2$. As $(U_2/W') \ll 1$ and the overall energy shift is of order U_2, these terms are negligible.

Thus, at least to order v^2, integrating out all particle and hole excitations at distances greater than $W'/2$ from the Fermi surface produces neither relevant new terms in the low-energy Hamiltonian nor significant constant contributions to the differential energy shift. As what remains is a fermionic theory at half filling, the result for $f(g)$ in Sec. 3.3 is unaffected by possible "high-energy" deviations from this condition, an important property if we wish to compare our predictions with empirical data. We would hope that a similarly universal solution for $f(g)$ could be found to higher orders in $(1-g)$. However, if the formula for the interdot conductance (recall Eq. 3.64) is correct to some non-leading order, such overall independence of the filling fraction must—as in the weak-coupling limit—come through cancellation of the separate filling-fraction dependences of the conductance and the energy shift when one is expressed in terms of the other. If this were shown to be true, we would see once again that the interdot conductance g and not the bare matrix element for tunneling or reflection is the correct parameter to achieve a universal description of the coupling dependence of a double-dot Coulomb blockade.

3.5 Conclusion

The present chapter substantially improves upon the results of Chapter 2 for the Coulomb-blockade peak splitting of two coupled quantum dots [51, 53, 54] and thereby makes an important contribution to the growing body of theoretical and experimental work on such coupled-dot systems [19, 43, 44, 45, 47, 48, 64, 65, 66, 87, 88, 89, 90, 91, 92, 93, 94]. By extending the weak-coupling theory to second order in g for arbitrary $N_{\rm ch}$, it has shown how the positive curvature with respect to g that is characteristic of the peak splitting for small $N_{\rm ch}$ crosses over to the negative curvature characteristic of

large $N_{\rm ch}$ as the number of channels is increased through $N_{\rm ch} \approx 10$. Furthermore, it has demonstrated that, at least for the leading two terms in the weak-coupling theory, the channel conductance g is the "correct" parameter to use in constructing a theory for the peak splitting that is universal in the sense that it is does not depend on the high-energy band structure. Finally, this chapter has made the $N_{\rm ch} = 2$ theory both stronger and broader—broader in that the subleading term is calculated; stronger in that the leading and subleading terms for strong-coupling are confirmed to be insensitive to the manner in which the high-energy cutoff is taken. Thus, the chapter has made more plausible efforts to connect weak- and strong-coupling behaviors and to compare theoretical results with the data from recent two-channel experiments [19, 43, 44, 48].

Acknowledgments

The authors are grateful for helpful conversations with F. R. Waugh, R. M. Westervelt, C. H. Crouch, C. Livermore, A. L. Moustakas, S. H. Simon, and S. Ramanathan. J. M. G. thanks the United States Air Force for financial support. The work for this chapter was also supported by the NSF through the Harvard Materials Research Science and Engineering Center, Grant No. DMR94-00396.

Chapter 4

Corrections for a Finite Barrier

4.1 Introduction

The opening of tunneling channels between two quantum dots leads to a transition from a Coulomb blockade characteristic of isolated dots to one characteristic of a single large composite dot [8, 13, 14, 15, 21]. For a pair of electrostatically identical quantum dots characterized by charging energies U much greater than their single-particle level spacings δ_{2D}, this transformation can be chronicled by tracking the splitting of the Coulomb blockade conductance peaks as a function of the conductance through the interdot tunneling channels [19, 43, 44, 45, 46]. If one assumes a single common value for the conductance in each tunneling channel (an assumption that is exactly fulfilled for a spin-symmetric system of only two channels, one for spin-up electrons and the other for spin-down electrons), one can divide the peak splitting by its saturation value and look for the relation between two dimensionless quantities [51, 52]: the fractional peak splitting f and the dimensionless channel conductance g [99].

For $g \ll 1$, Chapters 2 and 3 have treated the coupled-dot problem via a "transfer-Hamiltonian approach" [100], in which states localized on one dot are connected to those localized on the other by hopping matrix elements. (Here *localized* signifies that

a state is entirely restricted to one of the two dots.) The hopping matrix elements have been treated as constant, independent of the states connected, as they would be if the interdot barrier were a delta-function potential, having infinite height and zero width. For such a barrier, the leading small-g behavior of the fractional peak splitting is given by

$$f^{(1)}_{\xi=0} = \frac{2\ln 2}{\pi^2} N_{\rm ch} g, \tag{4.1}$$

where $N_{\rm ch}$ is the number of separate tunneling modes (spin-up and spin-down channels are counted separately). The superscript of $f^{(1)}_{\xi=0}$ tells us that this is the leading term in the weak-coupling limit. The subscript further specifies that this term is calculated for a tunneling barrier of effectively zero width ($\xi = 0$) and therefore, by implication, of infinite height. For $g \gtrsim 0.2$, subleading terms, which are higher-order in g, contribute significantly to the zero-width peak splitting, and the first set of these, which consists of terms proportional to g^2, has been calculated in Chapter 3.

In this chapter, we calculate a different correction to the $\xi = 0$, first-order in g result which arises from the fact that a realistic barrier possesses a finite height V_0 and a nonzero width ξ. For such a realistic barrier, the hopping matrix elements that move electrons between the dots are not independent of the states they connect, and, for small g, they depend exponentially on the energies of the states. As a result of this exponential dependence, in the weak-coupling ($g \ll 1$) limit, it can pay to tunnel to intermediate states with energies above the barrier, and the leading term in the fractional peak splitting then behaves as $(U/W)/|\ln g|$, where U is the interdot charging energy, which measures the capacitive energy cost of moving electrons between the dots, and W is the characteristic energy scale over which the hopping matrix elements change from their values at the Fermi energy $E_{\rm F}$.

We examine specifically the case of a finite-width interdot barrier that can be treated as parabolic near its peak. We find that, for such a barrier, the energy scale W is equal to $\hbar\omega/2\pi$, where ω is the harmonic oscillator frequency of the inverted parabolic

well. This frequency is proportional to the square root of the barrier curvature, which is proportional to V_0/ξ^2. It follows that the limit $\xi \to 0$ corresponds to the limit $U/W \to 0$. For $\xi \neq 0$, on the other hand, it is not always true that $U/W \ll 1$. In fact, in recent experiments by Waugh *et al.* [19, 43, 44], Crouch *et al.* [45], and Livermore *et al.* [46], it appears that U/W is roughly 1. Under such circumstances, we find that, for a given small value of the channel conductance ($g \ll 1$), the fractional peak splitting f is larger than the zero-width splitting, $f_{\xi=0}$, by a small but noticeable amount, and, in the extreme limit of $g \to 0$, the ratio of the finite-width peak splitting to the previously calculated zero-width peak splitting becomes very large. For intermediate values of g, on the other hand, the primary effect is a small increase in f accompanied by a reduction of the slope of the f-versus-g curve.

To find the leading term in the finite-width fractional peak splitting we adopt a *stationary-state approach* [100], in which the first step is to solve for the single-particle eigenstates of non-interacting electrons moving in the electrostatic potential of the coupled dots. The capacitive interactions between the electrons are then expressed in terms of these non-interacting double-dot eigenstates, and the off-diagonal elements of the interactions are treated perturbatively. The leading term in the finite-width fractional peak splitting, $f^{(1)}$, is determined by finding the value for $\rho = 1$ of a more general quantity $\tilde{f}^{(1)}(\rho)$, where ρ is a dimensionless parameter (defined by Eq. 4.2 below) which is a measure of the bias asymmetry between the dots [51, 52]. In the limit $U/W \to 0$, the zero-width result, $f^{(1)}_{\xi=0}$, is recovered. For finite U/W, an approximate analytic calculation demonstrates the limiting $1/|\ln g|$ behavior, which is confirmed by numerical results. For the particular choice $U/W = 1$, which corresponds to recent experiments [19, 43, 44, 45, 46], as well as for various other choices of the ratio U/W, the leading term in the fractional peak splitting is computed numerically as a function of g. It is confirmed that the condition $\xi \neq 0$ leads to an upward shift of the peak splitting for weakly coupled dots ($g \ll 1$). As g becomes larger, the effect of allowing

$\xi \neq 0$ becomes less dramatic, and, for $U/W \simeq 1$, previous predictions for the fractional peak splitting at intermediate values of g are essentially unaltered.

The structure of this chapter is as follows. Section 4.2 develops the stationary-state approach for calculating $\tilde{f}(\rho)$. Section 4.3 implements this approach for a parabolic interdot barrier, verifying the $1/|\ln g|$ behavior of the $g \to 0$ peak splitting that arises for $\xi \neq 0$ and putting the finite-width calculation in the context of earlier work. Section 4.4 summarizes the results and comments on the possible effects of $\xi \neq 0$ when the dots are strongly coupled ($g \simeq 1$).

4.2 The Stationary-State Approach

In order to solve for $\tilde{f}(\rho)$ via the stationary-state approach, we make the problem one-dimensional by considering a smooth, adiabatic interdot connection [see Fig. 4.1(a)] which, for simplicity, we presume to contain only one transverse orbital mode that lies near or below the Fermi energy $E_{\rm F}$ [34]. (The use of one orbital mode corresponds to the spin-symmetric $N_{\rm ch} = 2$ experiments of Waugh *et al.* [19, 43, 44], Crouch *et al.* [45], and Livermore *et al.* [46].) For such a single orbital-mode connection, the only parts of an electron wavefunction that can pass from dot to dot are those that overlap with the lowest transverse mode. Hence, in investigating the effect of the interdot connection, we can ignore all electrons but those in this lowest mode. We are left with a one-dimensional problem in which a representative electron with single-particle energy E moves in an effective potential $V(x) = E_{\rm tr}(x) + V_{\rm el}(x)$, where $E_{\rm tr}(x)$ is the spatially dependent energy of the lowest transverse mode and $V_{\rm el}(x)$ is the spatially dependent electrostatic energy. The characteristic length scale for the spatial variation of the effective potential is the barrier width ξ.

After adding hard boundaries at a distance $L_{\rm dot}$ from the barrier, we have a "box-like" double-dot system with a smoothly varying longitudinal potential [see Fig. 4.1(b)]. Of course, a realistic double-dot system will not look entirely like this. The assumption

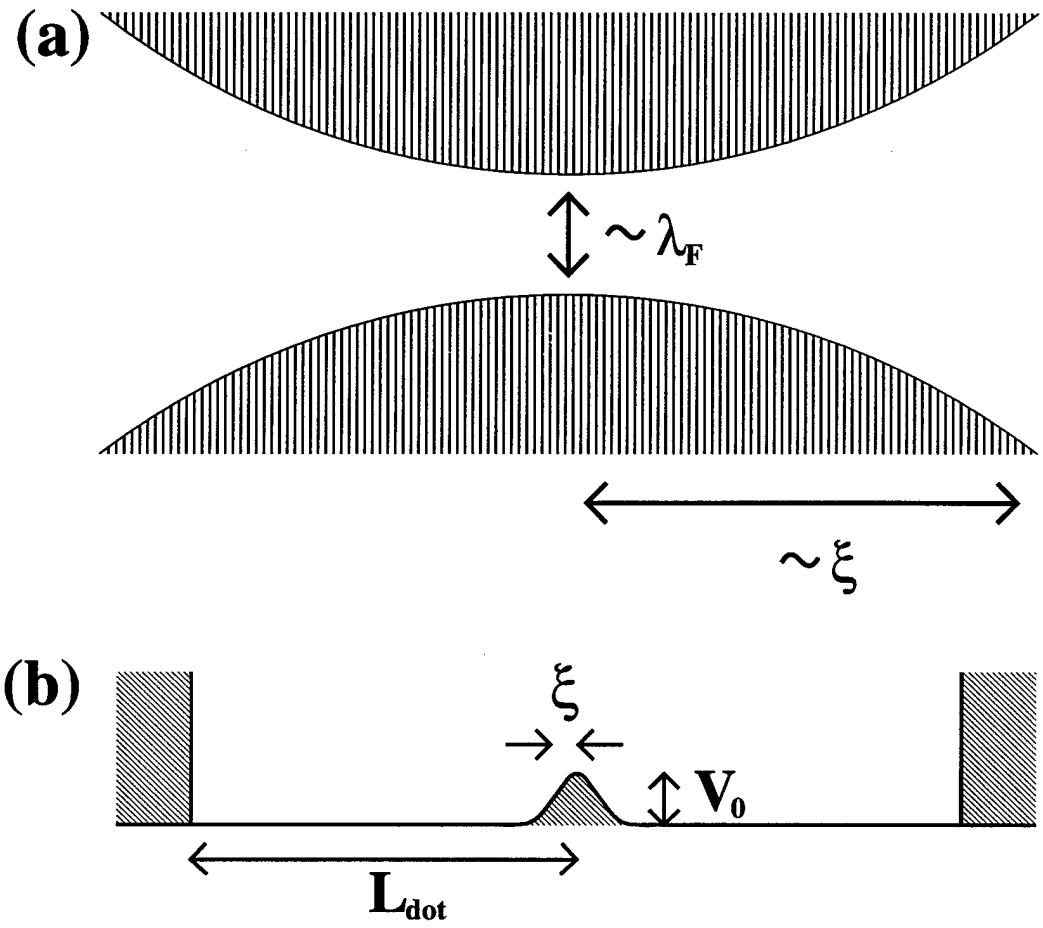

Figure 4.1: (a) Schematic diagram for a single orbital-mode connection between the two dots. Over a distance of order ξ, the connection narrows to a minimum width on the order of the Fermi wavelength λ_F. (b) "Box-like," 1D double-dot system with a central barrier. Hard confining walls are located at a distance L_{dot} from the barrier. The barrier is characterized by its height V_0 and half width ξ.

of adiabaticity will break down as the interdot channel widens, and disorder will then bring about a mixing of the various tranverse modes. However, so long as there is not a strong backscattering center in the vicinity of the connecting region, the breakdown of the adiabatic approximation will not significantly affect movement of particles between the dots—either with regard to the probability of transmission or with respect to the relevant phase shift upon passing through the barrier. Instead, in the large-dot limit in which the charging energy U is much greater than the two-dimensional dot level spacing δ_{2D}, the primary effect of the failure of adiabaticity will be a mere redistribution of the scattered particles among the various transverse modes in the wider "plains" of the dots. The probability of a particle in the lowest transverse mode entering or exiting the interdot connecting region is unaffected by such an external redistribution. Thus, the details of the "boundary conditions" away from the connecting region are unimportant, and, so long as the connecting region itself is sufficiently smooth, we can, for simplicity, extend the adiabatic approximation to the distance $L_{\rm dot}$, where hard boundaries are imposed to ensure the correct normalization within the one-dimensional channel [101].

Having arrived at our "box-like" model, we now concentrate on the Hamiltonian, which consists of two components. The first, H_0, is a diagonal term that gives the energies of the non-interacting, single-particle eigenstates, which form a discrete spectrum with an average level spacing proportional to $\hbar v_{\rm F}/L_{\rm dot}$ near the Fermi surface, where $v_{\rm F}$ is the Fermi velocity. The second, H_C, gives the capacitive energy cost of moving electrons from one side of the barrier to the other [51, 52]:

$$H_C = U(\hat{n} - \rho/2)^2. \tag{4.2}$$

Here, $\hat{n}$ counts the electrons transferred from dot 1 to dot 2 (assuming, for convenience, an even total number of electrons initially divided equally between the two dots). The dimensionless parameter ρ is a measure of the capacitively weighted bias and favors occupation of dot 2 when $\rho > 0$. (Note that U equals the quantity U_2 of Chapters 2 and 3.)

We need to solve for the dimensionless channel conductance g and the fractional peak splitting f. For the non-interacting electrons characterized by H_0, the dimensionless channel conductance g is simply the transmission probability for a particle incident on the barrier at the Fermi energy E_F [100]. f is not so easily determined because, in its evaluation, H_C is relevant. Thus, we must develop a means of dealing with $\hat{n}$, which is not diagonal in the basis of non-interacting single-particle eigenstates that is the cornerstone of our approach.

Our strategy is to switch to a basis that is simply related to the eigenstate basis but which renders $\hat{n}$ nearly diagonal at energies that are low compared to the barrier. We use the fact that, for a bound system containing two equal potential minima, the eigenstates come in well-defined, discrete pairs [102]. The states in these pairs have similar energies but opposite parities, the even-parity state having a lower energy than the odd-parity state. Thus, the non-interacting part of the Hamiltonian can be written in the form

$$H_0 = \sum_{\sigma,j} E_S(j) c^\dagger_{Sj\sigma} c_{Sj\sigma} + \sum_{\sigma,j} E_A(j) c^\dagger_{Aj\sigma} c_{Aj\sigma} , \tag{4.3}$$

where S and A are the even and odd parity indices, j is the pair index, and σ is the spin index (which might be more generally regarded as a channel index).

