Wavelet Analysis of Two Dimensional Quantum Scattering

A thesis presented
by

Jonathan Dillwyn Edwards

to
The Department of Physics
in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy
in the subject of

Physics

Harvard University
Cambridge, Massachusetts
June 1998

©1998 Jonathan Dillwyn Edwards
All rights reserved

For Nana and Papa

Abstract

This work is primarily concerned with the efficient numerical solution of the integral equations that describe quantum scattering in two dimensions from an array of one-dimensional Dirichlet boundaries. In particular we are interested in computing the diffractive scattering patterns that result when the boundaries have one or more sharply defined discontinuities, including edges, corners, and curvature mismatches. Conventional discretization methods applied to these equations do not yield efficient solutions because the global nature of their basis elements fails to capture the local nature of the scattering. We present an alternative discretization based on the discrete wavelet transformation, which uses basis elements that can readily adapt themselves to local structure at any scale. This method allows us to treat a number of scattering geometries that would have otherwise been numerically untenable. Foremost among these is the Westervelt gate, which consists of a subwavelength quantum point contact coupled to an open resonator. We are able to compute the electronic conductance of this device in its stable operating regime, and thus account for the fine-scale features seen in experimental conductances traces; as well as make predictions concerning its behavior in its unstable regime, including the existence of a set of conductance resonances supported by diffractive scattering. The existence of these resonances can be accounted for semiclassically by extending the conventional trace formula to include diffractive paths. Unlike previous calculations of diffractive contributions to the conductance of mesoscopic devices, the diffractive corrections in the Westervelt gate are of the same order in $\hbar$ as the classical contributions. The particular paths needed are given by the uniform extension of Keller's geometric theory of diffraction. Finally we suggest a method for imaging the wave function inside the gate using an atomic force microscope tip. We show that by measuring the shift in gate's conductance as a function of the tip position inside the gate, we should be able to map out the resonant wave functions.

Acknowledgements

The person perhaps most responsible for the particular content of this dissertation is John Doyle. After having immersed myself in the mathematical world of string theory for three years, John reminded me that the physical world is far more interesting than the cohomology of a Kähler manifold will ever be. This, however, would not have mattered had Rick Heller not been there to given me a second (and as it would later turn out a third) chance at physics. I would like to thank both of these men for teaching me more about the practice of physics than I ever thought possible. Then, of course, there is Lev Kaplan, Adam Lupu-Sax, and Jordan Katine. I thank Lev for his innate ability to know exactly when I have no idea what I'm talking about; Adam for always telling me how to do things, and then waiting patiently until I realize he's right; and Jordan for studying the gate that would become the focus of my dissertation. I should also thank five of my earlier teachers: my parents, for letting me do whatever I wanted to do; Richard Yannarelli, who first introduced me to science; Hugh Moore, for his very odd habit of making his students think; and Igor Klebanov, who helped me move from coursework to research. So thank you to all of these people, without whom none of this would have been possible, but more importantly to Melissa and Kate, without whom none of this would have mattered. I love you both more than I could ever say.

Citations to Previously Published Work

J. A. Katine, M. A. Eriksson, A. S. Adourian, R. M. Westervelt, J. D. Edwards, A. Lupu-Sax, E. J. Heller, K. L. Campman and A. C. Gossard, *Point Contact Conductance of an Open Resonator*, Phys. Rev. Lett., **79**(24), 4806-9(1997).

Contents

List of Figures

Chapter 1

Overview

This brief introductory chapter is intended to be a seamless overview of the ideas, methods, and results scattered throughout the chapters that follow. The first two sections introduce the underlying computational scheme: the boundary wall reformulation of two-dimensional quantum scattering in terms of integral equations, and the wavelet-based multiresolution analyses used to solve these equations numerically. These sections also provide additional mathematical and physical motivation for this scheme not found in subsequent chapters. The third section introduces the Westervelt gate, the principal mesoscopic device of interest in this work. It discusses what might naïvely be expected of the gate's conductance as a function of its dimensions, and what is actually found in numerical simulations performed using the boundary wall method. The novel semiclassical analysis necessary for a physical understanding of the gate's behavior is described in section four, along with a discussion of how to incorporate the effects of strong diffraction into the trace formula. Finally, section five provides an outline of what can be found in each chapter.

1.1 Boundary Wall Method

1.1.1 Mathematical class of problems

We wish to address the problem of two-dimensional quantum scattering at fixed energy, as described by the Helmholtz equation, from a curve Γ embedded in the plane. The scattering processes of interest to us are defined by requiring that the wave function satisfy a linear combination of zero and zero normal derivative boundary conditions along this curve. Γ should be well-behaved in the usual, vague physicist's sense, but is otherwise quite arbitrary; in particular it can consist of a number of disconnected segments, each

of which is allowed to have a finite number of discontinuities in its slope, curvature, or any of its higher derivatives. As with most partial differential equations (PDEs), the most difficult part of finding a solution in this general class of problems is imposing the boundary conditions. With this in mind we consider the following approach. The fundamental solution to the Helmholtz equation in two dimensions, $G_0(\mathbf{r}, \mathbf{r}'; E)$, is a known analytic function, $(-i/4)H_0^{(1)}(2\pi|\mathbf{r}-\mathbf{r}'|/\lambda)$, the zeroth order Hankel function of the first kind. If our problem were instead scattering from some potential V, we would thus naturally be led to a Lipmann-Schwinger description of the scattering [60]:

$$|\psi\rangle = |\psi\rangle_0 + \widehat{G}^0\widehat{V}|\psi\rangle, \tag{1.1}$$

where $|\psi >_0$ is a solution of the free problem satisfying suitable boundary conditions at infinity. [1] Now we ask the following question: can we construct a potential operator $\widehat{V}$, so that in some possibly singular limit, Eq. (1.1) reduces to our original problem complete with properly enforced boundary conditions?

To see that this can indeed be done, consider solving Schrödinger's equation in one dimension for a delta function potential $U(x) = U_0\delta(x - x_0)$. The solution is easily found to be [47]:

$$\psi(x) = \begin{cases} e^{ipx} - \frac{1}{1+\hbar^2 p/2imU_0}e^{-ipx} & x < 0 \\ \frac{1}{1+2imU_0/\hbar^2 p}e^{ipx} & x > 0 \end{cases}, \tag{1.2}$$

where for simplicity we have set $x_0 = 0$. For finite U_0 the transmission through this barrier remains nonzero. But in the limit $U_0 \to \infty$ the transmission goes to zero, and at the same time we force a zero in the wave function at x_0. Returning now to two dimensions, it should be immediately obvious how to construct the desired $\widehat{V}$ in the case of Dirichlet boundary conditions along Γ; we replace Γ with a potential of the form $U_0 \int ds\ \delta(\mathbf{r} - \mathbf{r}_\Gamma(s))$, where $\mathbf{r}_\Gamma(s)$ parametrizes Γ, and allow $U_0 \to \infty$. Neumann conditions are only slightly harder, requiring a potential built from the derivative of the delta function.

This procedure allows us to reformulate our original PDE, along with its boundary conditions, as a Fredholm integral equation of the first kind. Mathematicians also affect this reformulation, using the so-called direct method [30], which foregoes the physically appealing step of introducing partially transmissive surfaces. More commonly, however, they use the indirect method, which leads them to a Fredholm equation of the second

[1] More generally we can split the potential operator into a background and a perturbation, $\widehat{V} = \widehat{V}_B + \widehat{V}_\delta$, as long as the background problem is still exactly soluble. This will prove useful in Ch. 5 when we consider how to image the wave function inside the Westervelt gate.

kind [30]. The reason for this preference is that the integral operators associated with equations of the second kind are of order zero, [2] which makes them amenable to both theoretical and numerical analyses. In particular, the discretization of such operators using finite element methods leads to well-conditioned matrices.

In contrast the integral operators that appear in our equations are of order -1, and are known to be particularly ill-suited for numerical treatment. A naïve finite element discretization of such an operator yields a matrix whose condition number grows unboundedly with the level of discretization. If we work at finite precision, the information associated with the smallest eigenvalues of this matrix is quickly lost as we increase our resolution. Since these eigenvalues encode the fine-scale structure in the scattered wave function near discontinuities in Γ, the accuracy of our solution will thus be seriously compromised if these regions play a significant role in the scattering. Such adverse effects have been seen numerically by our group [50]. The failure of the original boundary wall method to compute the fine scale transmission spectrum of the Westervelt gate in the stable regime, and even to detect the presence of the anomalous conductance peaks in the unstable regime, is a direct consequence of this loss of accuracy. Nonetheless we shall proceed with our reformulation. As we shall see, one of the benefits of working with wavelet-based discretizations is that the condition numbers are automatically brought under control.

1.1.2 Physical motivation for this class of problems

With the ever increasing ability of experimentalists to fabricate nanoscale electronic devices on semiconductor heterostructures, it has become essential for theorists to provide a basis for understanding the transport properties of these devices. The length scales that characterize the behavior of the electrons in these new devices are quite different from those found in more conventional systems. The Fermi wavelength, λ_f, of the two-dimensional electron gas (2DEG) in the underlying heterostructure is typically smaller than, but within an order of, the overall dimension of the device, and both the electron mean free path and coherence length are many times this dimension. The electronic motion is thus ballistic and quantum mechanically coherent, as opposed to the classically diffusive behavior usually found; so in order to calculate the transport properties of these devices, we

[2] Operators of order zero do not change the order of differentiability of the function on which they operate. A complete discussion of the symbols and orders associated with pseudodifferential operators, of which integral operators are one example, can be found in [].

must solve for the full quantum mechanical scattering from the potentials associated with their defining gates. Since the exact form of these potentials is not known, we must consider how best to approximate them in order to carry out qualitatively, if not quantitatively, correct calculations.

One of the more remarkable aspects of the behavior of electrons in semiconductor heterostructures is that we are able to restrict our attention to the behavior of a two-dimensional gas of electrons in these otherwise three-dimensional systems. The existence of a 2DEG near the interface of two semiconductor crystals can be understood as follows. The motion of the electrons perpendicular to the interface is effectively described using the Hamiltonian for a particle in a potential well [62]. The potential arises due to charge migration between the two crystals induced by their different dopings. The parameters of the well are such that the energy separation between its ground and first excited states is large compared to the characteristic thermal energy of the electrons near the Fermi energy. These electrons are thus confined to the ground state; and since this state is spatially localized about the interface, we have a 2DEG. [3]

We are thus left to find two-dimensional potentials that effectively describe the gates defining our new devices. The scattering of electrons from one of these gates has been experimentally shown to consist primarily of specular reflection, with some small loss due to transmission through the gate [72, 37]. To a good approximation we are thus able to replace the gates defining these devices with infinitely hard one-dimensional barriers. If we are also able to approximate the background potential between the gates as being zero, then we are immediately led to the class of problems defined in the previous section. If the background is flat except for randomly scattered impurities, we may extend the method of the previous section by including additional s-wave scatterers to model the impurities.

[3] It is interesting to note that the same physics community that readily accepts the effective dimensionality of the electron gas being determined by quantum mechanics is amused with the notion of compactification from ten down to four dimensions in superstring theory. In both cases the mechanism behind the dimensional reduction is the same. The motion in one or more dimensions is quantized, either because of a confining potential, as in the heterostructure, or because the dimensions themselves are compact, as is true for the extra 6-dimensional manifold in superstring theory; and the energy spacing between the ground and first excited states is such that at the energies of interest, the system is confined to the ground state along these dimensions.

1.1.3 Limitations of the standard boundary wall method

As we noted above, the condition numbers of the matrices generated by the boundary wall method grow unboundedly with the increasing size of the discretization, and this growth results in a loss of fine scale structure in the scattered wave function. The limitations on the use of the method imposed by this loss are particularly troublesome for the Westervelt gate. The gate consists of a quantum point contact (QPC) coupled to an open resonator defined by the walls of the contact and a single curved mirror. For the purpose of comparison with experiment, we are interested in computing the conductance of the gate when the QPC is less than one quarter of a Fermi wavelength wide, and for comparison with the scattering theory approach to the conductance developed in [3], we are interested in the extreme limit of vanishingly small QPC. But while the QPC width is only a fraction of the Fermi wavelength, the radius of curvature of the mirror can be as much as twenty wavelengths; and the bounding walls that define the QPC must be of sufficient length to ensure that flux around their ends cannot interfere with the wave function inside the cavity. If we were to use a conventional finite element or finite difference method with a uniformly refined mesh, the number of discrete elements required by the Westervelt gate would be prohibitively large because of the disparate scales involved.

The effects of the loss of fine scale structure could not be seen in the initial test of the boundary wall method. The test was to place a numerical detector behind a straight wall and calculate the flux through the detector when the wall was illuminated by a plane wave. Because the detailed structure in the diffraction patterns from the ends of the wall did not extend down to the detector, the computation was unaffected. However, it is essential to get these patterns correct when trying to compute the narrow conductance resonances of the gate in the classically stable regime, and when investigating the anomalous conductance peaks in the unstable regime.

1.2 Wavelet Based Multiresolution Analysis

1.2.1 Failure of global analysis

We begin our discussion of wavelets with a simple example that illustrates the failure of global analysis to efficiently encode local structure. Imagine that we are given the score to Beethoven's Ninth Symphony along with a recording of it. Aesthetics aside, these

provide us with two equivalent representations of the symphony. [4] Conventional Fourier analysis of the recording, complete with phase information, provides us with yet a third in the form of a global spectral analysis. Given these three representations of the symphony, we consider the following question: if we listen to the recording, we note that at the thirtieth minute we do not hear an F#; how is this absence represented in the score and in the global spectrum? The latter is silent on this question unless we work a bit. Naïvely we might think that we should indeed hear an F#, not only at the thirtieth minute but throughout the recording, since there is a nonzero amplitude in the spectrum at the corresponding frequency. But we are leaving out the all important phase information. If we add up all the frequency components with not only their correct amplitudes but also their *phases*, we find that at the thirtieth minute of the recording, all of the other frequencies conspire to cancel out the F# that would otherwise have been there. The score answers our question in a more succinct manner: nobody played it!

Now let us consider how best to discretize the integral equation that describes scattering from a finite curve Γ embedded in the plane. Note that this problem is equivalent to that of two-dimensional signal representation, and is thus a generalization of our first example. The scattered wave function is determined entirely by the T-matrix, $T[\psi_0]$, defined along Γ. As we will see in Ch. 3 in our discussion of the boundary wall method, $T[\psi_0]$ is given implicitly by

$$-\int_\Gamma ds'\; G_0(s,s')T[\psi_0](s') = \psi_0(s). \tag{1.3}$$

Here s and s' are parameters along the curve Γ, G_0 is the point-to-point background Green's function, and ψ_0 is the incoming wave function. The scattered wave function $\psi(\mathbf{r})$ can be constructed from $T[\psi_0](s)$ by integrating the latter against $G_0\left(\mathbf{r},\mathbf{r}(s)\right)$ along the boundary. Since Γ is assumed to be finite, it has two sharply defined edges. The description of a sharp edge in Fourier space requires infinitely high frequencies—no matter how close we get to an edge, it still looks like an edge—so scattering from Γ will induce structure in $T[\psi_0](s)$ at all possible scales near the edges. This is the usual problem associated with diffractive scattering.

From a numerical standpoint we are only interested in the structure in $T[\psi_0]$ down to some finest scale, or equivalently in its Fourier components less than some cutoff frequency. The contributions from frequencies above this cutoff can be eliminated by pro-

[4] The score is most similar to the wavelet representation of the symphony in terms of the questions both can readily answer.

jecting Eq. (1.3) onto a finite dimensional space of functions. The projection used in the initial implementation of the boundary wall method was simply to replace the integral with a discrete sum over a finite set of points $\{s_i\}$, uniformly distributed along Γ. [5] The matrix representation of $G_0(s, s')$ that results from the discretization is dense, so the complexity of solving Eq. (1.3) scales as $O(N^3)$, where N is the size of the discretization. Moreover the matrix is poorly conditioned, so small scale structure in $T[\psi_0]$ is easily lost. Both of these difficulties arise from our choice of projection. The basis elements corresponding to our projection are localized in space and completely delocalized over the frequency interval of interest. Thus the interactions at *all* scales contribute to every matrix element. We might think that if we were to perform a discrete Fourier transform on $G_0(s_i, s_j)$, and thus separate the information into different frequencies, we would have a sparser, more stable representation of the scattering. However, because the Fourier basis functions are defined over the entire curve, the transformed matrix is still as poorly behaved. [6]

1.2.2 Multiresolution analysis

One of the more computationally efficient ways of constructing a function in an infinite dimensional Hilbert space is to begin with a coarse approximation to that function and then add on locally defined corrections that describe its structure at increasingly finer scales. A multiresolution analysis (MRA) is the framework within which this construction can be performed. For definiteness let us take our Hilbert space to be $L^2(\mathrm{R})$. The simplest example of an MRA of $L^2(\mathrm{R})$ consists of two elements: a sequence of nested subspaces $\{V_i : V_i \subset V_{i+1}, \forall\, i \in \mathrm{Z}\}$, which satisfy $\cap_i V_i = \{0\}$ and $\cup_i V_i = L^2(\mathrm{R})$, and the elements of which satisfy $f(x) \in V_j \leftrightarrow f(2^{-j}x) \in V_0$, $\forall j \in \mathrm{Z}$; and a single function $\varphi(x)$, called the scaling function, the integer translates of which form a basis for V_0. The fact that the basis of V_0 consists of the *discrete* translations of a single function, along with the relationship between the elements of V_0 and all other V_j, imply that each V_j contains only those functions with structure down to scale 2^{-j}. These spaces can thus serve as the initial approximation spaces in our proposed scheme. The local fine scale corrections are members of the detail spaces $\{W_i\}$, implicitly defined by $V_i = V_{i+1} \oplus W_{i+1}$. As with the V_i the basis elements

[5] The diagonal elements are infinite using this prescription, but since the kernel singularity is integrable, they can be defined as $G_0(s_i, s_i) = \int_{-d}^{d} ds\; G_0(0, s)$, for all s_i.

[6] The author has recently learned that it is possible to precondition the matrix representation of $G_0(s, s')$ using a windowed rather than global discrete Fourier transform [33].

of the W_i can all be derived from the integer translations and dyadic scalings of a single function ψ, called the "mother wavelet," which can in turn be written in terms of φ. The actual method by which useful MRAs are constructed can be found in Ch. 2. Here we are interested in discussing MRAs from the point of view of their associated tilings of the time-frequency plane.

Consider the basis elements of V_j for some fixed j. They can be used to represent functions with Fourier components less than some cutoff frequency proportional to 2^{-j}. Now as we will see in Ch. 2, φ can be chosen to have compact spatial support. Thus the basis of V_j tiles a strip of finite width of the time-frequency plane, as illustrated in Fig. 1.1(a). The problem with this tiling is the same as the problem with the initial basis used in the boundary wall method; the elements are completely delocalized over the frequency interval of interest. Rather than using the basis provided directly by V_j, let us consider the tilings generated by the mixed bases associated with the decompositions $V_j = V_{j-1} \oplus W_{j-1}$ and $V_j = V_{j-3} \oplus W_{j-3} \oplus W_{j-2} \oplus W_{j-1}$, as shown in Figs. 1.1(b,c). The upper rectangles in these figures represent the wavelets at the finest scale, while the lower rectangles are the scaling functions at the coarsest scale. Because the functions represented in these two figures are only partially localized in both time and frequency, they can be used to represent the *local* frequency content of a signal. The transformation from the representation in Fig. 1.1(a) to either Fig. 1.1(b) or (c) is referred to as a wavelet transformation. If we were to continue this transformation *ad infinitum*, we would find that the aspect ratios of the rectangles, defined as their heights divided by their widths, would approach zero as we approached zero frequency. This basis is thus optimal for representing signals whose global frequency content is centered about zero. However, a more general MRA construction allows us to move this optimal resolution to any frequency we choose, and even allows us to vary the optimum as a function of time. The transformations associated with these constructions are called wavelet packet transformations.

1.2.3 Function approximation and solution of integral equations

A large class of integral operators have nearly *sparse* matrix representations when represented in wavelet bases, meaning that the number of nonzero elements above some threshold scales as N^α, where N is the size of the matrix and $\alpha \gtrsim 1$. This allows us to use iterative solvers and reduces the complexity of solving discretized integral equations from

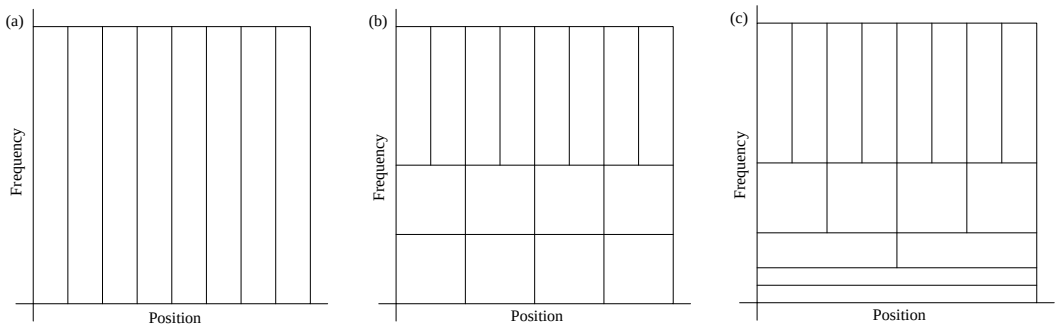

Figure 1.1: The tilings provided by the bases of (a) V_j, (b) $V_{j-1} \oplus W_{j-1}$, and (c) $V_{j-3} \oplus W_{j-3} \oplus W_{j-2} \oplus W_{j-1}$.

$\mathcal{O}(N^3)$ to $\mathcal{O}(N)$.

This successive refinement strategy outlined in the previous section is followed by the multigrid approach to solving pseudodifferential equations. Wavelets give rise to a natural, locally adaptive multigrid scheme. Unlike multigrid solvers based on finite element and finite difference schemes, where the details of the refinement step are strongly problem specific and give rise to additional difficulties of their own, the refinement step for the wavelet method is *universal*: it is independent of the problem being solved and the type of wavelets being used. For a particular class of operators, this method can be shown to be asymptotically optimal, meaning that the complexity of the algorithm is proportional to the size of the discretized system.

1.3 Phenomenology of the Westervelt Gate

Most of the theoretical effort in the mesoscopic community has been concentrated on the transport properties of closed devices in the presence of strong disorder. Random matrix theory and the closely associated supersymmetry techniques can be used to compute the *universal* properties of such devices by averaging over ensembles of disorder configurations. We will concern ourselves in this work with the behavior of open mesoscopic devices in the limit of zero disorder, and will thus be forced to deal with the full complexity of the scattering equations. Our efforts will be concentrated on the Westervelt gate, illustrated in Fig. 1.2. The gate consists of a subwavelength QPC coupled to an open resonant cavity, which is defined by the walls of the QPC itself and a single curved mirror. The conductance of the gate as a function of QPC to mirror separation was initially investigated by Jordan

Katine [40] of the Westervelt group here at Harvard. More recently other members of this group, including Mark Topinka and David Duncan, have begun efforts to confirm some of the numerical predictions made in this work.

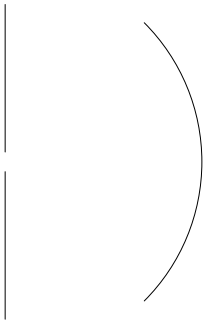

Figure 1.2: Schematic illustration of the Westervelt gate.

The classical dynamics of the cavity are naturally divided into two distinct regimes. When the QPC to mirror separation is less than the mirror's radius of curvature, or equivalently when the mirror's center of curvature lies outside the cavity, the cavity is classically stable in the following sense: if we ignore the possibility of backscattering through the QPC, then any trajectory beginning at the QPC and reflecting at least once from the mirror will be trapped between pairs of caustics running from either side of the QPC to the mirror. When the center of curvature lies inside the cavity, the gate is classically unstable: if again we ignore the QPC, then every trajectory, except for the one lying along the cavity's axis of symmetry, will eventually leave the cavity. The same division holds for the quantum dynamics. In the stable regime the cavity supports long-lived quasistates (the quantum dynamics can never be exactly stable because the wave function is allowed to tunnel out through the caustics that bound the classical motion), and in the unstable regime there are no such states.

The conductance behaves quite differently in these two regimes. In the stable regime we find a sequence of sharply defined peaks as we move the mirror toward the gate. These peaks arise from resonant conduction through the gate mediated by the quasistates of the cavity. As might be expected, the wave functions inside the cavity associated with these peaks have a well-defined number of radial and angular nodes, and resemble the eigenstates of the closed lemon billiard [32]. In the unstable regime, we find two classes of peaks. Those in the first class form a regularly spaced sequence with periodicity $\lambda_f/2$. The corresponding

wave functions lay along the axis of symmetry and again have a well-defined number of radial nodes. The peaks in the second class behave quite differently from those in the first and those in the stable regime. Their positions, widths, and even their number vary with the angular size of the mirror. And unlike the peaks in the first class, which persist for a number of wavelengths, these peaks can only be found when the mirror has been pulled out between one half and two wavelengths from the QPC. The associated wave functions are narrow strips of radial nodes that run from the QPC to the edges of the mirror, and are unlike anything that may have been expected. We will refer to these two classes of peaks as nonanomalous and anomalous, respectively.

1.4 Semiclassical Theory of Diffraction

The typical QPC to mirror separation of interest in the Westervelt gate is between $10\lambda_f$ and $20\lambda_f$, which puts us squarely in the semiclassical regime. One of the advantages of working in this regime is that the behavior of the system can be understood using a mixture of quantum and semiclassical reasoning; the former allows us to perform exact computations of physical properties, while the latter offers us some intuitive physical understanding in terms of the constructive and destructive interference of the wavefronts associated with classical trajectories. The nonanomalous peaks in the unstable regime can be associated with a kind of scarring along the periodic orbit lying along the symmetry axis of the gate. However, since this is the one remaining periodic orbit in the unstable regime, it is not immediately apparent how we can account semiclassically for the anomalous peaks.

Because the positions of the anomalous peaks vary as we change the angular size of the mirror, we suspect that they arise from scattering from the ends of the mirror. If we want to develop a semiclassical theory of these peaks, we thus need a description of the diffraction pattern from a sharp edge in terms of classical trajectories. At first glance, this would seem impossible; diffraction was one of the phenomena that could not be explained by the ray theory of light, and helped lead to the development of the wave theory. However, we must remember that even when the ray theory of wave propagation is applicable, for instance in the scattering of a plane wave from an infinite specular surface, it is only an effective description of the exact solution provided by the wave equation. When we say that we cannot account for diffraction using only classical rays, we mean that we cannot effectively describe the exact solution using rays that simply obey the law of reflection;

but this should not be surprising because this law was not derived in the presence of a sharp edge! There is nothing preventing us from introducing *new* rays, not obeying this law, in order to describe the diffraction pattern. For the case of a sharp edge, these new rays were introduced by Keller some fifty years ago in his geometric theory of diffraction (GTD) [41, 42].

Imagine a bundle of parallel trajectories incident on a sharp edge. Those trajectories that miss the edge continue to propagate in a straight line. Those that strike the edge inside its tip are scattered according to the law of reflection. But what happens to the trajectory that strikes the tip? Ordinarily its behavior would be undefined, but nothing prevents us from defining it in some useful manner. Now let us consider the asymptotic expansion of Sommerfeld's exact solution for the scattering of a two-dimensional plane wave from a half-infinite line [66]. Keller found that in this limit the scattered wave reduces almost everywhere to a weighted circular wave centered about the tip of the line. He then noted that this wave can effectively be described by a continuous distribution of outgoing radial trajectories originating from the tip, with an amplitude $d(\theta, \theta')$ assigned to each one. Here θ and θ' are the angles of the line and outgoing trajectory relative to a coordinate system centered at the tip. Keller referred to this function as the "coefficient of diffraction" for an edge. It is given by

$$d(\theta, \theta') = -\frac{\sin(\theta'/2)\sin(\theta/2)}{\cos\left((\theta' - \theta)/2\right)\cos\left((\theta' - \theta)/2\right)}. \tag{1.4}$$

Thus if we say that the trajectory that strikes the tip is scattered in all possible directions with an amplitude given by $d(\theta, \theta')$, then we are able to describe the diffraction pattern from a sharp edge in terms of classical trajectories.

The expression for $d(\theta, \theta')$ given in Eq. (1.4) fails to hold near the sharp reflection and transmission boundaries generated by geometric reflection. Imagine that our sharp edge is oriented along the negative y-axis, and our bundle of trajectories corresponds to a plane wave normally incident on the edge from the left. In this case the negative x-axis is a reflection boundary, and the positive x-axis is a transmission boundary. The wave function built from the classically scattered trajectories, neglecting Keller's diffractive trajectories, has a sharp discontinuity across both of these boundaries. Since the full solution to the wave equation must be continuous across this line, $d(\theta, \theta')$ must correct for this discontinuity. Unfortunately, we see from Eq. (1.4) that it is infinite along this line. The physical reason for this infinity will be discussed in Ch. 4. However, we should not have expected this

simple expression for $d(\theta, \theta')$ to be valid along this line. The amplitude of a circular wave is down by a factor of $\hbar^{1/2}$ from the incident wave, and thus the diffractive contribution is a higher order correction to the contribution from the usual classical trajectories. However, if the solution is to be continuous across the boundaries, it must be of the same order. A corrected coefficient of diffraction was introduced by Kouyoumjian and Pathak [46] ten years after Keller's original work. It incorporates the Fresnel integrals that correctly describe the diffraction pattern near the boundaries, and shows that there is a wedge-shaped region around each geometric boundary with an opening angle proportional to $\hbar^{1/2}$ in which the diffractive contribution is the same order as the classical.

The existence of diffractive corrections of the same order as the classical trajectories is essential to the existence of the anomalous conductance peaks in the Westervelt gate. As was noted in Sec. 1.3, the anomalous peaks exist only when the mirror's center of curvature is within one or two wavelengths of the QPC. In this case the line running from the QPC to the tip of the mirror is close to the reflection boundary of a bundle of trajectories that starts at the QPC and scatters from the edge. The resulting enhanced diffraction allows us to scatter enough amplitude back to the QPC in order to form a resonance inside the cavity.