At lower and lower energies relative to the barrier, the splitting within the pairs, $|E_A(j) - E_S(j)|$, approaches zero, but the spacing between pairs, $|E_S(j+1) - E_A(j)|$, remains approximately equal to δ_{1D}, where $\delta_{1D} = \pi\hbar v_F / L_{dot}$ (assuming we do not stray too far from the Fermi surface). It follows that, at low energies, one can form doublets of *quasi-localized* states—states that lie mostly on one of the two sides of the central barrier—from linear combinations of the symmetric and antisymmetric components of each eigenstate pair [102]. If $\phi_{Sj}(x)$ and $\phi_{Aj}(x)$ are the symmetric and antisymmetric eigenfunctions of the jth lowest-energy pair (with appropriately chosen overall phases),

the recipe for the quasi-localized wavefunctions is

$$\Omega_{j\alpha}(x) = \frac{1}{\sqrt{2}}\left[\phi_{Sj}(x) + (-1)^{\alpha+1}\phi_{Aj}(x)\right], \tag{4.4}$$

where the dot index α signifies that $\Omega_{j\alpha}(x)$ is primarily localized on the dot 1 side of the barrier if $\alpha = 1$ and on the dot 2 side if $\alpha = 2$.

At high energies relative to the barrier, we cannot form quasi-localized wavefunctions from combinations of just two states. Nevertheless, we continue to form the linear combinations analogous to those of Eq. 4.4. We refer to the full set of states $\Omega_{j\alpha}(x)$ as *semi-localized* to indicate that these states are sometimes quasi-localized (i.e., when they lie at low energies relative to the interdot barrier) and sometimes not.

The semi-localized states constitute the complete and orthogonal basis that we need to render $\hat{n}$ nearly diagonal at low energies. Their simple relation to the double-dot eigenstates translates into an equally simple relation between the corresponding creation and annihilation operators. The semi-localized annihilation operators are given by

$$a_{j\alpha\sigma} = \frac{1}{\sqrt{2}}\left[c_{Sj\sigma} + (-1)^{\alpha+1}c_{Aj\sigma}\right], \tag{4.5}$$

and the corresponding expression for H_0 is

$$H_0 = \sum_{\sigma,\alpha,j} E(j) a^\dagger_{j\alpha\sigma} a_{j\alpha\sigma} - \sum_{\sigma,j} t(j)(a^\dagger_{j2\sigma} a_{j1\sigma} + \text{H.c.}), \tag{4.6}$$

where $E(j)$ is the average energy of the pair and $t(j)$ is half the energy difference within the pair:

$$\begin{aligned} E(j) &= \frac{E_A(j) + E_S(j)}{2}, \\ t(j) &= \frac{E_A(j) - E_S(j)}{2}. \end{aligned} \tag{4.7}$$

It is important to note that, whereas $E(j)$ is in general on the order of the Fermi energy, $t(j)$ is no greater than the average level spacing $\delta_{1D} = \pi\hbar v_F/L_{dot}$ and becomes vanishingly small in the large-dot limit ($L_{dot} \to \infty$). The minuteness of $t(j)$ will permit us to ignore it in calculating the leading contribution to the fractional peak splitting.

We now write $\hat{n}$ in terms of the semi-localized operators. If dot 1 corresponds to the $x < 0$ side of the barrier and dot 2 corresponds to the $x > 0$ side, we have

$$\hat{n} = \frac{1}{2}\int dx\,[\Theta(x) - \Theta(-x)]\,\psi^\dagger(x)\psi(x), \tag{4.8}$$

where $\psi(x)$ is the position operator and $\Theta(x)$ is the Heaviside step function [103].

After writing $\psi(x)$ in terms of the semi-localized operators $a_{j\alpha\sigma}$, we see that $\hat{n} = \hat{n}_0 + \delta\hat{n}_C + \delta\hat{n}_T$, where $\hat{n}_0$ corresponds to the $\hat{n}$ we would have if the semi-localized states were truly localized, $\delta\hat{n}_C$ is the part of $\delta\hat{n} = (\hat{n} - \hat{n}_0)$ that does not transfer electrons from dot 1 to dot 2, and $\delta\hat{n}_T$ is the part of $\delta\hat{n}$ that does effect such a transfer:

$$\begin{aligned}
\hat{n}_0 &= \sum_{\sigma,\alpha,j} \frac{(-1)^\alpha}{2} a^\dagger_{j\alpha\sigma} a_{j\alpha\sigma}, \\
\delta\hat{n}_C &= \sum_{\sigma,\alpha,j_1,j_2} \left[B(j_2,\alpha;j_1,\alpha) - \frac{(-1)^\alpha}{2}\delta_{j_1,j_2} \right] a^\dagger_{j_2\alpha\sigma} a_{j_1\alpha\sigma}, \\
\delta\hat{n}_T &= \sum_{\sigma,\alpha,j_1,j_2} B(j_2,\bar{\alpha};j_1,\alpha) a^\dagger_{j_2\bar{\alpha}\sigma} a_{j_1\alpha\sigma}\,.
\end{aligned} \tag{4.9}$$

Here, $\bar{\alpha}$ means "not α" and

$$\begin{aligned}
B(j_2,\alpha_2;j_1,\alpha_1) &= \frac{1}{2}\int_0^{L_{\rm dot}} dx\ \left[(-1)^{\alpha_1+1}\phi^*_{Sj_2}(x)\,\phi_{Aj_1}(x)\right. \\
&\qquad \left. + (-1)^{\alpha_2+1}\phi^*_{Aj_2}(x)\,\phi_{Sj_1}(x)\right].
\end{aligned} \tag{4.10}$$

We have obtained the desired "semi-diagonal" form of $\hat{n}$. Using $\delta\hat{n} = \delta\hat{n}_C + \delta\hat{n}_T$ and assuming that g is small, we express the Hamiltonian in terms of one non-perturbative piece, H'_0, and two perturbative pieces, H'_T and H'_C:

$$\begin{aligned}
H'_0 &= \sum_{\sigma,\alpha,j} E(j) a^\dagger_{j\alpha\sigma} a_{j\alpha\sigma} + U(\hat{n}_0 - \rho/2)^2, \\
H'_T &= -\sum_{\sigma,j} t(j)(a^\dagger_{j2\sigma} a_{j1\sigma} + \text{H.c.}), \\
H'_C &= U(\hat{n}_0 - \rho/2)\delta\hat{n} + U\delta\hat{n}(\hat{n}_0 - \rho/2) + U(\delta\hat{n})^2.
\end{aligned} \tag{4.11}$$

As in Chapters 2 and 3, the fractional peak splitting is determined from $\tilde{f}(\rho)$, where

$$\tilde{f}(\rho) = \frac{\Delta(0) - \Delta(\rho)}{U/4} \tag{4.12}$$

and $\Delta(\rho)$ is the energy shift of the ground state of H_0' due to the perturbations H_T' and H_C' for the given value of ρ, where $0 \leq \rho < 1$ and the total number of particles in the double-dot system is even. The quantity $\left[\lim_{\rho \to 1} \tilde{f}(\rho)\right]$ equals the fractional peak splitting f.

Eq. 4.12 tells us that we are only interested in relative energy shifts. Consequently, we can ignore terms such as $\langle 0|U(\delta\hat{n})^2|0\rangle$ that are independent of ρ. (Here the brackets indicate an expectation value taken in the ground state of H_0'.) Another set of irrevelant terms are those of the form $\langle 0|U(\hat{n}_0 - \rho/2)\delta\hat{n}|0\rangle$, which are zero due to the symmetry of the ground state with respect to interchange of the two dots. Finally, terms that contain H_T' are also negligible because $t(j)$ goes to zero with the reciprocal of the system size and, unlike $\delta\hat{n}$, H_T' only connects each state to one other, rather than connecting each state to a manifold of others (see Chapter 2 for a similar situation with regard to odd orders in the transfer-Hamiltonian perturbation theory). After the above terms are omitted, it is apparent that the leading perturbative energy shift comes from the term that is second order in H_C'. To lowest order in $\delta\hat{n}$, this term is

$$\Delta^{(2)}(\rho) = -U^2 \left\langle 0 \left| \delta\hat{n}\, P_0 \frac{(\hat{n}_0 - \rho)^2}{H_0' - E_0'(\rho)} P_0\, \delta\hat{n} \right| 0 \right\rangle, \tag{4.13}$$

where $E_0'(\rho)$ is the energy of the ground state of H_0' and where P_0 is the operator that projects out the unperturbed ground state. $\Delta^{(2)}(\rho)$ can easily be seen to consist of two distinct parts: a term second-order in $\delta\hat{n}_C$, which involves hopping between states semi-localized on the same dot, and a term second-order in $\delta\hat{n}_T$, which involves hopping between states on different dots.

With Eq. 4.13, we have completed our tour of how to use the stationary-state approach to find both the interdot channel conductance g and the fractional peak splitting f. In order to progress further, we must adopt a model for the barrier that gives the energy dependence of the elements of $\delta\hat{n}$ (recall Eqs. 4.9 and 4.10).

4.3 Splitting and Conductance for a Parabolic Barrier

We assume that the barrier in the interdot tunneling channel can be reasonably modeled by a parabolic one. For an energy barrier with peak height V_0, such a model is plausible when $V_0 \simeq E_F \gg U$, which is the regime of experimental interest [104]. The formula for a parabolic potential $V(x)$ centered at the origin with half width ξ is the following:

$$V(x) = \begin{cases} V_0\left(1 - \frac{x^2}{2\xi^2}\right) & \text{if } |x| < \sqrt{2}\xi \\ 0 & \text{otherwise.} \end{cases} \tag{4.14}$$

A crucial energy scale for this barrier is the harmonic oscillator frequency ω of the inverted parabolic well. This frequency is given by the formula

$$\hbar\omega = \left(\frac{1}{\pi\sqrt{2}}\right)\frac{(2\pi\hbar)^2}{2m\lambda_V\xi}, \tag{4.15}$$

where $2\pi/\lambda_V = \sqrt{2mV_0/\hbar^2}$ and m is the effective mass of the electron.

The problem of transmission through and reflection from a parabolic barrier is well known and exactly solvable [105, 106]. The solutions are parabolic cylinder functions [107], and the dimensionless channel conductance is given by [106, 108]

$$g = \frac{1}{1 + e^{-2\pi y(E_F)}}, \tag{4.16}$$

where E_F is the Fermi energy and

$$y(E) = \frac{E - V_0}{\hbar\omega}. \tag{4.17}$$

From these equations, it follows that

$$\frac{(V_0 - E_F)}{\hbar\omega} = \frac{1}{2\pi}\ln\left(\frac{1-g}{g}\right), \tag{4.18}$$

and, for $g \ll 1$,

$$\frac{(V_0 - E_F)}{V_0} \simeq \frac{1}{2\pi^2\sqrt{2}}\left(\frac{\lambda_V}{\xi}\right)|\ln g|. \tag{4.19}$$

Equation 4.19 tells us that, even for experimental systems [19, 43, 44, 45] in which ξ is quite small ($\xi \simeq \lambda_F$), E_F is close to V_0 for $|\ln g| \ll 2\pi^2\sqrt{2}$. Thus, the assumption of a parabolic barrier appears reasonable for any measurable value of the interdot conductance.

We now consider the sizes of the various energies that appear in our peak splitting calculations. Equation 4.16 indicates that the energy scale W for the variation of transmission probabilities is $\hbar\omega/2\pi$. Recalling our discussion in Sec. 4.1, we have

$$\frac{U}{W} = \frac{2\pi U}{\hbar\omega}. \tag{4.20}$$

As observed in Chapter 2 and other previous work [45, 51], for symmetric dots, U equals $e^2/(C_\Sigma + 2C_{\rm int})$, where C_Σ is the total capacitance of one of the two dots and $C_{\rm int}$ is the interdot capacitance. The energy scale $\hbar\omega$ is, by comparison, only roughly known. From the fact that the barrier height V_0 is approximately equal to E_F, we know that $\lambda_V \simeq \lambda_F$. For ξ, we can use the "device resolution" d, which is the distance between the surface metallic gates and the two-dimensional electron gas (2DEG) and is typically on the order of 100 nm. The fact that the approximation $\xi \simeq d$ should be accurate safely within a factor of 2 can be surmised from calculations such as that of Davies and Nixon [109] in which they show that the potential profile induced in a 2DEG by a narrow line gate has a half width at half maximum that is approximately equal to d [110, 111]. In the AlGaAs/GaAs heterostructures of Waugh *et al.*, Crouch *et al.*, and Livermore *et al.* [19, 43, 44, 45, 46], where d is fairly small, about 50 nm (approximately one Fermi wavelength), further circumstantial evidence for $\xi \simeq d$ comes from the fact that the space between the gates that form the interdot barrier is about 100 nm (see Chapter 2). It follows that, for these experimental systems, $\hbar\omega$ is approximately $0.2E_F$. On the other hand, U is about $0.03E_F$, and, therefore, to within a factor of 2, $2\pi U/\hbar\omega \simeq 1$. For different systems in which the Fermi wavelength is still about 50 nm but the gates are further from the 2DEG [93, 112], the ratio $2\pi U/\hbar\omega$ is presumably even larger. Consequently, we expect it to be quite generally true that the ratio $U/W = 2\pi U/\hbar\omega$ is

greater than or approximately equal to 1.

On the other hand, since, in the sorts of experimental situations with which we are primarily concerned [19, 43, 44, 45, 46], both $W/E_{\rm F}$ and $U/E_{\rm F}$ are much less than 1, we are justified in linearizing the single-particle energy spectrum about the Fermi surface, taking $E(j) = E_{\rm F} + \hbar v_{\rm F}[k(j) - k_{\rm F}]$, where $k(j) = \sqrt{2mE(j)/\hbar^2}$.

We must now calculate $|B(j_2, \alpha_2; j_1, \alpha_1)|$ when $j_1 \neq j_2$. We avail ourselves of the exact, real solutions for the wavefunctions $\phi_{Pj}(x)$ in the presence of a parabolic potential [106, 107] (P is the parity index, which we set equal to 0 for symmetric wavefunctions and 1 for antisymmetric wavefunctions). Connecting these to the corresponding sinusoids, we find that, for $x > \sqrt{2}\xi$ and $L_{\rm dot} \gg \xi$, the eigenfunctions are approximately given by

$$\phi_{Pj}(x) = \frac{(-1)^P}{\sqrt{L_{\rm dot}}} \cos[k(y_{Pj})(x - \sqrt{2}\xi) + \gamma_P(y_{Pj})], \tag{4.21}$$

where $y_{Pj} = y(E_{Pj})$ and $k(y_{Pj}) = \sqrt{2mE_P(j)/\hbar^2}$. The hard-wall boundary condition then demands that there be an integer n such that the quantity in brackets equals $(2n+1)\pi/2$ when $x = L_{\rm dot}$.

As for the phase $\gamma_P(y)$ itself, it can be written in the following general form:

$$\gamma_P(y) = (-1)^P R(y) + D(y)\,, \tag{4.22}$$

If the connection to the sinusoids is made using the leading large-x forms for the parabolic cylinder functions [106, 107], $R(y)$ and $D(y)$ are given by

$$\begin{aligned} R(y) &= \frac{1}{2}\arctan(e^{\pi y})\,, \\ D(y) &= \frac{1}{2}[\arg\Gamma(1/2 - iy) + y\ln(4\pi\sqrt{2}\xi/\lambda_V)] + D_0\,, \end{aligned} \tag{4.23}$$

where D_0 is independent of y.

Returning to Eq. 4.10, we find that, if we restrict the integral to $x > \sqrt{2}\xi$, we have

$$\begin{aligned} B' &\simeq (-1)^{\alpha+1} \frac{\sin[D(y_2) - D(y_1)]\cos[R(y_2) + R(y_1)]}{2(k_2 - k_1)L_{\rm dot}}\,, \\ \bar{B}' &\simeq (-1)^{\alpha+1} \frac{\cos[D(y_2) - D(y_1)]\sin[R(y_2) + R(y_1)]}{2(k_2 - k_1)L_{\rm dot}}\,, \end{aligned} \tag{4.24}$$

where the bar of $\bar{B}'$ indicates that this is the term that moves electrons from dot α to dot $\bar{\alpha}$ and we have anticipated a continuum limit in replacing y_{Pj} and $k(y_{Pj})$ by $y_j = y[E(j)]$ and $k_i = k(y_j)$.

In the calculation of $\bar{B}'$ and B', we have neglected the integration over the region $|x| < \sqrt{2}\xi$. The results can therefore be expected to involve errors of order $\xi/L_{\rm dot}$ when compared to the actual values of B and $\bar{B}$. For $\bar{B}'$, this is not too much of a concern since, when both k_2 and k_1 approach the Fermi energy, the numerator of $\bar{B}'$ goes to $\sqrt{g}$ and the denominator goes to zero. Thus, for non-infinitesimal g, if we restrict our wave-vectors to a range about the Fermi surface such that $|k_i - k_F| \ll 1/\xi$ (in which case $\cos[D(y_2) - D(y_1)]$ can be approximated by 1), corrections to $\bar{B}'$ should be relatively small.

In contrast, the term B' is a bit more problematic, for its numerator goes to zero as $(k_2 - k_1)\xi$. Consequently, near the Fermi surface this term is of the order of the error, and to obtain a reliable result we must complete the integral numerically, using the parabolic cylinder functions in place of our sinusoids when $|x| < \sqrt{2}\xi$. We then find that the form for B' approximates the magnitude of the actual value of B if, after approximating $\cos[R(y_2) + R(y_1)]$ by 1, we replace $[D(y_2) - D(y_1)]$ with $\kappa(y_2 - y_1)$, where $\kappa \simeq 0.1$ for $g \sim 0.1$ and $\kappa \to 0$ as $g \to 0$. Our conclusion is that

$$\begin{aligned} B &\simeq (-1)^{\alpha+1} \frac{\sin[\kappa(y_2 - y_1)]}{2(k_2 - k_1)L_{\rm dot}}, \\ \bar{B} &\simeq (-1)^{\alpha+1} \frac{\sin[R(y_2) + R(y_1)]}{2(k_2 - k_1)L_{\rm dot}}. \end{aligned} \tag{4.25}$$

We can now calculate the leading parts of the energy shift $\Delta(\rho)$. The contribution from hopping between states on the same dot is given approximately by

$$\Delta_C^{(2)}(\rho) \simeq -\left(\frac{U\rho^2}{4}\right)\frac{N_{\rm ch}}{\pi^3}\left(\frac{2\pi U}{\hbar\omega}\right)\int_{-Y_1}^{0} dy_1 \int_0^{Y_2} dy_2 \, \frac{\sin^2[\kappa(y_2 - y_1)]}{(y_2 - y_1)^3}, \tag{4.26}$$

where the y_i's are now measured relative to $y_{\rm F}$ (i.e., $y_i \to y_i + y_{\rm F}$). The contribution

from hopping between the dots obeys

$$\Delta_T^{(2)}(\rho) = -\frac{N_{\rm ch}U^2}{4}\int_{-\Lambda_3}^{0}\frac{dk_1}{\pi}\int_{0}^{\Lambda_4}\frac{dk_2}{\pi}\,\frac{\sin^2[\tilde{R}(\hbar v_{\rm F}k_2/\hbar\omega)+\tilde{R}(\hbar v_{\rm F}k_1/\hbar\omega)]}{(k_2-k_1)^2} \times\left\{\frac{(1-\rho)^2}{\hbar v_F(k_2-k_1)+U(1-\rho)}+[\rho\to-\rho]\right\}, \quad (4.27)$$

where $\tilde{R}(y) = R[y(E_{\rm F})+y]$ and the bracketed expression $\rho \to -\rho$ stands for the quantity obtained by replacing ρ by $-\rho$ in the previous term. In Eqs. 4.26 and 4.27, ultraviolet cutoffs Y_r and Λ_r have been inserted in recognition of the fact that our formulas for the integrands break down at some distance Y_r or Λ_r from the Fermi surface.

For the same-dot-hopping shift $\Delta_C^{(2)}$, the presence of such cutoffs is essentially irrelevant since we find this term to be effectively negligible no matter what the choice of Y_r. In particular, even when the cutoffs are taken to infinity, this segment of the energy shift produces a contribution to $\tilde{f}(\rho)$ (recall Eq. 4.12) that is bounded by the following formula:

$$\tilde{f}_C(\rho) \lesssim \frac{N_{\rm ch}\rho^2}{200}\left(\frac{2\pi U}{\hbar\omega}\right). \quad (4.28)$$

The real contribution is perhaps substantially smaller than the bound because the integrand of Eq. 4.26 is systematically too large for the infinitesimal-transmission states that correspond to $|y_1| \gtrsim 1$.

In any case, it is clear that the contribution to $\tilde{f}(\rho)$ from same-dot hopping is essentially negligible. For $N_{\rm ch} \lesssim 2$ and $(2\pi U/\hbar\omega) \ll 10$, $\tilde{f}_C(\rho)$ is extremely small and essentially constant in g. Under such circumstances, it does not significantly affect even the quantitative results. When $(2\pi U/\hbar\omega) \gtrsim 10$, on the other hand, it can be relatively large. Nevertheless, it remains unimportant, for in this regime we can only obtain qualitatively good results for the value of $\Delta_T^{(2)}(\rho)$, and, qualitatively, the upward shift of $\tilde{f}(\rho)$ induced by same-dot hopping merely reinforces the effect from hopping between the dots.

We now consider the cutoffs Λ_r and their impact upon our understanding of the interdot-hopping result. Since the integrand in Eq. 4.27 is reliably precise only when

k_1 and k_2 are within $1/\xi$ of k_F, the ultraviolet cutoffs should be chosen such that $\Lambda_r \sim 1/\xi$. It follows that, to capture with quantitative precision the leading behavior of the peak splitting for $g \ll 1$, the set of wave-vectors within $1/\xi$ of the Fermi surface should encompass the range of energies in which the quantity $\tilde{R}(E)$ is rapidly growing. Consequently, the set of wave-vectors must extend at least to $k_F + k_0$, where $E(k_F + k_0) = V_0$. From Eq. 4.18, we see that the identity $k_0 = (V_0 - E_F)/\hbar v_F$ yields

$$k_0 \xi \simeq \frac{1}{2\pi\sqrt{2}} \ln\left(\frac{1-g}{g}\right) . \tag{4.29}$$

If we require that $k_0 \xi \lesssim 1$, we see that, for $g \ll 1$, we must have $|\ln g| \lesssim 2\pi\sqrt{2}$. Thus, we have a lower bound on the values of g for which our approximations are reliable. Fortunately, the lower bound is very small, and the requirement is only that $g \gtrsim 10^{-4}$.

We are now prepared to calculate $\Delta_T^{(2)}(\rho)$. After a switch to the dimensionless variables $x_r = (-1)^r \hbar v_F k_r / U$, Eq. 4.27 reduces to

$$\Delta_T^{(2)}(\rho) \;\cong\; -\frac{N_{ch} U}{4\pi^2} \int_0^{\bar{\chi}_1} dx_1 \int_0^{\bar{\chi}_2} dx_2 \, \frac{\tilde{T}(x_1, x_2)}{x_2 + x_1 + 1 - \rho} + [\rho \to -\rho] , \tag{4.30}$$

where $\bar{\chi}_r = \hbar v_F \Lambda_r / U$, the symbol $\cong$ signifies equality modulo terms that are independent of ρ, and the quantity $\tilde{T}(x_1, x_2)$ is given by

$$\tilde{T}(x_1, x_2) = \sin^2[\tilde{R}(U x_2/\hbar\omega) + \tilde{R}(-U x_1/\hbar\omega)] . \tag{4.31}$$

To obtain a result with negligible dependence on the cutoffs $\bar{\chi}_r$, we must have $\bar{\chi}_r \gg 1$. On the other hand, to ensure that the answer is quantitatively reliable, we need $\Lambda_r \lesssim 1/\xi$ or, equivalently, $\bar{\chi}_r \lesssim \hbar v_F / U\xi$. Thus, as promised, we can only expect Eq. 4.30 to give quantitatively reliable results for $U \ll \hbar v_F / \xi$; i.e., for $2\pi U/\hbar\omega \ll 2\pi\sqrt{2}$, where $2\pi\sqrt{2} \simeq 9$.

Having dealt with the issue of the ultraviolet cutoffs, we can now go about the business of evaluating the right side of Eq. 4.30. From the identity $\tilde{T}(0,0) = g$, it follows that the limit $2\pi U/\hbar\omega \to 0$ yields the zero-width ($\xi = 0$) linear-in-g equation for $\Delta^{(2)}(\rho)$ that was previously derived via a transfer-Hamiltonian approach [51, 52, 53, 54].

In contrast, in the limit $2\pi U/\hbar\omega \to \infty$, the energy shift given by Eq. 4.30 is independent of the interdot conductance for g a finite distance from both 0 and 1. The constancy of the shift follows from the fact that, except when g equals 0 or 1, $\tilde{T}(x_1, x_2)$ is always 0.5 within the bounds of integration, and the relevant parts of the energy shift are therefore the same as for $(2\pi U/\hbar\omega) = 0$ and $g = 0.5$. It should be re-emphasized, however, that such a result for the limit $2\pi U/\hbar\omega \to \infty$ can only be expected to be qualitatively correct.

What happens when the barrier width ξ is between 0 and ∞? By performing two partial integrations of the righthand side of Eq. 4.30 and dropping terms that go to zero as the cutoffs Λ_r become infinite, we find that

$$\begin{aligned}\tilde{f}^{(1)}(\rho) &= \frac{N_{\rm ch} g}{\pi^2}(1-\rho)\ln(1-\rho) \\ &\quad + \frac{N_{\rm ch}}{\pi^2}\int_0^{\bar{X}_1} dx_1 \left[\frac{\partial \tilde{T}(x_1,0)}{\partial x_1} h(\rho, x_1, 0)\right] \\ &\quad + \frac{N_{\rm ch}}{\pi^2}\int_0^{\bar{X}_2} dx_2 \left[\frac{\partial \tilde{T}(0,x_2)}{\partial x_2} h(\rho, 0, x_2)\right] \\ &\quad + \frac{N_{\rm ch}}{\pi^2}\int_0^{\bar{X}_1} dx_1 \int_0^{\bar{X}_2} dx_2 \frac{\partial^2 \tilde{T}(x_1,x_2)}{\partial x_1 \partial x_2} h(\rho, x_1, x_2) \\ &\quad + [\rho \to -\rho], \end{aligned} \tag{4.32}$$

where $h(\rho, x_1, x_2) = (x_2 + x_1 + 1 - \rho)\ln(x_2 + x_1 + 1 - \rho)$. The first term on the righthand side of Eq. 4.32 is the zero-width result. The other terms, which go to zero in the limit $\xi \to 0$, are the corrections due to a nonzero width.

Numerical evaluations of Eq. 4.32 are plotted in Fig. 4.2(a) for several values of the parameter $2\pi U/\hbar\omega$ in addition to the analytically derived results for the limits of zero-width and infinite-width barriers. A curious feature of these curves is that the corrections to the zero-width behavior are antisymmetric about $g = 0.5$, a property that can be demonstrated by considering what happens to the integrands under the transformations $g \leftrightarrow (1-g)$ and $x_1 \leftrightarrow x_2$. Though the antisymmetry is suggestive, it must be remembered that $\tilde{f}^{(1)}(\rho)$ is only the leading term in a perturbative expansion

about $g = 0$. The small positive contribution to $\tilde{f}(\rho)$ that comes from the formula for the same-dot-hopping shift $\Delta_C^{(2)}$ breaks this antisymmetry, and other higher-order corrections are likely to do the same. Nevertheless, some sort of rough antisymmetry about $g = 0.5$ is probably preserved, for, just as we find that, at small g, $\tilde{f}(\rho)$ is enhanced by hopping connections to states with large transmission amplitudes, so we can expect that, at large g, $\tilde{f}(\rho)$ is diminished by the fact that many of the occupied states from which one hops have transmission probabilities that are less than g.

Such musings aside, we can gain further insight into the nature of our result for $\tilde{f}^{(1)}(\rho)$ by making a rough analytic approximation to the righthand side of Eq. 4.32. To do this, it is best to return to Eqs. 4.12 and 4.30 and to derive the equivalent expression

$$\tilde{f}^{(1)}(\rho) = \frac{2N_{\rm ch}\rho^2}{\pi^2}\int_0^{\bar{x}_1} dx_1 \int_0^{\bar{x}_2} dx_2 \frac{\tilde{T}(x_1,x_2)}{(x_2+x_1+1)(x_2+x_1+1-\rho)(x_2+x_1+1+\rho)}. \tag{4.33}$$

We then postulate that, for small g, the magnitude of the $\tilde{f}^{(1)}(\rho)$ is largely determined by the portion of the integral that corresponds to $x_2 \geq x_0$, where $x_0 = \hbar v_{\rm F} k_0/U$ (recall k_0 from Eq. 4.29). For x_2 in this range, $\tilde{T}(x_1,x_2)$ is on the order of 1 and therefore much larger than $\tilde{T}(0,0)$ when $g \ll 1$. We label this *high-energy portion* of the double integral as $\tilde{f}^{(1)}_{\rm hep}(\rho)$. Since, in this part of the integral, $\tilde{T}(x_1,x_2)$ varies relatively slowly between 0.15 and 0.5, we approximate it by a constant $C_{\tilde{T}}$, where we take $C_{\tilde{T}} = 0.25$. For $x_0 \gtrsim 1$, we can drop the ρ's that appear in the integrand of Eq. 4.33. We then obtain

$$\tilde{f}^{(1)}_{\rm hep}(\rho) \sim \frac{N_{\rm ch}\rho^2}{4\pi^2(x_0+1)}. \tag{4.34}$$

From the identities $f = \tilde{f}(1)$ and $x_0 = (\hbar\omega/2\pi U)\ln[(1-g)/g]$, we conclude that

$$f^{(1)}_{\rm hep} \sim \left(\frac{N_{\rm ch}}{4\pi^2}\right)\frac{2\pi U/\hbar\omega}{|\ln g| + 2\pi U/\hbar\omega}. \tag{4.35}$$

This rough approximation to the leading behavior of the fractional peak splitting is only valid when $x_0 \gtrsim 1$ and $k_0 \lesssim 1/\xi$. The condition on x_0 is necessary to justify dropping the ρ's from the integrand in the high-energy portion of the double integral. The

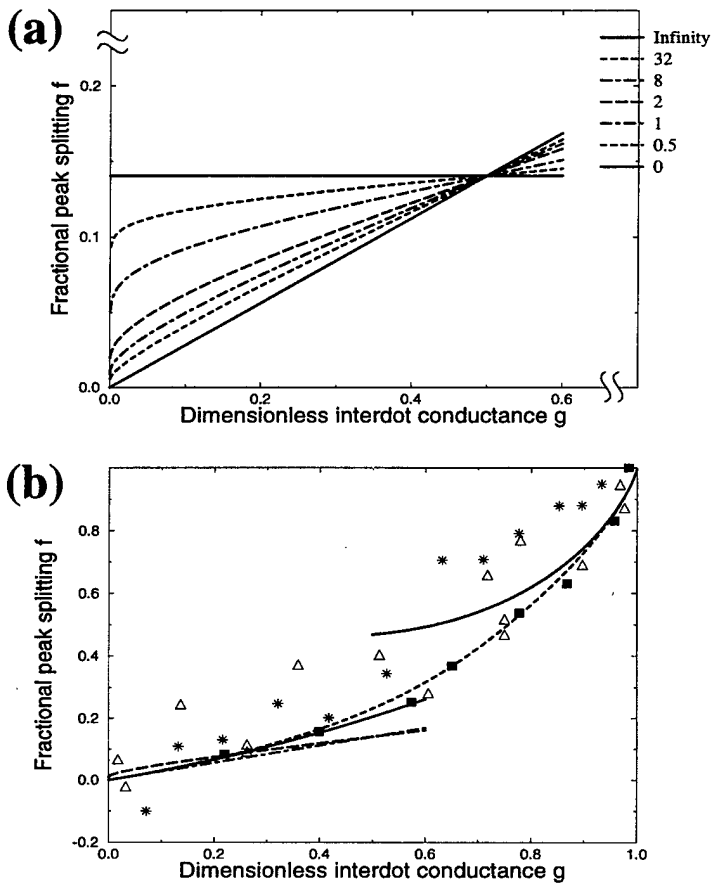

Figure 4.2: (a) Plots of the leading $g \to 0$ term of f, the fractional peak splitting, as a function of g, the dimensionless interdot channel conductance. Each curve corresponds to a different value of the quantity $2\pi U/\hbar\omega$ (see legend on right). All curves are for the case of two interdot tunneling channels, $N_{\rm ch} = 2$. The upward sloping solid line is the linear-in-g result that comes from considering an interdot barrier of effectively zero width ($2\pi U/\hbar\omega = 0$). The dashed and dot-dashed curves show the f-versus-g dependence for finite-width barriers with $2\pi U/\hbar\omega$ taking on values from 0.5 to 32. The horizontal solid line gives the leading term in the fractional peak splitting for an infinite-width barrier ($2\pi U/\hbar\omega \to \infty$). The curves can only be expected to be quantitatively accurate when $2\pi U/\hbar\omega \ll 10$. (b) f-versus-g results for the full domain of g when $N_{\rm ch} = 2$. The solid lines are the complete zero-width results in the weak and strong-coupling limits. These results contain both leading and subleading terms.[6] The plot for the leading zero-width term in the weak-coupling limit ($g \to 0$) is included as a dot-dashed curve. The small-dashed curve that extends from $(g, f) = (0, 0)$ to $(g, f) = (1, 1)$ is an interpolating curve that is derived from the zero-width results. The long-dashed line is the $2\pi U/\hbar\omega = 1$ curve from Fig. 4.2(a). The stars, triangles, and squares symbolize different sets of experimental data [43, 44, 46], the squares being the most recent [46].

condition on k_0 validates the approximate values for the magnitudes of $|A(j_2, \alpha_2; j_1, \alpha_1)|$ that we use throughout our calculation. The two restrictions together mean that the range of reliability for our approximation to $\tilde{f}^{(1)}_{\rm hep}$ is given by

$$\frac{2\pi U}{\hbar\omega} \lesssim |\ln g| \lesssim 2\pi\sqrt{2}. \tag{4.36}$$

For $2\pi U/\hbar\omega \simeq 1$, our approximations are good for g between a couple tenths and a few ten-thousandths.