1.5 Outline

Each chapter is written as though it were a journal article, including an abstract, introduction, and discussion. The author hopes that this format will make the overall work easier to read by making each chapter essentially self-contained. We begin in Ch. 2 with an exposition of wavelet analysis, including the continuous wavelet, discrete wavelet, and discrete wavelet packet transformations. We show how the discrete transformations can be used to generate sparse representations of integral operators, including those centered about a nonzero frequency. In Ch. 3 we derive the boundary wall equations and their wavelet-based discretizations, and apply them to the Westervelt gate operating in the stable regime, where we find a series of transmission resonances that can be associated with the stable modes of the cavity. The unstable regime of the Westervelt gate is treated in Ch. 4, where we find a sequence of anomalous conductance peaks associated with diffraction off the ends of the mirror. By extending the usual periodic orbit theory to include diffractive orbits, we are able to develop a semiclassical theory of these peaks. Finally in Ch. 5 we discuss the

possibility of imaging the wave function inside the gate at one of the conductance peaks using an AFM tip.

Chapter 2

A Physicist's Guide to Wavelets

This chapter provides an overview of wavelet analysis, with an emphasis on the discrete wavelet and wavelet packet transformations.

2.1 Introduction

2.1.1 Motivation behind wavelets

Many of the difficulties encountered when trying to represent continuous signals efficiently using Fourier analysis can be traced back to the spatially delocalized nature of the Fourier basis elements. While the Fourier transform is indeed efficient at encoding the global structure of a signal, it fails miserably with the local structure; the number of significant expansion coefficients required to represent spatially localized features increases unboundedly with the degree of localization. The windowed Fourier transformation is sometimes presented as a local alternative to the global transform. It uses the globally defined exponentials of conventional Fourier analysis convolved with the continuous translates of a fixed function with near compact support. The shortfall of this method is that it is only efficient at representing features on the order of the window width, or equivalently frequencies on the order of the inverse window width. Structure at frequencies higher that this inverse width must still be represented by the superposition of many basis functions. Thus when a signal has structure at a number of well-separated frequencies, only one of these frequencies will be well represented.

The continuous wavelet transformation (CWT) provides a true alternative to Fourier analysis that allows direct access to the local frequency content of a signal. Unlike

the spatially delocalized exponentials used in Fourier analysis, the basis elements of the CWT are localized in both space and frequency. They consist of all possible continuous scalings and translations of a single, spatially localized function referred to as the "mother wavelet." The advantage of having basis elements at all scales is that we are always guaranteed to find one that matches the local structure in a signal at any scale. The problem is that the basis is extremely overcomplete. Fortunately it is possible to remove all but a countable infinite number of these elements and still have a complete basis. The associated transformation is referred to as the discrete wavelet transformation (DWT). Moreover for the appropriate choice of mother wavelet, these discrete bases can be shown to be not only complete, but orthonormal.

The primary use of the wavelet transformation is the efficient representation of signals, and its most ardent users are members of the digital signal processing community. However, if we broaden our definition of signal, we find that there are many areas of numerical analysis that could potentially benefit from wavelets. One such area of particular interest to computational physicists is the solution of integral equations. The typical way of solving these equations is to introduce a discrete basis for the solution space of interest, and then rewrite the continuous equation as a matrix equation. The problem with this method is that the resulting matrices are very often dense, so the complexity of finding a solution scales as $\mathcal{O}(N^3)$. However, a matrix can be regarded as nothing more than a two-dimensional digital image, and for certain classes of operators, the wavelet transformed matrices can be shown to be sparse. This allows the use of iterative solvers, which reduces finding the solution to an $\mathcal{O}(N)$ process.

2.1.2 Outline

We begin our development of wavelet analysis in Sec. 2.2 with the definition of the CWT. We do not discuss the theory behind nor the applications of this transformation because it fails to yield efficient representations of integral operators. In Sec. 2.3 we proceed immediately to the more useful DWT. Rather than simply listing the formal properties of these transformations, we begin with an explicit construction of the transformation generated by the Haar wavelet. From this we are able to abstract away the defining properties of a multiresolution analysis (MRA), the general framework in which wavelet analysis is performed. We also discuss the interpretation of the wavelet transformation in terms

of high- and low-pass filters. The traditional wavelet constructions make extensive use of Fourier analysis, and thus are restricted to either translationally invariant or periodic spaces. In Sec. 2.4 we present the recently developed lifting scheme, which greatly simplifies the construction of wavelets with specific properties, and which in particular allows the construction of wavelets on bounded intervals. We finish in Sec. 2.5 with a generalization of the DWT called the discrete wavelet packet transformation, and its interpretation in terms of generalized filter banks.

2.2 Continuous wavelet transformation

Let ψ be a function with support localized near $x = 0$, and consider the continuous set of functions defined by all possible scalings of ψ,

$$\psi_a(x) = |a|^{-1/2}\psi\left(\frac{x}{a}\right), \text{ forall a} \in \mathrm{R}^+. \tag{2.1}$$

We would like to be able to say that the structure in an arbitrary function f at infinitely fine scales near $x = 0$ is encoded in the coefficients $f_a = < \psi_a, f > = \int dx\ \psi_a(x)f(x)$, albeit in a highly redundant manner. In order to represent the structure in f at nonzero x, we also introduce all possible translations of the ψ_a,

$$\psi_{a,b}(x) = |a|^{-1/2}\psi\left(\frac{x-b}{a}\right), \text{ forall a} \in \mathrm{R}^+, \mathrm{b} \in \mathrm{R}. \tag{2.2}$$

The fact that the doubly continuous set of coefficients $f_{a,b} = \int dx\ \psi_{a,b}(x)f(x)$ does, for certain functions ψ, contain all of the information in f is expressed by the following reconstruction formula: [1]

$$f(x) = \frac{1}{C_\psi}\int_0^\infty \int_{-\infty}^\infty \frac{da\ db}{a^2} < \psi_{a,b}, f > \psi_{a,b}(x), \tag{2.4}$$

where C_ψ is a normalization constant dependent only on ψ. The function ψ is called the "mother" wavelet, Eq. (2.2) is called the wavelet expansion of f, and the transformation from f to the coefficients $f_{a,b}$ is called the continuous wavelet transformation. Note that by construction the $\psi_{a,b}$ exist only for the unbounded space R, and are thus sometimes

[1] This formula holds in the weak (operator) sense: if $f, g \in L^2(\mathrm{R})$, then

$$< f, g >_{L^2} = \frac{1}{C_\psi}\int_0^\infty \int_{-\infty}^\infty \frac{da\ db}{a^2} < \psi_{a,b}, f >< \psi_{a,b}, g > . \tag{2.3}$$

referred to as "shift-invariant" wavelets. The simplest example of an admissible ψ is the "Mexican hat" wavelet $\psi(x) = (1 - x^2)\, e^{-x^2/2}$, shown in Fig. 2.1. A full exposition of the theory behind the continuous wavelet transformation can be found in [11, 36].

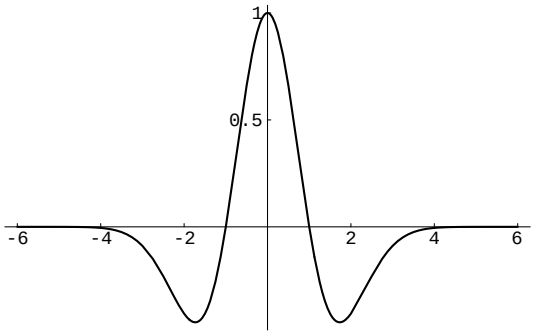

Figure 2.1: The Mexican hat wavelet.

2.3 Discrete wavelet transformation

2.3.1 Problems with the CWT

From the standpoint of efficient representations that can be used for numerical work, the CWT has the serious drawback of being incredibly redundant. This redundancy is in fact so large, it does not seem unreasonable that we could eliminate all but a countably infinite number of the $\psi_{a,b}$ and still have a complete set. With this in mind, let us choose some fixed scale parameter a_0, and restrict the allowed scalings of ψ to the integer powers of a_0, $\{a_0^n : n \in \mathrm{Z}\}$. If we are to construct a complete basis over the entire real line from discrete translations of these functions, the set of allowed translations for each $\psi_{a_0^n}$ must be properly adjusted to its scale. From Eq. (2.2) we see that the scale of $\psi_{a_0^n}$ is a_0^n times that of ψ. The minimal set of translations that would appear to ensure a sufficient sampling rate at this scale is thus $\{ma_0^n : m \in \mathrm{Z}\}$, so our proposed countable basis is

$$\psi_{n,m}(x) = |a_0|^{-n/2}\psi\left(\frac{x - ma_0^n}{a_0^n}\right), \text{ forall m, n} \in \mathrm{Z} \tag{2.5}$$

Even though Eq. (2.5) is well-motivated, it is still quite remarkable that we can find ψ such that these functions are complete without being overcomplete. But what is even more remarkable is that for $a_0 = 1/2$, and more generally for any rational a_0, ψ can

be chosen such that the $\psi_{m,n}$ provide an *orthonormal* basis for $L^2(\mathrm{R})$ [11, 36]. Two of the better known examples of such wavelets are the Meyer wavelet and the one-parameter family of Battle-Lemairé spline wavelets. The former is C^∞ and decays faster than any inverse polynomial, while the latter are C^k and decay exponentially [11]. Since these wavelets have noncompact support, they are not optimal for numerical work. The first example of a family of compactly supported wavelets that generated orthonormal bases was provided by the one parameter family of Daubechies [11]. The first two nontrivial members of this family are shown in Figs. 2.2. The lowest member of this family is just the Haar wavelet, which we discuss in the following section.

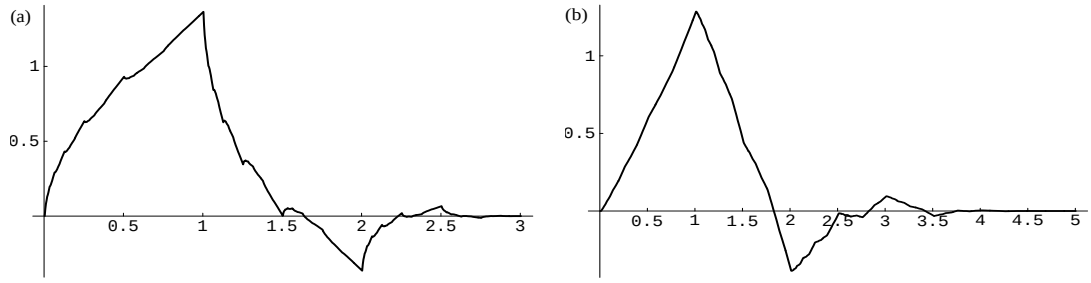

Figure 2.2: The (a) $N = 2$ and (b) $N = 4$ Daubechies wavelets.

2.3.2 The Haar wavelet

The simplest example of a compactly supported wavelet that gives rise to a countable, complete, orthonormal basis is the Haar wavelet,

$$\psi(x) = \begin{cases} 1, & 0 \le x < 1/2 \\ -1, & 1/2 \le x < 1 \\ 0, & \text{otherwise} \end{cases} . \tag{2.6}$$

This particular function was first used in 1910 to study the approximation properties of piecewise constant functions, but its place in the framework presented here was not realized until much more recently. In order to motivate the multiresolution framework presented in the next section, we will work through an informal, constructive proof showing that the Haar wavelet does indeed give rise to a well-behaved basis.

Before we begin with this proof, we need to discuss more carefully how to isolate the structure in a function f at a given scale. Since we are working with a discrete rather

than continuous set of scales, $\{a_0^n : n \in \mathrm{Z}\}$, it is possible to define the structure in f at scale a_0^n as the difference between two approximations to f, $f^{(n)}$ and $f^{(n-1)}$, containing all of the structure in f down to scales a_0^n and a_0^{n-1}, respectively. This, of course, leaves us with the question of how to construct such approximations, but this is a far more familiar problem. An intuitively simple, but computational inefficient way of doing this is to take the Fourier transform of f, filter out all of the frequency components above a certain cutoff, and transform back to real space. A more practical method is to replace f with its discrete samples on a uniform lattice. According to Nyquist's theorem, such a discrete approximation will contain all of the structure in f at frequencies up to twice the inverse of the lattice spacing. Thus we see that the notion of structure at a scale is fundamentally linked to the notion of the function approximation. In what follows we will always proceed by first defining a set of function approximation spaces, one associated to each scale a_0, and then construct the wavelet as the function that describes the differences between two such spaces.

Assume that we are given a function $f \in L^2(\mathrm{R})$ with support on the interval $[0, 1]$. This function can be arbitrarily well approximated by another function $f^{(n)}$, piecewise constant on the subintervals $[i \cdot 2^{-n}, (i+1) \cdot 2^{-n})$, with $i = 0, \ldots, 2^n - 1$, and $n \geq 0$. If we introduce the characteristic function on the unit interval,

$$\varphi(x) = \begin{cases} 1 & 0 \leq x < 1 \\ 0 & \text{otherwise} \end{cases}, \tag{2.7}$$

along with its dyadic scalings and translates as we did for ψ in Eq. (2.5), then we can write $f^{(n)} = \sum_i f_i^{(n)} \varphi_{n,i}$, where the expansion coefficients $f_i^{(n)}$ are computed by averaging f over the dyadic intervals $[i \cdot 2^{-n}, (i+1) \cdot 2^{-n})$,

$$f_i^{(n)} = \int dx \; f(x) \, \varphi_{n,i}(x) = 2^n \int_{i \cdot 2^{-n}}^{(i+1) \cdot 2^{-n}} dx \; f(x). \tag{2.8}$$

In this context, φ is called the Haar *scaling function*, and is used to generate the initial discrete approximation for f. Naïvely, all of the structure in $f^{(n)}$ is at scale 2^{-n}, but this is just an artifact of the representation. Even for constant f we can develop such an expansion; in that case, all of the coefficients $f_i^{(n)}$ would be equal. Thus we would like to construct a new representation of $f^{(n)}$ in which the structure present at different scales is immediately apparent.

We begin by isolating the structure at the finest scale, 2^{-n}. Let $f^{(n-1)}$ be a second approximation to f, piecewise constant on the coarser intervals $[i \; 2^{-n+1}, (i+1) \; 2^{-n+1})$.

The expansion coefficients for $f^{(n-1)}$ may be computed either by averaging f itself over the coarser intervals, or by simply averaging adjacent coefficients for $f^{(n)}$, $f_i^{(n-1)} = \left(f_{2i}^{(n)} + f_{2i+1}^{(n)}\right)/2$. By construction the difference between these two approximations should be just the structure in $f^{(n)}$ at scale 2^{-n}. For a given coarse interval, this difference may be expressed as

$$f_{2i}^{(n)} - f_i^{(n-1)} = f_{2i}^{(n)} - \frac{1}{2}\left(f_{2i}^{(n)} + f_{2i+1}^{(n)}\right) = -\frac{1}{2}\left(f_{2i+1}^{(n)} - f_{2i}^{(n)}\right) \tag{2.9}$$

on the first half of the interval, and

$$f_{2i+1}^{(n)} - f_i^{(n-1)} = f_{2i+1}^{(n)} - \frac{1}{2}\left(f_{2i}^{(n)} + f_{2i+1}^{(n)}\right) = \frac{1}{2}\left(f_{2i+1}^{(n)} - f_{2i}^{(n)}\right) \tag{2.10}$$

on the second half. If we define the function $\psi_{n-1,i} = (\varphi_{n,2i} - \varphi_{n,2i+1})/2$, then we can express this difference as $d^{(n-1)} = \sum_i d_i^{(n-1)}\psi_{n-1,i}$, where $d_i^{(n-1)} = f_{2i+1}^{(n)} - f_{2i}^{(n)}$. Since this $\psi_{n-1,i}$ coincides exactly with the function defined in Eq. (2.6), we see that the Haar wavelets can indeed be used to express the structure in a piecewise continuous function at a given scale.

If we continue this process, decomposing $f^{(n-1)}$ into $f^{(n-2)} + d^{(n-2)}$, etc., then we can write $f^{(n)} = f^{(0)} + \sum_{j=0}^{n-1} d^{(j)}$, where $f^{(0)}$ describes the structure in $f^{(n)}$ at the coarsest scale, and $d^{(j)}$ the structure at scale 2^{j-1}. The transformation from the initial set of coefficients $\{f_i^{(n)}\}$ to $\{f_i^{(0)}, d_i^{(j)}\}$ defines the DWT. As is shown in [11], this expansion converges uniformly to $f^{(n)}$ as $j \to -\infty$. Since any function in $L^2(\mathrm{R})$ can be arbitrarily well approximated by a piecewise constant function, we see that the Haar wavelets $\psi_{j,i}$ provide a complete basis for $L^2(\mathrm{R})$. The orthonormality of these functions is trivial to prove.

2.3.3 Abstract wavelets

Multiresolution scheme

We wish to abstract from our analysis of the the Haar system a general framework that will allow us to construct more complicated scaling functions and mother wavelets, so we need to place the Haar construction in a more formal setting. We begin by introducing a sequence of function approximation spaces,

$$V_j = \left\{f : f|_{[k\cdot 2^j,(k+1)\cdot 2^j]} = \text{const}, \forall \mathrm{k} \in \mathrm{Z}\right\}, \tag{2.11}$$

where V_j consists of all those piecewise continuous functions that have structure at scale 2^j or larger. It immediately follows that these subspaces are nested,

$$\cdots \subset V_{-2} \subset V_{-1} \subset V_0 \subset V_1 \subset V_2 \subset \cdots, \tag{2.12}$$

and that their infinite intersection and union are given by $\cap_{j\in \mathrm{Z}} V_j = \{0\}$ and $\cup_{j\in \mathrm{Z}} V_j = L^2(\mathrm{R})$, respectively. A function $f(x) \in V_j$ also satisfies $f(2^j x) \in V_0$, so the V_j should be thought of as scaled versions of the single space V_0; and if a function $f(x) \in V_0$, then so are the translated functions $f(x-n)$ for all $n \in \mathrm{Z}$. These last two properties are a restatement of the fact that the dyadic scalings and translations of the Haar scaling function provide bases for the V_j.

We now make the following definition: any set of nested subspaces of $L^2(\mathrm{R})$ whose infinite intersection and union are given by $\{0\}$ and $L^2(\mathrm{R})$ itself, respectively, along with a function φ whose dyadic scalings and translations can be used to construct bases for these spaces, is said to form a multiresolution analysis (MRA) of $L^2(\mathrm{R})$. Because of the nestedness of the V_j and their union property, we can think of them as successively better approximation spaces for $L^2(\mathrm{R})$. Despite the very general nature of this definition, we can write down a single equation that completely characterizes the scaling function φ, up to a multiplicative constant. We know that $\varphi \in V_0 \subset V_1$ and that the $\varphi_{1,n}$ are a basis for V_1. Thus we must have

$$\varphi(x) = \sum_n h_n \varphi_{1,n}(x) = \sqrt{2} \sum_n h_n \varphi(2x-n), \tag{2.13}$$

for some set of real coefficients h_n. This equation is referred to as the *two-scale* relation for φ. The h_n are called low-pass filter coefficients, for reasons that will become apparent in the following section.

Having generalized the piecewise constant approximation spaces of the Haar construction, we need to generalize the wavelet that represented the difference between two such spaces. We begin by formally defining the orthogonal complement of V_{j-1} in V_j to be the space W_{j-1}: $V_j = V_{j-1} \oplus W_{j-1}$. It follows immediately from this definition that the W_j are mutually orthogonal. The proof is simple: if $j' < j$, then $W_{j'} \subset V_{j'+1} \subseteq V_j$; combining this with $V_j \perp W_j$, we find $W_{j'} \perp W_j$. The bases of these spaces are provided by the dyadic scalings and translates $\psi_{j,k}$ of a single function ψ. It is assumed that the $\psi_{0,k}$ form an orthonormal for W_0; and since $W_{j'} \perp W_j$ for all $j' \neq j$, the $\psi_{j,k}$ provide an orthonormal

basis for $L^2(\mathrm{R})$. The wavelet ψ satisfies a two-scale relation similar to Eq. (2.13) for the scaling function:

$$\psi(x) = \sum_n g_n \varphi_{1,n}(x) = \sqrt{2} \sum_n g_n \varphi(2x - n), \tag{2.14}$$

for some g_n, called high-pass filter coefficients. This equation seems to fit quite naturally into our framework: since we began by constructing the function approximation spaces, and only later defined the wavelet to describe the difference between two such spaces, the scaling functions appear to be the more fundamental object. However, it is possible to reverse the process: if we know there is a certain structure present in the function, and we would like to represent this structure using a single wavelet, then we can ask what kind of approximation space will give rise to the desired wavelet. This approach will be discussed more thoroughly in Sec. 2.4.

One of the simplest and most useful criteria we can impose on our MRA is that both φ and ψ be compact, which implies that only a finite number of the filter coefficients are nonzero. In particular if we are interested in representing the function defined on finite curves rather than the entire real line, noncompact wavelets are of no use to us. Another simple criterion is that φ be able to reproduce polynomials up to some order. In almost all cases of interest, the function being represented is locally polynomial, and we would like to be able to represent this with the smallest number of coefficients possible. A third and more pragmatic criterion is that the initial construction of the approximation and the implementation of the DWT should be simple. This will be discussed further in Secs. 2.3.3 and 2.4.

Discrete wavelet transformation

Assume now that we have constructed a useful MRA, and we have the projection $f^{(j)}$ of a function f into V_j,

$$f^{(j)}(x) = \sum_k < \varphi_{j,k}, f > \varphi_{j,k}(x) \equiv \sum_k \lambda_{j,k} \varphi_{j,k}(x); \tag{2.15}$$

how to actually evaluate the inner products involved in this projection will be discussed in Sec. 2.3.3. We are interested in decomposing $f^{(j)}$ into $f^{(j-1)} + d^{(j-1)}$, and thus in computing the coefficients $\lambda_{j-1,k} \equiv < \psi_{j-1,k}, f >$ and $\gamma_{j-1,k} \equiv < \varphi_{j-1,k}, f >$. Using Eq. (2.14) we can write

$$\psi_{j,k}(x) \quad = \quad 2^{j/2} \psi(2^j x - k)$$

$$\begin{aligned} &= 2^{j/2}\sqrt{2}\sum_n g_n \varphi(2^{j+1}x - 2k - n) \\ &= \sum_n g_n 2^{(j+1)/2}\phi(2^{(j+1)}x - (2k+n)) \\ &= \sum_n g_n \phi_{j+1,2k+n}(x), \end{aligned} \tag{2.16}$$

and thus

$$\gamma_{j-1,k} = \sum_n g_{n-2k}\lambda_{j,n}. \tag{2.17}$$

Similarly, using Eq. (2.13) we can write

$$\begin{aligned} \varphi_{j,k}(x) &= 2^{j/2}\varphi(2^j x - k) \\ &= 2^{j/2}\sqrt{2}\sum_n h_n \varphi(2^{j+1}x - 2k - n) \\ &= \sum_n g_n 2^{(j+1)/2}\phi(2^{(j+1)}x - (2k+n)) \\ &= \sum_n g_n \phi_{j+1,2k+n}(x), \end{aligned} \tag{2.18}$$

and thus

$$\lambda_{j-1,k} = \sum_n h_{n-2k}\lambda_{j,n}. \tag{2.19}$$

Eqs. (2.17) and (2.19) provide us with the desired map from $f^{(j)}$ to $f^{(j-1)} + d^{(j-1)}$. A single iteration of these equations defines the abstract DWT. By construction this mapping is transformation is invertible. The equation for inverse is given by

$$\lambda_{j,l} = \sum_n h_{l-2n}\lambda_{j-1,l} + \sum_n g_{l-2n}\gamma_{j-1,l} \tag{2.20}$$

We will discuss the proper data structure to use when implementing the DWT in Sec. 2.4.

It is interesting to consider the tilings of the space-frequency plane generated by each successive wavelet transformation. The scaling functions used to represent the signal at the finest level are localized in space and completely delocalized over the frequency interval represented by our sampling. The corresponding tiling of the time frequency plane is shown in Fig. 2.3(a). For the Daubechies and other compactly supported bases, the vertical lines denote the actual boundaries of the scaling functions. The vertical extent of these functions is not so clearly defined, and the horizontal boundaries are only meant to denote their characteristic extent in frequency space.

After one iteration of the wavelet transformation, we have the tiling shown in Fig. 2.3(b). Pairs of adjacent rectangles from the initial tiling have been replaced by rectangles of twice the width and half the height; the upper rectangles represent the wavelets,

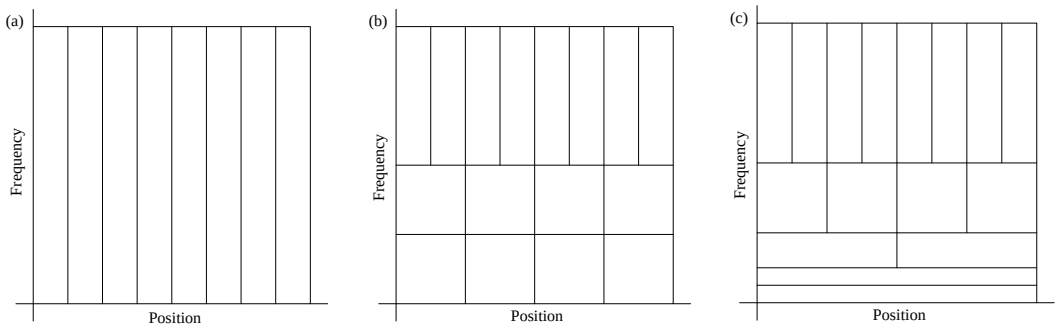

Figure 2.3: Tilings induced by the bases of (a) V_j, (b) $V_{j-1} \oplus W_{j-1}$, and (c) $V_{j-3} \oplus W_{j-3} \oplus W_{j-2} \oplus W_{j-1}$.

and the lower the scaling functions. If we continue this process, always combining two of the lower rectangles, then after four iterations, we have the tiling shown in Fig. 2.3(c). From these figures we see that the coefficients g_n and h_n can be interpreted as the coefficients of high-pass and low-pass filters, respectively. We can also see that wavelets are best at representing signals whose frequency spectrum is concentrated near zero frequency. Generalized tilings with optimal resolution about nonzero frequencies will be considered in Sec. 2.5.

Construction of the expansion

In order to compute the expansion coefficients $\lambda_{j,k}$, we need to know the form of $\varphi_{j,k}(x)$. However, there are no analytic expressions available for most scaling functions of interest; the best we can do is numerically compute the value of $\varphi_{j,k}(x)$ at the dyadic rationals. This is done using the "cascade algorithm" as follows. A trivial but useful way of characterizing φ is through its expansion coefficients in $V_0 \oplus_{j=1}^{\infty} W_j$: $< \varphi, \varphi_{0,k} >= \delta_{0,k}$ and $< \varphi, \psi_{j,k} >= 0$, for all $j \geq 0$. If we apply the inverse DWT to these coefficients, that at each level we find

$$< \varphi, \varphi_{j,k} >= \sum_n h_{k-2n} < \varphi, \varphi_{j-1,n} > . \tag{2.21}$$

Now it can be shown that $\varphi_{-j,k}(x) \to \delta(x - 2^{-j}k)$ as $j \to \infty$, and that the convergence is exponential [11]. Thus the inverse DWT converges to the values of the scaling function at the dyadic rationals. Since the wavelets can also be expressed in terms of φ, this algorithm alo allows us to compute ψ at the dyadic rationals.

There is an alternative to numerical quadrature if the dual scaling function has a nonzero number of vanishing moments. If φ satisfies $\int dx\, x^m \varphi(x) = 0$ for $m = 1, 2, \ldots, N-$

1,[2] then we have a simple one-point quadrature formula:

$$\begin{aligned} < \varphi_{j,k}, f > &= 2^{j/2} \int dx\ f(x)\varphi(2^j x - k) \\ &= 2^{j/2} \int dx\ f(x)\varphi\left(2^{j/2}(x - 2^j k)\right) \\ &= 2^{j/2} \int dx\ f\left(x + 2^{-j}k\right)\varphi(2^j x) \\ &= 2^{j/2} \sum_{n=0}^{\infty} \frac{1}{n!} \int dx\ f^{(n)}\left(2^{-j}k\right) x^n \varphi\left(2^j x\right) \\ &= 2^{j/2} f\left(2^{-j}k\right) + \mathcal{O}\left(2^{-j(N+1/2)}\right) \end{aligned} \tag{2.22}$$

The $\lambda_{j,l}$ computed using this formula are referred to as the "empirical" expansion coefficients, while those computed by some higher order quadrature formula are the "theoretical" coefficients. The practical difference between these is found in the behavior of the projected function $f^{(j)} = \sum_l \lambda_{j,l}\varphi_{j,l}(x)$. If we use theoretical coefficients, then we may sample $f^{(j)}(x)$ at scales finer than the distribution of scaling functions. If we use the empirical, we may not.

Biorthogonal wavelets

One of the problems with an MRA built from a single function φ is that if the wavelet transformation is to be invertible, then the scaling function and wavelet cannot be symmetric, except in the trivial case of the Haar wavelet [11]. Thus when we project a function f into V_j, we introduce an asymmetry into the function that was previously not there. We can avoid this by using pairs of scalings functions and wavelets $\{\varphi, \tilde{\varphi}, \psi, \tilde{\psi}\}$, along with their associated filters $\{h, \tilde{h}, g, \tilde{g}\}$, in order to construct bases for our approximation spaces [9]. The dyadic scalings and translates of φ and ψ still provide complete bases for the $\{V_j\}$ and $\{W_j\}$, but they are no longer orthonormal. Rather we have $< \tilde{\varphi}_{j,k}, \varphi_{j',k'} >= \delta_{j,j'}\delta_{k,k'}$, as well as $< \tilde{\psi}_{j,k}, \psi_{j',k'} >= \delta_{j,j'}\delta_{k,k'}$; i.e., φ and $\tilde{\varphi}$ are biorthogonal, as are ψ and $\tilde{\psi}$. The projection operators into V_j and W_j are now written as $\mathcal{P}_j(\cdot) = \sum_l < \tilde{\varphi}_{j,l}, \cdot > \varphi_{j,l}$ and $\mathcal{Q}_j(\cdot) = \sum_l < \tilde{\psi}_{j,l}, \cdot > \psi_{j,l}$. The functions $\tilde{\varphi}$ and $\tilde{\psi}$ are referred to as dual functions. Note that $\tilde{\varphi}$ and $\tilde{\psi}$ can be used as bases for a second sequence of approximation spaces $\{\tilde{V}_j\}$ and $\{\tilde{W}_j\}$. In this case φ and ψ are used to build the projection operators.

[2] The scaling function must satisfy $\int dx\ \varphi(x) = 0$.