It is instructive to compare our result for $f^{(1)}_{\rm hep}$ with the zero-width fractional peak splitting, $f^{(1)}_{\xi=0}$. From Eqs. 4.1 and 4.35, we see that, for $2\pi U/\hbar\omega = 1$, the ratio $f^{(1)}_{\rm hep}/f^{(1)}_{\xi=0}$ is about 0.6 when $g = 0.1$ and about 25 when $g = 0.001$. For very weak coupling ($g \ll 0.1$), the correction to the $\xi = 0$ result is proportionately very large, and, as $g \to 0$, it dominates the behavior of the peak splitting. On the other hand, as g assumes more intermediate values ($g \sim 0.1$), the results for $\xi = 0$ and $\xi \neq 0$ converge.

A direct comparison of our results for $f^{(1)}_{\rm hep}$ with the full numerical results for $f^{(1)}$, which are plotted in Fig. 4.2(a), confirms that $f^{(1)}_{\rm hep}$ does indeed capture the essential f-versus-g behavior, particularly as $2\pi U/\hbar\omega$ becomes larger and the exponential enhancement of the tunneling amplitudes becomes more important. The sharp increase in slope as $g \to 0$ can now be understood as resulting from the fact that, in this limit, the high-energy portion of the peak splitting is proportional to $(2\pi U/\hbar\omega)/|\ln g|$. This proportionality also explains why the increase in slope as $g \to 0$ becomes less dramatic as $2\pi U/\hbar\omega$ decreases. The success of our rough analytic approximation supports the supposition that, for small g, a substantial portion of the peak splitting comes from tunneling into virtual states lying near or above the top of the barrier.

Turning to Fig. 4.2(b), we now examine the significance of the calculated finite-width corrections in the context of what we know about the f-versus-g curve in the entire range from $g = 0$ to $g = 1$, and we consider the implications of these corrections for the relevant experiments [19, 43, 44, 45, 46]. The long-dashed curve in Fig. 4.2(b) is the curve from Fig. 4.2(a) for the value $2\pi U/\hbar\omega = 1$, which we believe to be appropriate for the cited

experiments. The dot-dashed line is the leading-order-in-g, zero-width curve, which also appears in Fig. 4.2(a). The small-dashed curve in Fig. 4.2(b) is an interpolation for the entire zero-width f-versus-g curve. This interpolation has been designed to match both the second-order-in-g calculation of the fractional peak splitting for weak coupling ($g \simeq 0$) and also the two-term calculation for strong coupling ($g \simeq 1$), which were obtained in Chapter 3 and are shown as solid curves in Fig. 4.2(b). The stars, triangles, and squares represent different sets of experimental data.

For the particular value of $2\pi U/\hbar\omega$ that is illustrated in Fig. 4.2(b), we see that, although the finite-width correction to f changes the answer by a large factor in the region of small g, the correction is small on an absolute scale. The difference between the dashed curve and the dot-dashed curve never exceeds 0.02 and therefore causes only a small correction to the overall shape of the f-versus-g curve. Qualitatively, the correction due to the finite thickness of the barrier is quite similar to adding a small constant to f near $g = 0$ and then decreasing the slope of the f-versus-g curve at small g. This qualitative similarity follows from the fact that the region where f drops rapidly to zero, at very small g, is almost invisible in the plot. Consequently, the correction to the zero-width curve might be hard to distinguish from the effects of a small interdot capacitance, which have already been included in analyzing the data. We therefore conclude that introduction of the finite thickness correction has little effect on the agreement between theory and the existing experimental data, for which $2\pi U/\hbar\omega \simeq 1$. Nevertheless, such corrections may be important in future experiments.

4.4 Conclusion

By developing a new approach to the coupled-dot problem that relies upon the non-interacting, single-particle eigenstates of the full coupled-dot system, we solve for the leading correction to zero-width, weak-coupling results that were derived in Chapters 2 and 3. The nonzero barrier width ξ and finite barrier height V_0 mean that the off-

diagonal "hopping terms" vary exponentially with the energies of the states they connect. For a small interdot channel conductance ($g \ll 1$), the resulting enhancement of tunneling to "high-energy" states above the barrier leads to an increase in the magnitude of the fractional peak splitting f observed at a given value of g. For a parabolic barrier, the magnitude of this increase grows with the value of the ratio $2\pi U/\hbar\omega$, where U is the interdot charging energy and ω is the frequency of the inverted parabolic well. Except in a very small region near $g = 0$ where f behaves like $(2\pi U/\hbar\omega)/|\ln g|$, the increase in f is accompanied by a decrease in the slope of the f-versus-g curve. The effect upon the overall shape of the f-versus-g curve is not very substantial for experiments in which $(2\pi U/\hbar\omega) \simeq 1$ but could be crucial in interpreting experiments involving significantly wider barriers.

One might worry that the finite-width corrections to higher-order terms in the weak-coupling expansion could lead to a more dramatic alteration of the f-versus-g curve. However, the corrections to such "large-g" terms should be muted by the fact that, as g increases, there is less difference between tunneling amplitudes between states at the Fermi energy and tunneling amplitudes between a state at the Fermi energy and a state lying above the barrier.

A more vital source of concern might be the treatment of the electron-electron interactions in the vicinity of the barrier. Clearly, the use of a sharp step function in the equation for $\hat{n}$ (recall Eq. 4.8) is an artifice. A more realistic model would account for the fact that, though electrons in and about the interdot channel still repel one another locally, their interactions with the rest of the electrons in the system are screened by the surface gates.

Finally, one might wonder whether higher-order corrections to f preserve at least a rough antisymmetry about $g \simeq 0.5$. We have seen that the leading small-g correction, when directly extended to $g = 1$, changes sign and becomes negative for $g > 0.5$. Although a proper calculation of the behavior at such large values of g requires consider-

ation of higher-order diagrams, which we have neglected, we believe that the negativity of the correction to f at large values of g is a generally right physical feature. When g is large and the reflection probability at the Fermi energy is therefore small, the energy dependence of the reflection amplitude, for $\xi \neq 0$, should lead to a decrease in f as a result of the enhanced reflection coefficient for occupied states lying below the barrier.

Acknowledgments

The authors thank C. Livermore, C. H. Crouch, R. M. Westervelt, and I. E. Smolyarenko for helpful discussions. The work for this chapter was supported by the NSF through the Harvard Materials Research Science and Engineering Center, Grant No. DMR94-00396.

Appendix A

Weak-Coupling Details

As described in Chapter 3, the procedure in evaluating the fourth-order energy shift is to calculate the $(N_{\rm ch})^2$ and $N_{\rm ch}$ terms separately. Calculation of the $(N_{\rm ch})^2$ terms is facilitated by rewriting them in terms of two energy variables instead of four. Calculation of the $N_{\rm ch}$ terms is made easier by differentiating twice with respect to ρ while performing the integrations over energy and then integrating twice with respect to ρ at the end. Terms that are constant or linear with respect to ρ cancel in the final result, the relative energy shift $(\Delta_0 - \Delta_\rho)$, so we have not lost useful information as a result of the double differentiation.

The "wrinkle" in these computations, the appearance of integrals of the form

$$\mathcal{P}\int_0^{R\psi} dx\, \frac{\ln(x+B)}{x+A}\,,$$

is resolved by Taylor-expanding the logarithm about $(x+A)$ for $(B-A) < |x+A|$ and about $(B-A)$ for $(B-A) > |x+A|$. For $A < 0$, one first breaks the integral into the intervals $(0, |A| - \epsilon)$ and $(|A| + \epsilon, R\psi)$. After this, one proceeds as usual. The results

are

$$\mathcal{P}\int_0^{R\psi} dx\,\frac{\ln(x+B)}{x+A} = \frac{1}{2}\ln^2(R\psi+A)-\frac{1}{2}\ln^2 A + \sum_{n=1}^{\infty}\frac{(-1)^{n+1}}{n^2}\left[\left(\frac{B-A}{A}\right)^n-\left(\frac{B-A}{R\psi+A}\right)^n\right] \quad \text{for } 0<(B-A)<A<(R\psi+A),$$

$$= \frac{1}{2}\ln^2(R\psi+A)-\ln(B-A)\ln A+\frac{1}{2}\ln^2(B-A)+\frac{\pi^2}{6} - \sum_{n=1}^{\infty}\frac{(-1)^{n+1}}{n^2}\left[\left(\frac{A}{B-A}\right)^n+\left(\frac{B-A}{R\psi+A}\right)^n\right] \quad \text{for } 0<A<(B-A)<(R\psi+A),$$

$$= \ln(B-A)[\ln(R\psi+A)-\ln A] + \sum_{n=1}^{\infty}\frac{(-1)^{n+1}}{n^2}\left[\left(\frac{R\psi+A}{B-A}\right)^n-\left(\frac{A}{B-A}\right)^n\right] \quad \text{for } 0<A<(R\psi+A)<(B-A),$$

$$= \frac{1}{2}\ln^2(R\psi-|A|)-\ln(B+|A|)\ln|A|+\frac{1}{2}\ln^2(B+|A|)+\frac{\pi^2}{6} + \sum_{n=1}^{\infty}\frac{1}{n^2}\left(\frac{|A|}{B+|A|}\right)^n-\sum_{n=1}^{\infty}\frac{(-1)^{n+1}}{n^2}\left(\frac{B+|A|}{R\psi-|A|}\right)^n \quad \text{for } A<0<(B+|A|)<(R\psi-|A|),$$

$$= \ln(B+|A|)[\ln(R\psi-|A|)-\ln|A|] + \sum_{n=1}^{\infty}\frac{1}{n^2}\left(\frac{|A|}{B+|A|}\right)^n+\sum_{n=1}^{\infty}\frac{(-1)^{n+1}}{n^2}\left(\frac{R\psi-|A|}{B+|A|}\right)^n \quad \text{for } A<0<(R\psi-|A|)<(B+|A|). \quad \text{(A.1)}$$

These five integrals are all we need. In confirming that the solution for $(B-A)<(R\psi+A)$ evolves continuously into that for $(R\psi+A)<(B-A)$, it is useful to recognize [85]

that

$$\sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n^2} = \frac{\pi^2}{12}. \tag{A.2}$$

Having equipped ourselves to smooth the "wrinkles," we can proceed with a fuller description of calculation of the fourth-order terms. The $(N_{\rm ch})^2$ calculation is reviewed first. An illustrative segment of the $N_{\rm ch}$ calculation follows.

In Chapter 3, it was remarked that each of the $(N_{\rm ch})^2$ terms could be written in terms of two energy variables ($\epsilon_I = \epsilon_3 - \epsilon_1$, $\epsilon_4 - \epsilon_2$) instead of the original four. The "cost" of this conversion is the appearance of a nontrivial density of states:

$$\int_0^{\epsilon_F} d\epsilon_1 \int_0^{\epsilon_F} d\epsilon_2 \int_{\epsilon_F}^{W} d\epsilon_3 \int_{\epsilon_F}^{W} d\epsilon_4\, h(\epsilon_3 - \epsilon_1, \epsilon_4 - \epsilon_2) = \int_0^{W} d\epsilon_{\rm I}\, \nu(\epsilon_{\rm I}) \int_0^{W} d\epsilon_{\rm II}\, \nu(\epsilon_{\rm II})\, h(\epsilon_{\rm I}, \epsilon_{\rm II}), \tag{A.3}$$

where $\nu(\epsilon)$ is the density of states. For a system at or below half filling,

$$\nu(\epsilon) = \begin{cases} \epsilon & \text{for } 0 \le \epsilon < \epsilon_F, \\ \epsilon_F & \text{for } \epsilon_F \le \epsilon < (W - \epsilon_F), \\ (W - \epsilon) & \text{for } (W - \epsilon_F) \le \epsilon < W. \end{cases} \tag{A.4}$$

(We need not worry about a system above half filling as such as system can be mapped to one below half filling through an exchange of particles and holes.)

Using the new variables $\epsilon_{\rm I}$ and $\epsilon_{\rm II}$, we can sum the integrands for the $(N_{\rm ch})^2$ terms shown in Fig. 3.1 (the others are obtained by taking $\rho \to -\rho$). If we drop the common factor $-N_{\rm ch}(t/\delta)^4 U_2$, the result is the following:

$$\begin{aligned} h_{\rm tot} \;=\; & \frac{-2}{[\epsilon_{\rm I} + U_2(1-\rho)]^2[\epsilon_{\rm II} + U_2(1-\rho)][\epsilon_{\rm II} + \epsilon_{\rm I} + U_2(4-2\rho)]} \\ & + \frac{2}{[\epsilon_{\rm I} + U_2(1-\rho)]^2[\epsilon_{\rm II} + U_2(1+\rho)][\epsilon_{\rm II} + \epsilon_{\rm I}]}. \end{aligned} \tag{A.5}$$

It is not hard to find relations such as

$$\int_0^{\epsilon_F} d\epsilon_{\rm I}\, \nu(\epsilon_{\rm I}) \int_{\epsilon_F}^{W} d\epsilon_{\rm II}\, \nu(\epsilon_{\rm II}) [h_{\rm tot}(\epsilon_{\rm I}, \epsilon_{\rm II}, \rho) - h_{\rm tot}(\epsilon_{\rm I}, \epsilon_{\rm II}, 0)] = 0$$

in the limit $\psi = W/U_2 \to \infty$. Accordingly, we need only calculate

$$\int_0^{\epsilon_F} d\epsilon_{\rm I}\, \nu(\epsilon_{\rm I}) \int_0^{\epsilon_F} d\epsilon_{\rm II}\, \nu(\epsilon_{\rm II}) h_{\rm tot}(\epsilon_{\rm I}, \epsilon_{\rm II}, \rho).$$

The process of evaluating this double integral is lengthy but straightforward. The only "wrinkles" that appear—integrals of the form of Eq. A.1—are no longer problematic. The end result is Eq. 3.9.

We now move to consideration of the fourth-order terms linear in the number of conducting channels. Recall that the $(N_{\rm ch})^2$ terms were added before the integrations over energy were performed. This order of tasks is reversed for the $N_{\rm ch}$ terms, the computation of which revolves primarily about finding a favorable permutation of the operations of differentiating and integrating with respect to ρ, integrating with respect to the ith energy variable, and integrating by parts. Consequently, perhaps the best way to describe the derivation of the $N_{\rm ch}$ contribution is to walk through the computation of a single illustrative term. After seeing the methodology employed in calculating this term, the tireless reader should have little difficulty in computing the rest.

The representative term we choose is that corresponding to diagram 2 of Fig. 3.1(b). This term involves an exchange of a pair of electrons and, consequently, picks up an "exchange" minus sign. The diagram depicts the following sequence of events:

I. Electron 1 tunnels from dot 1 to dot 2, going from a single-particle state with kinetic energy ϵ_1 to one with kinetic energy ϵ_3. The energy of the resulting double-dot state relative to that of the unperturbed ground state is $[\epsilon_3 - \epsilon_1 + U_2(1-\rho)]$.

II. Electron 2 tunnels from dot 1 to dot 2, going from a single-particle state with kinetic energy ϵ_2 to one with kinetic energy ϵ_4. The system's energy relative to the unperturbed ground state is now $[\epsilon_4 + \epsilon_3 - \epsilon_2 - \epsilon_1 + 2U_2(2-\rho)]$.

III. Electron 2 tunnels back to dot 1, settling into the initial single-particle state of Electron 1. The ensuing relative system energy is $[\epsilon_3 - \epsilon_2 + U_2(1-\rho)]$.

IV. Electron 1 tunnels back to dot 1, settling into Electron 2's initial single-particle state. The unperturbed ground state has been recovered.

With all the intermediate-state energies known, it is easy to write down the contribution to the fourth-order energy shift:

$$\Delta^{(4)}_{N_{\rm ch},2}(\rho) = t^4 \sum_{\sigma} \sum_{\epsilon_1,\epsilon_2} \sum_{\epsilon_3,\epsilon_4} \frac{1}{[\epsilon_3 - \epsilon_2 + U_2(1-\rho)]} \times \frac{1}{[\epsilon_4 + \epsilon_3 - \epsilon_2 - \epsilon_1 + 2U_2(2-\rho)]} \times \frac{1}{[\epsilon_3 - \epsilon_1 + U_2(1-\rho)]} \quad \text{(A.6)}$$

The sums over ϵ_1 and ϵ_2 extend from 0 to the Fermi energy ϵ_F. Those for ϵ_3 and ϵ_4 go from ϵ_F to the bandwidth W. The sum over the channel index σ results from the fact that Electrons 1 and 2 can share any one of the $N_{\rm ch}$ tunneling channels. Though the formula contains such unphysical terms as that for which $\epsilon_1 = \epsilon_2$, such terms are down by factors of the level spacing δ divided by FW or $(1-F)W$, and their inclusion has no effect in the limit $W/\delta \to \infty$.

Accordingly, we can cease worrying about these terms, for we assume that $\delta \ll U_2 \ll W$, a postulate that permits us to work in the continuum limit, replacing the sums in Eq. A.6 by integrals:

$$\Delta^{(4)}_{N_{\rm ch},2}(\rho) = N_{\rm ch} \left(\frac{t}{\delta}\right)^4 \int_0^{\epsilon_F} d\epsilon_1 \int_0^{\epsilon_F} d\epsilon_2 \int_{\epsilon_F}^{W} d\epsilon_3 \int_{\epsilon_F}^{W} d\epsilon_4 \frac{1}{[\epsilon_3 - \epsilon_2 + U_2(1-\rho)]} \times \frac{1}{[\epsilon_4 + \epsilon_3 - \epsilon_2 - \epsilon_1 + 2U_2(2-\rho)]} \times \frac{1}{[\epsilon_3 - \epsilon_1 + U_2(1-\rho)]} \quad \text{(A.7)}$$

These integrals can be rewritten in terms of dimensionless variables x_i:

$$x_i = \begin{cases} \frac{\epsilon_F - \epsilon_i}{U_2} & \text{for } i = 1 \text{ or } 2 \\ \frac{\epsilon_i - \epsilon_F}{U_2} & \text{for } i = 3 \text{ or } 4\,. \end{cases} \quad \text{(A.8)}$$

With this choice of integration variables, it becomes clear that $\Delta^{(4)}_{N_{\rm ch},2}(\rho)$ is linear in U_2.

Specifically, we find that

$$\begin{aligned}\Delta^{(4)}_{N_{\rm ch},2}(\rho) &= N_{\rm ch}\left(\frac{t}{\delta}\right)^4 U_2 \times I(\rho,F,\psi)\,,\\ I(\rho,F,\psi) &= \int_0^{F\psi}dx_1\int_0^{F\psi}dx_2\int_0^{(1-F)\psi}dx_3\int_0^{(1-F)\psi}dx_4\,\frac{1}{[x_3+x_2+1-\rho]}\\ &\quad\times\frac{1}{[x_4+x_3+x_2+x_1+2(2-\rho)]}\times\frac{1}{[x_3+x_1+1-\rho]}\,. \end{aligned}\tag{A.9}$$

All the shuffling of notatation still leaves us confronted with a quadruple integral. Opting to postpone a frontal assault, we try a sidestepping movement, computing the partial derivative with respect to ρ:

$$I_\rho = \int_0^{F\psi}dx_1\int_0^{F\psi}dx_2\int_0^{(1-F)\psi}dx_3\int_0^{(1-F)\psi}dx_4\left(\frac{1}{[\ ]^2[\ ][\ ]}+\frac{2}{[\ ][\ ]^2[\ ]}+\frac{1}{[\ ][\ ][\ ]^2}\right),\tag{A.10}$$

where the subscript ρ signifies that I_ρ is the partial derivative of I with respect to ρ and the brackets on the right-hand side of the equation have the same contents in the same order as those in Eq. A.9. As the third term in the integrand of Eq. A.10 differs from the first only by an exchange of the indices 1 and 2, we can drop the third term and double the first. When the enhanced first term is integrated by parts with respect to x_2, the products are two triple-integral terms and a quadruple-integral term that exactly cancels the second term of Eq. A.10. With the definitions $A_0 = 0$ and $A_1 = F\psi$, we have

$$\begin{aligned}I_\rho &= 2\sum_{p=0}^{1}(-1)^p\int_0^{F\psi}dx_1\int_0^{(1-F)\psi}dx_3\int_0^{(1-F)\psi}dx_4\,\frac{1}{[x_3+A_p+1-\rho]}\\ &\quad\times\frac{1}{[x_4+x_3+x_1+A_p+2(2-\rho)]}\times\frac{1}{[x_3+x_1+1-\rho]}\,.\end{aligned}\tag{A.11}$$

Having benefited once from differentiation with respect to ρ, we try it again. The second derivative of I with respect to ρ has the following form:

$$I_{\rho\rho} = 2\int_0^{F\psi}dx_1\int_0^{(1-F)\psi}dx_3\int_0^{(1-F)\psi}dx_4\left(\frac{1}{[\ ]^2[\ ][\ ]}+\frac{2}{[\ ][\ ]^2[\ ]}+\frac{1}{[\ ][\ ][\ ]^2}\right),\tag{A.12}$$

where the bracket contents correspond—in order of appearance— to those of Eq. A.11. $I_{\rho\rho}$ lacks the convenient symmetry between first and third terms that was so handy before. Nevertheless, integration of the first term by parts with respect to x_3 still helps. The triple integrals that result cancel the third term and half the middle term, leaving

$$\begin{aligned} I_{\rho\rho} &= 2\sum_{p=0}^{1}\sum_{q=0}^{1}\frac{(-1)^{p+q}}{A_p+B_q+1-\rho}\int_0^{F\psi}dx_1\int_0^{(1-F)\psi}dx_4 \\ &\quad\times\frac{1}{[x_4+x_1+A_p+B_q+2(2-\rho)][x_1+B_q+1-\rho]} \\ &\quad+2\sum_{p=0}^{1}(-1)^p\int_0^{F\psi}dx_1\int_0^{(1-F)\psi}dx_3\int_0^{(1-F)\psi}dx_4\,\frac{1}{[x_3+A_p+1-\rho]} \\ &\quad\times\frac{1}{[x_4+x_3+x_1+A_p+2(2-\rho)]^2[x_3+x_1+1-\rho]}\,, \end{aligned} \tag{A.13}$$

where $B_0=0$ and $B_1=(1-F)\psi$.

We now straightforwardly integrate over x_4, using the relation

$$\frac{1}{(x+a)(x+b)}=\frac{1}{b-a}\left(\frac{1}{x+a}-\frac{1}{x+b}\right).$$

The result is that

$$\begin{aligned} I_{\rho\rho} &= 2\sum_{p=0}^{1}\sum_{q=0}^{1}\sum_{r=0}^{1}\frac{(-1)^{p+q+r+1}}{A_p+B_q+1-\rho}\int_0^{F\psi}dx\,\frac{\ln[x+A_p+B_q+B_r+2(2-\rho)]}{x+B_q+1-\rho} \\ &\quad+2\sum_{p=0}^{1}\sum_{q=0}^{1}\frac{(-1)^{p+q}}{A_p+B_q+3-\rho}\int_0^{F\psi}dx_1\int_0^{(1-F)\psi}dx_3\,\frac{1}{[x_3+A_p+1-\rho]} \\ &\quad\times\left(\frac{1}{[x_3+x_1+1-\rho]}-\frac{1}{[x_3+x_1+A_p+B_q+2(2-\rho)]}\right). \end{aligned} \tag{A.14}$$

Recalling Eq. A.1, we see that, as $\psi\to\infty$, the leading part of the first term in Eq. A.14 behaves like $[\ln^2\psi/(A_p+B_q+1-\rho)]$ and therefore goes to zero unless $p=q=0$. The same is true for the second term—which upon integration over x_1 will have a form like that of the first term. Hence, we can eliminate the sums over p and q and, after

integrating the second term over x_1, have

$$I_{\rho\rho} = \frac{2}{1-\rho}\sum_{r=0}^{1}(-1)^{r+1}\int_0^{F\psi} dx\,\frac{\ln[x+B_r+2(2-\rho)]}{x+1-\rho} + \frac{2}{3-\rho}\sum_{r=0}^{1}(-1)^{r+1}\int_0^{(1-F)\psi} dx\,\frac{\ln\left[\frac{x+A_r+1-\rho}{x+A_r+2(2-\rho)}\right]}{x+1-\rho}. \qquad \text{(A.15)}$$

We recognize that the second term is nontrivial only for $r=0$ and apply Eq. A.1 to do the remaining integrations over x. After dropping terms that go to zero as $\psi\to\infty$, we arrive at the "final" formula for $I_{\rho\rho}$:

$$I_{\rho\rho} = I_{\rho\rho}^{(1)} + I_{\rho\rho}^{(2)},$$

$$\left(\frac{1-\rho}{2}\right) I_{\rho\rho}^{(1)} = \ln([1-F]\psi)\ln(F\psi) - \frac{1}{2}\ln^2(F\psi) + \sum_{n=1}^{\infty}\frac{(-1)^{n+1}}{n^2}\left(\frac{F}{1-F}\right)^n - \frac{\pi^2}{6} - \ln([1-F]\psi)\ln(1-\rho) - \frac{1}{2}\ln^2(3-\rho) + \ln(3-\rho)\ln(1-\rho) + \sum_{n=1}^{\infty}\frac{(-1)^{n+1}}{n^2}\left(\frac{1-\rho}{3-\rho}\right)^n \quad \text{for } F\le(1-F),$$

$$= \frac{1}{2}\ln^2([1-F]\psi) - \sum_{n=1}^{\infty}\frac{(-1)^{n+1}}{n^2}\left(\frac{1-F}{F}\right)^n - \ln([1-F]\psi)\ln(1-\rho) - \frac{1}{2}\ln^2(3-\rho) + \ln(3-\rho)\ln(1-\rho) + \sum_{n=1}^{\infty}\frac{(-1)^{n+1}}{n^2}\left(\frac{1-\rho}{3-\rho}\right)^n \quad \text{for } F>(1-F);$$

$$\left(\frac{3-\rho}{2}\right) I_{\rho\rho}^{(2)} = \frac{\pi^2}{6} + \frac{1}{2}\ln^2\left(\frac{1-\rho}{3-\rho}\right) - \sum_{n=1}^{\infty}\frac{(-1)^{n+1}}{n^2}\left(\frac{1-\rho}{3-\rho}\right)^n. \qquad \text{(A.16)}$$

Before undoing the differentiations with respect to ρ, we pause to remark on the meaning that can be attached to the derivatives $I_{\rho\rho}$ and I_ρ. The second derivative $I_{\rho\rho}$ can be interpreted physically [after multiplication by $N_{\rm ch}U_2(t/\delta)^4$] as reflecting a change in the effective differential charging energy $U_{\rm eff} = 2[\partial^2 E_g^{(0)}(\rho)/\partial\rho^2]_{\rho=0}$, where $E_g^{(0)}(\rho)$

is the ground-state energy as a function of ρ for a given value of the dimensionless channel conductance g. (One might choose to speak of an effective differential capacitance [67] $C_{\text{eff}} = e^2/2U_{\text{eff}}$.) Similarly, up to a proportionality factor, the first derivative I_ρ can be understood as a tunneling-induced correction to an effective value for $\hat{n}$ (recall Eq. 3.11) [67, 72, 73].

What is desired here, however, is I itself, I being proportional to contribution of diagram 2 to the fourth-order energy shift (recall Eq. A.9). Integrating $I_{\rho\rho}$ twice with respect to ρ gives us I up to additive terms that are constant or linear with respect to ρ:

$$I(\rho, F, \psi) = a_0 + a_1\rho + \int_0^\rho dx_1 \int_0^{x_1} dx_2\, I_{\rho\rho}(x_2, F, \psi)\,. \tag{A.17}$$

As mentioned in Chapter 3 and at the beginning of this appendix, the unknown terms $(a_0 + a_1\rho)$ are not relevant to our result. The $a_1\rho$ term is negligible due to the existence of the mirror image of diagram 2, in which the roles of dots 1 and 2 are exchanged. Such a switch of $\hat{n}_1$ and $\hat{n}_2$ is equivalent —in calculating energies—to taking $\rho \to -\rho$. Consequently, when the total fourth-order shift is calculated, the $a_1\rho$ in Eq. A.17 cancels with the $-a_1\rho$ from the mirror image. Likewise, the a_0 part drops from the final result as we are only concerned with the difference between the energy shifts for arbitrary ρ and $\rho = 0$.

The irrelevance of the a_0 and $a_1\rho$ terms tells us that we need only calculate I modulo terms constant or linear with respect to ρ. In other words, we need only find an equivalence class

$$I(\rho, F, \psi) \cong \int_0^\rho dx_1 \int_0^{x_1} dx_2\, I_{\rho\rho}(x_2, F, \psi)\,, \tag{A.18}$$

where the congruence symbol indicates equivalence up to additive terms that are constant or linear with respect to ρ. We are therefore free to drop any constant or linear terms that crop up on the right side of Eq. A.18.

Confident that we have figured out what we wish to do, we can return to the pedestrian business of doing it. We observe that the ρ-dependent sum in Eq. A.16 can be written in a more integrable form:

$$\sum_{n=1}^{\infty}\frac{(-1)^{n+1}}{n^2}\left(\frac{1-\rho}{3-\rho}\right)^n = \sum_{n=1}^{\infty}\frac{(-1)^{n+1}}{n^2}\left(\frac{1}{3}\right)^n + \ln 2[\ln(1-\rho)-\ln(3-\rho)+\ln 3] + \int_0^{\rho} dx\,\ln\left(\frac{2-x}{3-x}\right)\left(\frac{-1}{1-x}+\frac{1}{3-x}\right). \tag{A.19}$$

Integration of $I^{(1)}_{\rho\rho}$ and $I^{(2)}_{\rho\rho}$ with respect to ρ gives

$$\begin{aligned}\frac{1}{2}I^{(1)}_{\rho} &= -\ln(1-\rho)\left[\ln([1-F]\psi)\ln(F\psi)-\frac{1}{2}\ln^2(F\psi)+\sum_{n=1}^{\infty}\frac{(-1)^{n+1}}{n^2}\left(\frac{F}{1-F}\right)^n\right.\\ &\qquad \left.-\frac{\pi^2}{6}+\sum_{n=1}^{\infty}\frac{(-1)^{n+1}}{n^2}\left(\frac{1}{3}\right)^n+\ln 2\ln 3\right]\\ &\quad +\frac{1}{2}\ln^2(1-\rho)[\ln([1-F]\psi)-\ln 2]-\frac{1}{2}\int_0^{\rho}dx\,\frac{\ln^2(3-x)}{1-x}\\ &\quad +\int_0^{\rho}dx\,\frac{\ln(3-x)\ln(1-x)}{1-x}-\ln 2\int_0^{\rho}dx\,(\rho-x)\frac{\ln(3-x)}{1-x}\\ &\quad +\int_0^{\rho}dx_1\,\frac{\rho-x_1}{1-x_1}\int_0^{x_1}dx_2\,\ln\left(\frac{2-x_2}{3-x_2}\right)\left(\frac{-1}{1-x_2}+\frac{1}{3-x_2}\right)\\ &\qquad \text{for } F\le(1-F),\\ &= -\ln(1-\rho)\left[\frac{1}{2}\ln^2([1-F]\psi)-\sum_{n=1}^{\infty}\frac{(-1)^{n+1}}{n^2}\left(\frac{1-F}{F}\right)^n\right.\\ &\qquad \left.+\sum_{n=1}^{\infty}\frac{(-1)^{n+1}}{n^2}\left(\frac{1}{3}\right)^n+\ln 2\ln 3\right]+\ldots\\ &\qquad \text{for } F>(1-F);\end{aligned} \tag{A.20}$$

$$\begin{aligned}
\frac{1}{2}I_\rho^{(2)} &= [-\ln(3-\rho)+\ln 3]\left[\frac{\pi^2}{6}-\sum_{n=1}^{\infty}\frac{(-1)^{n+1}}{n^2}\left(\frac{1}{3}\right)^n-\ln 2\ln 3\right] \\
&\quad -\frac{1}{6}[\ln^3(3-\rho)-\ln^3 3]-\frac{1}{2}\ln 2[\ln^2(3-\rho)-\ln^2 3] \\
&\quad +\frac{1}{2}\int_0^\rho dx\,\frac{\ln^2(1-x)}{3-x}-\int_0^\rho dx\,\frac{\ln(3-x)\ln(1-x)}{3-x} \\
&\quad -\ln 2\int_0^\rho dx\,\frac{\ln(1-x)}{3-x} \\
&\quad -\int_0^\rho dx_1\,\frac{1}{3-x_1}\int_0^{x_1} dx_2\,\ln\left(\frac{2-x_2}{3-x_2}\right)\left(\frac{-1}{1-x_2}+\frac{1}{3-x_2}\right). \qquad \text{(A.21)}
\end{aligned}$$

The ellipsis in the second equation for $I_\rho^{(1)}$ indicates that the remainder of $I_\rho^{(1)}$ for the system above half filling is the same as the corresponding remainder for the system below or at half filling.