Sparse function/operator representation

Once we have the single-scale decomposition of a function f in terms of the $\lambda_{j,l}$, we may construct a multiscale representation using the DWT. The wavelet coefficients $\gamma_{j',l}$ for $j' < j$ encode the detail structure in f at scale $2^{-j'}$, so if we are to have an efficient representation of f, we want most of these to vanish. Now the $\gamma_{j',l}$ may also be computed directly as $\gamma_{j',l} = \int dx\ f(x)\tilde{\psi}_{j',l}(x)$. If we assume that $\int dx\ x^m\tilde{\psi}(x) = 0$ for $m = 0, 1, 2, \ldots, N-1$, then for functions that are locally well approximated by a polynomial, most of these coefficients will be negligible. Another way of saying this is that φ and its translates exactly reproduce polynomials up to order $N-1$ at all scales. N is called the order of the wavelet family.

We also want our wavelets to satisfy $\int dx\ x^m\psi(x) = 0$ for $m = 0, 1, 2, \ldots, \tilde{N}-1$. This does not increase the sparsity of the representation, but it does have the following consequence. A function f can be represented as $\sum_l \lambda_{j,l}\varphi_{j,l} + \sum_l \gamma_{j,l}\psi_{j,l}$ for any j. If the first $\tilde{N}$ moments of ψ vanish, then the first $\tilde{N}$ moments of f are encoded in the average coefficients. This is in keeping with the interpretation of the V_j as containing the average global properties of f, while the W_j contain the local details.

2.4 Second generation wavelets

2.4.1 General lifting scheme

As we saw in Sec. 2.3 the wavelets, scaling functions, and both their duals can all be generated using the cascade algorithm once the filter coefficients are known. The construction of a set of biorthogonal basis functions with specific properties can thus be reduced to the construction of the associated filters. We now present the "lifting scheme," recently discovered by Sweldens [69, 70, 10], which allows the iterative construction and application of any set of biorthogonal filters beginning with the trivial filters defined by the "lazy wavelet" transformation. In particular we can use this scheme to construct wavelets and dual wavelets that have an arbitrary number of vanishing moments.

In order to simplify our discussion of the lifting scheme, we introduce the complex number representation of discrete signals commonly found in the signal processing literature [11]. Let $\{f_n : n \in Z\}$ be either the elements of a discrete signal or the discrete samples of a continuous signal f. We can combine these elements into a single object defined by

$f(\omega) = \sum_n f_n e^{in\omega}$. This induces a mapping of the space of discrete signals onto the ring of Laurent polynomials defined over the unit circle in the complex plane. To see why this is useful, we first note that the filters can be given a similar representation: $h(\omega) = \sum_n h_n e^{-in\omega}$, $g(\omega) = \sum_n g_n e^{-in\omega}$, etc. We then note that using this notation, the application of a filter to a signal, usually written as a convolution product, reduces to a simple multiplication.

A single "lifting" transformation is defined as follows [69]. Given an initial set of finite, biorthogonal filters, $\{h_0, \tilde{h}_0, g_0, \tilde{g}_0\}$ with scaling function φ_0, we may construct a new set $\{h_0, \tilde{h}, g, \tilde{g}_0\}$ having the same scaling function with the following transformation:

$$\tilde{h}(\omega) = \tilde{h}_0(\omega) + \tilde{g}_0(\omega)\overline{s(2\omega)}, \tag{2.23}$$

$$g(\omega) = g_0(\omega) - h_0(\omega)s(2\omega), \tag{2.24}$$

where $s(\omega)$ is an arbitrary trigonometric polynomial. More generally it can be shown that all filters with the same scaling function may be derived in this manner for some $s(\omega)$. If $\{\varphi_0, \tilde{\varphi}_0, \psi_0, \tilde{\psi}_0\}$ is the initial set of biorthogonal functions, then the functions $\{\varphi_0, \tilde{\varphi}, \psi, \tilde{\psi}_0\}$ associated with the new filters are given by

$$\varphi(x) = \varphi_0(x), \tag{2.25}$$

$$\psi(x) = \psi_0(x) - \sum_k s_k \varphi(x-k), \tag{2.26}$$

$$\tilde{\varphi}(x) = \sum_k (\tilde{h}_0)_k \tilde{\varphi}(2x-k) + \sum_k s_{-k} \tilde{\psi}(x-k), \tag{2.27}$$

$$\tilde{\psi}(x) = \sum_k \tilde{g}_k \tilde{\varphi}(2x-k). \tag{2.28}$$

Eq. (2.25) simply tells us that, as promised, the scaling function is unchanged. Eqs. (2.27) and (2.28) define the new dual functions, but are not of much use because they are only implicit definitions. Eq. (2.26), however, is quite useful since it defines the new wavelet in terms of the initial scaling function and wavelet. This allows us complete control over the shape of the new wavelet. If we switch all of the functions and filters in Eqs. (2.23)–(2.28) with their dual counterparts, the resulting equations define the "dual lifting" transformation, which can be used to construct all possible filters that have the same dual scaling function as the initial set. The lifting scheme is the construction of a new set of biorthogonal filters beginning with some initial set through a succession of lifting and dual lifting steps.

As we mentioned above, the initial filters are usually taken to be those associated with the lazy wavelet, or polyphase, transformation. These filters, which we will denote

as $h_{\rm lazy}$ and $g_{\rm lazy}$, have an all but trivial action: they subsample a signal into its even and odd components. However, as was recently shown in [10], any set of biorthogonal filters can be written as a sequence of lifting and dual lifting steps, along with a scaling operation, applied to these filters. [3] A single dual lifting step followed by a single lifting step is sufficient to generate the biorthogonal Haar transformation; and combining these with a second pair of dual and direct lifting steps, we are able to generate the entire $(2n, 2n')$ family of biorthogonal wavelets described in Sec. 2.3.3. Since the former provides the simplest example of a wavelet and the latter are examples of interpolating filters, and are thus of considerable numerical interest, we will outline their construction and application below. If we think of the even and odd samples as the output of low and high pass filters, then Nyquist's theorem tells us that these filters will be subject to a tremendous amount of aliasing. The lifting and dual lifting steps can be thought of as a means of reducing the aliasing in the low pass and high pass filters, respectively.

The construction of the polynomials associated with the dual and direct lifting steps is often best done by considering the action of the transformation on the signal itself. As shown in [69], the wavelet transformation generated by an initial set of filters $\{h_0, \tilde{h}_0, g_0, \tilde{g}_0\}$, followed by a single dual and single direct lifting described by $\tilde{s}(\omega)$ and $s(\omega)$, can be written in three steps. The first transformation is just that generated by the initial filters:

$$\lambda^0_{j,l} = \sum_k (\tilde{h}_0)_{k-2l} \lambda_{j+1,k} \quad \text{and} \quad \gamma^0_{\rm j,l} = \sum_{\rm k} (\tilde{\rm g}_0)_{\rm k-2l} \lambda_{\rm j+1,k}. \tag{2.29}$$

The second and third steps are

$$\gamma_{j,l} -= \sum_k \tilde{s}_{l-k} \lambda^0_{j,k}, \tag{2.30}$$

and

$$\lambda_{j,l} += \sum_k s_{l-k} \gamma_{j,k}. \tag{2.31}$$

Note that we have used the C-style operators $-=$ and $+=$; this was done intentionally to show how easy it is to implement these transformations. In this context the dual lifting step is also called the "predict" phase of the transformation, and the lifting step is the "update." The reasons for these will explained below. The inverse transformation is found

[3] If we write our filters as a 2×2 matrix of Laurent polynomials, where the entries of the matrix specify the action of the filters on the even and odd parts of the signal, then the proof of this statement reduces to the Euclidean algorithm on the ring $SL(2, R[z, z^{-1}])$.

by reversing the order of Eqs. (2.30) and (2.31), switching += with −=, and performing the conventional inverse of Eq. (2.29).

2.4.2 Examples of lifting

Haar transformation

We begin with the Haar transformation. The lifting polynomials are $\tilde{s}(\omega) = e^{i\omega}$ and $s(\omega) = -e^{i\omega}/2$. Eqs. (2.29)–(2.31) become $\lambda_{l,j+1} = \lambda_{2l,j}$ and $\gamma_{l,j+1} = \lambda_{2l+1,j}$; $\gamma_{l,j+1}-= \lambda_{l,j+1}$; and $\lambda_{l,j+1}+= \gamma_{l,j+1}$. If we combine these equations, we find the expected

$$\lambda_{l,j+1} = \frac{1}{2}\left(\lambda_{2l,j} + \lambda_{2l+1,j}\right), \tag{2.32}$$

$$\gamma_{l,j+1} = \lambda_{2l+1,j} - \lambda_{2l,j}. \tag{2.33}$$

Even though the Haar transformation is simple to begin with, the lifting formulation offers an extra advantage: the transformation can be done in place; no additional storage needs to be allocated.

Family of $(2n, 2n')$ filters

We begin with the vanishing moments. The condition that $\int dx\ x^k \tilde{\psi}(x) = 0$ for $k = 0, 1, \ldots, 2n-1$ is equivalent to φ and its translates being able to reproduce polynomials up to order $2n-1$. In particular if the coefficients $\lambda_{j,l}$ are the discrete samples of an order $2n-1$ polynomial, then in such a basis all of the detail coefficients $\gamma_{j+1,l}$ should vanish. We see from Eq. (2.30) that the $\gamma_{j+1,l}$ are computed by subtracting the values of the surrounding average coefficients weighted by the s_k from the initial $\gamma^0_{j+1,l}$. Thus in order for these coefficients to vanish, we must be able to predict the initial values exactly using only the average coefficients. Since we know that the initial samples $\lambda_{j,l}$ are derived from an order $2n-1$ polynomial, this prediction can be done using Neville's algorithm for polynomial interpolation on $2n$ of the surrounding $\lambda^0_{j+1,l}$. The actual procedure by which the s_k are computed using this algorithm can be found in [22].

We now turn to the dual moments. Using Eq. (2.26) the condition $\int dx\ x^k \psi(x) = 0$ for $k = 0, 1, \ldots, 2n'-1$ can be rewritten as

$$\int dx\ x^k \psi^{pr}(x) = \sum_l s_l \int dx\ x^k \varphi^{pr}(x-k), \tag{2.34}$$

where φ^{pr} and ψ^{pr} are the scaling function and wavelet that results from the predict phase. Using the duals of Eqs. (2.25–2.28), we can rewrite this as a $2n' \times 2n'$ system of equations that can be solved for the s_k.

2.4.3 Wavelets on the interval

Most wavelet constructions published to date rely explicitly on Fourier techniques. The reason is that scaling and translation reduce to simple algebraic operations in Fourier space: the Fourier transformation of $f(ax)$ is $a^{-1}\hat{f}(x)$, and the transform of $f(x-b)$ is $e^{ikb}\hat{f}(x)$. However, this reliance makes the extension from infinite and periodic domains to bounded intervals quite difficult. The lifting scheme has no such dependence; it is performed entirely in ordinary space. The construction of the family of $(2n, 2n')$ wavelets, can be easily extended to bounded domains by requiring that the number of vanishing (dual) moments remains the same even for those wavelets affected by the boundaries. However, the loss of translational invariance does come at a price; both the prediction and update coefficients become spatially dependent, and the latter also become scale dependent. Again the details of the implementation can be found in [22].

2.5 Wavelet packets

Traditionally the n-th iteration of a multilevel wavelet transformation is applied to the output of the low-pass filter from the $(n-1)$-th iteration, thus splitting it into two finer frequency bands. Because of this bias, the highest resolution in frequency space will always be found at zero frequency. For a signal dominated by a nonzero frequency, the wavelet representation will thus be far from optimal. We can remove this bias by sometimes applying the transformation to the output of the high pass filter. This allows us to shift the optimal resolution to any desired frequency.

For a sequence of n successive transformations, the different possible splittings generate a binary tree with 2^n leaf nodes. The ordinary wavelet transformation is simply the leaf node reached by always splitting on the low-pass output. If our signal contains a single dominant frequency, then the optimal sequence will correspond to a single path through this tree, ending at a particular leaf node. If there are multiple frequencies, then there will be multiple paths, each ending at a node. The basis functions associated with

such transformations can be written as linear combinations of ordinary wavelets, and are thus called wavelet packets.

When the dominant frequency is known, it is easy to find the optimal path through the tree. If not, then there is an extension of the conventional wavelet transformation that tries all possible sequences, and then decides which is optimal based on the minimization of some additive entropy function. The additional cost of this search is minimal: the complexity of the transformation increases from $\mathcal{O}(N)$ to $\mathcal{O}(N \log N)$. The details of this transformation can be found in [76].

As with the ordinary DWT, the wavelet packet transformation can be thought of in terms of the tiling of the space-frequency plane it generates. A sample tiling is shown in Fig. 2.4. Note that the narrow frequency bands, corresponding to basis elements with the finest resolution in frequency space, can be shifted to an arbitrary, nonzero frequency.

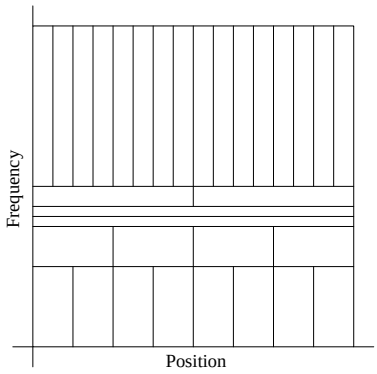

Figure 2.4: Tiling of the space-frequency plane generated by a discrete wavelet packet transformation.

Chapter 3

Wavelet Analysis of Two Dimensional Quantum Scattering

A physically intuitive way of treating two-dimensional quantum scattering from an arbitrary one-dimensional boundary is to reformulate the scattering equations for the wave function as a set of coupled integral equations of the first kind for the associated T-matrix. A naïve discretization of these equations leads to a dense linear system, which requires the use of an $\mathcal{O}(N^3)$ solution algorithm, where N is the number of basis elements in the discretization. Moreover the resulting system is poorly conditioned, so fine scale features of the wave function are easily lost. Here we present a multiscale discretization of the coupled integral equations using compactly supported, biorthogonal wavelets. This results in an almost *sparse* system with *bounded* condition number that can be iteratively solved, thus avoiding the expense of a direct solver. The complexity of the solver scales as $\mathcal{O}(N^\alpha)$, where $1 \lesssim \alpha < 2$. Because of its numerical stability, our method is sensitive to the details of the scattering at all length scales. As an illustrative example, we compute the electronic conductance of a quantum point contact coupled to an open resonator as the dimensions of the resonator are varied. In this example the bounding walls defining the contact are each $40\lambda_\mathrm{f}$ long, where λ_f is the Fermi wavelengths, while the contact itself is only $\lambda_\mathrm{f}/4$ wide.

3.1 Introduction

3.1.1 Motivation

In a recent paper da Luz *et al.* [13] suggest that a physically intuitive and computationally efficient way of treating n-dimensional quantum scattering from an arbitrary

collection of $(n-1)$-dimensional Dirichlet boundaries is to recast the differential equation for the wave function as a set of coupled integral equations for the associated T-matrix. These equations can then be discretized and solved using standard matrix inversion techniques, and the full two-dimensional wave function can be reconstructed from a one-dimensional sum over sources. Da Luz *et al.* refer to their method as "boundary wall" for reasons that will become clear in Sec. 3.2. While this method is indeed physically appealing, allowing them to treat both the interior and exterior problems for closed surfaces simultaneously, as well as barriers of finite penetrability, it can become numerically unstable for sufficiently fine discretizations of the barrier surfaces. This is particularly troublesome when there are a number of disparate length scales present, as there often are in problems of interest to the mesoscopic community.

The discretization used in [13] consists of replacing the continuous boundaries with uniform one-dimensional lattices, and approximating the continuous functions and operators defined on these boundaries by their values on the lattice points. [1] A more formal description of this method in terms of a projection operator onto a finite-dimensional function approximation space is as follows: the basis of the approximation space consists of N elements, where N is the number of lattice points; each basis element is equal to 1 at one of the lattice points, and 0 on all of the others; the dual basis elements used to build the projection operator into this space are N δ-functions, one located at each scatterer. While the use of δ-functions as projection operators in and of itself does not give rise to any numerical difficulties, [2] the uniformity of their distribution does. If we are interested in scattering from boundaries that have structure on scales fine compared to their overall extent, or we want to compute the near field wave function close to any one of the boundaries, then the fine sampling needed in some small region will lead to a fine discretization everywhere, and thus to potentially large linear systems. Since the complexity of the matrix solver used in [13] scales as $O(N^3)$, where N is the number of points in the

[1] The authors of [13] also use a discretization based on dividing each boundary Γ into a finite number of disjoint segments Γ_i, and then replacing the continuous functions f on Γ with the discontinuous functions $\sum_i \left(\int_{\Gamma_i} ds\ f(s) \right) \chi_{\Gamma_i}$, where χ_{Γ_i} is the characteristic function on Γ_i. This method shares the same difficulties as the lattice-based discretization, so we will restrict our attention to the latter. It is worth noting, however, that this method corresponds to using the projection operator associated with the lowest order member of the family of average interpolating wavelets [15, 71]. The projection operators built from higher order members of this family are quite useful in implementing the fully adaptive wavelet packet transformation [18].

[2] We shall see later that δ-functions provide an exponentially convergent approximation to the dual scaling functions associated with interpolating wavelet families.

discretization, time and memory constraints immediately restrict the kinds of scattering geometries we can consider.

An alternative to the use of direct solvers is the use of iterative methods. These methods generate a sequence of vectors $\{x_n\}$ that converge to the solution of the system $A \cdot x = b$ in the sense that they minimize $||A \cdot x - b||$ in some suitably defined norm. The algorithm used to generate each x_n requires only that we be able to compute the product of A (and sometimes A^T) with a vector. Thus if the number of significant elements of A is of order $\mathcal{O}(N^\alpha)$ with $1 \leq \alpha < 2$, these methods can be more efficient than direct solvers. The problem with these methods is that they are extremely sensitive to the conditioning of A. Now the integral operators that describe scattering in two dimensions are operators of order -1. In general if $\mathcal{O}_n$ is an operator of order n, then $\mathcal{O}_n\psi$ has n fewer continuous derivatives than does ψ. When operators of negative order are discretized using a single-scale basis as described above, the resulting stiffness matrices are usually poorly conditioned, with the condition number growing unboundedly with the size of the discretization. Because of the finite precision at which computers work, this leads to a loss of information in the final two-dimensional wave function. Thus we are left in the uncomfortable position that as we try to include fine-scale scattering effects, and thus increase the overall size of our linear system, this instability defeats our efforts.[3]

A general way of avoiding a loss of information is through the black art of preconditioning [27]. This involves replacing our initial system, $A \cdot x = b$, with a secondary system, $(\tilde{A}^{-1} \cdot A) \cdot x = \tilde{A}^{-1} \cdot b$, where $\tilde{A}$ is chosen so that $\tilde{A}^{-1} \cdot A$ is close to the identity. $\tilde{A}$ must, of course, be chosen so that its inverse can be readily computed, and it itself should be well conditioned. At best $\tilde{A}$ can be derived from A through some physically well-motivated process, such as eliminating matrix elements describing fine-scale interactions; at worst $\tilde{A}$ can simply be chosen to have the inverses of the diagonal elements of A along its diagonal and zeroes everywhere else. One of the more recent conditioning techniques to have been developed is based upon transformation to a *multiscale* basis. The basis described above is an example of a single-scale basis; the same elements are used to describe structure on both coarse and fine scales A multiscale basis uses different elements to describe structure at different scales. This segregation of information into different scales gives rise to linear systems that are automatically well conditioned. A prominent example of this technique is

[3]Direct solvers that are stable in the presence of near zero eigenvalues, such as the Gauss-Jacobi method with full pivoting [27], are not affected by this loss.

found in the multigrid solvers for partial differential equations [58].

Here we present a wavelet-based multiscale discretization of the boundary wall equations, which allows us to avoid both the numerical instability and the $\mathcal{O}(N^3)$ complexity of the original discretization. In their simplest incarnation, wavelets are defined as the dyadic scalings and translates of a single function, ψ, appropriately called the "mother wavelet." If ψ is chosen properly, the complete set of wavelets can be shown to provide complete, orthonormal bases for many function approximation spaces of interest. The most important characteristic of wavelets is that they are localized in both space and frequency. Their aspect ratios in the space-frequency plane can be perfectly adapted to representing the local frequency content of a signal. The operator describing wave propagation in two dimensions has an almost *sparse* representation when discretized in such a basis; the number of matrix elements containing most of the information scales as $O(N^\alpha)$, where $1 \lesssim \alpha < 2$. Moreover because of the multiscale nature of the basis, the representation is well conditioned; in fact, the condition number is uniformly bounded independent of the size of the discretization. These two conditions allow us to use iterative solvers for our linear systems.

3.1.2 Outline

We begin in Sec. 3.2 with a review of the boundary wall method and a discussion of why operators of negative order are poorly behaved with respect to conventional discretizations. Sec. 3.3 outlines the properties of wavelets needed in this paper: Sec. 3.3.1 gives a brief exposition of wavelet analysis in general; Sec. 3.3.2 shows how to project a continuous function onto a wavelet basis; Sec. 3.3.3 discusses why wavelets provide excellent bases for both function and operator approximation; and finally Sec. 3.3.4 offers a more detailed description of the particular families of wavelets used in this work, as well as the actual numerical method used to implement the associated wavelet transformations. Rather than couching wavelet theory purely in the language of functional analysis, we will follow an approach borrowed from the digital signal processing community that represents wavelets and wavelet packets as different tilings of the position-frequency plane. This approach should be more intuitively appealing to physicists, who are used to thinking in terms of phase space. In Sec. 3.4 we outline the application of wavelets to the boundary wall equations. We test our method in Sec. 3.5 against a few canonical problems whose analytic solutions are at least partially known, as well as a more complicated example that illustrates the power

of our method: we calculate the conductance of a subwavelength quantum point contact (QPC) coupled to an open resonator.[4] We conclude in Sec. 3.6 with suggestions for future work, including the use of wavelet packets in order to achieve sparser representations of the scattering matrix and making the method fully adaptive.

3.2 Boundary wall method

3.2.1 Theory

The general problem we wish to address is two-dimensional quantum scattering at fixed energy from a one-dimensional boundary, Γ, embedded in the plane. The scattering from the boundary is defined by requiring the wave function, $\psi(\mathbf{r})$, to satisfy the Helmholtz equation away from the boundary, and the Dirichlet condition, $\psi(\mathbf{r})|_{\mathbf{r}\in\Gamma} = 0$, on Γ. More generally we could impose a linear combination of Dirichlet and Neumann boundary conditions [13] along Γ, but we will restrict ourselves here to the simpler case, except for a few remarks in 3.6 on the potential use of Neumann conditions in mesoscopic transport calculations. The class of admissible curves, Γ, will also be restricted to curves well-behaved in the usual physicist's sense; in particular, we assume that Γ does not have structure on arbitrarily small scales, although it is allowed to have a finite number of discontinuities in any of its derivatives. It may be possible to incorporate curves with self-similar structure, that is fractals, into the paradigm developed here, but we will not consider this possibility.[5] Finally we will assume for the moment that Γ is connected. This has no real effect on the theory; we simply wish to treat the disconnected case more explicitly.

Before describing how the boundary wall method solves this problem, we consider the physical motivation for its study. Recent advances in electron beam lithography now allow experimentalists to pattern gates on GaAs/Al_xGa_{1-x}As heterostructures down to the nanoscale level. Applying a bias voltage to these gates depletes the two-dimensional electron gas (2DEG) underneath them, making them behave essentially as specular electron mirrors. The outlines of these gates in the 2DEG are sharply defined on the order of λ_f, so we may not ignore diffraction from their ends. The AlGaAs wafers themselves have also become clean enough that in the low temperature limit, their mean free path and phase

[4] The results of this section were first presented in [40].

[5] The techniques needed to generalize the boundary wall method to fractal curves are similar to those behind wavelet-based, fractal image compression schemes [12].

coherence lengths are many times longer than the Fermi wavelength, λ_f, of the 2DEG. Thus if we neglect impurity, scattering and we assume the depletion region is sharply defined, we are actually in a position to study experimentally the abstract system proposed above. The degree to which our assumptions are satisfied and how they affect our results will be discussed in later sections.

As with most partial differential equations (PDEs), the most difficult part of finding the solution to our problem is imposing the boundary conditions. With the actual physical systems described above in mind, we consider the following approach. The fundamental solution to the Helmholtz equation in two dimensions is a known analytic function, the zeroth order Hankel function of the first kind $H_0^{(1)}$. Thus if our problem were instead scattering from some potential V, we would naturally be led to a Lipmann-Schwinger description of the scattering [60]:

$$|\psi\rangle = |\psi\rangle_0 + \widehat{G}_0 \widehat{V} |\psi\rangle, \tag{3.1}$$

where $|\psi\rangle_0$ is a solution of the free problem satisfying some suitable boundary condition at infinity, $\widehat{G}_0$ is the free Green's function, which in position representation is just $G_0(\mathbf{r_1}, \mathbf{r_2}) = -(i/4)H_0^{(1)}(2\pi|\mathbf{r_1} - \mathbf{r_2}|/\lambda)$, and $\widehat{V}$ is the potential operator. Now we ask whether we can construct a potential $\widehat{V}$ so that in some possibly singular limit, Eq. (3.1) reduces to our original problem, complete with properly enforced boundary conditions.

The required potential, in position representation, is provided by

$$V(\mathbf{r}) = \lambda \int_\Gamma ds\ \delta(\mathbf{r} - \mathbf{r}_\Gamma(s)), \tag{3.2}$$

a δ-function wall parametrized by $\mathbf{r}_\Gamma$. As in one dimension, this wall is semitransparent for finite λ, but in the limit $\lambda \to \infty$ it becomes impenetrable, and forces a zero in $\psi(\mathbf{r})$ along Γ. Substituting this potential into Eq. (3.1), we find that the wave function is given by

$$\psi(\mathbf{r}) = \psi_0(\mathbf{r}) + \int_\Gamma ds\ G_0\left(\mathbf{r}, \mathbf{r}(s)\right) \lambda \psi(\mathbf{r}(s)). \tag{3.3}$$

We have thus reduced our problem to finding an equation for $\lambda\psi(\mathbf{r})$ along Γ. This is provided by evaluating Eq. (3.3) on Γ itself. With an obvious change of notation from $\mathbf{r}(s)$ to s, we have

$$\psi(s) = \psi_0(s) + \int_\Gamma ds'\ G_0(s, s') \lambda \psi(s'), \tag{3.4}$$

which we will write as

$$\int_\Gamma ds'\ \left(\frac{1}{\lambda}\delta(s, s') - G_0(s, s')\right) \lambda \psi(s') = \psi_0(s). \tag{3.5}$$

Written in this way, it is easy to take the $\lambda \to \infty$ limit. [6] The first term on the left vanishes, and we are left with

$$-\int_\Gamma ds'\; G_0(s,s')\lambda\psi(s') = \psi_0(s). \tag{3.6}$$

Note that as $\lambda \to \infty$, $\psi(s') \to 0$ in such a way that their product remains finite; this must be true on purely physical grounds since their product is the weighting function that appears in Eq. (3.3).

The product $\lambda\psi(s')$ is simply the potential operator applied to the wave function and evaluated on the boundary. In conventional treatments of scattering theory [60], this product is defined to be the T-matrix $T[\psi_0](s')$. Thus, as promised, we are left with a Fredholm integral equation of the *first* kind for the scattering T-matrix,

$$-\int_\Gamma ds'\; G_0(s,s')T[\psi_0](s') = \psi_0(s). \tag{3.7}$$

Substituting this back into Eq. (3.3), we find that $\psi(\mathbf{r})$ in the plane is given by

$$\psi(\mathbf{r}) = \psi_0(\mathbf{r}) + \int_\Gamma ds\; G_0\left(\mathbf{r}, \mathbf{r}(s)\right) T[\psi_0](\mathbf{r}(s)). \tag{3.8}$$

The explicit generalization of Eq. (3.7) to the case where Γ is a sum of connected curves $\Gamma = \cup_{i=1}^n \Gamma_i$ is

$$\begin{pmatrix} \int_{\Gamma_1} ds_1'\; G_0(s_1,s_1') & \cdots & \int_{\Gamma_n} ds_n'\; G_0(s_1,s_n') \\ \vdots & \ddots & \vdots \\ \int_{\Gamma_1} ds_1'\; G_0(s_n,s_1') & \cdots & \int_{\Gamma_n} ds_n'\; G_0(s_n,s_n') \end{pmatrix} \cdot \begin{pmatrix} T[\psi_0](s_1') \\ \vdots \\ T[\psi_0](s_n') \end{pmatrix} = \begin{pmatrix} \psi_0(s_1) \\ \vdots \\ \psi_0(s_n) \end{pmatrix}, \tag{3.9}$$

where s_i parametrizes Γ_i.

3.2.2 Problems with the method

Let us consider how best to solve Eq. (3.7). If an analytic solution were possible, it would contain information about the structure in $T[\psi_0]$ at infinitely fine scales. From a numerical standpoint we are only interested in the structure of $T[\psi_0]$ down to some finest scale, or equivalently in Fourier components less than some cutoff frequency. The contributions from frequencies above this cutoff are easily eliminated by projecting Eq. (3.7) onto a finite dimensional function approximation space. For the purposes of this discussion we will use the simple projection described in the introduction, which consists of replacing

[6] It is possible to keep λ finite; the resulting walls are semitransparent. See [13] for applications.

the continuous curves with their one-dimensional lattice equivalents. In practice the lattice spacing is determined by how close to Γ we wish to evaluate the scattered wave function. However, we must be careful; according to Nyquist's theorem, all of the information in frequencies above one half the discrete sampling rate is aliased into the low frequency components. Thus there is some minimal sampling rate that ensures that this aliasing is negligible, to the numerical precision at which we work.

Now let us consider what happens to this sampling rate when Γ has a discontinuity. This could be as simple as the discontinuity associated with a sharp edge; it could be the discontinuity in the slope at a sharp corner; or it could even be a discontinuity in the curvature, as occurs at the matching points between the endcaps and straight segments in a stadium billiard. The description of a discontinuity in Fourier space requires infinitely high frequencies—no matter how close we get to the discontinuity, it still looks like discontinuous—so scattering from Γ will induce structure in $T[\psi_0](s)$ at all possible scales near the discontinuity. This is the usual problem associated with diffractive scattering. The sampling rate will thus be greatly increased near the discontinuity, and since we are using a uniform discretization, this will increase the sampling rate globally.

The matrix representation of $G_0(s, s')$ that results from this discretization is dense, and thus the complexity of solving Eq. (3.7) scales as $O(N^3)$, where N is the size of the discretization. When Γ has one or more discontinuities, this severely limits the size of the scattering system that we can handle numerically. Morever the matrix is poorly conditioned, so fine-scale structure in $T[\psi_0]$ is easily lost. The density of the matrix is a result of the extent of the basis elements in the space-frequency plane. They are exactly localized in space and completely delocalized in frequency, and thus the interactions at all scales are uniformly distributed over all of the matrix elements. The poor-conditioning is primarily due to the nature of the operator G_0 itself. G_0 is of order -1, meaning that it increases the number of continuous derivatives of the function on which it operates, i.e., it is a smoothing operator. If we express $T[\psi_0]$ in terms of the eigenvectors of $G_0(s_i, s_j)$, then those eigenvectors corresponding to the fine-scale structure of $T[\psi_0]$ will have near zero eigenvalues.[7] But again the extent of the basis elements does not help us. Most preconditioners work by transforming a matrix to form close to the identity, and because the matrix representation

[7] Note that operators of negative order reverse the usual correspondence between the size of the eigenvalues and the characteristic scale of the structure in their corresponding eigenvectors. This means that the relaxation techniques commonly used in multigrid methods cannot be applied to these operators [6].

of G_0 is dense, it is initially quite far from the identity. We might think that if we were to perform a discrete Fourier transform on $G_0(s, s')$, and thus separate the information into different frequencies, we would have a sparser, more stable representation of the scattering. However, because the Fourier basis functions are defined over the entire curve, the transformed matrix is still as poorly behaved.[8]

3.3 Wavelet Analysis

3.3.1 Multiresolution analysis

The lattice basis used above is an example of a single-scale basis; all of the basis elements have the same extent in the space-frequency plane, up to spatial translation. In keeping with the standard conventions of wavelet analysis, we will assume that there are 2^j such basis elements per unit spatial interval in our initial approximation space, and we will refer to this space as V_j. The projection $f^{(j)} = \mathcal{P}_j f$ of a continuous function f into V_j contains all of the structure in f down to scale 2^{-j}. The problem is that all of this structure is encoded in terms of basis elements at scale 2^{-j}, independent of the scale at which the structure naturally lives. We would like to decompose V_j into $V_{j-1} \oplus W_{j-1}$ and $f^{(j)}$ into $f^{(j-1)} + d^{(j)}$, where $f^{(j-1)}$ contains the structure in f down to the scale 2^{-j+1} and $d^{(j)}$ is just the structure at scale 2^{-j}. Furthermore we would like to be able to iterate this decomposition, and decompose V_j into $V_{j'} \oplus W_{j'} \oplus \cdots \oplus W_{-j+1}$ and $f^{(j)}$ into $f^{(j')} + d^{(j')} + \cdots + d^{(-j+1)}$, where j' defines some coarsest scale.

A multiresolution analysis (MRA) is the framework that allows us to do this. We begin by defining the properties of an MRA of the space of functions $L^2(\mathrm{R})$. An MRA consists of a sequence of nested subspaces $\{V_j : \forall j \in \mathrm{Z}\}$ and a function $\varphi \in V_0$, called the *scaling function*, satisfying the following properties: (1) $V_j \subset V_{j+1}$ for all $j \in \mathrm{Z}$; (2) $\cap_{j\in\mathrm{Z}} V_j = \{0\}$ and $\cup_{j\in\mathrm{Z}} V_j = L^2(\mathrm{R})$; (3) if $f(x) \in V_j$, then $f(2^{-j}x) \in V_0$; (4) if $f(x) \in V_0$, then $f(x-n) \in V_0$ for all $n \in \mathrm{Z}$; and (5) the integer translates of $\varphi(x)$, $\{\varphi(x-n) : n \in \mathrm{Z}\}$, form an *orthonormal* basis of V_0. The V_j should be thought of as successively refined approximation spaces for the full space of functions, $L^2(\mathrm{R})$ in this case. Property (3) shows that these spaces are dyadic scalings of one another, and properties (3)–(5) imply that the sets of dyadic scalings and translates of φ given by $\{\varphi_{j,k}(x) = 2^{j/2}\varphi(2^j x - k) : k \in \mathrm{Z}\}$

[8] The author has recently learned that it is possible to precondition the matrix representation of $G_0(s, s')$ using a windowed rather than global discrete Fourier transform [33].

provide orthonormal bases for the V_j—thus the multiresolution aspect of the analysis. We also define an auxiliary sequence of subspaces, $\{W_j : j \in \mathrm{Z}\}$, where W_j is the orthogonal complement of V_j in V_{j+1}. Note that the definition of the W_j implies that they are mutually orthogonal. These spaces contain the structure at scale 2^{-j} that cannot be written in terms of basis elements at scale 2^{-j+1}. As with the V_j, orthonormal bases for these spaces can be written as the dyadic scalings and translations of a single function ψ, called the "mother wavelet;" and since $W_j \subset V_{j+1}$, ψ can also be written in terms of φ. Because of their physical meanings, the V_j are called the *average* spaces, and the W_j are the *detail* spaces. The coefficients of a function f projected into one of these spaces are given either by $\lambda_{j,l} =< \varphi_{j,l}, f >$ or $\gamma_{j,l} =< \psi_{j,l}, f >$.

The simplest example of an MRA is the Haar MRA. The spaces V_j^{Haar} consist of all functions piecewise continuous on the intervals $[l \cdot 2^{-j}, (l+1) \cdot 2^{-j})$, and the scaling function and wavelet are given by

$$\varphi^{\mathrm{Haar}}(x) = \begin{cases} 1 & 0 \le x < 1 \\ 0 & \text{otherwise} \end{cases} \quad \text{and} \quad \psi^{\mathrm{Haar}}(x) = \begin{cases} 1, & 0 \le x < 1/2 \\ -1, & 1/2 \le x < 1 \\ 0, & \text{otherwise} \end{cases} . \tag{3.10}$$

These functions were actually introduced in 1910 to study the approximation properties of piecewise continuous functions, but their place in the more general setting of multiresolution analysis was not realized until much more recently.

Now let $f^{(j)}$ be the projection of a function f into the space V_j. We would like to decompose $f^{(j)}$ into a projection $f^{(j')}$ at some coarser scale j', and a sum of corrections $d^{(j')} + \cdots + d^{(j-1)}$, one in each intermediate detail space. To do this, we need to know the relationship between the dyadic scalings of φ at adjacent scales, as well as the expression for ψ in terms of φ. Since $V_0 \subset V_1$, the scaling function can be written as

$$\varphi(x) = \sum_n h_n \varphi_{1,n}(x) = \sqrt{2} \sum_n h_n \varphi(2x - n), \tag{3.11}$$

for some set of real coefficients $\{h_n\}$. This equation is referred to as the *two-scale* relation for φ, and it determines φ up to a constant. We also have $W_0 \subset V_1$, so the mother wavelet can be written as

$$\psi(x) = \sum_n g_n \varphi_{1,n}(x) = \sqrt{2} \sum_n g_n \varphi(2x - n), \tag{3.12}$$

for some set $\{g_n\}$. As with φ, this equation determines ψ up to a constant. Using these two expression, we find that the average and detail coefficients of $f^{(j)}$ in V_{j-1} and W_{j-1} are

related to its average coefficients in V_j as

$$\lambda_{j-1,l} = \sum_n h_{n-2l}\lambda_{j,l}, \tag{3.13}$$

$$\gamma_{j-1,l} = \sum_n g_{n-2l}\lambda_{j,l}. \tag{3.14}$$

Eqs. (3.13) and (3.14) provide us with the desired map from $f^{(j)}$ to $f^{(j-1)} + d^{(j-1)}$. A single iteration of these equations defines the discrete wavelet transformation (DWT). By construction this transformation is invertible. The equation for the inverse is given by

$$\lambda_{j,l} = \sum_n h_{l-2n}\lambda_{j-1,l} + \sum_n g_{l-2n}\gamma_{j-1,l}. \tag{3.15}$$

Let us now consider the tilings of the space-frequency plane induced by these different decompositions of $f^{(j)}$. When we initially project a function f into V_j, $f^{(j)}$ is expressed in terms of basis functions with support approximately equal to a rectangle of width 2^{-j} and height 2^j in the space-frequency plane. After one iteration of the DWT, $f^{(j)}$ is expressed in terms of functions that are twice as wide and half as high. These tilings are illustrated in Figs. 3.1(a,b). In the case of compactly supported bases, the vertical lines are exact, but the horizontal lines are still only approximate since a function cannot be compactly supported in both space and frequency [11]. If we continue this decomposition, then after four iterations of the DWT, we have tiling in Fig. 3.1(c). From these figures we see that the coefficients $\{g_n\}$ and $\{h_n\}$ can be interpreted as the coefficients of high-pass and low-pass filters g and h, respectively. This interpretation will prove useful in Sec. 3.3.4 when we try to extend wavelets to bounded domains.

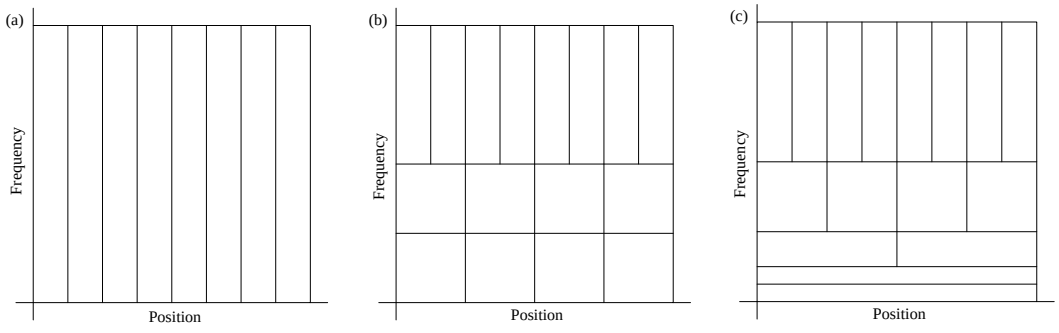

Figure 3.1: Tilings of the space-frequency plane induced by the bases of (a) V_j, (b) $V_{j-1} \oplus W_{j-1}$, and (c) $V_{j-3} \oplus W_{j-3} \oplus W_{j-2} \oplus W_{j-1}$.

One of the problems with an MRA built from a single function φ is that if the wavelet transformation is to be invertible, then the scaling function and wavelet cannot be symmetric, except in the trivial case of the Haar wavelet [11]. Thus when we project a function f into V_j, we introduce an asymmetry into the function that was previously not there. We can avoid this by using pairs of scalings functions and wavelets $\{\varphi, \tilde{\varphi}, \psi, \tilde{\psi}\}$, along with their associated filters $\{h, \tilde{h}, g, \tilde{g}\}$, in order to construct bases for our approximation spaces [9]. The dyadic scalings and translates of φ and ψ still provide complete bases for the $\{V_j\}$ and $\{W_j\}$, but they are no longer orthonormal. Rather we have $< \tilde{\varphi}_{j,k}, \varphi_{j',k'} >= \delta_{j,j'}\delta_{k,k'}$, as well as $< \tilde{\psi}_{j,k}, \psi_{j',k'} >= \delta_{j,j'}\delta_{k,k'}$; i.e., φ and $\tilde{\varphi}$ are biorthogonal, as are ψ and $\tilde{\psi}$. The projection operators into V_j and W_j are now written as $\mathcal{P}_j(\cdot) = \sum_l < \tilde{\varphi}_{j,l}, \cdot > \varphi_{j,l}$ and $\mathcal{Q}_j(\cdot) = \sum_l < \tilde{\psi}_{j,l}, \cdot > \psi_{j,l}$. The functions $\tilde{\varphi}$ and $\tilde{\psi}$ are referred to as dual functions. Note that $\tilde{\varphi}$ and $\tilde{\psi}$ can be used as bases for a second sequence of approximation spaces $\{\tilde{V}_j\}$ and $\{\tilde{W}_j\}$. In this case φ and ψ are used to build the projection operators.

The simplest example of a biorthogonal MRA is provided by a trivial generalization of the Haar MRA. The V_j remain the same, $\tilde{V}_j = V_j$ for all j, and $\tilde{\varphi} = \varphi$ and $\tilde{\psi} = \psi$. The actual MRAs used in this work are described in Sec. 3.3.4.

3.3.2 Calculation of wavelet coefficients

We now discuss how to compute the initial projection of a function f into V_j. There is, of course, a straightforward recipe for this projection: compute all of the integrals, $\lambda_{j,l} = 2^{j/2} \int dx\ f(x)\varphi(2^j - l)$. However, most scaling functions do not have closed analytic expressions, so these integrals must be computed numerically. The numerical values of φ at the dyadic rationals can be computed using the "cascade algorithm" [11], but if we are interested in the projections of two-dimensional operators, then the computational resources needed to compute these integrals numerically can dominate the problem. There is an alternative to numerical quadrature if the dual scaling function has a nonzero number of vanishing moments. If $\tilde{\varphi}$ satisfies $\int dx\ x^m \tilde{\varphi}(x) = 0$ for $m = 1, 2, \ldots, N-1$,[9] then we have a simple one-point quadrature formula,

$$\lambda_{j,l} = 2^{j/2} \int dx\ f(x)\tilde{\varphi}_{j,l}(x) = 2^{j/2} f\left(2^{-j}l\right) + \mathcal{O}\left(2^{-j(N+1/2)}\right). \tag{3.16}$$

The second equality follows by shifting $x \to x + 2^{-j}l$ and then Taylor expanding f around $x = 2^{-j}l$. The $\lambda_{j,l}$ computed using this formula are referred to as the "empirical" expan-

[9] The scaling function must satisfy $\int dx\ \varphi(x) = 0$.

sion coefficients, while those computed by some higher order quadrature formula are the "theoretical" coefficients. The practical difference between these is found in the behavior of the projected function $f^{(j)} = \sum_l \lambda_{j,l}\varphi_{j,l}(x)$. If we use theoretical coefficients, then we may sample $f^{(j)}(x)$ at scales finer than the distribution of scaling functions. If we use the empirical, we may not.

3.3.3 Sparse realization of functions

Once we have the single-scale decomposition of a function f in terms of the $\lambda_{j,l}$, we may construct a multiscale representation using the DWT. The wavelet coefficients $\gamma_{j',l}$ for $j' < j$ encode the detail structure in f at scale $2^{-j'}$, so if we are to have an efficient representation of f, we want most of these to vanish. Now the $\gamma_{j',l}$ may also be computed directly as $\gamma_{j',l} = \int dx\ f(x)\tilde{\psi}_{j',l}(x)$. If we assume that $\int dx\ x^m\tilde{\psi}(x) = 0$ for $m = 0, 1, 2, \ldots, N-1$, then for functions that are locally well approximated by a polynomial, most of these coefficients will be negligible. Another way of saying this is that φ and its translates exactly reproduce polynomials up to order $N-1$ at all scales. N is called the order of the wavelet family.

We also want our wavelets to satisfy $\int dx\ x^m\psi(x) = 0$ for $m = 0, 1, 2, \ldots, \tilde{N}-1$. This does not increase the sparsity of the representation, but it does have the following consequence. A function f can be represented as $\sum_l \lambda_{j,l}\varphi_{j,l} + \sum_l \gamma_{j,l}\psi_{j,l}$ for any j. If the first $\tilde{N}$ moments of ψ vanish, then the first $\tilde{N}$ moments of f are encoded in the average coefficients. This is in keeping with the interpretation of the V_j as containing the average global properties of f, while the W_j contain the local details.

3.3.4 Lifting scheme

The "lifting scheme," discovered recently by Sweldens [69, 70, 10], allows the iterative construction of any set of finite, biorthogonal filters beginning with the trivial filters defined by the "Lazy wavelet" transformation. Since we know that the scaling functions, wavelets, and their duals are all determined up to constants once the complete set of filters $\{h, g, \tilde{h}, \tilde{g}\}$ is known, we can use this scheme to construct wavelets and dual wavelets with an arbitrary number of vanishing moments and dual moments, respectively.

The construction of filters in the lifting scheme is done by considering the action of the DWT on the average coefficients of some projected function. The filters defined

by the Lazy wavelet have an all but trivial action: they subsample these coefficients into their even and odd members. The problem with these filters is that, according to Nyquist's theorem, they are subject to a tremendous amount of aliasing. We thus want to augment their action to reduce this aliasing. We know from Sec. 3.3.3 that if the average coefficients alone are able to reproduce polynomials up to order $N-1$, the high-pass filter should be well defined. Let us assume that the coefficients $\lambda_{j,l}$ are the discrete samples of an order $2n-1$ polynomial. After one iteration of the Lazy wavelet transformation, the detail coefficients $\gamma_{j-1,l}$ are all nonzero. However, we know that their values can be predicted using Neville's algorithm for polynomial interpolation on $2n$ of the surrounding $\lambda_{j-1,l}$. Thus if we subtract the $\lambda_{j-1,l}$ from each of the $\gamma_{j-1,l}$ weighted by some set of coefficients $\{\tilde{s}_k\}$, which can depend on the particular $\gamma_{j-1,l}$, we can zero out all of the $\gamma_{j-1,l}$. [10] When the $\lambda_{j,l}$ are the samples of some more general function, the new $\gamma_{j-1,l}$ will still be nonzero, but will represent only that part of the function that cannot be locally represented as a polynomial of order $2n-1$. This is subtraction is called the *predict* phase of the construction; we write the new detail coefficients as $\gamma^{\mathrm{pr}}_{j,l}$. We also know that the low-pass filter should be well defined if the first $2n'$ moments of a projected function can be written in terms of its average coefficients. After the initial Lazy wavelet transformation, this cannot be true because we lose half of the average coefficients. However, it is possible to restore the moments by updating each $\lambda_{j-1,l}$ with contributions from $2n'$ of its surrounding $\gamma^{\mathrm{pr}}_{j-1,l}$ weighted by a second set of coefficients $\{s_k\}$, which again can depend on the particular $\lambda_{j-1,l}$. This is the *update* phase of the construction.

The actual filters generated by these simple transformations can be constructed using the equations in [69]; it is found that they correspond to the $(N,\tilde{N}) = (2n, 2n')$ family of compact wavelets first discovered in [9]. However, there is no need for this extra step. We have managed to decompose a transformation that usually would have to be performed as a convolution product into a series of steps that can performed *in place*. Moreover in this form, many of the characteristics of the transform are immediately obvious: the transformation is invertible—we just reverse the predict and update steps, and change the signs of all of the coefficients; if both n and n' are finite, the basis functions are compact; and if the coefficients used to predict (update) the $\gamma_{j,l}$ ($\lambda_{j,l}$) are symmetrically distributed, the basis functions are symmetric. This construction is also easily extended to finite domains: when

[10] The procedure for constructing the required $\{\tilde{s}_k\}$ can be found in [22].

we are near one of the boundaries, we simply shift our choice of coefficients used for the predict and update steps.

In this work we will restrict ourselves to these families. The complete set of functions for the $(2,2)$, $(4,2)$, and $(6,2)$ families are shown in Figs. 3.2 - 3.4.

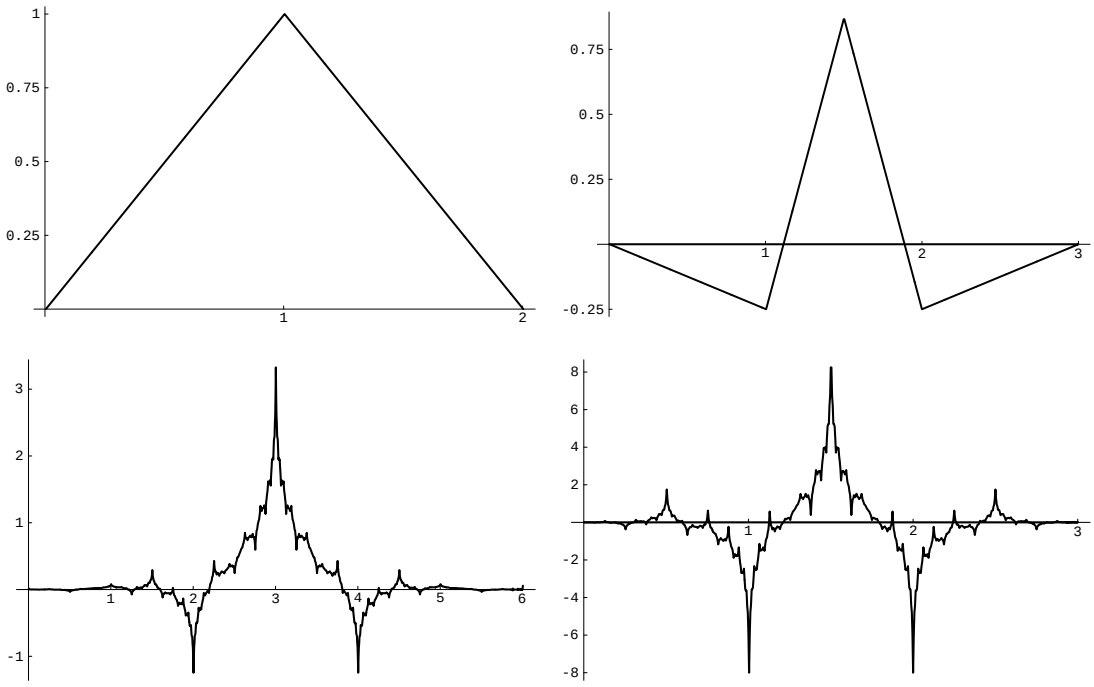

Figure 3.2: The functions (a) φ, (b) ψ, (c) $\tilde{\varphi}$, and (d) $\tilde{\psi}$ for the $(2,2)$ family of wavelets.

3.4 Solving the boundary wall equations

3.4.1 One-dimensional T-matrix

We begin by projecting the elements of Eq. (3.7) onto the basis of scaling functions at some finest resolution.

$$G_0(s,s') = \sum_{l,l'} \tilde{\varphi}_{j,l}(s)\,(G_0)_{j,l;j,l'}\,\tilde{\varphi}_{j,l'}(s'), \tag{3.17}$$

$$T[\psi_0](s) = \sum_{l} (T[\psi_0])_{j,l}\,\varphi_{j,l}(s), \tag{3.18}$$

$$\psi_0(s) = \sum_{l} (\psi_0)_{j,l}\,\tilde{\varphi}_{j,l}(s). \tag{3.19}$$

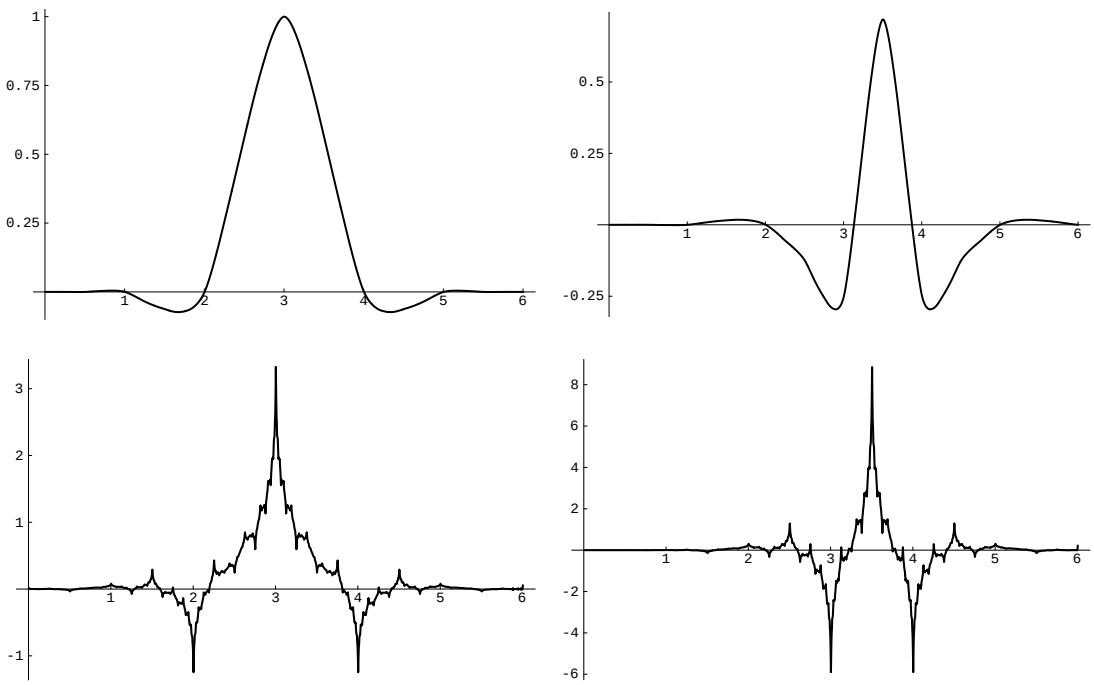

Figure 3.3: The functions (a) φ, (b) ψ, (c) $\tilde{\varphi}$, and (d) $\tilde{\psi}$ for the $(4,2)$ family of wavelets.

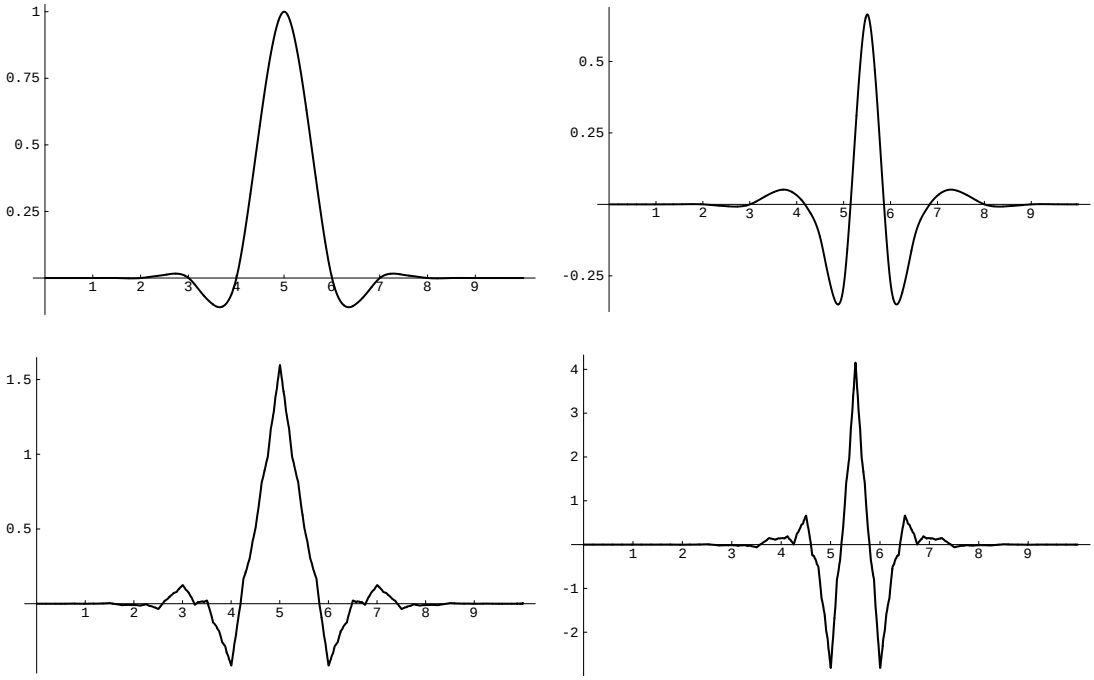

Figure 3.4: The functions (a) φ, (b) ψ, (c) $\tilde{\varphi}$, and (d) $\tilde{\psi}$ for the $(6,4)$ family of wavelets.

The actual resolution used is found by requiring that the wave function, or some physical quantity derived from it, does not change to some fixed precision as we increase the size of the discretization. Inserting these expansions into Eq. (3.7), and using the biorthogonality condition on φ and $\tilde{\varphi}$, we find the linear system

$$\sum_{l'} \left(G_0\right)_{j,l;j,l'} \left(T[\psi_0]\right)_{j,l'} = \left(\psi_0\right)_{j,l} . \tag{3.20}$$

Note that we have chosen to expand $T[\psi_0]$ in φ. As can be seen from Figs. 3.2–3.4, φ is quite regular compared to $\tilde{\varphi}$. For reasons to be discussed in Sec. 3.5.4, we need to sample $T[\psi_0]$ at a resolution finer than the distribution of scaling functions, so we would rather expand it in terms of regular functions. This also makes the computation of the coefficients of G_0 and ψ_0 easier; since we are not able to use the one-point quadrature formula, we would rather numerically integrate against φ than $\tilde{\varphi}$.

We chose to work with the $(2,2)$ family of biorthogonal wavelets since these have the smallest support. We checked our results by performing some of the computations using the higher order $(4,2)$ and $(6,4)$ families. In the case of the QPC, to be discussed in Sec. 3.5, we found that the conductance did not change at the level of precision to which it was initially computed. The one-dimensional integrals determining the projection of ψ_0, as well as the two-dimensional integrals determining the projection of G_0 away from the diagonal, were computed using Richardson extrapolation based on the closed trapezoidal rule [58]. The two-dimensional integrals along the diagonal were also computed using extrapolation, but with an open trapezoidal rule to avoid the diagonal singularity in G_0. In both cases the number of refinements used by the Richardson algorithms was determined by demanding that the integrals converge to a fixed precision. As with the choice of wavelet family, the actual precision was chosen by demanding that an increase in precision did not change the conductance of the QPC at the initial precision.

Once we have the projection of Eq. (3.20) onto the scaling basis, we perform the maximum allowed number of discrete wavelet transformations. The resulting linear system is still dense, but many if not most of its coefficients are near zero. Following [5], we define a threshold parameter, $\tau = \alpha \cdot \max_i \sum_j |(G_0)_{i,j}|/N$, where N is the size of the discretization, and zero out all of the elements less than τ. The free parameter α can be used to vary the fraction of coefficients retained. As will be seen in Sec. 3.5, as many as 60% of the coefficients can be zeroed out when calculating the conductance of a QPC. After eliminating as many elements as possible, the resulting system can be solved using the routines provided by the

numerical library QMRPACK [23].