In deriving Eqs. A.20 and A.21, we eliminated a number of integrals over x_i's by using an identity [86] that is easily derived for double integrals:

$$\int_0^\rho dx_1\int_0^{x_1} dx_2\,f(x_2)=\int_0^\rho dx\,(\rho-x)f(x)\,. \qquad \text{(A.22)}$$

Nonetheless, in the final terms of $I_\rho^{(1)}$ and $I_\rho^{(2)}$, double integrals remain. These can be reduced to single-integral form with a little extra work. Defining $L^{(m)}$ to be the last term of $(1/2)I_\rho^{(m)}$, we discover that

$$\begin{aligned}
L^{(1)} &= \int_0^\rho dx_2\,\ln\left(\frac{2-x_2}{3-x_2}\right)\left(\frac{-1}{1-x_2}+\frac{1}{3-x_2}\right)\int_{x_2}^\rho dx_1\left(1-\frac{1-\rho}{1-x_1}\right) \\
&= \int_0^\rho dx\,\ln\left(\frac{2-x}{3-x}\right)\left(\frac{-1}{1-x}+\frac{1}{3-x}\right) \\
&\qquad \times\left[(\rho-x)+(1-\rho)\ln\left(\frac{1-\rho}{1-x}\right)\right], \\
L^{(2)} &= -\int_0^\rho dx\,\ln\left(\frac{2-x}{3-x}\right)\left(\frac{-1}{1-x}+\frac{1}{3-x}\right) \\
&\qquad \times\left[(\rho-x)+(3-\rho)\ln\left(\frac{3-\rho}{3-x}\right)\right]. \qquad \text{(A.23)}
\end{aligned}$$

We perform the final integration over ρ to derive $I^{(1)}$ and $I^{(2)}$, where $I \cong I^{(1)} + I^{(2)}$.

$$\begin{aligned}
\frac{1}{2}I^{(1)} \cong\ & (1-\rho)\ln(1-\rho)\left[\ln([1-F]\psi)\ln(F\psi) - \frac{1}{2}\ln^2(F\psi) + \sum_{n=1}^{\infty}\frac{(-1)^{n+1}}{n^2}\left(\frac{F}{1-F}\right)^n\right. \\
& \left. -\frac{\pi^2}{6} + \sum_{n=1}^{\infty}\frac{(-1)^{n+1}}{n^2}\left(\frac{1}{3}\right)^n + \ln 2\ln 3 + \ln([1-F]\psi) - \ln 2\right] \\
& - \frac{1}{2}(1-\rho)\ln^2(1-\rho)[\ln([1-F]\psi) - \ln 2] - \frac{1}{2}\int_0^{\rho} dx\,(\rho-x)\frac{\ln^2(3-x)}{1-x} \\
& + \int_0^{\rho} dx\,(\rho-x)\frac{\ln(3-x)\ln(1-x)}{1-x} - \ln 2\int_0^{\rho} dx\,(\rho-x)\frac{\ln(3-x)}{1-x} \\
& + (1-\rho)\int_0^{\rho} dx\,\ln\left(\frac{1-\rho}{1-x}\right)\ln\left(\frac{2-x}{3-x}\right)\left(\frac{-1}{1-x}+\frac{1}{3-x}\right) \\
& + \int_0^{\rho} dx\,(\rho-x)\ln\left(\frac{2-x}{3-x}\right)\left(\frac{-1}{1-x}+\frac{1}{3-x}\right) \\
& \qquad \text{for}\quad F \le (1-F), \\
\cong\ & (1-\rho)\ln(1-\rho)\left[\frac{1}{2}\ln^2([1-F]\psi) - \sum_{n=1}^{\infty}\frac{(-1)^{n+1}}{n^2}\left(\frac{1-F}{F}\right)^n\right. \\
& \left. + \sum_{n=1}^{\infty}\frac{(-1)^{n+1}}{n^2}\left(\frac{1}{3}\right)^n + \ln 2\ln 3 + \ln([1-F]\psi) - \ln 2\right] + \ldots \\
& \qquad \text{for}\quad F > (1-F); \qquad\qquad (A.24)
\end{aligned}$$

$$\begin{aligned}
\frac{1}{2}I^{(2)} \cong\ & [(3-\rho)\ln(3-\rho) - 3\ln 3]\left[\frac{\pi^2}{6} - \sum_{n=1}^{\infty}\frac{(-1)^{n+1}}{n^2}\left(\frac{1}{3}\right)^n - \ln 2\ln 3\right] \\
& + \frac{1}{6}\left[(3-\rho)\ln^3(3-\rho) - 3(3-\rho)\ln^2(3-\rho) + 6(3-\rho)\ln(3-\rho)\right. \\
& \qquad \left. -3\ln^3 3 + 9\ln^2 3 - 18\ln 3\right] \\
& + \frac{1}{2}\ln 2[(3-\rho)\ln^2(3-\rho) - 2(3-\rho)\ln(3-\rho) - 3\ln^2 3 + 6\ln 3] \\
& + \frac{1}{2}\int_0^{\rho} dx\,(\rho-x)\frac{\ln^2(1-x)}{3-x} - \int_0^{\rho} dx\,(\rho-x)\frac{\ln(3-x)\ln(1-x)}{3-x} \\
& - \ln 2\int_0^{\rho} dx\,(\rho-x)\frac{\ln(1-x)}{3-x} \\
& - (3-\rho)\int_0^{\rho} dx\,\ln\left(\frac{3-\rho}{3-x}\right)\ln\left(\frac{2-x}{3-x}\right)\left(\frac{-1}{1-x}+\frac{1}{3-x}\right) \\
& - \int_0^{\rho} dx\,(\rho-x)\ln\left(\frac{2-x}{3-x}\right)\left(\frac{-1}{1-x}+\frac{1}{3-x}\right). \qquad\qquad (A.25)
\end{aligned}$$

We are essentially done. Upon multiplying the sum of $I^{(1)}$ and $I^{(2)}$ by $N_{\rm ch}U_2(t/\delta)^4$, we have the relevant contribution from diagram 2 to the fourth-order energy shift. After so much work, one might wonder whether we have achieved anything more. Providentially, the answer is that, yes, we have. As explained earlier, we have also solved for the contribution from the corresponding mirror-image diagram, which is obtained by replacing ρ with $-\rho$ in Eqs. A.24 and A.25. Perhaps more surprisingly, we have solved for the contributions from another pair of mirror-image terms. A swap of F and $(1-F)$ in Eq. A.9 turns it into the formula for the contribution from diagram 3 of Fig. 3.1(b). Thus, exchanging F and $(1-F)$ in Eqs. A.24 and A.25 yields the contribution from diagram 3. A further replacement of ρ with $-\rho$ gives the contribution from mirror image of diagram 3. The cost of calculating diagram 2 is high, but at least we benefit from a package deal—4 for the price of 1.

Appendix B

Strong-Coupling Details

This appendix consists of three parts presenting various calculations described or cited in Chapter 3. The first part computes $S_{\rm b}^{(1)}$, $S_{\rm b}^{(2)}$, $K_{\rm c}(\tau)$, and $K_{\rm c}(0)$, thereby producing the results quoted in Eq. 3.29 and making explicit the origin of the factor e^{γ} that appears in the prefactors of Eqs. 3.32 and 3.36. The second part of the appendix provides the derivation of the first strong-coupling energy correction (see Eq. 3.35). The third part derives the second strong-coupling correction (see Eq. 3.36).

B.1 Calculation of Charge-Channel Averages

The leap from Eq. 3.28 to Eq. 3.29 requires evaluation of the expectation values

$$\begin{aligned} D_1 &= \left\langle \cos\left[\sqrt{\pi}\theta_{\rm c}(\tau) + \frac{\pi\rho}{2}\right]\right\rangle_{\rm c}, \\ D_2 &= \left\langle \cos\left[\sqrt{\pi}\theta_{\rm c}(\tau_1) + \frac{\pi\rho}{2}\right] \cos\left[\sqrt{\pi}\theta_{\rm c}(\tau_2) + \frac{\pi\rho}{2}\right]\right\rangle_{\rm c}. \end{aligned} \tag{B.1}$$

(Recall that time-ordering is implicit in the path-integral definition of $\langle \hat{A} \rangle_{\rm c}$ in Eq. 3.26.) The cosines and products of cosines can be written as linear combinations of terms of the form $e^{\hat{Z}}$, where $\hat{Z}$ is linear in the charge displacement operators $\theta_{\rm c}(\tau)$ and the charge displacement operators are themselves linear in boson creation and annihilation operators (see, for example, Haldane [57]). Therefore, one can apply a standard relation for

the expectation value of the exponential of a linear combination of boson operators [55]

$$\langle e^{\hat{Z}} \rangle = e^{\frac{1}{2}\langle \hat{Z}^2 \rangle}, \tag{B.2}$$

which can easily be shown to hold for our charge-integration brackets with implicit time-ordering.

Using Eq. B.2, we discover that

$$\begin{aligned} D_1 &= e^{-\frac{\pi}{2}K_c(0)} \cos\left(\frac{\pi\rho}{2}\right), \\ D_2 &= \frac{1}{2} e^{-\pi[K_c(0)+K_c(\tau_1-\tau_2)]} \cos(\pi\rho) + \frac{1}{2} e^{-\pi[K_c(0)-K_c(\tau_1-\tau_2)]}, \end{aligned} \tag{B.3}$$

where $K_c(\tau)$ is the charge-charge correlation function,

$$K_c(\tau) = \langle \theta_c(\tau)\theta_c(0) \rangle_c . \tag{B.4}$$

Eq. 3.29 of Chapter 3 follows immediately. We find $S_b^{(1)}$ by replacing $\cos[\sqrt{\pi}\theta_c(\tau)+\pi\rho/2]$ in S_b with D_1. For $S_b^{(2)}$, we recall that

$$\langle [S_b - \langle S_b \rangle_c]^2 \rangle_c = \langle S_b^2 \rangle_c - \langle S_b \rangle_c^2$$

and apply the formulas for D_1 and D_2 accordingly.

To get the formula for $K_c(\tau)$ (Eq. 3.30), we must labor a bit more. Because the unperturbed action $S_0^{(c)}$ is quadratic in charge displacement operators $\tilde{\theta}_c(\omega_m)$, $S_0^{(c)}$ fits exactly the form for the canonical action of a real scalar field [95]. Consequently,

$$\left\langle \tilde{\theta}_c(\omega_m)\tilde{\theta}_c(-\omega_n) \right\rangle_c = \frac{\beta}{|\omega_m| + \frac{2U_2}{\pi}} \delta_{\omega_m,\omega_n} . \tag{B.5}$$

From this identity and the relation between $\theta_c(\tau)$ and its Fourier transform (recall Eq. 3.20), we construct a summation formula for $K_c(\tau)$:

$$K_c(\tau) = \frac{1}{\beta} \sum_{\omega_m} \frac{e^{-i\omega_m\tau}}{|\omega_m| + \frac{2U_2}{\pi}}. \tag{B.6}$$

In the zero-temperature ($\beta \to \infty$) limit, we may safely transform this sum into an integral. Before doing so, however, we should note that, unless the ω_m possess an

ultraviolet cutoff, $K_c(0)$ diverges logarithmically. The standard means of imposing such a cutoff in Luttinger liquid theory [55] is to insert a factor of $e^{-|\omega_m|/W}$ on the right side of Eq. 3.20. This insertion generates a factor of $e^{-2|\omega_m|/W}$ in Eq. B.6, yielding

$$K_c(\tau) = \int_{-\infty}^{\infty} \frac{d\omega}{2\pi} \frac{e^{-i\omega\tau} e^{-\frac{2|\omega|}{W}}}{|\omega| + \frac{2U_2}{\pi}}, \tag{B.7}$$

which is equivalent to Eq. 3.30 in Chapter 3.

The way is clear for evaluation of the same-time corrrelation function $K_c(0)$. After setting $\tau = 0$ in Eq. B.7, we integrate by parts and convert to the dimensionless integration variable $x = 2\omega/W$. The result is that

$$\begin{aligned} K_c(0) &= -\frac{1}{\pi} \ln\left(\frac{4U_2}{\pi W}\right) + \frac{1}{\pi} \int_0^{\infty} dx\, e^{-x} \ln\left(x + \frac{4U_2}{\pi W}\right) \\ &= -\frac{1}{\pi} \ln\left(\frac{4U_2}{\pi W}\right) + \frac{1}{\pi} e^{(4U_2/\pi W)} \left(\int_0^{\infty} dx\, e^{-x} \ln x - \int_0^{4U_2/\pi W} dx\, e^{-x} \ln x \right) \end{aligned} \tag{B.8}$$

The first integral in the parentheses equals the negative of γ, the Euler-Mascheroni constant [85]. The second integral goes to zero as we take the limit $W/U_2 \to \infty$. In this limit, the exponential factor multiplying the integrals goes to 1. The final result is the following:

$$K_c(0) = -\frac{1}{\pi} \ln\left(\frac{4e^{\gamma} U_2}{\pi W}\right). \tag{B.9}$$

The derivation of Eq. B.9 shows that the coefficient e^{γ} comes from exponentiating a secondary part of $\langle \theta_c(0)\theta_c(0) \rangle$. One might be concerned that Luttinger liquid theory does not faithfully capture such subsidiary dependences [96]. However, Chapter 3 presents evidence that these coefficients are general and independent of the high-energy band structure.

B.2 The First Strong-Coupling Correction

As stated in Chapter 3, in the limit of strong coupling ($g \to 1$), the first correction (see Eq. 3.35) to the open-channel ($g = 0$) ground-state energy is obtained by diagonalizing

the Hamiltonian $H_{\rm New}$ (see Eq. 3.31). This diagonalization can be accomplished through another version of the "debosonization" procedure used by Matveev [34]. As we wish to "debosonize" the action $S_{\rm New} = S_0^{(\rm s)} + S_{\rm b}^{(1)}$ (recall Eqs. 3.23 and 3.29), it is useful to observe that $S_0^{(\rm s)}$ corresponds to the Euclidean action for non-interacting fermions on a semi-infinite lattice ending at $x = 0$ [62]. For these fermions, we take $\theta_{\rm s}(\tau)$ to correspond to the $x = 0$ value of the phase field, $\phi_f(\tau) = \Phi_f(0,\tau)$, rather than the $x = 0$ value of the charge displacement field $\theta_f(\tau) = \Theta_f(0,\tau)$. Making $\theta_f(\tau) = 0$ the boundary condition at the edge, we find that the properly normalized creation operator for a fermion at $x = 0$ is given by

$$\psi_f^\dagger(0,\tau) = \sqrt{\frac{W}{4\pi\hbar v_F}}\, e^{i\sqrt{\pi}\phi_f(\tau)}\,, \tag{B.10}$$

where, as usual, W is the bandwidth and v_F is the Fermi velocity [62, 55]. $\psi_f^\dagger(0,\tau)$ can be expressed in terms of reciprocal-space creation operators:

$$\psi_f^\dagger(0,\tau) = \frac{1}{\sqrt{2\pi}} \int_{-\Lambda}^{\Lambda} dk\, f_k^\dagger\,. \tag{B.11}$$

The fermionic energies are cut off in the usual way at $W/2$, the corresponding wavevector cutoff being $\Lambda = W/2\hbar v_F$.

After these machinations, "refermionization" proceeds apace. Since the unperturbed action $S_0^{(\rm s)}$ is an action for non-interacting fermions, the unperturbed Hamiltonian $H_0^{(\rm s)}$ is simply the sum of the single-particle energies of those fermions. On the other hand, the perturbation $H_{\rm b}^{(1)}$ that corresponds to $S_{\rm b}^{(1)}$ is a term linear in fermion creation and annihilation operators. In particular, using Eq. B.9 to determine $e^{-\frac{\pi}{2}K_{\rm c}(0)}$, we obtain

$$\begin{aligned} H_0^{(\rm s)} &= \int_{-\Lambda}^{\Lambda} dk\, \xi_k\, f_k^\dagger f_k\,, \\ H_{\rm b}^{(1)} &= \tilde{V} \cos\!\left(\frac{\pi\rho}{2}\right) \sqrt{\frac{2e^\gamma \hbar v_F U_2}{\pi^3}} \int_{-\Lambda}^{\Lambda} dk\, (f_k^\dagger + f_k)\,, \end{aligned} \tag{B.12}$$

Not being quadratic in fermion creation and annihilation operators, the fermionic Hamiltonian we have derived is not yet in an easily diagonalizable form. To make it so,

we follow Matveev [34] in defining a new set of fermion operators such that

$$f_k = (d + d^\dagger)d_k \,. \tag{B.13}$$

Plugging this equivalence into Eq. B.12 yields Eq. 3.32.