Keeping α in the definition of τ fixed and varying N, we find that the number of significant elements in the scattering matrix scales as $\mathcal{O}(N^\alpha)$ with $1 \lesssim \alpha < 2$. Now it was shown in [5] that the wavelet representation for operators of the form $K(x,y) = C/(x-y)^\beta$ for some $\beta > 0$, the so called Calderón-Zygmund operators, is sparse; i.e., the number of significant elements above a given threshold scales as $\mathcal{O}(N)$. Unfortunately our operator is not of this form. In the asymptotic regime,

$$G_0\left(\mathbf{r}(s), \mathbf{r}(s')\right) \sim \sqrt{\frac{2}{\pi}} \frac{e^{ik|\mathbf{r}(s)-\mathbf{r}(s')|-i\pi/4}}{\sqrt{k|\mathbf{r}(s)-\mathbf{r}(s')|}}. \tag{3.21}$$

The presence of a nonzero global frequency in G_0 invalidates the proof in [5]. The reason for this breakdown can be seen in Fig. 3.1: wavelets are only efficient at representing global structure near zero frequency. In Sec. 3.6 we suggest that it may be possible to achieve better sparsity if we use wavelet *packets* instead of pure wavelets.

3.4.2 Two-dimensional wave function

Once the solution for $T[\psi_0]$ is found in the wavelet basis, it can be reexpressed in the single-scale basis using the inverse DWT. Since we did not use the one-point quadrature formula, the numerically computed scaling functions can then used to construct an expression for $T[\psi_0]$ in a point-like basis at a resolution finer than the initial distribution of basis elements. The values of the two-dimensional wavefunction can be computed using a simple discretization of Eq. (3.8)

$$\psi(\mathbf{r}) = \psi_0(\mathbf{r}) + \sum_{i\in\Gamma} G_0\left(\mathbf{r}, \mathbf{r}(s_i)\right) T[\psi_0](s_i), \tag{3.22}$$

where the sum is performed using Richardson extrapolation based on the closed trapezoidal rule. As a general rule we should not compute, or at least should not believe the computed values of, the two-dimensional wavefunction on lattice points closer to one of the boundaries than the distance between points along the boundary where we compute $T[\psi_0]$. The probability current associated with $\psi(\mathbf{r})$ is given by $\mathcal{J}(\mathbf{r}) = \Im\{\psi(\mathbf{r})^\dagger \nabla_{\mathbf{r}} \psi(\mathbf{r})\}$. Since the derivative of $G_0\left(\mathbf{r}, \mathbf{r}(s_i)\right)$ with respect to $\mathbf{r}$ is simply $(i/4)H_1^{(1)}(2\pi|\mathbf{r}-\mathbf{r}(s_i)|/\lambda)$, $\nabla_{\mathbf{r}}\psi(\mathbf{r})$ can be computed in the same way as $\psi(\mathbf{r})$ itself.

Note that there are two different length scales involved in this reconstruction: the distance between scaling functions and the distance between points at which we compute

$T[\psi_0]$. When we perform our initial projection of Eq. (3.7) onto the scaling function basis, we are effectively taking into account all interactions down to the distance between basis elements. We should thus expect that our approximate solution contains all of the structure in the exact solution down to this scale. Since we perform our initial projection using nontrivial quadrature rules rather than the one-point formula, we may sample this solution below this scale. This is not to say that the increased sampling rate will give us any new information at scales not present in the initial projection, but just that we can sample a continuous function at any scale we choose. For example we may sample the function $\cos\omega$ at a rate finer than ω^{-1}, but this will not reveal the presence of any other frequencies.

3.5 Numerical Simulations

In this section we present the results of the boundary wall method applied to a number of scattering geometries.

3.5.1 Straight walls

The modulus of the wave function describing the scattering of a plane wave from an (effectively) half-infinite line is shown in Fig. 3.5. This wave function was computed by keeping only 75% of the coefficients in the scattering matrix.

3.5.2 Arc segments

Figs. 3.6 shows the moduli of the wave functions describing the scattering of a plane wave from half circles of two different radii. In both cases, the wave functions were computed keeping only 60% of the coefficients in the scattering matrix. The pattern seen near the horizontal axis in each of these figures is described by the Pearcey function [4, 68]. Because of spherical aberration, the parallel classical trajectories corresponding to the incoming plane wave do not focus to single point, but rather form a cusp [4]. The Pearcey function describes the universal part of the diffractive correction to the semiclassical wave function near the cusp.

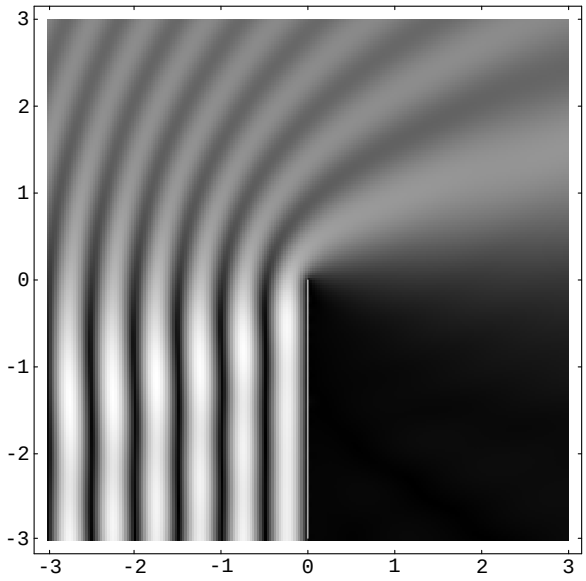

Figure 3.5: The modulus of the wave function describing the scattering of a plane wave normally incident on a half-infinite line. The wavefunction satisfies a Dirichlet boundary condition along the line. Units along both axes are measured in wavelengths.

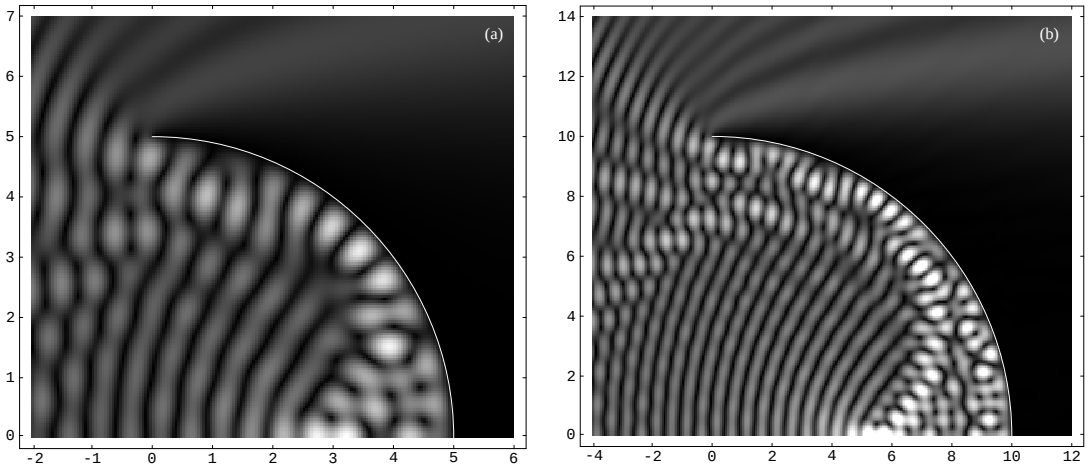

Figure 3.6: The modulus of the wave function describing the scattering of a plane wave incident on a half circle of radius (a) 5λ and (b) 10λ. Units along both axes are measured in wavelengths.

3.5.3 QPC

Fig. 3.7 shows the conductance of the QPC as a function of width computed both numerically and using the exact expansion of the wavefunction in terms of Mathieu functions [3]. The numerical conductance curve in this figure was computed using a semicircular "detector" of radius 1.5λ centered at the QPC. The conductance of such a detector is defined as the integral of the normal component of the current over the detector. The two curves agree quite well, so we can be confident that our method for computing the conductance is sound.

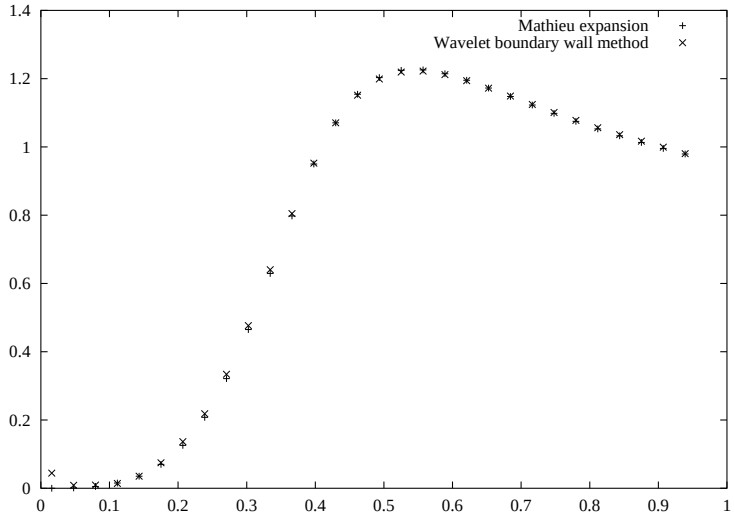

Figure 3.7: Comparison of the conductance of a QPC as a function of width computed using the boundary wall method and using the exact expansion of the wave function in terms of Mathieu functions. The width of the QPC is measured in wavelengths, and the conductance is the ratio of the quantum mechanical conductance to the classical conductance.

3.5.4 Westervelt gate

The Westervelt gate is illustrated in Fig. 3.8. It consists of a QPC coupled to an open resonator formed by the walls that define the QPC and a quarter circle mirror. In all of our simulations, the mirror's radius of curvature is taken to be 5λ, the QPC is $\lambda/4$ wide, and the walls that define the QPC are each 40λ long. In Figs. 3.9 we show partial numerical conductance traces computed using the semicircular detector described in the

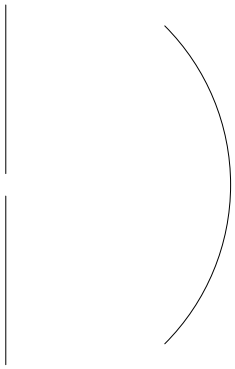

Figure 3.8: Schematic illustration of the Westervelt gate.

previous section. Each peak was fit with a function of the form

$$\frac{A + B(E - E_n)}{(E - E_n)^2 + (\Gamma_n/2)^2} + C \tag{3.23}$$

where the linear term in the denominator represents the interference between the usual resonance term $1/(E - E_n + i\Gamma_n/2)$ and a complex background phase shift. The sharp

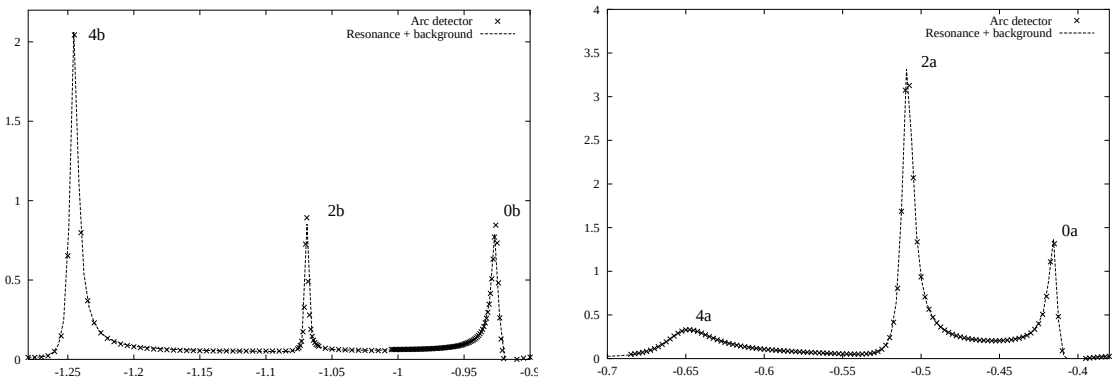

Figure 3.9: Partial conductance traces for the Westervelt gate as function of the separation between the QPC and the mirror's center of curvature. The separation is measured in units of wavelength, and the conductance is normalized to the conductance of the QPC with no mirror present.

peaks seen in Fig. 3.9 correspond to resonant conductance through the device, mediated by one of the quasistates of the resonant cavity.

In Fig. 3.11 we compare the conductance traces near the 2_a peak computed with a linear detector in the QPC itself and the 1.5λ radius arc detector. The close agreement

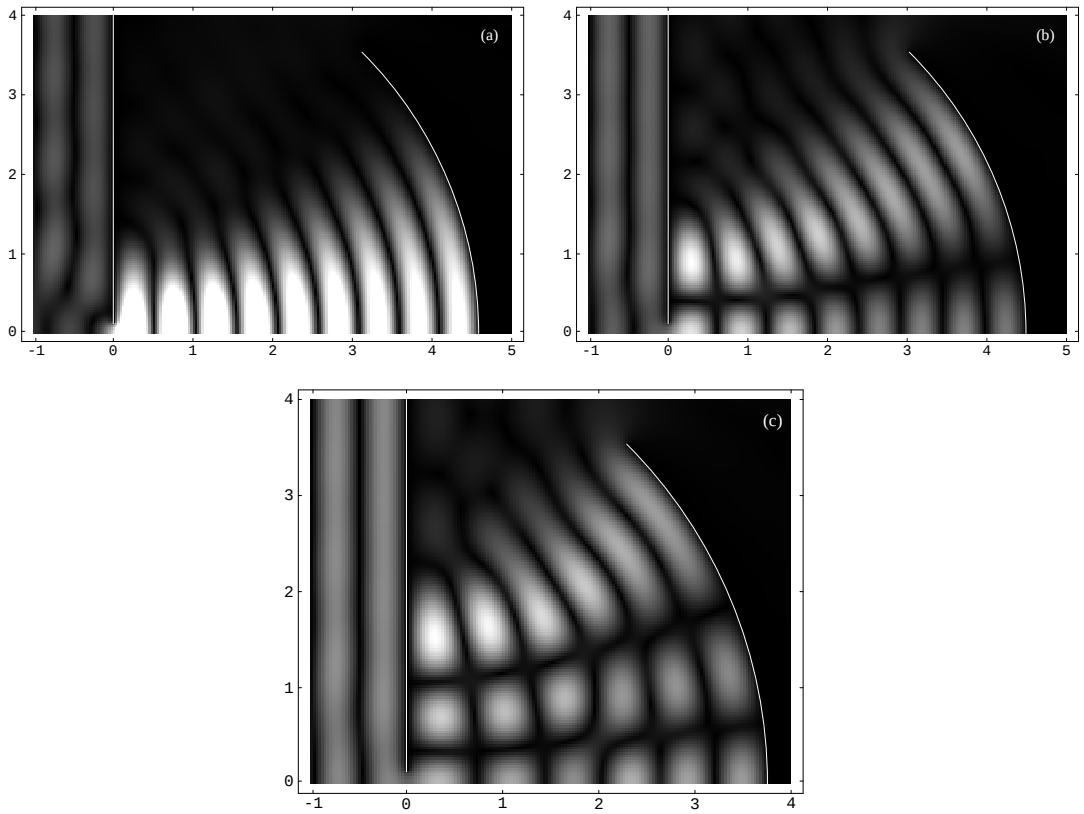

Figure 3.10: Moduli of the wave functions associated with peaks (a) 0_a, (b) 2_a, and (c) 4_b in Fig. 3.9.

between the curves in these two figures is a numerical demonstration of both the unitarity of our method, and its ability to compute fine scale interactions in the presence of much larger scales.

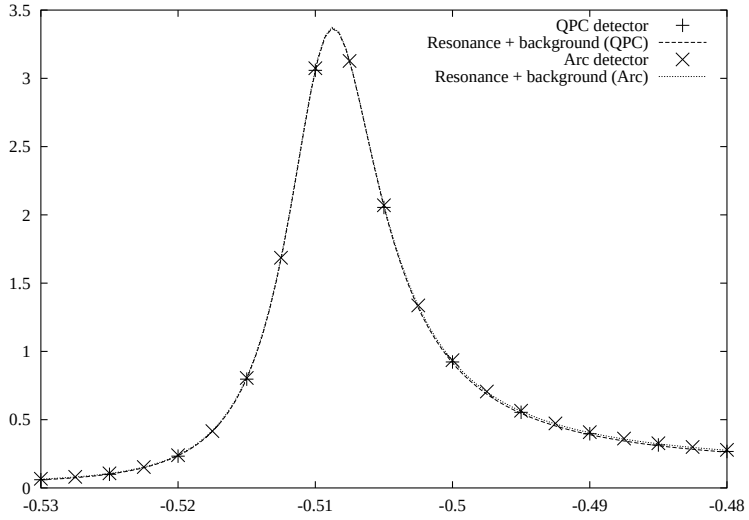

Figure 3.11: Comparison between linear detector in the QPC and arc detector.

Finally in Fig. 3.12 we see the effect on the conductance of different thresholds on the scattering matrix. We are able to discard all but 40% of the elements and still compute the conductance two within 1% of its exact value.

3.6 Future Work

3.6.1 Extensions

In Sec. 3.4 we noted that because of the nonzero global frequency in G_0, it is impossible to achieve strictly scattering sparse matrices using wavelets. Wavelets were initially developed to represent the local frequency content of a function. It was always assumed that the any global structure present would be at zero frequency. This is reflected in the structure of the DWT. Conventionally the n-th iteration of a multilevel wavelet transformation is applied to the output of the low pass filter from the $(n-1)$-th iteration. Because of this bias, the highest resolution in frequency space will always be found near zero frequency. For a signal dominated by a nonzero frequency, the wavelet representation will thus be far from optimal. We can remove this bias by sometimes applying the transformation

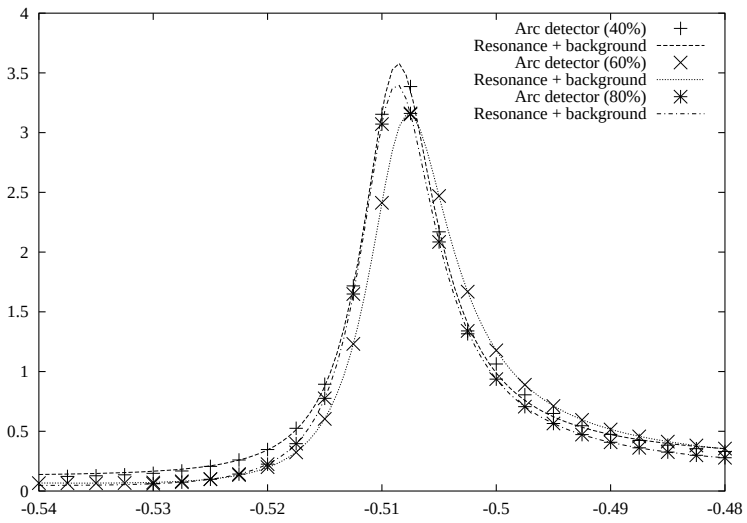

Figure 3.12: Conductance curves computed for various thresholds.

to the output of the high pass filter. This allows us to shift the optimal resolution to any desired frequency. The basis functions associated with this modified transformation can be written as linear combinations of ordinary wavelets, and are thus called wavelet packets. A recent study of two-dimensional electromagnetic scattering from obstacles with sharp edges using wavelet packet compression found that the number of nonzero elements above a given threshold parameter scaled as $\mathcal{O}(N^\alpha)$, with $\alpha \sim 1.3$ [14]. As a general rule it is impossible to achieve $\alpha = 1$ in the presence of the sharp corners even using wavelet packets, because the dominant frequency changes as we approach the corners.

Besides the implementation of wavelet packets, there are a number of other improvements that we are making to the method [18]. The method described in this paper allows us to construct a representation of the scattering matrix having only $\mathcal{O}(N^\alpha)$ significant elements; but because we construct this matrix in a single-scale basis first, the overall complexity of the construction still scales as $\mathcal{O}(N^2)$. As might be expected, the initial filling of the dense matrix currently dominates the computational resources. However this is because each element requires on average a 33×33-point numerical quadrature. Preliminary tests indicate that the conductance of the Westervelt gate as measured with a semicircular detector should remain the same if we instead use the one-point formula. This would effectively eliminate the $\mathcal{O}(N^2)$ bottleneck in computational time, but there would still be a memory bottleneck. If we want to scatter from a device with a total boundary length

of 1000λ, and we need 10 scaling functions per wavelength, then the scattering matrix will have 10^8 double precision complex entries, and will require 3.2 gigabytes of memory. Such a device size is not unreasonable: in order compute the one millionth eigenstate of the stadium billiard, which is the current upper bound in stadium technology, we require a device of this size.

It would be to our great advantage if we could build the scattering matrix directly in the sparse representation. For the Calderón-Zygmund operators discussed in Sec. 3.4.1, this is indeed possible [5]. We can define an "oracle" function that gives us an upper bound on the size of each matrix element in the wavelet representation without ever computing them. The particular bounds derived in this paper do not hold for operators with global structure at nonzero frequencies, but it should be possible to derive new bounds for arbitrary wavelet packet representations. Besides eliminating the memory bottleneck, an oracle function would allow us to make our method adaptive. Currently, we begin our computation at some finest scale and work toward coarser scales. An adaptive method would begin by computing the significant matrix elements at the coarsest scale, and then add on detail elements only where needed. For a device such as the Westervelt gate, the coarse elements would initially eliminate the incident wave from behind the bounding walls, along with the waves diffracted around the outer edges of these walls; and then the fine-scale elements near the QPC would give rise to the detailed structure of the wave function inside the resonator. An adaptive method has the added advantage that it can used as part of a convergence loop. For example if we want to compute the conductance of the Westervelt gate, then we can compute it after each addition of fine-scale matrix elements, and continue to add corrections until it converges to some specified precision.

3.6.2 Physical interpretation

It would be interesting to develop a physical interpretation of the structure of the scattering matrix in the wavelet representation that would allow us to understand the importance of interactions at a given scale and between scales. Such an understanding might be helpful in implementing the adaptive scheme outlined above. It might also prove useful in developing a theory of object multiple scattering, in which ordinary multiple scattering theory is applied to the matrices that describe scattering from entire boundaries. It is known that the fine-scale wavelet coefficients near singularities have universal forms [34], so

we may be able to develop a *universal* form for the scattering matrices for boundaries with particular features, such as sharp edges, corners, or even higher order discontinuities. Along these same lines, it may even be possible to develop approximate, semiclassical forms for these matrices. Semiclassical theories of diffraction, such as the uniform extension of Keller's geometric theory of diffraction [42], describe the scattering from geometric singularities in terms of a number of universal "coefficients of diffraction." These coefficients are derived by considering the asymptotic expansion of the exact solutions for a small number of canonical problems, such as scattering from a wedge. The problem with these theories is that there are only a finite number of exactly soluble problems, and they do not always cover the cases of interest. Using wavelets, it may be possible to derive numerical forms for these coefficients.

Chapter 4

Diffractive Resonances in Open Mesoscopic Cavities

We present a computational and theoretical analysis of the electronic transport properties of a quantum point contact (QPC) coupled to an open resonator operated in its classically *unstable* regime. The dynamics of the resonator are such that it supports a single isolated, unstable periodic orbit, so naïvely we would expect to find a single periodic sequence of resonant conduction peaks associated with this orbit. However, we also find multiple sequences of anomalous peaks that cannot be associated with any known periodic orbits. An extension of the conventional semiclassical theory of transport in microstructures shows that these anomalous peaks are supported by resonances associated with "diffractive" periodic orbits—orbits beginning and ending at the QPC, constrained to have one or more of their points pass through either end of the mirror defining the resonator, but otherwise obeying Fermat's principle. Our analysis differs from other recent papers that have attempted to incorporate diffractive corrections into the trace formula in that here the corrections are of the same order in $\hbar$ as the classical contributions themselves. The transport properties of this device operated in its classically stable regime were also the focus of a recent combined experimental and computational study by Katine *et al.* (PRL **79** (24), 4806 (1997)). As in that paper, the numerical simulations presented here were performed using a wavelet packet based multiresolution scheme, which allowed us to deal properly with the sharp features and widely varying length scales present in the device.

4.1 Introduction

4.1.1 Motivation

With recent advances in the production of semiconductor heterostructures having ultralow impurity concentrations ($\sigma < x \times 10^{-y}\mathrm{cm}^2$), experimentalists are now able to study the behavior of two-dimensional electron gases (2DEGs) with mean free paths, l_{mean}, and phase coherence lengths, l_ϕ, many times their Fermi wavelengths, λ_f, at cryongenic temperatures. Unlike the diffusive transport found in most solid state devices, transport in these 2DEGs is dominated by coherent, ballistic electron motion; it is thus meaningful for them to speak about *complex amplitudes* rather than simple probabilities for propagation along well-defined classical paths, as well as interference phenomena among these paths. Along with the low impurity, high mobility samples available to them, advances in electron beam lithography have also allowed them to pattern nanoscale voltage gates onto these samples, with features sharp on the order of a wavelength. These gates can be arranged to form two-dimensional resonators with overall characteristic dimensions $l_{\mathrm{x,y}}$ satisfying $\lambda_\mathrm{f} < l_{\mathrm{x,y}} < l_{\mathrm{mean},\phi}$; that is, resonators that are many wavelengths across, in which coherence is still maintained after many traversals. The first inequality puts us squarely into the regime where $\hbar$ is small, and with the second we find ourselves in a position to study semiclassical approximations to the fully quantum theory of transport in these devices.

Most of the efforts, both experimental and theoretical, to explore this new found regime have concentrated on closed billiards coupled to the rest of the system through narrow, well-defined leads. The focus of these studies has been to check semiclassical predictions derived from the Gutzwiller trace formula [29] and its many relatives, with a particular concern for the effects of classical chaos on the quantum behavior. Two-dimensional billiards provide an especially convenient framework in which to study the latter question, since the nature of their classical dynamics is determined solely by the shapes of their boundaries. The commonly used circular quantum dot and the Bunimovitch stadium provide the simplest examples of integrable and chaotic systems, respectively.

But despite all of the advances that have made this regime accessible, it is still generally impossible to study the contributions of the classical orbits that contribute to a single quantum state. A step toward studying isolated states was taken in a recent combined experimental and computational study by Katine *et al.* [38] (see also [19]), in which the usual closed billiard was replaced by a Westervelt gate—a subwavelength quantum point contact

(QPC) coupled to an open resonator defined by the walls of the QPC itself and a quarter circle mirror (see Fig. 4.1); a sort of open lemon billiard [32]. This system is classically stable when the QPC to mirror separation is less than the mirror's radius of curvature ρ; otherwise it is unstable. In the stable regime it supports long-lived quasistates.

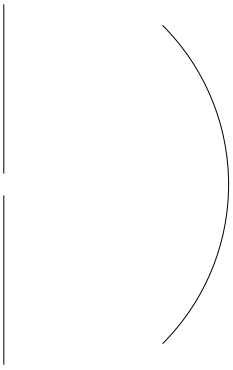

Figure 4.1: Schematic of the Westervelt gate.

The authors of [38] consider the behavior of the conductance of the gate in its stable regime. The first advantage of this device is that because of the strong coupling to the continuum on either side of the mirror, quasistates with only a finite number of angular nodes can be supported. The density of quasistates with widths sufficiently narrow to be considered states thus becomes effectively *one*-dimensional, decreasing the number of states that can potentially be mixed. The second advantage is that the existence of these states depends on the focusing ability the mirror, and this focusing is stable against the small perturbations to the mirror's shape that will undoubtedly be present because of imperfect fabrication. [1] This stability along with the small number of states allowed the authors of [38] to begin to see the signatures of individual classical orbits.

In the present study we consider what happens to the electronic transport properties of the Westervelt gate in the classically unstable regime. A quick analysis of the dynamics in this regime shows that there remains only a single isolated, unstable periodic orbit along the axis of symmetry. Thus we might expect the conductance to have a single sequence of periodically spaced peaks, associated with a sort of scarring along these orbits,

[1] To be more precise, because of spherical aberration the rays reflected from the mirror will not come to a pointlike focus, but rather will form a type of caustic known as a cusp; such caustics are stable against small perturbations of their defining surfaces. The exact form and positions of these caustics play an important role in structure of the quasistates in the unstable regime, as will be seen in Sec. 4.4.

and nothing more. In numerical simulations, we do indeed find these peaks, but only as the simplest part of a much richer spectrum. There are additional sequences of *anomalous* peaks, with heights and widths comparable to the expected peaks, whose positions and even number changes as we vary the angular size of the mirror. Perhaps even more surprising than the peaks themselves are the spatial wave functions associated with these peaks: they exhibit bands of enhanced amplitude running directly from the QPC to the very edges of the mirror. The existence of such states would not have been predicted from a naïve semiclassical analysis.

These anomalous peaks can be associated with a class of paths topologically distinct from paths found in conventional semiclassical analysis: paths that begin at the QPC, make some number of specular reflections before striking one of the ends of the mirror, and are then reflected back to the QPC. These new paths can be found from an ordinary stationary phase analysis with the additional constraint that one or more of their points pass through either end of the mirror. The existence of such paths has been known for quite some time in the electromagnetic scattering literature [41, 42], but their contributions have only recently been incorporated into semiclassical calculations [73, 59, 74, 75, 54, 65, 7]. In general paths that are constrained to pass through points where the classical dynamics exhibits a sharp discontinuity can be used to construct a geometric description of the the diffraction associated with the discontinuity. This is indeed their function here. The anomalous conduction peaks exist because strong diffraction off the ends of the mirror allows amplitude to return to the QPC, and thus a resonance to be established. Semiclassically this amplitude returns along the paths just described.

The peculiar nature of the spatial wave functions associated with these conduction resonances is due to the openness of the device, and will also be explained below.

4.1.2 Outline

We begin in Sec. 4.2 by simply presenting the numerical conductance traces in both the stable and unstable regimes of the Westervelt gate, and the wave functions associated with the nonanomalous and anomalous peaks in the latter regime. In Sec. 4.3 we show how the Landauer formula for the conductance of a closed mesoscopic device with a number of finite-width leads attached can be adapted to the Westervelt gate, and discuss the semiclassical limit of this formula. In Sec. 4.4 we study the classical dynamics of the gate

in the unstable regime needed to understand the nonanomalous peaks and their associated wave functions: we derive the stability indices of the one remaining periodic orbit in this regime, which allow us to understand the widths of the peaks, and recursion relations for the infinite sequence of cusps formed by reflection from the curved mirror, which allow us to understand the global structure of the wave functions. In Sec. 4.5 we provide a description of diffraction from a sharp edge in terms of classical paths, and show that near geometric reflection and transmission boundaries, the amplitudes associated with these paths can be of the same order as those for paths specularly reflected near the edge. In Sec. 4.6 we use these diffractive paths to develop a semiclassicaly understanding of the anomalous conductance peaks in the unstable regime. Finally in Sec. 4.7 we discuss whether the anomalous peaks are observable experimentally, and how this work fits in with other recent papers that incorporate the effects of strong diffraction.

4.2 Numerical Experiments

4.2.1 Numerical method

The phenomena described in this paper occur only when the QPC is less than half a wavelength wide. In this case the wave function transmitted by the QPC is essentially independent of the incoming source, up to normalization, so for simplicity we took the incoming wave function to be a normally incident plane wave. Following the method described in [13, 19], the physical gates defining the Westervelt gate were represented as one-dimensional Dirichlet boundaries, and the scattering equations were written as a system of coupled integral equations for the T-matrices along these boundaries. These equations were discretized using the $(2, 2)$ family of biorthogonal wavelets described in edw:scattering, which allowed us to deal with the multiple lengths scales and sharp edges of the Westervelt gate, and the resulting linear system was solved using an iterative solver. The two-dimensional wave function and its gradient were then found by integrating the T-matrix against the free Green's function and its derivative, respectively. The conductance of the gate was computed by integrating the normal component of the probability current $\mathbf{j} = \Im(\psi \nabla \psi)$ across a suitably defined numerical detector. Two types of detectors were used: a simple straight line across the QPC, and a semicircle of radius 1.5λ centered at the QPC. The agreement between the resulting curves ensured that our solutions were unitary.

4.2.2 Behavior in the stable regime

In Fig. 4.2 we see a sample numerical conductance trace from the stable regime. No additional damping was included in the computation of this trace, so the observed peak widths represent the coupling of the quasistates to the continuum. This coupling is purely quantum mechanical in nature; the quasistates can either tunnel out through the QPC or across the caustics that bound the classical motion inside the resonator. The wave functions

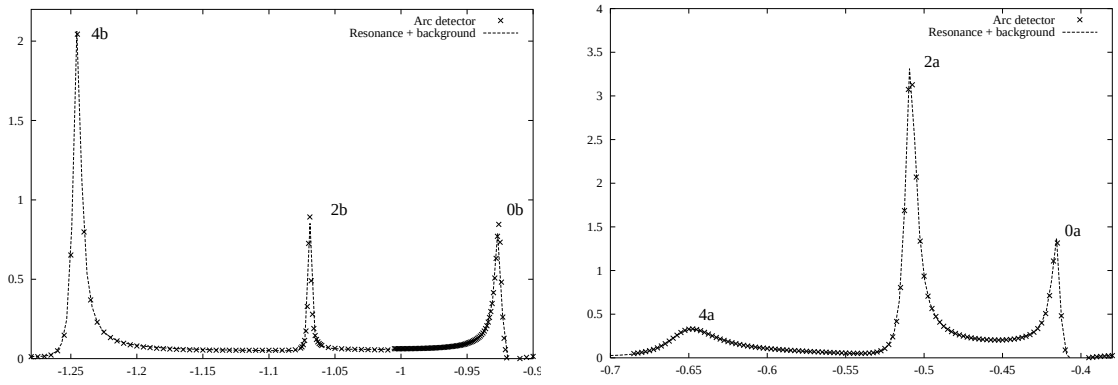

Figure 4.2: Partial conductance traces in the stable regime. The mirror's radius of curvature is 10λ, and its angular size is $\pi/2$. The horizontal axis measures the separation between the QPC and the mirror's center of curvature, in units of wavelength, and the conductance is normalized to that of the QPC alone.

associated with the sequence of peaks labeled 0_x have zero angular nodes, those with the sequence 2_x have two angular nodes, and so on. [2] Images of these quasistates can be found in [19]. The periodicity of the 0_x sequence is simply $\lambda_f/2$, as might have been expected from the one-dimensional character of the associated wave functions. The absolute position of the first peak in this sequence also follows from one-dimensional arguments, as long as we remember to include the Maslov phase associated with the cusp formed after reflection from the mirror. The periodicities and absolute starting positions of the higher order sequences can be computed from more complicated semiclassical arguments [17].

[2] Quasistates with an odd number of angular nodes are not excited in our device because they have zero overlap with the QPC.

4.2.3 Behavior in the unstable regime

Conductance curves

In Fig. 4.3 we see numerical conductance traces from the unstable regime for gates defined by mirrors of various angular sizes. Unlike those from the stable regime, these traces

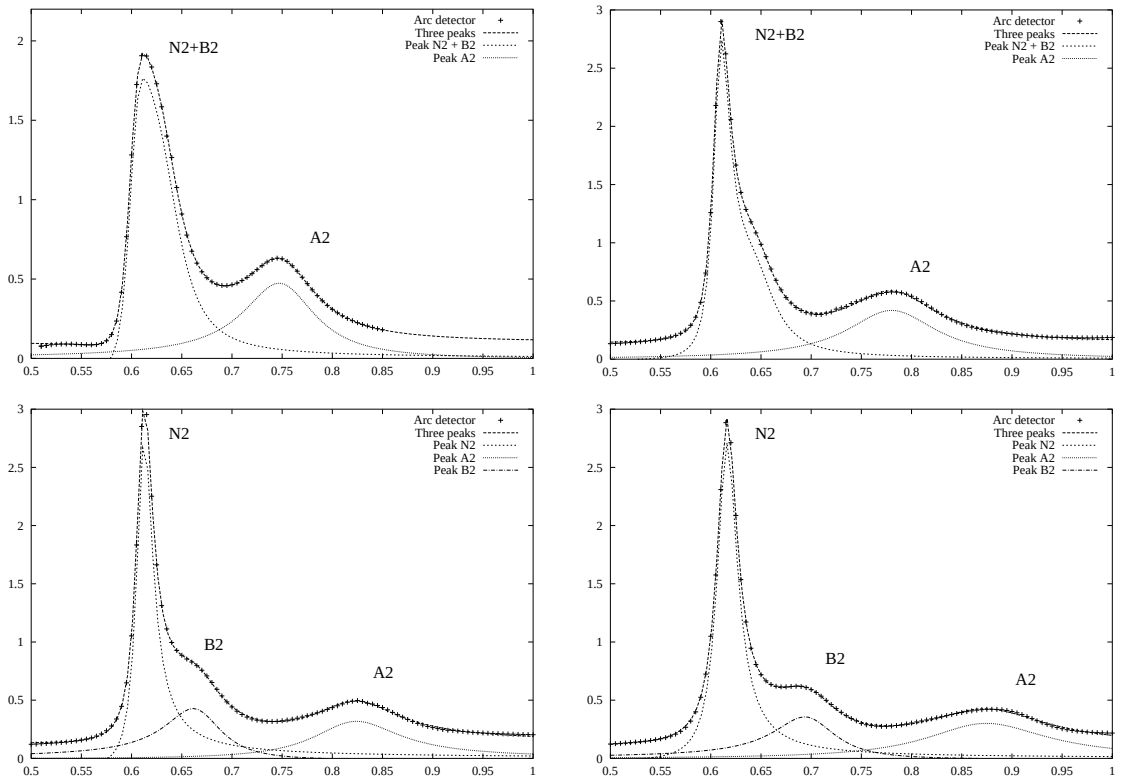

Figure 4.3: Partial conductance traces in the unstable regime. Four different mirrors are used, having angular sizes (a) $36\pi/64$, (b) $38\pi/64$, (c) $40\pi/64$, and (d) $42\pi/64$. In all cases the radius of curvature is 10λ.

are characterized by two types of peaks with very different behaviors. The peaks labeled N_x persist over the entire range of QPC to mirror separations tested here, and their positions and widths do not change appreciably as we vary the size of the mirror. The peaks labeled A_x, B_x, and C_x exist only in a small window of separations; and their positions, widths, and even number vary with the angular size of the mirror. We will refer to the former as nonanomalous peaks, and the latter as anomalous. If we ignore the presence of the QPC, there is one remaining isolated, unstable periodic orbit, $\gamma_{\rm cl}$, along the symmetry axis of the device. The nonanomalous peaks can be associated with a type of scarring along $\gamma_{\rm cl}$ [31], and

form an unbroken sequence along with the 0_x sequence in the stable regime. Naïvely there are no other periodic orbits with which we can associate the anomalous peaks. However, we are neglecting the effects of the edges. It is well known that sharp discontinuities give rise to strong diffractive effects in the scattered wave function. What is not well know is that the local wave fronts associated with the diffraction can be described using classical ray paths.

Wave functions

The wave functions associated with the first two nonanomalous peaks are shown in Figs. 4.4. They generally resemble the Gaussian beam modes of confocal resonators; and as we will see in Sec. 4.4.2, the positions of their "waists" depend on the long time classical dynamics in the neighborhood of γ_{cl}. The wave functions associated with the anomalous

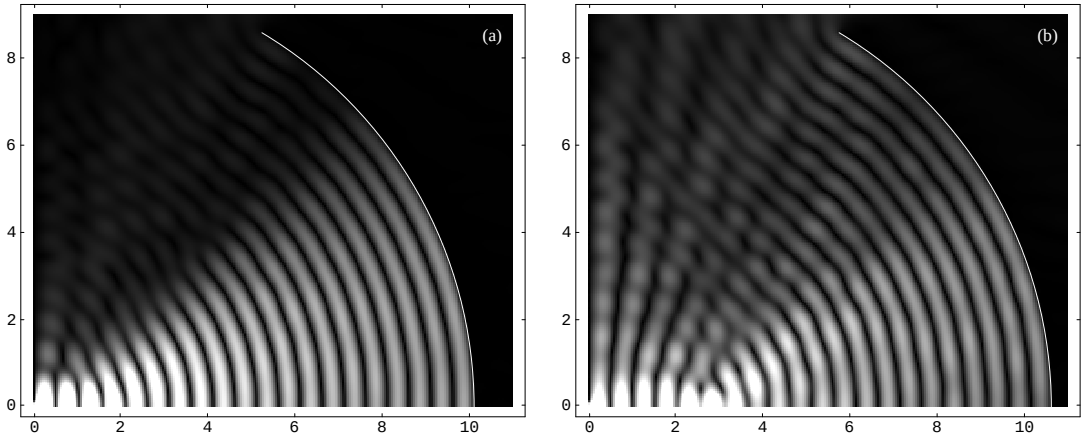

Figure 4.4: Moduli of the wave functions associated with peaks (a) N_1 and (b) N_2 of Fig. 4.3(d).

peaks A_1, B_1, and C_1 are shown in Figs. 4.5. It might initially be guessed that the edge state in Fig. 4.5(a) arises from diffraction, but this cannot explain the state in Figs. 4.5(b) that lives well inside the edge. Moreover, the anomalous concentration of probability near the edges also occurs away from the conductance peaks, as is shown in Fig. 4.6. As we will see in Sec. 4.6.4, these states have more to do with the openness of the device than with diffraction.

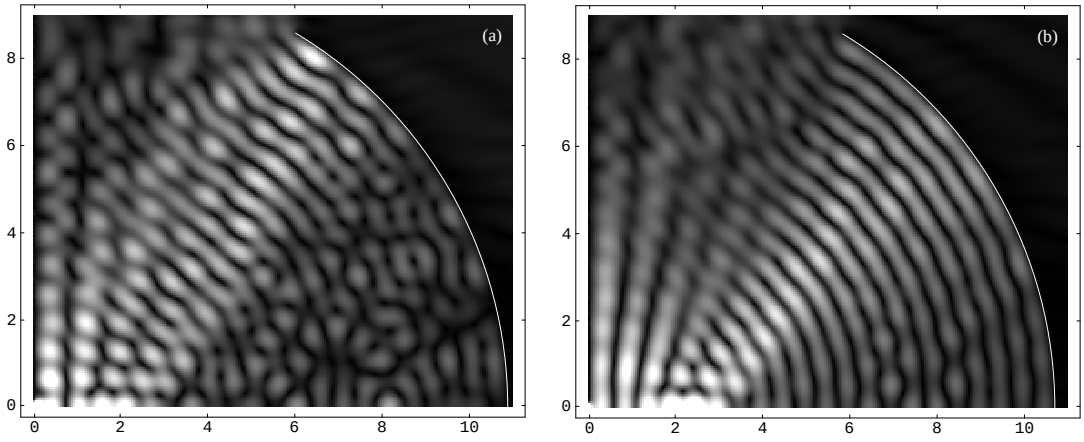

Figure 4.5: Moduli of the wave functions associated with peaks (a) A_2 and (b) B_2 of Fig. 4.3(d).

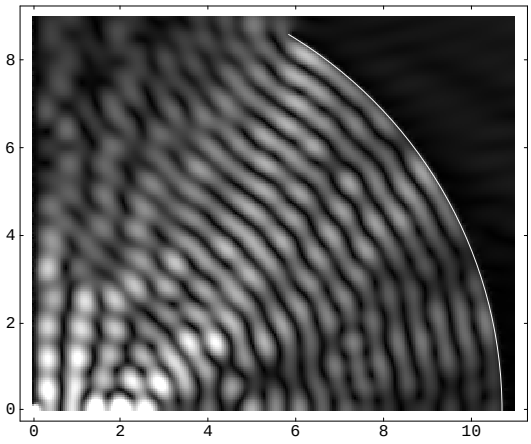

Figure 4.6: Modulus of the wave function for a QPC to center of curvature separation of 0.77λ, intermediate between peaks B_2 and A_2.

4.3 Quantum and semiclassical theories of conductance

4.3.1 The Landauer formula for conductance

Because of the work of Landauer [48] and others who followed him [8, 2], it is understood that the electronic transport properties of a mesoscopic device are properly computed by solving the full quantum scattering problem defined by that device. One of the more commonly used results to emerge from this conceptual framework is an expression for the conductance of a closed billiard with $N \geq 2$ leads attached to it. Assuming that all of the leads have finite transverse widths, the conductance g_{ji} between leads i and j is given by the "corrected Landauer formula" [8, 2],

$$g_{ji} = \frac{2e^2}{h} \sum_{n,m} \left| t_{mn}^{(ji)} \right|^2 , \tag{4.1}$$

where $t_{mn}^{(ji)}$ is the transition amplitude from the n-th transverse mode of lead i to the m-th mode of j, and the sum goes over the open modes in each lead. The $t_{mn}^{(ji)}$ can be computed in one of two ways: their quantum values can be found by numerically solving the scattering problem defined by the billiard plus leads; or their semiclassical values can be found by first approximating how the lead modes scatter into the billiard, and then propagating this scattered amplitude inside the billiard using a semiclassical expansion of the Green's function.

As written, Eq. (4.1) can only be applied to devices with well-defined leads. However, its range can be extended to the Westervelt gate if we think of the left and right halves of the infinite plane as leads, and we replace the sum over modes with an expansion over a complete orthonormal basis. The particular basis we will use in our computations is motivated by our treatment of the QPC. Since we are concerned with subwavelength contacts, we would like to approximate the QPC as having zero width. This amounts to neglecting the fine-scale structure of the wave function inside the QPC, which although important for determining the exact field near the contact, should be smoothed out in the far field where we compute flux integrals. This neglect forces a zero in the wave function over the entire y-axis, so we can describe the scattered wave function using the basis set $\{\sqrt{\frac{m}{\hbar}} H_l^{(1)}(kr)\cos(l\phi) : l = 1, 3, 5, \ldots\}$; these functions are solutions of the Helmholtz equation in free space and automatically satisfy the boundary conditions. Note that we have normalized the states to have unit integrated flux over a semicircle in either the right or left

half plane; this will be important in the following section. This basis is particularly well suited to the Westervelt gate because the exact field scattered by the QPC in the zero-width limit is proportional to $H_1^{(1)}(kr)\cos(\phi)$ [3]. It also convenient if we are interested in computing corrections to the zero-width approximation. The exact solution for the scattering of a plane wave from a slit of arbitrary width can be expressed as an infinite expansion in Mathieu functions mf:book,barn, and for subwavelength slits, the leading corrections can be reexpressed in terms of the lowest order Hankel functions.

4.3.2 Exact transition amplitudes

In order to make the expression for the S-matrix elements as concise as possible, we define the generalized basis element $|\eta, l; \mathrm{in(out)}>$. The subscript η takes on the values $+/-$, and denotes whether the basis function lies in the right or left half-plane; l is the same index as in $H_l^{(1)}(kr)\cos(l\phi)$; and "in(out)" denotes whether the function is incoming or outgoing. If we assume that the incoming wave is incident from the left in the l-th channel, then the full wave function can be written as

$$|\psi_{-,l}^{\mathrm{full}}>= |-, l; \mathrm{in}> + \sum_{\eta,\mathrm{m}} \mathrm{t}_{\mathrm{ml}}^{\eta,-} |\eta, \mathrm{m}; \mathrm{out}> . \tag{4.2}$$

These amplitudes can be written as operator matrix elements as follows. Let $\widehat{\mathcal{J}}(\mathbf{r_0}) = \frac{1}{2m}\left(\widehat{\mathbf{p}}\,\delta(\widehat{\mathbf{r}}-\mathbf{r_0}) + \delta(\widehat{\mathbf{r}}-\mathbf{r_0})\,\widehat{\mathbf{p}}\right)$ be the usual probability current operator, and define a pair of contours Γ_η, one in either half-plane, beginning and ending at the walls on opposite sides of the QPC. Note we have extended our η-notation to denote in which half-plane each contour lives. In general the Γ_η can be arbitrary curves, but for simplicity we will assume that they are semicircles centered at the QPC. If we define a pair of integrated current operators $\widehat{\mathcal{J}}_\eta = \int_{\Gamma_\eta} \widehat{\mathcal{J}} \cdot \hat{\mathbf{n}}$, where $\hat{\mathbf{n}}$ is the outward pointing normal along Γ_η, then as shown in [61], the transition amplitudes can be written as matrix elements of these operators:

$$\begin{aligned} t_{ml}^{\eta,-} &= \langle \eta, m; \mathrm{out}|\, \widehat{\mathcal{J}}_\eta \,|\psi_{-,\mathrm{l}}^{\mathrm{full}}\rangle \\ &= -\frac{i\hbar}{2m}\int_{-\pi/2}^{\pi/2} r d\phi \, \left\{ (\psi_{\eta,m}^{out}(r,\phi))^* \frac{\partial}{\partial r}\psi_{-,l}^{\mathrm{full}}(r,\phi) - (\psi_{-,l}^{\mathrm{full}}(r,\phi))^* \frac{\partial}{\partial r}\psi_{\eta,m}^{out}(r,\phi) \right\} \end{aligned} \tag{4.3}$$

Since flux is conserved, the radius r used in Eq. (4.3) can be chosen to be sufficiently large that we can use the asymptotic expansions for ψ^{out}. Note that in order to use this expression, we need to solve the full scattering problem defined by the gate.

In order to illustrate the use of Eq. (4.3), we now derive an expression for the conductance of the Westervelt gate in the limit of vanishing QPC width, parametrized by the p-wave phase shift. This calculation also serves to raise a number of issues associated with a pointlike QPC that we will again encounter in our semiclassical calculations. Our final expression for the conductance has also been derived from a first principles scattering theory approach in barn. However, the derivation presented in that paper does not have an obvious semiclassical generalization.

We begin with a plane wave,

$$\psi^{in}_{-,\theta}(\mathbf{r}) = \langle \mathbf{r}|-,\theta;\mathrm{in}\rangle = \sqrt{\frac{\mathrm{m}}{\hbar}}\mathrm{e}^{\mathrm{i}\mathbf{k}\cdot\mathbf{r}}, \tag{4.4}$$

incident at an angle θ with respect to the x-axis. The normalization of these states is such that the probability flux is measured in units of $\hbar/m$. Note that we are concerned with fixed energy scattering, so the label θ rather than $\mathbf{k}$ is unambiguous. We have also adapted our generalized bracket to plane waves and replaced l with θ. If initially we neglect scattering from the QPC, the full wave function in the left half-plane is

$$\psi^{full}_{-,\theta}(\mathbf{r}) = \psi^{in}_{-,\theta}(\mathbf{r}) - \psi^{in}_{-,\pi-\theta}(\mathbf{r}) = 4i\sqrt{\frac{m}{\hbar}}J_1(kr)\cos\phi\cos\theta + \mathrm{higherpartialwaves}, \tag{4.5}$$

where the relative minus sign enforces the Dirichlet boundary condition on the y-axis. Here we have used the expansion $e^{i\mathbf{k}\cdot\mathbf{r}} = \sum_{l=-\infty}^{\infty} i^l J_l(kr)e^{il\theta} = J_0 + iJ_1e^{i\theta} + iJ_1e^{-i\theta} + \cdots$ to express the plane waves in terms of Bessel functions, and kept only the lowest order nonvanishing term. The combination $J_1(kr)\cos\phi$ is, of course, just the p-wave. In the limit of vanishing QPC width, only the p-wave is scattered, so all of the higher partial waves can be ignored.

The function $J_1(kr)$ can be written as $((H_1^{(1)}(kr))^* + H_1^{(1)}(kr))/2$, the sum of an incoming and outgoing Hankel function, and the effect of the QPC can be summarized by a complex phase shift, $H_1^{(1)}(kr) \to e^{i\delta_p}H_1^{(1)}(kr)$. We thus write

$$J_1(kr) \to \frac{H_1^{(1)}(kr))^* + e^{i\delta_p}H_1^{(1)}(kr)}{2} = J_1(kr) + \frac{e^{i\delta_p}-1}{2}H_1^{(1)}(kr). \tag{4.6}$$

At one extreme, if $\delta_p = 0$, then the second term vanishes and the QPC has no effect; at the other extreme, if $\delta_p = i\infty$, the incoming wave is completely absorbed by the QPC, and the outgoing wave is completely removed from Eq. (4.6). Thus the effect of the scattering may be summarized by saying that the incident plane wave induces an effective p-wave source at the QPC.

This effective source picture is useful for computing the conductance of the Westervelt gate. The actual wave function in the right half-plane is quite complicated, even in the asymptotic region, because of the mirror. However, for the purpose of computing transmission amplitudes, we can ignore the mirror and pretend the wave function is simply that radiated by our p-wave source. In this case the only nonvanishing amplitude is

$$\begin{aligned} \langle \psi^{out}_{+,1} | \mathcal{J}_\eta | \psi^{full}_{+,\theta} \rangle &= -\frac{i\hbar}{2m} 2i \cos\theta (e^{2i\delta_p} - 1) \frac{m}{\hbar} \frac{2}{\pi k} \int_{-\pi/2}^{\pi/2} d\phi \left(e^{-ikr} \partial_r \frac{e^{ikr}}{\sqrt{r}} - e^{ikr} \partial_r \frac{e^{-ikr}}{\sqrt{r}} \right) \cos^2\phi \\ &= 2i \cos\theta \left(e^{2i\delta_p} - 1 \right), \end{aligned} \tag{4.7}$$

where we have evaluated the basis functions in the asymptotic regime. If we divide by $2i\cos\theta$ to normalize the incident flux in the p-wave channel to 1, we find that the conductance is given by

$$\frac{2e^2}{h} \left| e^{\delta_p} - 1 \right|^2, \tag{4.8}$$

in agreement with the result in [3]. If we then set $\delta_p = i\infty$, this reduces to its maximum value of $2e^2/h$. But this is just as expected, since we already saw that $\delta_p = i\infty$ implies that there is no reflected wave.

4.3.3 Semiclassical transmission amplitudes

We would like to develop a semiclassical approximation for the expression in Eq. (4.3) that would perhaps allow us to understand the physical origin of the anomalous conductance peaks. As we already noted in the previous section, the wave function in the right half-plane is quite complicated because of the mirror. It would thus be much easier to work with reflection rather than transmission amplitudes. Since there are only two "leads" for the Westervelt gate, the latter can easily be derived from the former. Thus we are interested in the amplitudes $t^{-,-}_{QPC,\theta} = < -, \mathrm{QPC}; \mathrm{out} | \, \widehat{\mathcal{J}}_- \, | -, \theta; \mathrm{in} >$, where $|-, \mathrm{QPC}; \mathrm{out}>$ is the state transmitted through the QPC. As with the basis states $|-, l; \mathrm{out}>$, this state is normalized to have unit integrated flux over a semicircle in the left half-plane. Since we have taken the QPC to be effectively pointlike, this state is independent of the source. In practice its position space representation can be found from the numerical solution for the scattering of a plane wave from a slit, or by projecting the known, exact solution for a narrow slit in terms of Mathieu functions onto the $|-, l; \mathrm{out}>$.

The semiclassical approximation to $< \mathbf{r} | \psi >$ can be computed as follows. The state initially transmitted through the QPC is $< \mathbf{r} | +, \mathrm{QPC}; \mathrm{out} >$, up to a flux dependent

normalization. The value of $< \mathbf{r} | \psi >$ at any point $\mathbf{r}$ in the right half-plane can be computed using standard semiclassical propagation, but the scattering back through the QPC cannot be described semiclassically. Thus we would rather perform the azimuthal integral in the right half-plane. Since flux is conserved, we can use Green's theorem to deform the integration contour through the QPC into the right half-plane. We are thus left with

$$t^{--}_{QPC,\theta} = \frac{-i\hbar}{2m} \int_{-\pi/2}^{\pi/2} r\, d\phi \left\{ \left(\psi^{in}_{+,QPC}(r,\phi)\right)^{\dagger} \partial_r \psi^{\mathrm{sc}}_{+,\theta}(r,\phi) - \left(\psi^{\mathrm{sc}}_{+,\theta}(r,\phi)\right)^{\dagger} \partial_r \psi^{in}_{+,QPC}(r,\phi) \right\}, \tag{4.9}$$

where $\psi^{\mathrm{in}}_{+,\mathrm{QPC}} = \left(\psi^{\mathrm{out}}_{+,\mathrm{QPC}}\right)^{\dagger}$. The actual classical paths used to compute $\psi^{\mathrm{sc}}_{+,\theta}$ will be discussed in Sec. 4.6.1.

4.4 Nonanomalous conductance peaks

4.4.1 Classical dynamics of the billiard

The classical dynamics of a point particle in a two-dimensional, hard-walled billiard are particularly easy to describe: when the particle strikes the boundary it undergoes specular reflection, and between reflections it propagates along a straight line. The path γ_n between the n and $(n+1)$-th bounces is completely determined by the position of the n-th bounce along the boundary and the angle between the local normal and the outgoing path. The original two-dimensional system can thus be reduced to a one-dimensional map $M : (q_n, p_n) \mapsto (q_{n+1}, p_{n+1})$ defined on the boundary. The standard phase space coordinates used to describe such bounce maps are the Birkhoff coordinates $(q,p) = (s, \sin\phi)$, where s is the arc length relative to some fixed position, and ϕ is the angle between the local normal and the outgoing path. Since the magnitude of the momentum is constant, $\sin\phi$ is just the projection of the momentum onto the boundary up to an unimportant scaling factor.

The map M provides us with a convenient representation of the global dynamics of the billiard. However, to compute semiclassical propagation amplitudes, we need a description of the local dynamics of the bundle of trajectories around γ_n. This is provided by the Jacobian matrix J_{γ_n}, the linearization of M around γ_n. A convenient parametrization of the neighborhood of γ_n is $(\delta z, \delta\phi)$, where δz is the displacement *perpendicular* to the path, and $\delta\phi$ is the angular displacement relative to ϕ_n. The linearized equations of motion have a particularly simple form in this coordinate system: $d(\delta z)/dt = \delta\phi$ and $d(\delta\phi)/dt = 0$.

Let t be a parameter along the classical path, $t_n^{\pm} = t_n \pm \epsilon$ be the values of this

parameter immediately before and after the n-th bounce, and $\delta z_n^{\pm} = \delta z_n(t_n^{\pm})$ and $\delta\phi_n^{\pm} = \delta\phi_n(t_n^{\pm})$ be the corresponding displacements. If we integrate the equations for $(\delta z, \delta\phi)$ from t_n^+ to t_{n+1}^-, we find

$$\delta z_{n+1}^- = \delta z_n^+ + \tau_{n+1}\delta\phi_n^+ \tag{4.10}$$

$$\delta\phi_{n+1}^- = \delta\phi_n^+, \tag{4.11}$$

where $\tau_{n+1} = t_{n+1} - t_n$. The Jacobian for free propagation is the matrix that takes us from $(\delta z_n^+, \delta\phi_n^+)$ to $(\delta z_{n+1}^-, \delta\phi_{n+1}^-)$, and is thus

$$J_{\tau_{n+1}} = \begin{pmatrix} 1 & \tau_{n+1} \\ 0 & 1 \end{pmatrix}. \tag{4.12}$$

The Jacobian describing a bounce is only slightly harder to derive. The perpendicular displacement simply changes sign following a bounce, $\delta z_n^+ = -\delta z_n^-$, and if the local curvature of the boundary vanishes at s_n, then the angular displacement also undergoes a simple reversal of sign. However, if the radius of curvature ρ_n does not vanish at s_n, then there is an additional term arising from the variation of the local normal. From Fig. 4.7 we see that the perpendicular displacement δz_n^- projects down to an arc of length $\delta z_n^-/\cos\phi_n$ along the boundary. The local normal varies by an angle $\delta\phi_{\rm norm} = -\delta z_n^-/\rho_n\cos\phi_n$ over this distance, where the sign of ρ_n is defined to be positive if the center of curvature is on the same side as the incoming path. The angle of the incident ray relative to the fixed normal at s_n is $-\phi_n^+$; relative to the local normal at $s_n + \delta s_n$, it is $-(\phi_n + \delta\phi_{norm})$. The angle of the incident displaced ray relative to this local normal is $-\phi_n^- - \delta\phi_{\rm norm} - \delta\phi_n^-$. Following the scattering the outgoing displaced ray makes an angle $\phi_n^- + \delta\phi_{\rm norm} + \delta\phi_n^-$, and $\phi_n^- + 2\delta\phi_{\rm norm} + \delta\phi_n^-$ relative to the fixed normal at s_n. Since this angle is defined as $\phi_n^+ - \delta\phi_{\rm norm}^+$, we have $\delta\phi_n^+ = -\delta\phi_n^- - \delta\phi_{\rm norm}^-$. Thus the equations describing a bounce are

$$\delta z_n^+ = -\delta z_n^-, \tag{4.13}$$

$$\delta\phi_n^+ = -\delta\phi_n^- + \frac{2\delta z_n^-}{\rho_n\cos\phi_n}, \tag{4.14}$$

and the Jacobian is

$$J_{\rho_n} = -\begin{pmatrix} 1 & 0 \\ -2/(\rho_n\cos\phi_n) & 1 \end{pmatrix}. \tag{4.15}$$

Using these two elementary Jacobians we can construct the full Jacobian for the periodic orbit γ_{cl} that starts at the QPC, reflects from the center of the mirror, and then

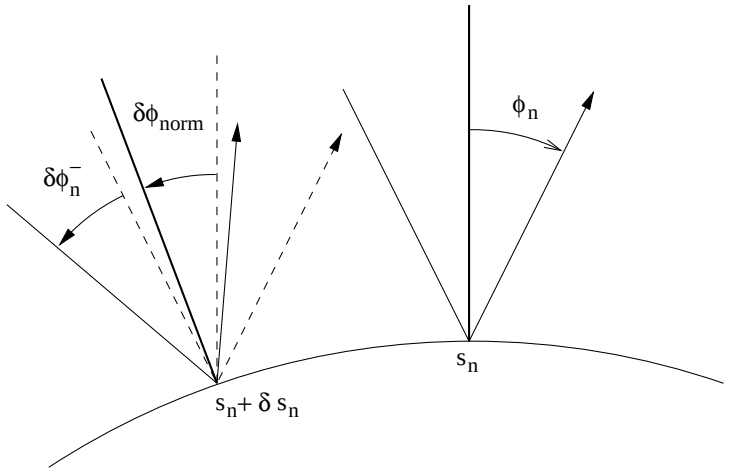

Figure 4.7: The effect of a specular bounce on the displacement coordinates $(\delta z, \delta\theta)$. The heavy solid lines denote the local normals at s_n and $s_n + \delta s_n$. The solid lines denote the reflection of γ_{n-1} into γ_n and the reflection of the associated displaced path. The dashed lines denote the parallel transports of γ_{n-1} and the local normal at s_n to $s_n + \delta s_n$.

returns to the QPC.

$$J_{\gamma_{cl}} = (-1)^2 \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} 1 & R \\ 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} 1 & 0 \\ -2/\rho & 1 \end{pmatrix} \cdot \begin{pmatrix} 1 & R \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 - 4R/\rho & 2R \\ -2/\rho & 1 \end{pmatrix}, \quad (4.16)$$

where R is the QPC/mirror separation and ρ is radius of curvature of the mirror. The eigenvalues of this matrix, the so-called stability indices, are $\lambda_\pm = 1 - 2R/\rho \pm 2\sqrt{(R/\rho)^2 - R/\rho}$. If $R \leq \rho$ it is easy to check that the eigenvalues form a complex conjugate pair of modulus one, as they must since $R \leq \rho$ defines the stable regime. If $R \geq \rho$, then we have one positive and one negative eigenvalue, and the dynamics of the local bundle are unstable, as they should be since we are in the unstable regime. In the limit that $R \to \rho$ from above, $\lambda_- \simeq 1 - 2R/\rho$, so for R sufficiently close to ρ, the resonator is barely unstable.