One can now perform the Bogoliubov transformation that produces Eq. 3.33. To find the correction to the open-channel energy, one notes that $H_{\rm b}^{(1)}$ of Eq. 3.32 has an expectation value of zero in the ground state of $H_0^{(s)}$, which is the open-channel ($\tilde{V} = 0$) part of $H_{\rm New}$. Therefore, if the ground state of $H_0^{(s)}$ is represented by the ket $|0\rangle$, $\langle 0|H_{\rm New}|0\rangle = E_0$, where E_0 is the ground-state energy for $H_0^{(s)}$. From the diagonalized form of $H_{\rm New}$ (see Eq. 3.33), it is then deduced that the equation for $(E_{\rm New} - E_0)$ is the following:

$$\Delta_{\rm str}^{(1)}(\rho) = -\int_0^\Lambda dk\, \xi_k \langle 0|C_k^\dagger C_k + \tilde{C}_k^\dagger \tilde{C}_k|0\rangle \,. \tag{B.14}$$

At this point, it is necessary to know the exact equations for $\tilde{C}_k$ and C_k. As found by Matveev [34], for $\Gamma = \tilde{V}^2[8e^\gamma U_2/\pi^2]\cos^2(\pi\rho/2)$, they are

$$\begin{aligned}
\tilde{C}_k &= \frac{d_k + d_{-k}^\dagger}{\sqrt{2}}\,, \\
C_k &= \frac{\xi_k}{\sqrt{\xi_k^2 + \Gamma^2}} \frac{d_k - d_{-k}^\dagger}{\sqrt{2}} - \sqrt{\frac{\hbar v_F \Gamma}{2\pi(\xi_k^2 + \Gamma^2)}}(d + d^\dagger) \\
&\quad + \frac{\Gamma}{\pi\sqrt{\xi_k^2 + \Gamma^2}} \mathcal{P}\int_{-\Lambda}^{\Lambda} \frac{d\xi_{k'}}{\xi_k - \xi_{k'}} \frac{d_{k'} - d_{-k'}^\dagger}{\sqrt{2}}\,,
\end{aligned} \tag{B.15}$$

As before, the symbol $\mathcal{P}$ indicates that only the principal value of the integral is computed.

With the explicit equations for $\tilde{C}_k$ and C_k before us, it is clear that, for $k > 0$, $\tilde{C}_k|0\rangle = 0$, and

$$\Delta_{\rm str}^{(1)}(\rho) = -\int_0^\Lambda dk\, \xi_k \langle 0|C_k^\dagger C_k|0\rangle. \tag{B.16}$$

Concentrating on what remains, we see that, for $k > 0$, both the first term of C_k and the $k' > 0$ part of the third term of C_k annihilate the $H_0^{(s)}$ ground state. Hence,

$$\begin{aligned}\langle 0|C_k^\dagger C_k|0\rangle &= \frac{\hbar v_F \Gamma}{2\pi(\xi_k^2+\Gamma^2)}\left[1+\frac{\Gamma}{\pi\hbar v_F}\int_0^\Lambda \frac{dk'}{k+k'}\int_0^\Lambda \frac{dk''}{k+k''}\langle 0|(d_{-k'}^\dagger + d_{k'})(d_{-k''}+d_{k''}^\dagger)|0\rangle\right] \\ &= \frac{\hbar v_F \Gamma}{2\pi(\xi_k^2+\Gamma^2)} + \frac{\Gamma^2}{\pi^2(\xi_k^2+\Gamma^2)}\left(\frac{1}{k}-\frac{1}{k+\Lambda}\right). \qquad \text{(B.17)}\end{aligned}$$

Plugging into Eq. B.16, we find that

$$\begin{aligned}\Delta_{\rm str}^{(1)}(\rho) &= -\frac{\Gamma}{2\pi}\int_0^{W/2}\frac{\xi_k\,d\xi_k}{\xi_k^2+\Gamma^2} - \frac{\Gamma^2}{\pi^2}\int_0^{W/2}\frac{d\xi_k}{\xi_k^2+\Gamma^2} \\ &\qquad + \frac{\Gamma^2}{\pi^2}\int_0^{W/2}\frac{\xi_k\,d\xi_k}{(\xi_k+\frac{W}{2})(\xi_k^2+\Gamma^2)} \\ &= -\frac{\Gamma}{4\pi}\ln\left(\frac{W^2}{4\Gamma^2}+1\right) - \frac{\Gamma}{2\pi} \\ &= -\frac{\Gamma}{2\pi}\left[\ln\left(\frac{W}{2\Gamma}\right)+1\right]. \qquad \text{(B.18)}\end{aligned}$$

Here we have dropped terms that vanish in the limit $W/U_2 \to \infty$. Application of the identity $\Gamma = \tilde{V}^2[8e^\gamma U_2/\pi^2]\cos^2(\pi\rho/2)$ yields Eq. 3.35.

B.3 The Second Strong-Coupling Correction

The second correction term in the strong-coupling limit (see Eq. 3.36) is derived by treating $S_{\rm b}^{(2)}$ (see Eq. 3.29) as a perturbation to the system described by $H_{\rm New}$ of Eq. 3.31. Using the standard formula for the grand-canonical potential in the finite-temperature path-integral approach [97],

$$\Omega - \Omega_0 = -\frac{1}{\beta}\sum(\text{All connected graphs}), \qquad \text{(B.19)}$$

we see that the lowest-order correction to the ground-state energy of $H_{\rm New}$ is given by

$$\Delta_{\rm str}^{(2)}(\rho) = \lim_{\beta\to\infty}\frac{1}{\beta}\langle{\rm New}|S_{\rm b}^{(2)}|{\rm New}\rangle, \qquad \text{(B.20)}$$

where $|{\rm New}\rangle$ is the ground-state ket for $H_{\rm New}$. The minus sign in Eq. B.19 has been canceled by the minus sign that arises from the fact that this leading term from $S_{\rm b}^{(2)}$ corresponds to a first-order graph and therefore carries a factor of -1 [97].

Recalling Eq. 3.29 and observing that the parts of $S_{\rm b}^{(2)}$ that are independent of ρ are irrelevant to calculation of the fractional peak splitting f, our immediate task is to evaluate the quantity

$$X(\tau_1,\tau_2) = \left(\frac{\tilde{V}W}{\pi}\right)^2 e^{-\pi K_{\rm c}(0)} \cos^2\!\left(\frac{\pi\rho}{2}\right) \langle \text{New}|\cos[\sqrt{\pi}\theta_{\rm s}(\tau_1)]\cos[\sqrt{\pi}\theta_{\rm s}(\tau_2)]|\text{New}\rangle\,. \tag{B.21}$$

Under "debosonization" (see Part 2 of this appendix), this becomes

$$\begin{aligned} X(\tau_1,\tau_2) &= \lambda^2 \int_{-\Lambda}^{\Lambda} dk_1 \int_{-\Lambda}^{\Lambda} dk_2 \,\langle \text{New}|\left[d_{k_1}^\dagger(d+d^\dagger)+(d+d^\dagger)d_{k_1}\right]_{\tau_1} \\ &\quad \times \left[d_{k_2}^\dagger(d+d^\dagger)+(d+d^\dagger)d_{k_2}\right]_{\tau_2} |\text{New}\rangle\,, \end{aligned} \tag{B.22}$$

where the bracket subscripts indicate that the enclosed operators are evaluated at imaginary times τ_1 and τ_2, respectively, and we have used

$$\lambda = \tilde{V}\cos(\pi\rho/2)\sqrt{2e^\gamma \hbar v_F U_2/\pi^3}\,. \tag{B.23}$$

We are now within hailing distance of Eq. 3.36. Using the truncated equations for C_k and $\tilde{C}_k$ (recall Eq. 3.34), we express the d_k's in terms of these operators. The subleading terms in this transformation are negligible as, in the end result, they take us beyond second order in $\tilde{V}$. Similarly, the time-dependence of the operator sum $(d+d^\dagger)$ is subleading as $(d+d^\dagger)$ first appears in the expansion of the diagonalizing operators at order $\tilde{V}$. Accordingly, $(d+d^\dagger)$ commutes with $H_{\rm New}$ to zeroth order and can be considered time-independent. In contrast, from Eq. 3.33, we know that $C_k(\tau) = C_k e^{-\xi_k \tau}$ and $C_k^\dagger(\tau) = C_k^\dagger e^{\xi_k \tau}$. Application of these insights to Eq. B.22 gives

$$\begin{aligned} X(\tau_1,\tau_2) &= 2\lambda^2 \int_0^{\Lambda} dk_1 \int_0^{\Lambda} dk_2 \,\langle \text{New}|C_{k_1}(\tau_1)C_{k_2}^\dagger(\tau_2)|\text{New}\rangle \\ &= \frac{2\lambda^2}{\hbar v_F}\int_0^{W/2} d\xi\, e^{-(\tau_1-\tau_2)\xi} \\ &= \frac{2\lambda^2}{\hbar v_F}\,\frac{1-e^{-(\tau_1-\tau_2)W/2}}{\tau_1-\tau_2}\,. \end{aligned} \tag{B.24}$$

We now return to Eqs. 3.29 and B.20. Switching to dimensionless variables $x_i = \tau_i W/2$ and substituting for λ, we obtain

$$\Delta^{(2)}_{\rm str}(\rho) = \tilde{V}^2 \cos^2\Big(\frac{\pi\rho}{2}\Big) \frac{8e^{\gamma}U_2}{\pi^3\beta W} \int_0^{\beta W/2} dx_1 \int_0^{x_1} dx_2 \times \Big(1 - e^{-\pi K_c[2(x_1-x_2)/W]}\Big) \frac{1 - e^{-(x_1-x_2)}}{x_1 - x_2} . \quad \text{(B.25)}$$

We eliminate one of the integrations by expressing the integrand in terms of $x = (x_1 - x_2)$ and observing that in the double-integral the density of states for a given value of x is $(\beta W/2 - x)$:

$$\Delta^{(2)}_{\rm str}(\rho) = \tilde{V}^2 \cos^2\Big(\frac{\pi\rho}{2}\Big) \frac{8e^{\gamma}U_2}{\pi^3\beta W} \int_0^{\beta W/2} dx \left(\frac{\beta W}{2} - x\right) \Big[1 - e^{-\pi K_c(2x/W)}\Big] \frac{1 - e^{-x}}{x} \quad \text{(B.26)}$$

Transformation of Eq. B.26 into Eq. 3.36 follows recognition of the fact that, for x on the order of $\beta W/2$, the integrand is effectively zero. This is known from the identity

$$K_c\left(\frac{2x}{W}\right) = -\frac{1}{\pi}\mathrm{Re}\left\{e^{(4U_2/\pi W)(1+ix)}\,\mathrm{Ei}\left[-(4U_2/\pi W)(1+ix)\right]\right\} , \quad \text{(B.27)}$$

where $\mathrm{Ei}[-z]$ is the first exponential integral function [98]. For $z \gg 1$, $\mathrm{Ei}[-z]$ goes as e^{-z}/z. Therefore, the integrand goes to zero as $1/x^2$ for $x > \pi W/4U_2$, and the region $x \gg W/U_2$ makes a comparatively negligible contribution to the integral. This conclusion corroborates the statement made in Chapter 3 that the factor $[1 - e^{-\pi K_c(2x/W)}]$ furnishes an ultraviolet cutoff on the order of $\psi = W/U_2$. Since we calculate in the limit $\beta \to \infty$, we know that $\beta W/2 \gg W/U_2$ and, hence, that the integrand is effectively zero for x on the order of $\beta W/2$. We can approximate the weight function $(\beta W/2 - x)$ by $(\beta W/2)$. The result is Eq. 3.36.

Bibliography

[1] BERNARD LE BOVIER DE FONTENELLE, *Conversations on the Plurality of Worlds*, translated by H. A. Hargreaves (University of California Press, Berkeley, 1990), p. 22.

[2] C. W. J. BEENAKKER and H. VAN HOUTEN, Solid State Phyics **44**, 1 (1991).

[3] Y. IMRY, in *Directions in Condensed Matter Physics*, edited by G. Grinstein and G. Mazenko (World Scientific, Singapore, 1986).

[4] S. WASHBURN and R. A. WEBB, Rep. Prog. Phys. **55**, 1311 (1992).

[5] Y. IMRY, *Introduction to Mesoscopic Physics* (Oxford, New York, 1997).

[6] K. K. LIKHAREV, IBM J. Res. Develop. **32**, 144 (1988).

[7] U. MEIRAV, M. A. KASTNER, and S. J. WIND, Phys. Rev. Lett. **65**, 771 (1990).

[8] D. V. AVERIN and K. K. LIKHAREV, in *Mesoscopic Phenomena in Solids*, edited by B. L. Altshuler, P. A. Lee, and R. A. Webb (North Holland, Amsterdam, 1991).

[9] C. W. J. BEENAKKER, Phys. Rev. B **44**, 1646 (1991).

[10] P. LAFARGE, H. POTHIER, E. R. WILLIAMS, D. ESTEVE, C. URBINA, and M. H. DEVORET, Z. Phys B **85**, 327 (1991).

[11] H. GRABERT, Z. Phys. B **85**, 319 (1991).

[12] M. H. DEVORET, D. ESTEVE, and C. URBINA, Nature **360**, 547 (1992).

[13] M. DEVORET and H. GRABERT, in *Single Charge Tunneling*, Vol. 294 of *NATO Advanced Study Institute Series B Physics*, edited by H. Grabert and M. H. Devoret (Plenum, New York, 1992).

[14] H. VAN HOUTEN, C. W. J. BEENAKKER, and A. A. M. STARING, in *Single Charge Tunneling*, Vol. 294 of *NATO Advanced Study Institute Series B Physics*, edited by H. Grabert and M. H. Devoret (Plenum, New York, 1992).

[15] M. A. KASTNER, Rev. Mod. Phys. **64**, 849 (1992).

[16] M. TINKHAM, *Introduction to Superconductivity* (McGraw-Hill, New York, 1996), ch. 7.

[17] F. STERN, in *Physics of Low-Dimensional Semiconductor Structures*, edited by P. Butcher, N. H. March, and M. P. Tosi (Plenum, New York, 1993).

[18] E. B. FOXMAN, U. MEIRAV, P. L. MCEUEN, M. A. KASTNER, O. KLEIN, P. A. BELK, D. M. ABUSCH, and S. J. WIND, Phys. Rev. B **50**, 14 193 (1994).

[19] F. R. WAUGH, Ph.D. thesis, Harvard University, 1994.

[20] R. C. ASHOORI, Nature **379**, 413 (1996).

[21] U. MEIRAV and E. B. FOXMAN, Semicond. Sci. Technol. **11**, 255 (1996).

[22] C. H. CROUCH, Ph.D. thesis, Harvard University, 1996.

[23] PLATO, in *Plato*, translated by B. Jowett, vol. 7, *Great Books of the Western World*, edited by R. M. Hutchins (Encyclopaedia Britannica, Chicago, 1978), p. 231.

[24] D. C. GLATTLI, C. PASQUIER, U. MEIRAV, F. I. B. WILLIAMS, Y. JIN, and B. ETIENNE, Z. Phys. B **85**, 375 (1991).

[25] J. H. DAVIES and J. A. NIXON, Phys. Rev. B **39**, 3423 (1989).

[26] J. M. GOLDEN and B. I. HALPERIN, unpublished (cond-mat/9611173).

[27] N. GARCÍA and L. ESCAPA, Appl. Phys. Lett. **54**, 1418 (1989).

[28] L. I. GLAZMAN, G. B. LESOVIK, D. E. KHMEL'NITSKII, and R. I. SHEKHTER, JETP Lett. **48**, 238 (1988).

[29] G. KIRCZENOW, Solid State Commun. **68**, 715 (1988).

[30] A. SZAFER and A. D. STONE, Phys. Rev. Lett. **62**, 300 (1989).

[31] B. J. VAN WEES, L. P. KOUWENHOVEN, E. M. M. WILLEMS, C. J. P. M. HARMANS, J. E. MOOIJ, H. VAN HOUTEN, C. W. J. BEENAKKER, J. G. WILLIAMSON, and C. T. FOXON, Phys. Rev. B **43**, 12 431 (1991).

[32] D. A. WHARAM, T. J. THORTON, R. NEWBURY, M. PEPPER, H. AHMED, J. E. F. FROST, D. G. HASKO, D. C. PEACOCK, D. A. RITCHIE, and G. A. C. JONES, J. Phys. C **21**, L209 (1988).

[33] KARSTEN FLENSBERG, Phys. Rev. B **48**, 11 156 (1993); Physica B **203**, 432 (1994).

[34] K. A. MATVEEV, Physica B **203**, 404 (1994); Phys. Rev. B **51**, 1743 (1995).

[35] R. BROWN and E. SIMANEK, Phys. Rev. B **34**, 2957 (1986).

[36] A. GROSHEV, Phys. Rev. B **44**, 1502 (1991).

[37] L. E. HENRICKSON, A. J. GLICK, G. W. BRYANT, and D. F. BARBE, Phys. Rev. B **50**, 4482 (1994).

[38] W. ZWERGER and M. SCHARPF, Z. Phys. B **85**, 421 (1991).

[39] W. ZWERGER, Z. Phys. B **93**, 333 (1994).

[40] D. J. THOULESS, Phys. Rev. Lett. **39**, 1167 (1977).

[41] M. STOPA, Phys. Rev. B **54**, 13 767 (1996).