4.4.2 Wave functions

Observed wave functions

In Figs. 4.4 we saw the wave functions corresponding to the first three expected peaks. We would like to account for the global structure of these wave functions, in particular the location of their "waists." Since we are in the unstable regime, it might be expected that the waist should coincide with the focus of the rays after a single specular reflection

from the mirror. But when the QPC to mirror separation is only slightly larger than the mirror's radius of curvature, the long time dynamics cannot be ignored. Thus we must take into account multiple reflections from the mirror.

Exact equations for the first cusp

We begin with a point source of rays located at the QPC. In the paraxial approximation, these rays would be refocused to a single point after reflecting from the mirror. However, because of spherical aberration, this focus will be stretched over some finite distance, as illustrated in Fig. 4.8. The envelope of the reflected rays forms a curve known as a *cusp* [1, 57]. Such curves are stable against small perturbations of the surfaces that define them, in this case the mirror. This stability is essential to the experimental robustness of the conductance peaks.

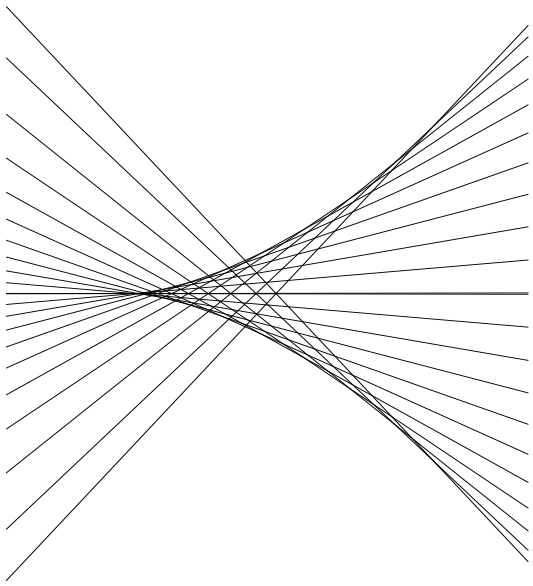

Figure 4.8: Cusp formed after a single reflection from the quarter circle mirror of the Westervelt gate.

A derivation of the exact equations for the cusp is as follows. Consider the diagram in Fig. 4.9. Note that choice of the variables θ_0^+ and θ_1^- to describe the rays at the walls defining the QPC is meant to remind us that the Westervelt gate is just an open billiard, so we can use Birkhoff coordinates to describe its dynamics. If x and y are the coordinates

of the reflected ray relative to the center of curvature, then we have

$$\tan\theta_1^- = \frac{\rho\sin\phi - y}{\rho\cos\phi - x} \tag{4.17}$$

The angle θ_1^- can be written in terms of ψ, the angle of reflection relative to the mirror's local normal, using $\theta_1^- = \phi + \psi$; ψ itself can in turn be eliminated in favor of θ_0^+ using $\theta_0^+ + \psi = \phi$; and θ_0^+ can then be written in in terms of ϕ, ρ, and r using the defining equation of the incident ray, $\tan\theta_0^+ = \rho\sin\phi/(r + \rho\cos\phi)$. The final expression for the left hand side of Eq. (4.17) is

$$f_1(\phi) \equiv \tan\left(2\phi - \theta_0^+\right) = \frac{(1 + (r/\rho)\sec\phi)\tan 2\phi - \tan\phi}{(1 + (r/\rho)\sec\phi) + \tan 2\phi\tan\phi}. \tag{4.18}$$

Eq. (4.17) may be neatly written as $F(x, y; \phi) = 0$, and should be thought of as an implicit equation for the reflected ray corresponding to each value of the parameter ϕ.

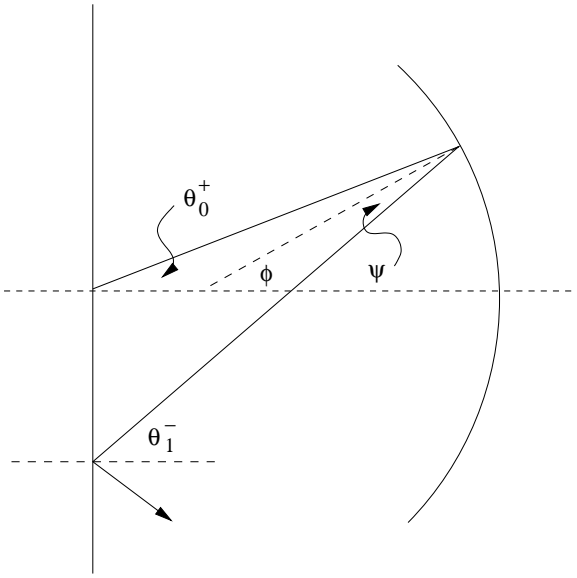

Figure 4.9: Coordinates used in the derivation of the parametric equations for the first cusp.

The procedure for finding the envelope of the curves defined by such an equation is well known in classical algebraic geometry. In App. 4.8 we show that the envelope satisfies not only $F = 0$, but also $\partial F/\partial\phi = 0$. Combining these two equations, we find the following parametric equations for the envelope:

$$x = \rho\cos\phi - \frac{\rho}{f_1'(\phi)}(\cos\phi + f_1(\phi)\sin\phi), \tag{4.19}$$

$$y = \rho\sin\phi - \frac{\rho f_1(\phi)}{f_1'(\phi)}(\cos\phi + f_1(\phi)\sin\phi), \tag{4.20}$$

where $f'(\phi)$ denotes the derivate with respect to ϕ.

The equations for x and y can be expanded as $x = c_0 + c_2\phi^2 + O(\phi^4)$ and $y = c_3\phi^3 + O(\phi^5)$. If we eliminate ϕ from these equations, we find $x \simeq c_0 + c_2 c_3^{-2/3} y^{2/3}$. This equation defines the *universal* part of the cusp. We will write it as $x = x_1^{\rm cusp} + x_1^{\rm curv} y^{2/3}$, where $x_1^{\rm cusp}$ is the position of the tip of the cusp, while $x_1^{\rm curv}$ describes how fast the cusp curves away from the x-axis.

Recursion relations for the n-th cusp

The cusp formed after a single reflection from the mirror is only the first in an infinite sequence of cusps formed inside the resonator. It is impossible to derive the exact equations for the n-th cusp in any useful closed form, but it is possible to find recursion relations for the coefficients of its universal piece. The difficulty arises because after a single bounce from the mirror, there is a nontrivial relation $s_1(\theta_1^+)$ between the positions of the returning rays along this wall and their outgoing directions. We need to find generalizations of the equations in the previous section that allow an arbitrary initial distribution of rays $s_n(\theta_n^+)$, where the subscripts denote the positions and outgoing directions following the n-th bounce from the mirror. From Fig. 4.10 we see that the equation for the n-th reflected ray is given by

$$f_n(\phi) \equiv \tan(2\phi - \theta_n^+) = \frac{\rho\sin\phi - x}{\rho\cos\phi - y}, \tag{4.21}$$

where $\theta_n^+(\phi)$ is now implicitly given by $\tan\theta_n^+ = (-s_n(\theta_n^+) + \rho\sin\phi)/(r + \rho\cos\phi)$. If we write $s_n(\theta_n^+) \simeq s_n^{(1)}\theta_n^+$ and perform a small angle expansion of this equation, keeping only the lowest order terms in both θ_n^+ and ϕ, we find $\theta_n^+ \simeq \left(\rho/(r + \rho + s_n^{(1)})\right)\phi$. Using this first order expansion in the defining equation for $f_n(\phi)$, and then the expansion of $f_n(\phi)$ in Eq. (4.19), we find that the tip of the cusp generated by the initial distribution of rays $s_n(\theta_n^+) \simeq s_n^{(1)}\theta_n^+$ is $x_n^{\rm cusp} = \rho(r + s_n^{(1)})/(2r + \rho + 2s_n^{(1)})$.

Having found the position of the n-th cusp, we need to find the distribution $s_{n+1}(\theta_{n+1}^+)$ of rays that will generate the $(n+1)$-th cusp. Setting $x = -r$ and $y = s_{n+1}(\theta_{n+1}^-)$ in Eq. (4.21), we find

$$s_{n+1}(\theta_{n+1}^-) = \rho\sin\phi - \tan\theta_{n+1}^-(\rho\cos\phi + r), \tag{4.22}$$

where we have used the relation $\theta_{n+1}^- = 2\phi - \theta_n^+(\phi)$. Using this relation again, and inverting

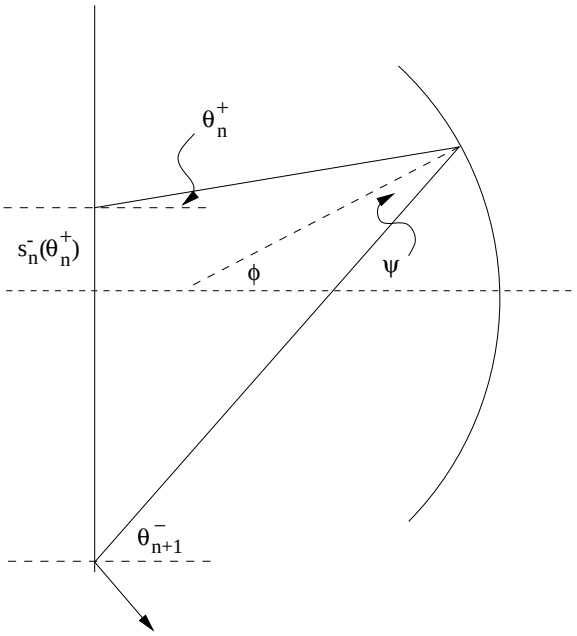

Figure 4.10: Coordinates used in the derivation of the parametric equations for the n-th cusp.

our previous relation between θ_n^+ and ϕ, a first order expansion of Eq. (4.22) gives

$$s_{n+1}(\theta_{n+1}^+) = (x_{cusp} + r)\,\theta_{n+1}^+, \tag{4.23}$$

where we have used the trivial relation for a straight wall, $\theta_n^+ = -\theta_n^-$. The physical meaning of this relation is that to lowest order, the cusp reduces to a pointlike focus. Combining this with the above equation for $x_n^{\rm cusp}$, we find the desired recursion relation for the positions of the cusps:

$$x_{n+1}^{\rm cusp} = \rho\,\frac{2r + x_n^{\rm cusp}}{4r + \rho + 2x_n^{\rm cusp}}, \tag{4.24}$$

where we have defined $x_0^{\rm cusp} = 0$ for convenience.

In order to derive a similar relation for $x_n^{\rm curv}$, we need to expand $s_n(\theta_n^+)$ up to third order, $s_n(\theta_n^+) \simeq s_n^{(1)}\theta_n^+ + s_n^{(3)}\,(\theta_n^+)^3$. Going through the same steps as before, but keeping terms up to third order in both θ_n^+ and ϕ, we find that the distributions of rays that results from the n-th reflection is

$$s_n(\theta_n^+) \simeq x_n^{\rm cusp}\theta_n^+ + \left(\frac{1}{3}x_n^{\rm cusp} + \frac{4}{27}\,(x_n^{\rm curv})^3\right)(\theta_n^+)^3\,, \tag{4.25}$$

and the curvature coefficients are given by

$$\left(x_{n+1}^{\mathrm{curv}}\right)^3 = \rho\,\frac{27\left(x_n^{\mathrm{cusp}}+r\right)^2\left(x_n^{\mathrm{cusp}}+r+\rho\right)^2+4\left(x_n^{\mathrm{curv}}\right)^3\rho^3}{4\left(2\rho+r+2x_n^{\mathrm{cusp}}\right)^4}. \tag{4.26}$$

Limiting cusp and the surface of section

Numerically the x_n^{cusp} are found to approach a finite limit. This limit can be found analytically from Eq. (4.24) by setting $x_n^{\mathrm{cusp}} = x_{n+1}^{\mathrm{cusp}} = x_\infty^{\mathrm{cusp}}$. Solving for x_∞^{cusp}, we find

$$x_\infty^{\mathrm{cusp}} = -r + \sqrt{r(r+\rho)}. \tag{4.27}$$

A similar expression for x_∞^{curv} can be found from Eq. (4.26), but will not be given here because of its complicated and generally unenlightening nature.

In Fig. 4.11 we see that the waists of the "scarred" wave functions coincide with the limiting cusps. This shows that the long time dynamics affect the structure of the wave function. This, however, should be expected; for all three wave functions shown in Fig. 4.11 the cavity is barely unstable, so the stability index of the unstable manifold is still quite small.

Although it may seem physically obvious that the cusps must posses a limit point within the resonator, it is worthwhile asking whether this has an interpretation in terms of classical dynamics. In Sec. 4.4.1 we reduced the two-dimensional billiard to a one-dimensional map by observing that all of the nontrivial action occurred at the boundary. This construction is a particular example of a standard dimensional reduction technique. Given some $2n$-dimensional phase space, rather than discuss the behavior of trajectories in the full space, we consider their intersection with a surface of codimension 2, called the "surface of section." The $s_n(\theta_n^+)$ can thus be thought of as snapshots of the evolution of the initial distribution of rays at the QPC each time they pass through the section defined by the boundary.

The point $s = \theta = 0$ is what remains of the periodic orbit γ_{cl} in this reduced phase space. Because γ_{cl} is an unstable orbit, the flow about this point is hyperbolic, with $\lambda_\pm$ being the indices of stability of the stable and unstable manifolds. In Fig. 4.12 we see a surface of section plot taken at the flat mirror. The vertical line is the initial distribution localized at the QPC, the dashed line is a local approximation to the unstable manifold, and the remaining lines are the first five iterations of the map. We can thus see the collapse of the initial distribution onto the invariant manifold with increasing number of reflections.

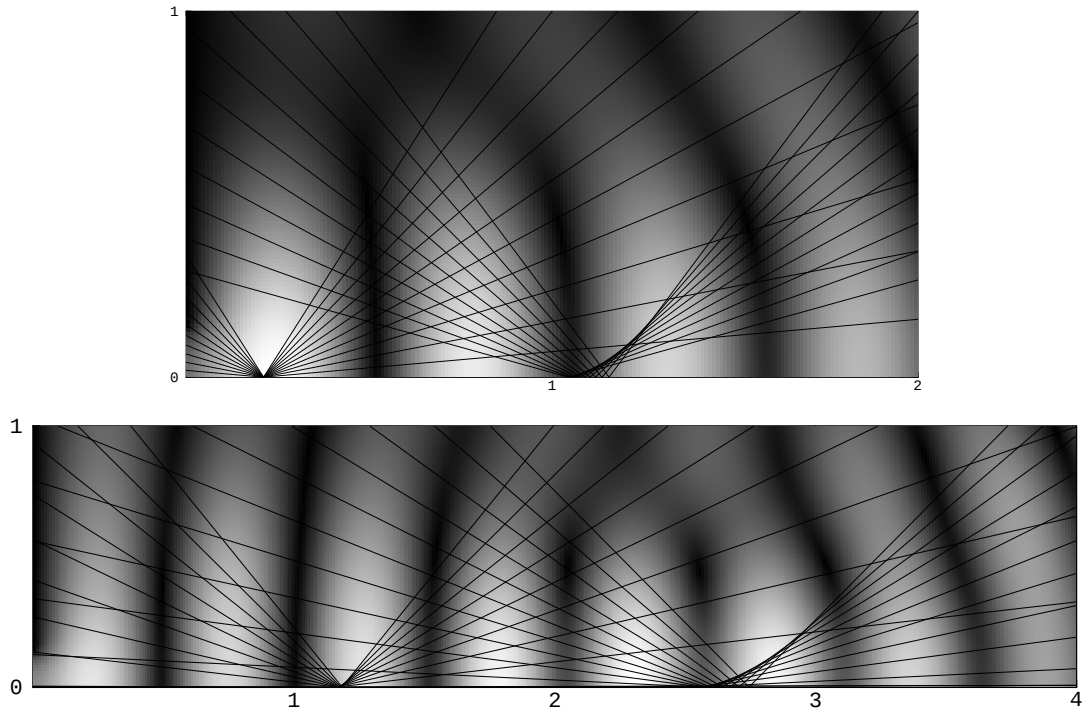

Figure 4.11: Detail of the wave functions in Fig. 4.4 showing the positions of the first and limiting cusps.

Since the coefficients of the n-th distribution, $s_n(\theta_n^+)$, are given in terms of $x_n^{\rm cusp}$ and $x_n^{\rm curv}$, their limits are simply a local parametrization of the unstable manifold. In fact, the dashed line in Fig. 4.12 was computed using $x_\infty^{\rm cusp}$ and $x_\infty^{\rm curv}$

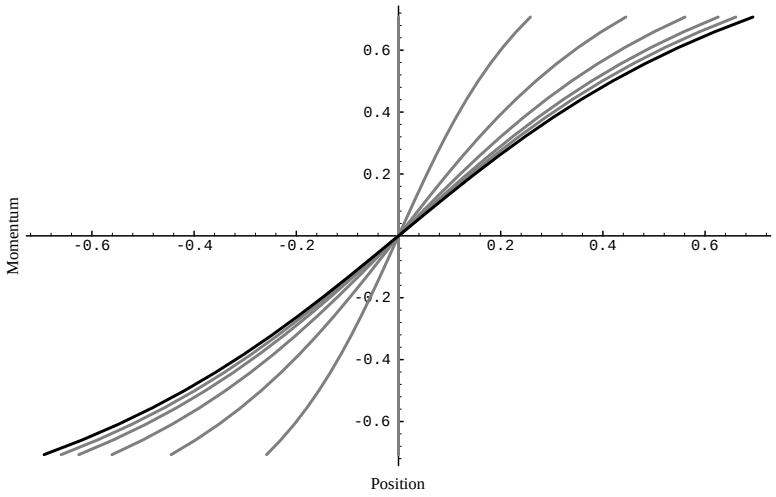

Figure 4.12: Surface of section centered at the QPC.

4.5 Geometric Theory of Diffraction

4.5.1 Keller's original theory

Consider the general problem of how best to describe quantum scattering from a one-dimensional curve in two dimensions. In the limit that the wavelength becomes small compared to the characteristic length scales of the curve, loosely referred to as the $k \to \infty$ limit, it becomes possible to describe the incident and scattered wave fronts in terms of their local normals, and the point-to-point scattering amplitudes in terms of these normals. An extreme example of this is the scattering of a point source from a straight wall: if we replace the circular wave fronts of the source by rays centered at the source point, their behavior at the mirror is described by the law of reflection. This is, of course, nothing more than geometrical optics if rather than thinking of quantum wave functions we think of electromagnetic waves.

Now let us consider scattering from a curve with a sharp corner. Near the corner the curve possess structure at infinitely fine length scales; no matter how closely we approach the corner it is always sharp. The scattered wave function from such a curve has features

that cannot be accounted for solely in terms of rays obeying the law of reflection, such as the appearance of amplitude in regions where there are no reflected rays; such features are said to arise from diffraction. As a rule it is assumed that when a curve has structure at a scale on the order of a wavelength, it is impossible to describe the scattered wave fronts in terms of rays. But beginning with the work of Keller [41, 42] and continuing with the work of others interested in computationally efficient methods of computing high frequency electromagnetic scattering from sharp objects, it has been understood how to describe diffraction using rays. Keller showed that the far-field diffraction pattern of a point source scattered from a half-infinite line could be described as the sum of the wave fronts associated with the ordinary rays reflected from the half-line and those associated with an effective family of radially outgoing rays beginning at the endpoint. He referred to them as "diffractive rays," and the function that determined how much amplitude was to be propagated along each one was the "coefficient of diffraction." This coefficient was extracted from the asymptotic expansion of Sommerfeld's exact solution for the half-line.

Besides simply providing an effective description of the known solution, these diffractive rays satisfy a generalization of Fermat's principle. Just as the lengths of the ordinary geometric rays are stationary with respect to local deformations, the diffractive rays are stationary with the added constraint that they pass through the endpoint of the half-line. The fact that these diffractive rays can be found from a stationarity principle suggests that they are more than just a convenience. Following the path integral approach to quantum mechanics, the amplitude for propagating from the source point to any other point is given by the sum over *all* possible paths weighted by $e^{iS_{path}}$, where S_{path} is essentially the length of the path. In the semiclassical limit, only those paths in a narrow bundle around the stationary paths contribute; all the remaining paths interfere destructively with one another. But now let us ask what happens if the sum over paths is cut off, as it is when the curve ends. The amplitudes from those paths that scatter from the end of the curve will not be canceled out, because the paths that would have provided the destructive interference are no longer present! When we write this sum as an integral over the half-line, we find that Keller's diffractive rays are exactly the stationary phase paths associated with the finite limit of integration.

We now turn to a derivation of the diffraction coefficient for the half-infinite line. Since we are interested in incorporating diffraction into the semiclassical trace formula, we will phrase our discussion of Keller's theory in terms of asymptotic expansions of Green's

functions. As with Keller's original derivation, we begin with Sommerfeld's exact solution. The exact Green's function can be given the following contour integral representation [66]:

$$G_{\rm line}(\mathbf{r}',\mathbf{r}) = G^{-}_{\rm line}(r',r,\theta'-\theta) - G^{+}_{\rm line}(r',r,\theta'+\theta), \tag{4.28}$$

where

$$G^{\pm}_{\rm line}(r',r,\phi_{\pm}) = \frac{-i}{16\pi}\int_{\Gamma_A+\Gamma_B} dz\ \frac{H_0^{(1)}\left(k\sqrt{r^2+r'^2-2rr'\cos z}\right)}{1-e^{-i(z-\phi_{\pm})/2}}, \tag{4.29}$$

and we have introduced the convenient shorthand $\phi_{\pm} = \theta' - \theta$. The coordinate system used in Eq. (4.29) is centered about the endpoint of the half-line, and the line itself coincides with the $\theta = 0$ axis. For the purposes of our derivation, we will orient the line along the negative y-axis, and measure θ and θ' in a clockwise fashion, as illustrated in Fig. 4.13. The

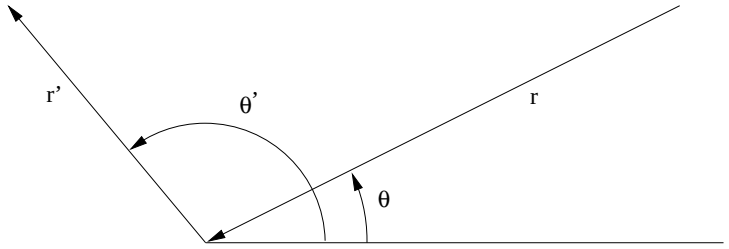

Figure 4.13: Coordinate system used in the definition of $G_{\rm line}$.

contours Γ_A and Γ_B are shown in Fig. 4.14, along with the branch structure and asymptotic behavior of the integrand as a function of z. The poles of the integrand are all simple, and are located at $\{\phi_{\pm}+4\pi n : n \in Z\}$. We would like to evaluate Eq. (4.29) in the limit $k \to \infty$. Using Cauchy's theorem we can replace the integral over $\Gamma_A \cup \Gamma_B$ by that over $\Gamma_C \cup \Gamma_D$ plus $2\pi i$ times the sum of the residues of the poles between $-\pi$ and π. As usual we may deform Γ_C and Γ_D as long as we avoid the branch cuts and remain within one of the shaded zones as we approach $\pm i\infty$. Since we are interested in the asymptotic behavior of $G_{\rm line}$, we can replace $H_0^{(1)}(z)$ with its asymptotic expansion, $\sqrt{2/\pi z}\ e^{iz-i\pi/4}$. We then immediately see that the integrand has stationary phase points at $\{n\pi : n \in Z\}$. Thus if we choose Γ_C and Γ_D to be paths of steepest descent through the stationary points at $-\pi$ and π, respectively, we should be able to compute everything explicitly.

We begin with the residues. Since $G_{\rm line}$ is symmetric with respect to $(\theta,\theta') \mapsto (2\pi-\theta, 2\pi-\theta')$, we can restrict θ to the interval $[0,\pi]$ without loss of generality. The only

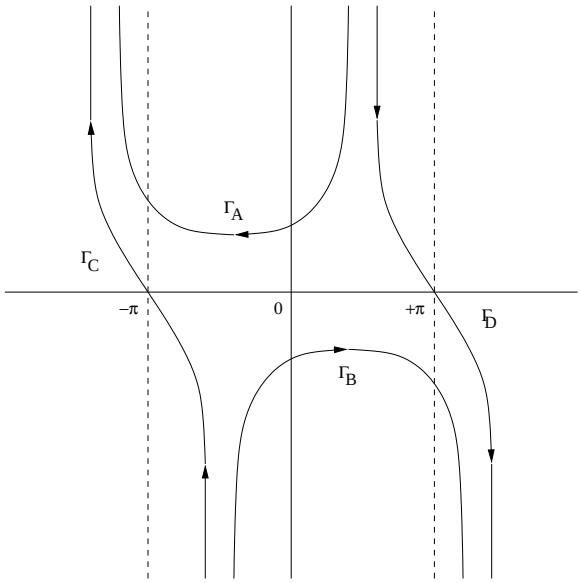

Figure 4.14: Contour used in the evaluation of Eq. 4.29

possible poles in $[-\pi, \pi]$ are then $z = \phi_\pm$. Using the Heaviside step function Φ, we can write their contributions as

$$\begin{aligned} G_{geom}(\mathbf{r}';\mathbf{r}) &= -\frac{i}{4}H_0^{(1)}\left(k\sqrt{r^2+r'^2-2rr'\cos\phi_-}\right)\Phi(\pi-\phi_-) \\ &\quad +\frac{i}{4}H_0^{(1)}\left(k\sqrt{r^2+r'^2-2rr'\cos\phi_+}\right)\Phi(\pi-\phi_+) \\ &= G_0(r',\theta';r,\theta)\,\Phi(\pi-\phi_-) - G_0(r',2\pi-\theta';r,\theta)\,\Phi(\pi-\phi_+), \quad (4.30) \end{aligned}$$

where we have noted that the combination $(-i/4)H_0(k|\mathbf{r}'-\mathbf{r}|)$ is simply the free space propagator. The physical interpretation of each of these terms is simple: $G_0(r',\theta';r,\theta)$ is the amplitude to propagate directly from $\mathbf{r}$ to $\mathbf{r}'$, and $G_0(r',2\pi-\theta';r,\theta)$ is the amplitude to go from $\mathbf{r}$ to $\mathbf{r}'$ after specular reflection from the line (see Fig. 4.13). The second term picks up a relative minus sign upon reflection. Thus we have recovered the standard geometrical optics approximation. What remains must be the diffractive contribution.

The contributions from the stationary points are only a bit more difficult to compute. If we perform the change of coordinates $z \to z - \pi$ while shifting Γ_C to the right by π, along with $z \to -z + \pi$ while shifting Γ_D to the left by π, then the sum of the integrals along Γ_C and Γ_D may be written as

$$\frac{1}{16\pi}\int_{\epsilon-i\infty}^{-\epsilon+i\infty} dz\,\frac{H_0^{(1)}\left(k\sqrt{r^2+r'^2+2rr'\cos z}\right)}{\cos\left((z+\phi_\pm)/2\right)}, \quad (4.31)$$

where the contour passes through the stationary phase point at $z = 0$. If we now replace $H_0^{(1)}(z)$ with its asymptotic expansion, we can use the standard result for evaluation along a path of steepest descent to write this as

$$G_{diff}(\mathbf{r}';\mathbf{r}) = -\frac{1}{4\pi}\frac{\sin(\theta'/2)\sin(\theta/2)}{\cos(\phi_-/2)\cos(\phi_+/2)}\frac{e^{ik(r+r')+i\pi/2}}{k\sqrt{rr'}}. \tag{4.32}$$

This can in turn be written as

$$G_{diff}(\mathbf{r}';\mathbf{r}) = G_0^{sc}(\mathbf{r}',\mathbf{r_0})d(\theta',\theta)G_0^{sc}(\mathbf{r_0},\mathbf{r}), \tag{4.33}$$

where

$$d(\theta',\theta) = -\frac{\sin(\theta'/2)\sin(\theta/2)}{\cos\left((\theta'-\theta)/2\right)\cos\left((\theta'-\theta)/2\right)} \tag{4.34}$$

is Keller's diffraction coefficient. Thus we see that diffractive contribution can indeed be thought of as coming from rays that scatter off the endpoint with an angular dependent amplitude.