[42] Y. MEIR, N. S. WINGREEN, and P. A. LEE, Phys. Rev. Lett. **66**, 3048 (1991).

[43] F. R. WAUGH, M. J. BERRY, D. J. MAR, R. M. WESTERVELT, K. L. CAMPMAN, and A. C. GOSSARD, Phys. Rev. Lett. **75**, 705 (1995).

[44] F. R. WAUGH, M. J. BERRY, C. H. CROUCH, C. LIVERMORE, D. J. MAR, R. M. WESTERVELT, K. L. CAMPMAN, and A. C. GOSSARD, Phys. Rev. B **53**, 1413 (1996).

[45] C. H. CROUCH, C. LIVERMORE, F. R. WAUGH, R. M. WESTERVELT, K. L. CAMPMAN, and A. C. GOSSARD, Surf. Sci. **361-362**, 631 (1996).

[46] C. LIVERMORE, C. H. CROUCH, R. M. WESTERVELT, K. L. CAMPMAN, and A. C. GOSSARD, Science, **274**, 1332 (1996).

[47] I. M. RUZIN, V. CHANDRASEKHAR, E. I. LEVIN, and L. I. GLAZMAN, Phys. Rev. B **45**, 13 469 (1992); L. I. GLAZMAN and V. CHANDRASEKHAR, Europhys. Lett. **19**, 623 (1992).

[48] L. W. MOLENKAMP, KARSTEN FLENSBERG, and M. KEMERINK, Phys. Rev. Lett. **75**, 4282 (1995).

[49] C. H. CROUCH and J. M. GOLDEN, unpublished work.

[50] If one approximates the two coplanar regions of charge by parallel semi-infinite sheets, one has an effectively two-dimensional system, and the intercapacitance can be shown, via a conformal transformation, to go as $L \ln(L/d)$, where L is the characteristic size of the coplanar regions and d is the distance between them at nearest approach. This intercapacitance diverges in the limit $d \to 0$. However, a

more realistic model of the charge regions—in which there is an isolated point of nearest approach, rather than an extended line of such points—results in this logarithmic divergence being cut off, the distance d of nearest approach being effectively replaced by some larger and finite quantity that reflects a sort of average distance between the two regions.

[51] J. M. GOLDEN and B. I. HALPERIN, Phys. Rev. B **53**, 3893 (1996).

[52] J. M. GOLDEN and B. I. HALPERIN, Phys. Rev. B **54**, 16 757 (1996).

[53] K. A. MATVEEV, L. I. GLAZMAN, and H. U. BARANGER, Phys. Rev. B **53**, 1034 (1996).

[54] K. A. MATVEEV, L. I. GLAZMAN, and H. U. BARANGER, Phys. Rev. B **54**, 5637 (1996).

[55] V. J. EMERY, "Theory of the One-Dimensional Electron Gas," in *Highly Conducting One-Dimensional Solids*, edited by J. T. Devreese, R. P. Evrard, and V. E. van Doren (Plenum, New York, 1979), 247-303.

[56] R. HEIDENREICH, R. SEILER, and D. A. UHLENBROCK, J. Stat. Phys. **22**, 27 (1980).

[57] F. D. M. HALDANE, Phys. Rev. Lett. **47**, 1840 (1981); J. Phys. C **14**, 2585 (1981).

[58] E. FRADKIN, *Field Theories of Condensed Matter Systems* (Addison-Wesley, Reading, 1991), pp. 74-88.

[59] H. J. SCHULZ, in *Strongly Correlated Electronic Materials: The Los Alamos Symposium 1993*, edited by K. Bedell *et al.* (Addison-Wesley, Reading, Massachusetts, 1994), p. 187

[60] H. J. SCHULZ, in *Mesoscopic Quantum Physics, Les Houches, Session LXI, 1994*, edited by E. Akkermans, G. Montambaux, J.-L. Pichard, and J. Zinn-Justin (Elsevier, Amsterdam, 1995), p. 533.

[61] R. SHANKAR, Acta Phys. Pol. B **26**, 1835 (1996).

[62] C. L. KANE and M. P. A. FISHER, Phys. Rev. Lett. **68**, 1220 (1992); Phys. Rev. B **46**, 7268 (1992). *ibid.* **46**, 15 233 (1992).

[63] A. FURUSAKI and N. NAGAOSA, Phys. Rev. B **47**, 4631 (1993).

[64] C. A. STAFFORD and S. DAS SARMA, Phys. Rev. Lett. **72**, 3590 (1994).

[65] C. A. STAFFORD and S. DAS SARMA, unpublished.

[66] G. KLIMECK, GUANLONG CHEN, and S. DATTA, Phys. Rev. B **50**, 2316 (1994); GUANLONG CHEN, G. KLIMECK, S. DATTA, GUANHA CHEN, and W. A. GODDARD III, *ibid.* **50**, 8035 (1994).

[67] D. S. GOLUBEV and A. D. ZAIKIN, Phys. Rev. B **50**, 8736 (1994).

[68] S. V. PANYUKOV and A. D. ZAIKIN, Phys. Rev. Lett. **67**, 3168 (1991); Physics Letters A **183**, 115 (1993).

[69] A. D. ZAIKIN, D. S. GOLUBEV, and S. V. PANYUKOV, Physica B **203**, 417 (1994).

[70] G. FALCI, J. HEINS, GERD SCHÖN, and GERGELY T. ZIMANYI, Physica B **203**, 409 (1994); G. FALCI, GERD SCHÖN, and GERGELY T. ZIMANYI, Phys. Rev. Lett. **74**, 3257 (1995).

[71] HERBERT SCHOELLER and GERD SCHÖN, Phys. Rev. B **50**, 18 436 (1994).

[72] HERMANN GRABERT, Phys. Rev. B **50**, 17 364 (1994).

[73] L. I. GLAZMAN and K. A. MATVEEV, Zh. Éksp. Teor. Fiz. **98**, 1834 (1990) [Sov. Phys. JETP **71**, 1031 (1990)]; K. A. MATVEEV, *ibid.* **99**, 1598 (1991) [**72**, 892 (1991)]

[74] H. O. FROTA and KARSTEN FLENSBERG, Phys. Rev. B **46**, 15 207 (1992).

[75] R. I. SHEKHTER, Zh. Éksp. Teor. Fiz. **63**, 1410 (1972) [Sov. Phys. JETP **36**, 747 (1973)]; I. O. KULIK and R. I. SHEKHTER, *ibid.* **68**, 623 (1975) [**41**, 308 (1975)].

[76] D. V. AVERIN and K. K. LIKHAREV, Zh. Éksp. Teor. Fiz. **90**, 733 (1986) [Sov. Phys. JETP **63**, 427 (1986)]; J. Low Temp. Phys. **62**, 345 (1986).

[77] In the experiments of Waugh *et al.* [19, 43, 44], a channel between the dots is opened when the voltage on the gate tips framing the channel is changed by a relatively small percent. As the separation between the gate tips is approximately 100 nm, the above suggests that the potential landscape created by the applied voltages varies substantially over distances on the order of 50 nm that are approximately equal to a Fermi wavelength ($\lambda_F \simeq 40$ nm).

[78] For accounts of the formal theory of scattering, see, for example, PAUL ROMAN's *Advanced Quantum Theory* (Addison-Wesley, Reading, MA, 1965) or MICHAEL D. SCADRON's *Advanced Quantum Theory* (Springer-Verlag, Berlin, 1991).

[79] Matveev's result in Ref. [34] arises from exponentiating $\ln(\pi W/N_{ch}e^{\gamma}U_2)$, where W is the Luttinger liquid's ultraviolet cut-off. Luttinger liquid theory's unreliability in calculating the constant factors in the arguments of such logarithms is well documented. See, for example, J. SOLYOM, Adv. Phys. **28**, 201 (1979).

[80] D. V. AVERIN and YU V. NAZAROV, in *Single Charge Tunneling*, Vol. 294 of *NATO Advanced Study Institute Series B Physics*, edited by H. Grabert and M. H. Devoret (Plenum, New York, 1992), p. 217.

[81] F. GUINEA and G. SCHÖN, Europhys. Lett. **1**, 585 (1986); J. Low Temp. Phys. **69**, 219 (1987).

[82] In the limit $\rho \to 1$, calculation of such derivative quantities by a perturbative expansion in g breaks down as a result of divergences with respect to $(1 - \rho)$. In dealing with these divergences, it is useful to map the problem onto that of a two-level system. The divergent behavior is then found to be analogous to that of a Kondo model and can be treated by similar renormalization-group techniques (see Refs. [70] and [73]). Though still providing insights with regard to the form of f_ρ, the two-level system approach is not so useful in computing the fractional peak splitting since f_ρ, though singular, is not divergent as $|\rho| \to 1$. Thus, the non-singular parts of f_ρ cannot be neglected compared to the singular terms that could be obtained from consideration of a two-level system.

[83] GORDON BAYM, *Lectures on Quantum Mechanics* (Addison-Wesley, Redwood City, 1973), p. 113 ff.

[84] M. P. A. FISHER and WILHELM ZWERGER, Phys. Rev. B **32**, 6190 (1985).

[85] MURRAY R. SPEIGEL, *Mathematical Handbook of Formulas and Tables*, Schaum's Outline Series (McGraw-Hill, New York, 1992).

[86] W. H. PRESS, S. A. TEUKOLSKY, W. T. VETTERLING, and B. P. FLANNERY, *Numerical Recipes in C: The Art of Scientific Computing*, 2nd ed. (Cambridge University Press, Cambridge, 1992), 161.

[87] C. J. B. FORD, P. J. SIMPSON, M. PEPPER, D. KERN, J. E. F. FROST, D. A. RITCHIE, and G. A. C. JONES, Nanostructured Materials **3**, 283 (1993).

[88] M. KEMERINK and L. W. MOLENKAMP, Appl. Phys. Lett. **65**, 1012 (1994).

[89] N. TSUKADA, M. GOTODA, and M. NUNOSHITA, Phys. Rev. B **50**, 5764 (1994).

[90] T. SAKAMOTO, S. W. HWANG, F. NIHEY, Y. NAKAMURA, and K. NAKAMURA, Jpn. J. Appl. Phys. **33**, 4876 (1994); Superlattices and Microstructures **16**, 291 (1994).

[91] F. HOFMANN, T. HEINZEL, D. A. WHARAM, J. P. KOTTHAUS, G. BÖHM, W. KLEIN, G. TRÄNKLE, and G. WEIMANN, Phys. Rev. B **51**, 13 872 (1995).

[92] R. J. HAUG, R. H. BLICK, and T. SCHMIDT, Physica B **212**, 207 (1995).

[93] N. C. VAN DER VAART, S. F. GODIJN, Y. V. NAZAROV, C. J. P. M. HARMANS, J. E. MOOIJ, L. W. MOLENKAMP, and C. T. FOXON, Phys. Rev. Lett. **74**, 4702 (1995).

[94] P. PALS and A. MACKINNON, J. Phys.: Condens. Matter **8**, 3177 (1996); **8**, 5401 (1996).

[95] D. J. AMIT, *Field Theory, the Renormalization Group, and Critical Phenomena*, 2nd ed. (World Scientific, Singapore, 1984).

[96] J. SOLYOM, Adv. Phys. **28**, 201 (1979).

[97] J. W. NEGELE and H. ORLAND, *Quantum Many-Particle Systems* (Addison-Wesley, Redwood City, 1988).

[98] I. S. GRADSHTEYN and I. M. RYZHIK, *Table of Integrals, Series, and Products*, 5th ed., edited by Alan Jeffrey (Academic, Boston, 1994), 933ff.

[99] The *dimensionless conductance* g is the "dimensionful" channel conductance divided by the conductance quantum e^2/h. *Tunneling channel* refers to any distinct orbital *or* spin tunneling mode. Thus, in the spin-1/2, zero-magnetic-field systems with which we are primarily concerned, the number of tunneling channels $N_{\rm ch}$ is twice the number of orbital tunneling modes.

[100] C. B. DUKE, *Tunneling in Solids.* Solid State Physics, Supplement 10, edited by H. Ehrenreich and D. Turnbull (Academic, New York, 1969).

[101] As described in the text, an adiabatic double-dot system is effectively one-dimensional and therefore characterized by the 1D level spacing $\delta_{1D} = \pi\hbar v_F/L_{dot}$ near the Fermi surface. A realistic system, on the other hand, will not be characterized by such a clean separation between the transverse and longitudinal degrees of freedom. In general, each two-dimensional eigenstate in the dots will have some overlap with the lowest transverse modes, and, for states near the Fermi surface, the proportion of this overlap is, on average, equal to $2/k_F L_{dot}$. At the same time, the two-dimensional dot states are characterized by the level-spacing $\delta_{2D} = 2\pi\hbar^2/mL_{dot}^2$. Combining the transverse mode overlap with the 2D level spacing we see that the correct normalization of the 1D channel can be described by an "effective level spacing" $\delta^{eff} = (k_F L_{dot}/2)\delta_{2D}$, which precisely equals the 1D level spacing δ_{1D} above. Of course, since the overlap factor $2/k_F L_{dot}$ is only true on average, it is important that we are working in the large-dot limit where $U \gg \delta_{2D}$ and, therefore, the energy of the double-dot system is, to first approximation, reflective of the average behavior of the single-particle eigenstates.

[102] Recognition of the existence of parity pairs and of the usefulness of forming quasi-localized states from pair components has a long history. See, for example, F. HUND, Z. Phys. **43**, 805 (1927); L. D. LANDAU and E. M. LIFSHITZ, *Quantum Mechanics: Non-Relativistic Theory*, Vol. 3 of Course of Theoretical Physics, 4th ed. (Pergamon, Oxford, 1991), p. 183; R. E. PRANGE, Phys. Rev. **131**, 1083 (1963). The fact that the alternation between even-parity and odd-parity states continues without interruption as one considers progressively larger energies can be proven by following the arguments in Ch. 5 of E. MERZBACHER, *Quantum Mechanics*, 2nd ed. (Wiley, New York, 1970).

[103] Use of a sharp step function is somewhat unrealistic. However, a "smeared-out" step function such as $[1 + \tanh(x/\xi)]/2$ creates a region of characteristic width ξ about $x = 0$ where electrons are neither in dot 1 nor dot 2 from the point of view of $\hat{n}$ (recall Eq. 4.8). Under such circumstances, the value of $\hat{n}$ can be changed simply by having a net movement of electrons toward the barrier—without any tunneling between the dots. Clearly, such a picture misses the crucial fact that electrons in the vicinity of the barrier do, in fact, repel one another. Until a better model of the electron repulsion in the barrier region is developed, it is best to stick with a sharp step and thereby avoid inventing a region near the barrier in which electrons do not interact.

[104] K. A. MATVEEV and L. I. GLAZMAN, Phys. Rev. B **54**, 10 339 (1996).

[105] E. GUTH and C. J. MULLIN, Phys. Rev. **59**, 575 (1941); C. HERRING and M. H. NICHOLS, Rev. Mod. Phys. **21**, 185 (1949); D. W. JUENKER, G. S. COLLADAY, and E. A. COOMES, Phys. Rev. **90**, 772 (1953); K. W. FORD, D. L. HILL, M. WAKANO, and J. A. WHEELER, Ann. Phys. (USA) **7**, 239 (1959).

[106] J. N. L. CONNOR, Molecular Physics, **15**, 37 (1968).

[107] J. C. P. MILLER in *Handbook of Mathematical Functions*, edited by M. Abramowitz and I. A. Stegun, (National Bureau of Standards, Washington, D.C., 1964), Appl. Math. Ser. 55, p. 685.

[108] E. C. KEMBLE, Phys. Rev. **48**, 549 (1935); L. D. LANDAU and E. M. LIFSHITZ, *Quantum Mechanics* (Ref. [102]), p. 184.

[109] J. H. DAVIES and J. A. NIXON, Phys. Rev. B **39**, 3423 (1989).

[110] For a zero-width gate, the half width equals d. For broader gates, the half width is larger, but it grows relatively slowly with the gate width w, only reaching about

$1.4d$ for $w = 2d$. For much larger values of w, the potential profile becomes rather flat beneath the gate, and a parabolic model is not useful.

[111] Of course, in considering tunneling between the dots, we are not dealing with a line gate but two narrow split gates separated by a gap of length ℓ. However, for weakly coupled dots, the potential profile in the gap is presumably reasonably well approximated by that of a line gate with an effective width w_{eff}. This should remain true even when electron-electron interactions are included because, near the barrier peak and for $g \simeq 0$, the corrections due to electronic screening should be relatively small. In any case, these corrections tend to counteract the broadening due to $w_{\text{eff}} \neq 0$.

[112] N. C. VAN DER VAART, A. T. JOHNSON, L. P. KOUWENHOVEN, D. J. MAAS, W. DE JONG, M. P. DE RUYTER VAN STEVENINCK, A. VAN DER ENDEN, and C. J. P. M. HARMANS, Physica B **189**, 99 (1993).