The importance of including diffractive periodic orbits in semiclassical calculations was first illustrated by Wirzba [77] in a study of the resonances of the two disk system. Using a semiclassical expansion of the exact solution found by Rice and Gaspard [26], he showed that subleading resonances could be attributed to creeping orbits, and he was able to extract their correct prefactors, stability indices, and phases. In a later paper, Vattay *et al.* [73] (see also [59]) noted that such orbits could be directly incorporated into the trace formula and spectral determinant without knowing the exact solution by calculating the semiclassical Green's function in the presence of diffractive sources according to Keller's prescription. More recent papers have focused on semiclassical spectra of both open and closed billiards when the billiard contains one or more sharp corners. A semiclassical calculation of the quasistates supported by the diffractive periodic orbit between the vertices of two infinite wedges was performed in [75]; and semiclassical eigenspectra of a number of closed billiards have been computed, including the cardiod [7], triangles [54], and the sector [65]. The calculation performed for the sector is unique in that it involves calculating the *uniform* diffraction coefficient, which we discuss in the following section.

4.5.2 Uniform extension of Keller's theory

Consider now the special case of a plane wave normally incident on a half-infinite line. As before we assume that the half-line lies along the negative y-axis, and that the

wave is incident from the left. In the limit that $k \to \infty$, geometrical optics predicts that the intensity is discontinuous across the x-axis: above the axis the intensity is uniformly equal to one, while below it is equal to either four or zero, depending on whether we are in front of or behind the line. The x-axis in this example is referred to as a geometric shadow boundary. This is not what we find when we include the effects of diffraction by solving the full scattering problem. For $x < 0$ the intensity develops oscillations along any line perpendicular to the shadow boundary. The magnitude of these oscillations vanishes as $|y| \to \infty$, but near the shadow boundary, they are sufficient to ensure that the intensity is continuous. Similarly for $x > 0$ the intensity develops oscillations above the boundary that match continuously onto a decreasing exponential below the axis. This is, of course, nothing but the Fresnel diffraction pattern. What is important to note is that the diffractive corrections to the geometrical optics prediction are of the same order as the geometrical contributions themselves.

Let us now ask whether it is possible to compute these diffractive corrections using Keller's theory. From Eq. (4.33) we can immediately see this is impossible. In the asymptotic limit, each of the Green's functions in this expression contributes a factor of $1/\sqrt{k}$, so the diffractive terms are down by $1/\sqrt{k}$ compared to the geometric terms. An even more serious problem is that the diffraction coefficient blows up along the x-axis, because along this axis we have $\phi_\pm = \pi$. Since $d(\theta', \theta)$ was found from an asymptotic expansion of the exact solution, which does not blow up, there must be a problem with our derivation of Eq. (4.33). The problem is that in the limit that $\phi_\pm \to \pi$, the pole that gives rise to the geometric optics term approaches the stationary phase point that gives rise to the diffractive term. When this happens, the method used to evaluate the integral along the path of steepest descent is no longer valid. Physically speaking, it is impossible to separate the contributions from two different paths as these paths coalesce.

A corrected diffraction coefficient that gives the proper value of $G_{diff}(\mathbf{r}', \mathbf{r})$ along the shadow boundary was introduced by Kouyoumjian and Pathak [46]. Their correction incorporates Fresnel integrals directly into $d(\theta', \theta)$, and allows them to produce the rather nonuniform distribution of rays required near the boundary. As we shall see in Sec. 4.6, their corrected $d(\theta', \theta)$ is needed to account for the anomalous conductance peaks semiclassically. Thus we now return to our derivation to deal properly with the stationary phase integral in the presence of a pole.

We begin with the asymptotic expansion of Eq. (4.31):

$$\frac{1}{16\pi}\int dz\,\sqrt{\frac{2}{\pi k\sqrt{r^2+r'^2+2rr'\cos z}}}\frac{e^{ik\sqrt{r^2+r'^2+2rr'\cos z}-i\pi/4}}{\cos\left((z+\phi_\pm)/2\right)}. \tag{4.35}$$

This equation is of the general form defined in Eq. (4.39) with $z_s = 0$, $z_p(\phi_\pm) = \eta\pi - \phi_\pm$ and $\phi_0 = \eta\pi$ where $\eta = \pm 1$. Applying Eq. (4.41) we find

$$G_{diff}(\mathbf{r}',\mathbf{r}) = \frac{1}{4}\frac{e^{ik(r+r')+i\pi/4}}{\sqrt{\pi k(r+r')}}\frac{|d(\eta,\phi_-)|}{\cos(\phi_-/2)}K\left(|d(\eta,\phi_-)|\sqrt{\frac{krr'}{r+r'}}\right) - (\phi_- \to \phi_+) \tag{4.36}$$

where $d(\eta,\phi_\pm) = \sqrt{2}\sin(\eta\pi/2 - \phi_\pm)$.

4.5.3 Multiple diffraction

In the example of scattering from a half-infinite line, there was never the possibility of multiple diffractive scattering. However, the Westervelt gate has a number of diffractive scattering centers, so we need to know how to compute the amplitude associated with a classical path with multiple diffractive segments. Naïvely we simply multiply together the coefficients of diffraction for each such scattering event, but this is not the right prescription if one of these scatterers lies along a geometric shadow boundary of another.

To see why this is so, let us reconsider the general construction for finding coefficients of diffraction: begin with exact expression for the Green's function in some simple diffractive geometry, the so-called canonical problem; take its asymptotic expansion, being careful along geometric shadow boundaries; and factor out the incoming and outgoing semiclassical Green's functions. The remaining term is $d(\theta',\theta)$. What is important to note is that the Green's function results from the scattering of a *uniform* circular wave.

Now consider a particular stationary path that involves two successive diffractive scatterings at $\mathbf{r_1}$ and $\mathbf{r_2}$. Away from any geometric shadows, the diffracted component of the wave scattered at $\mathbf{r_1}$ is essentially uniform circular. Thus we can simply multiply the amplitude of this wave by $d(\theta',\theta)$ of the second scatterer to get the total amplitude for this path. However, if the second scatterer lies along a geometric shadow of the first, it will be illuminated by a highly nonuniform wave. Since this violates the assumption of uniform illumination implicit in the construction above, we cannot simply multiply the two $d(\theta',\theta)$ together. When this happens we have to solve a new canonical problem involving double diffraction.

4.6 Anomalous conductance peaks

4.6.1 Diffractive resonances

We note from Fig. 4.3 that the anomalous conductance peaks are only seen when the mirror is no more than one or two wavelengths into the unstable regime. This can now be understand in terms of the enhanced diffraction that occurs near geometric shadow boundaries. When the mirror's center of curvature lies very close to the QPC, the path running from the QPC to one of the mirror's edges coincides almost exactly with the geometric reflection boundary of a wave beginning at the QPC and scattering off the edge. The enhanced diffraction near this boundary provides a mechanism to scatter sufficient amplitude back at the QPC so a resonance can be established. As the mirror moves further into the unstable regime, the diffractive paths no longer lie close enough to the reflection boundary to have any appreciable effect.

4.6.2 Diffractive paths for the Westervelt gate

The diffractive orbits that contribute to the amplitude in Eq. (4.9) begin and end at the center of the QPC and include n diffractive scatterings from the edges of the mirror. They can be written in terms of paths that begin at the QPC and end at one of mirror's edges and their inverses. If we denote these simpler paths as $\gamma_n^{\pm}$, where $n = 0, 1, 2, \ldots, \infty$ is the number of specular bounces from the mirror and $\pm$ denotes whether the path strikes the upper or lower edge, the required orbits are $\prod_{i=1}^{N} \gamma_{n_i'}^{\epsilon_i - 1} \cdot \gamma_{n_i}^{\epsilon_i}$, for $N = 1, 2, 3, \ldots, \infty$ and all sequences $\{n_i, n_i', \epsilon_i\}_{i=1}^{N}$. Note that once the wave has entered the cavity, we treat the QPC as a specular reflector, so there are no extra diffractive paths arising from the QPC.

As is usually the case with semiclassical analysis, we would like to associate each anomalous resonance with only a small subset of these orbits. Now in the stable regime, $\gamma_0^{\pm}$ are the only two diffractive paths. The observed absence of anomalous peaks in the stable regime suggests that multiple retracings of a single diffractive path, in this case $\left(\gamma_0^{\epsilon_i - 1} \cdot \gamma_0^{\epsilon_i}\right)^N$, cannot produce a conductance resonance. The periodicity of the anomalous sequences in the unstable regime further suggests that more than one path is needed for a diffractive resonance. As we move from peak A_2 to A_3, the lengths of the first few paths, $l(\gamma_0)$, $l(\gamma_1)$, $l(\gamma_2)$ do not change by $\lambda_f/2$, as would be expected if this peak were associated only with the retracings of one of these orbits. However, the difference $l(\gamma_1) - l(\gamma_0)$ does

change by one wavelength, suggesting that the A_x sequence arises from an interference among the combinations of these two paths.

There is a computational difficulty if the anomalous peaks are associated with multiple retracings of a small number of orbits. Such orbits necessarily involve multiple diffractive scatterings, and as we noted in Sec. 4.5, there is no simple semiclassical description of multiple diffraction. The semiclassical amplitude for propagating along a path with n diffractive events involves n-fold Fresnel integrals, because the pattern resulting from all but the first scattering results from illumination by a Fresnel pattern rather than a uniform source. However, if the cavity is large enough, the rapid oscillations of the Fresnel pattern may be sufficiently stretched out between scatterings that we can again approximate the illumination as uniform, and simply multiply the diffraction coefficients.

4.6.3 Semiclassical conductance curves

The semiclassical conductance curves can be found in the published version of this chapter.

4.6.4 Wave functions

Wave function patterns similar to those seen in Figs. 4.5 can be produced by scattering two or more plane waves with nearly coincident wave vectors from a straight wall. Thus we would like to say that the anomalous wave functions are a result of interference between the local wave fronts associated with the different $\gamma_n^{\pm}$. However, we may then immediately ask why such patterns are not seen in closed billiards. To see why open and closed billiards are inherently different, we consider the distribution of diffractive paths in the Westervelt gate.

We begin by dividing the $\gamma_n^{\pm}$ into two general classes: those paths that make only a small number of specular reflections from the mirror before scattering from the edge, and all of the remaining paths, which make arbitrarily many reflections before diffractively scattering. The motivation for this division can be found in Fig. 4.15: the positions of the final specular reflections of the paths in the first class are apparently uncorrelated, as are the angles at which the scatter from the edge; the positions of those in the second tend toward a limit point, as do the angles at which they scatter from the edge. The exact number of paths included in the first class is somewhat arbitrary and depends on the angular size

of the mirror. For the gate shown in Fig. 4.15 we will include $\gamma_0^{\pm}$ through $\gamma_4^{\pm}$ in the first class. Since the paths in the second class all strike the edge at essentially the same angle, the local wave fronts associated with these paths will contribute to the overall background amplitude of the wave functions in the cavity. The remaining 5 paths in the first class then interfere to produce the patterns seen in Fig. 4.15. Such a division is generally impossible in closed billiards.

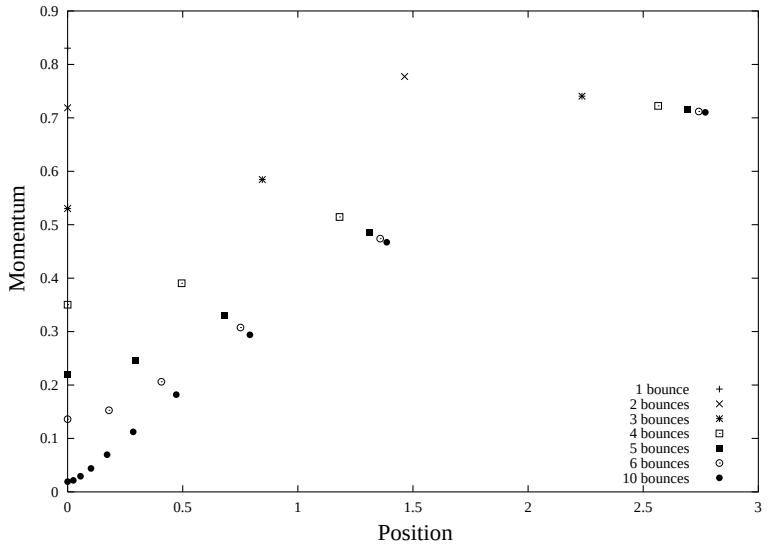

Figure 4.15: Surface of section showing the sequence of specular reflections from the straight walls for the $\gamma_n^{\pm}$ with $n = 1, 2, 3, 4, 5, 6, 10$.

This argument, however, would imply that the anomalous patterns would be seen for any QPC to mirror separation, not just those corresponding to the anomalous resonance peaks. In fact they are, although away from the peaks the bands are not as pronounced nor well defined as those shown in Fig. 4.5. But assuming that we are correct in saying that only a small number of the γ_n are responsible for the peaks, then the multiple traversals along these paths should be sufficient to produce the enhanced patterns. Again the semiclassical wave functions can be found in the published version of this chapter.

4.7 Discussion

4.7.1 Relation to other work

Two other recent series of papers by Delos *et al.* [55, 56, 25] and Sieber *et al.* [63, 64] are also concerned with how to treat the semiclassical approximation uniformly near singularities in the classical dynamics. The Delos papers discuss what to do near a cylindrical cusp, and the Sieber papers show what to do near period n-tupling bifurcations. In both cases the authors proceed by solving canonical problems, much as in Keller's approach to diffraction. It is also interesting that Sieber notes that if two bifurcations occur too close together in parameter space, we must solve a new canonical problem, which takes into account both events simultaneously. This is much like the case with double diffraction; if the second diffractive scatterer lies too near a geometric shadow boundary of the first, you must first solve the two scatterer problem exactly.

4.7.2 Possible experimental realization

An exact physical realization of our numerical gate would correspond to a device with walls that were both infinitely hard and infinitely thin—that is to say unphysical. Thus we must ask to what degree our simulations represent realizable experiments. The device used in the original study of the stable regime was made by patterning copper gates onto a GaAs/Al_xGa_{1-x}As heterostructure. When a negative bias voltage was applied to these gates, the 2DEG underneath them was depleted, making them essentially specular reflectors for electrons. Our numerically one-dimensional mirrors thus begin with a nontrivial width on the order of the gate width. In order to make the mirror move toward the QPC, the voltage on the quarter circle gate was increased, thus increasing the depletion zone within the 2DEG, and effectively moving the gate in toward the QPC. But this, of course, increased the depletion in all directions, and thus our ideal mirrors became even more two-dimensional.

Despite this obvious difference between our numerical device and its experimental realization, their associated conductance curves agree quite well in the stable regime [38]. There is a simple reasons for this: the lemon billiard eigenstates associated with the peaks in the stable regime see only one side of the mirror near the axis of symmetry. The natural widths of these states do depend on the angular extent of the mirror, but since the widths of the experimental peaks are dominated by impurity scattering, even the overall size not

matter. The positions of the anomalous peaks, however, do depend entirely on the positions of the edges of the mirror; and since the angular size of the mirror increases as the mirror moves in toward the QPC, the traces generated here would not correspond to any experimental trace. But if this experiment were ever performed, it would simple enough to perform simulations in which the angular size of the mirror varied linearly with the QPC to mirror separation. More importantly the anomalous peaks depend on the mirror having well-defined edges so that the electrons can diffractively scatter from them; this requires the depletion regions to have decay lengths on the order of a wavelength. There is experimental evidence that the edges are indeed sharp enough for diffractive scattering to occur [39].

4.8 Computing envelopes

In Sec. 4.4.2 we found an equation of the form $F(x, y; \phi) = 0$ that describes how rays reflect from the mirror in the Westervelt gate. For each value of the parameter ϕ, this equation implicitly defines a ray path γ_ϕ in the (x, y) plane. We want to find the equation for the envelope E of these paths.

Let $(x, y) = (x_\phi, y_\phi)$ be the common point of tangency of E and γ_ϕ. We define two auxiliary functions of ϕ whose values are the x and y coordinates of this common point, $t^x(\phi) = x_\phi$ and $t^y(\phi) = y_\phi$. Since the tangent of E at (x_ϕ, y_ϕ) is perpendicular to the normal derivative of the curve defined by the parameter ϕ, we have

$$F_x t^x_\phi + F_y t^y_\phi = 0, \tag{4.37}$$

where the subscripts denote partial derivatives. On the other hand, if we simply take the total derivative of $F = 0$ with respect to ϕ, we find

$$F_x t^x_\phi + F_y t^y_\phi + F_\phi = 0. \tag{4.38}$$

Combining these two equations, we find that the envelope satisfies $F_\phi = 0$.

4.9 Steepest descent near a simple pole

We are interested in the asymptotic evaluation of integrals of the general form

$$\int_\Gamma dz\ F(z)\ e^{ikS(z)} = \int_\Gamma dz\ \frac{f(z)}{z - z_p(\phi)}\ e^{ikS(z)}, \tag{4.39}$$

where $S(z)$ has a first-order stationary phase point at $z = z_s$, Γ is the path of steepest descent through z_s, and $z_p(\phi) \to z_s$ as $\phi \to \phi_0$. We implicitly define a new coordinate t by the equation $S(z) = S(z_s) + it^2$. This maps Γ on the real axis, the stationary point z_s onto $t = 0$, the pole $z_p(\phi)$ onto $t_p^2(\phi) = i\,(S(z_s) - S(z_p(\phi)))$, and the integral onto

$$\int_{-\infty}^{\infty} dt\; G(t)\; e^{ikS(z_s)-kt^2} \text{ with } \mathrm{G(t)} = \mathrm{F(z(t))}\frac{\mathrm{dz}}{\mathrm{dt}}. \tag{4.40}$$

If we expand $G(t) = (t - t_p(\phi))^{-1} \sum_{n=0}^{\infty} g_n t^n$ and retain only the lowest order nonvanishing term in t, we are left with

$$g_0\; t_p(\phi) \int_{-\infty}^{\infty} dt\; \frac{e^{-kt^2}}{t^2 - t_p^2(\phi)} = g_0\sqrt{i\tau} \int_{-\infty}^{\infty} dt\; \frac{e^{-kt^2}}{t^2 - i\tau^2} = g_0\sqrt{i\tau}\; \frac{2\pi}{\tau}\; e^{i\pi/4} K(\sqrt{k\tau^2}), \tag{4.41}$$

where $K(z)$ is the modified Fresnel function,

$$K(z) = \frac{e^{-iz^2+i\pi/4}}{\sqrt{\pi}} \int_z^{\infty} du\; e^{iu^2} \sim \frac{e^{i\pi/4}}{2z\sqrt{\pi}} + o\left(\frac{1}{z^3}\right), \tag{4.42}$$

and for convenience we have defined $t_p^2(\phi) = i\tau^2$.

Chapter 5

Imaging the Single Electron Wave Function in Mesoscopic Structures

The results of a recent study by Katine *et al.* (PRL **79** (24), 4806 (1997)) imply that it should be experimentally possible to realize isolated quantum states in an open mesoscopic device consisting of a quantum point contact coupled to an open resonator. We propose a method for (statistically) imaging the wave functions of these states using an atomic force microscope (AFM). As the dimensions of the cavity are tuned, the conductance of the device exhibits a sequence of sharply defined peaks, each corresponding to resonant conduction through one of the quasistates of the cavity. Our proposed method is to tune the cavity to one of these peaks, scan the AFM tip over the inside of the resonator, and measure either the absolute shift in conductance or the peak shift as a function of tip position. For a tip sufficiently small that it behaves as a perturbative s-wave scatterer, we show that the former is proportional to $\psi(\mathbf{r}_{AFM})$, with a spatially varying coefficient that depends only on the details of the cavity geometry, while the latter is proportional to $|\psi(\mathbf{r}_{AFM})|^2$.

5.1 Introduction

5.1.1 Motivation

Physicists have long wondered whether it will ever be experimentally possible to image a quantum system in a pure state. Even the remarkable images provided by scanning tunneling microscopes (STMs) of individual atoms on atomically flat surfaces, arranged to form everything from quantum corrals to the letters "IBM," do not fit this criterion. The STM is only capable of imaging the local density of states, so these images are the result

of a statistical mixture of the different possible quantum states of the atoms.

The experiments of Sridhar mapping the electromagnetic eigenstates of planar microwave cavities [67] illustrate the kind of imaging we would like to perform for quantum systems. The cavities used in these experiments are defined by one or more strips of conducting material set normally between two closely spaced, parallel conducting plates. Their resonant frequencies are experimentally determined by measuring the transmission spectrum between a pair of wire loops attached to their boundaries. These cavities support a family of modes for which the variation of the magnetic field transverse to the plates is fixed and uniform over the cavity, while the variation of the remaining planar field is governed by the two-dimensional scalar Helmholtz equation. Those modes for which only the lowest mode transverse to the plates can be excited are effectively two dimensional. The spatial variation of these modes in the plane can be mapped by introducing a small conducting sphere into the cavity and measuring the shift in the associated resonant frequency as a function of the position of the sphere. For a sufficiently small sphere, the shift in frequency is proportional to the absolute square of the field strength at the center of the sphere [35].

The motivation behind Sridhar's experiments was to provide an experimental setting to test the predictions of scarring in classically chaotic systems. Since the planar magnetic field satisfies the same equation as the two-dimensional wave function in free space, these cavities provide us with a means of simulating the behavior of a two-dimensional quantum billiard with a purely classical system. In order to imitate Sridhar's experiments in a true quantum system, we require two things: a two-dimensional quantum system in which we can create long-lived, quasi-resonant states, and an analogous means to image them. We propose that a suitable system is provided by the Westervelt gate, illustrated in Fig. 5.1, and that the imaging can be performed using an atomic force microscope (AFM).

Both the quantum and semiclassical transport properties of the Westervelt gate have been the focus of recent studies [38, 19, 16]. The gate itself consists of a subwavelength quantum point contact (QPC) coupled to an open resonator, and is defined by voltage gates patterned on a GaAs/$Al_xGa_{1-x}As$ heterostructure. As shown in [38], it should be experimentally possible to isolate well-defined quasistates in this device by tuning it to one of its conductance resonances, analogous to the method of Sridhar. The motivation behind the use of an AFM tip stems from another recent experiment by Eriksson *et al.* [21] that shows it is possible to use a scanned probe microscope (SPM) to characterize the electronic current flow through a microscale gate. The SPM tip is used to perturb the two-dimensional

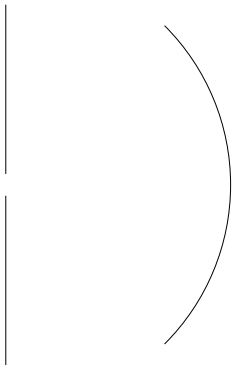

Figure 5.1: Schematic illustration of the Westervelt gate.

electron gas (2DEG) locally within the gate, and the conductance is measured as the tip is scanned over the opening. In these experiments the size of the gate is large compared with the Fermi wavelength, so the conductance may be thought of in terms of classical transport through the gate. However, for gates small enough that the electronic wave function is coherent across the opening, and an SPM tip small enough that it behaves as an s-wave scatterer, the perturbation to the current will be proportional to value of the wave function at the center of the tip, $\psi(\mathbf{r}_{SPM})$, with a spatially varying coefficient of proportionality that depends only on the geometry of the gate.

5.1.2 Outline

We begin in Sec. 5.2 with a review of the electronic conductance profile of the Westervelt gate as a function of QPC-mirror separation, and a discussion of how the resonant structure in this profile should allow us to isolate well-defined quantum states within the cavity. In Sec. 5.3 we consider two possible ways of using the AFM tip to image one of these states: measuring the shift of the corresponding resonant peak, analogous to the procedure used by Sridhar, and measuring the absolute change in the conductance as we move off the peak. In Sec. 5.4 we show a number of simulated AFM scans using the first of these methods, and finally in Sec. 5.5 we discuss whether either of our proposed methods is experimentally feasible.

5.2 Isolated states in the Westervelt gate

A sample numerical conductance trace for the Westervelt gate is shown in Fig. 5.2. Each peak in this trace arises from resonant transport through the gate mediated by one of the quasistates of the cavity. The wave functions associated with the peaks labeled 0_a

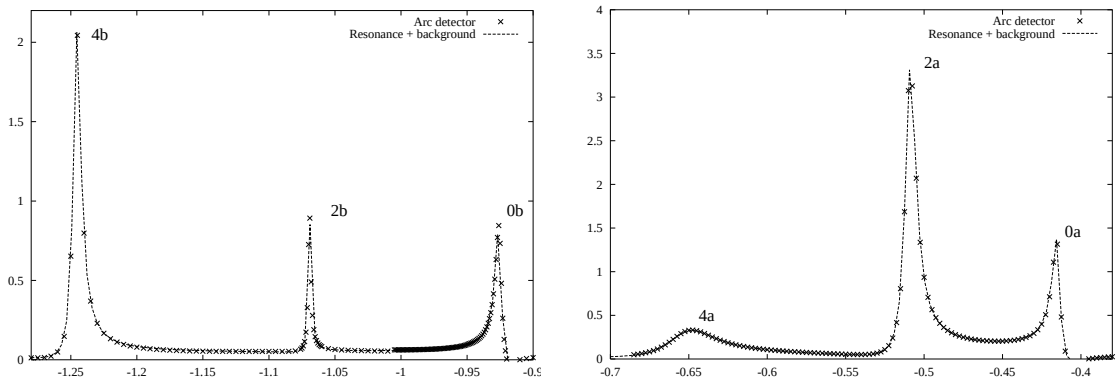

Figure 5.2: Partial conductance traces. The horizontal axis measures the separation between the QPC and the mirror's radius of curvature, in units of Fermi wavelength, and the conductance is normalized to that of the QPC alone. The mirror's radius of curvature is 10λ.

through 0_c have zero angular nodes, those with peaks 2_a through 2_c have two angular nodes, and so on. There are no peaks corresponding to wave functions with an odd number of nodes because of the symmetric placement of the QPC in the cavity. Because of the finite extent of the mirror defining the cavity, the gate can only support quasistates with a small number of the angular nodes. This has the pleasant consequence of making the density of states inside the cavity more one than two dimensional. This one-dimensional character manifests itself in the wide gaps that divide the conductance resonances into clusters. The numerically computed wave functions for 0_a, 2_a, and 4_a are shown in Figs. 5.3. Note that they are quite similar in appearance to the eigenmodes of the *closed* lemon billiard illustrated in [32].

The trace in Fig. 5.2 was computed without any additional damping to model the impurity scattering present in the experimental devices. The widths of the peaks are thus due to the coupling of the quasistates to the continuum via tunneling, both through the QPC and across the caustics that bound the classical motion inside the cavity. The experimental traces have quite a different character: each peak is broadened, and the clusters of narrow peaks seen in Fig. 5.2 are replaced by a single periodic series of peaks [38].

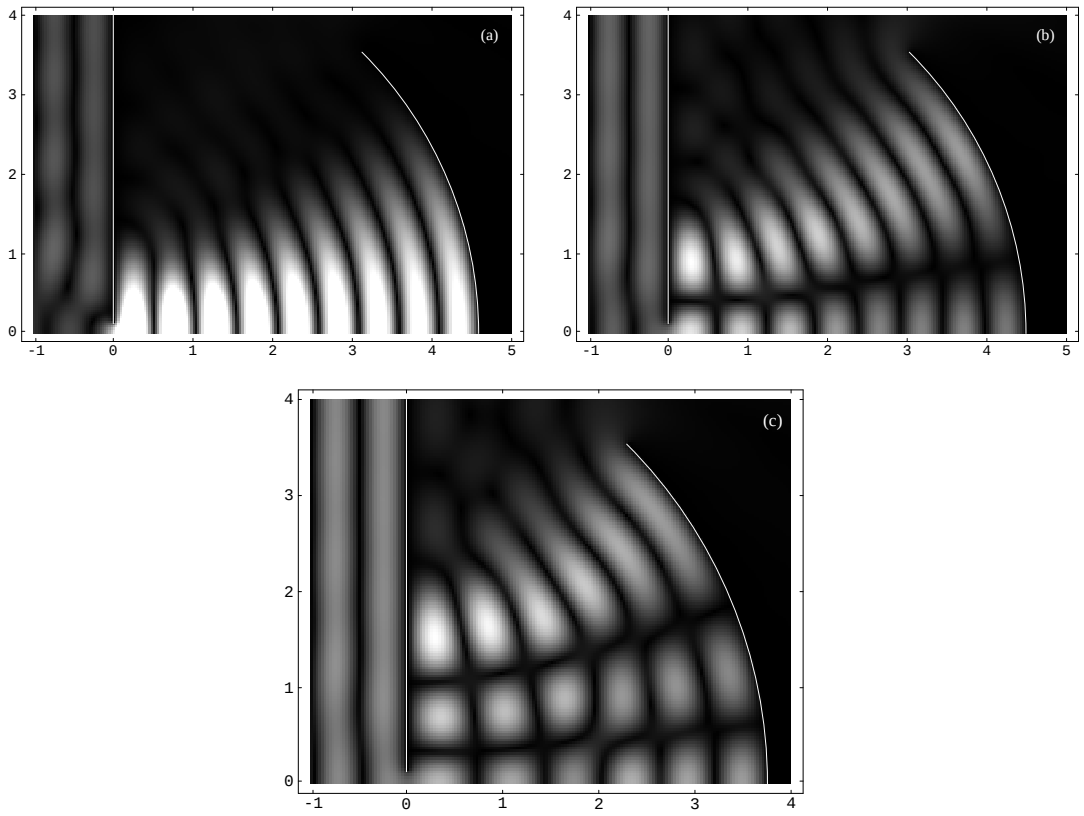

Figure 5.3: Moduli of the wave functions associated with peaks (a) 0_a, (b) 2_a, and (c) 4_a in Fig. 5.2.

Thus, it is currently impossible to isolate individual quantum states in the gate. However, AlGaAs samples with mobilities 10 times higher than those used in [38] have recently become available, and it is expected that with these new materials, we should be able to resolve individual conductance peaks. For the remainder of this paper we will assume that it is experimentally possible to realize single mode quasistates inside the gate.

5.3 Proposed Imaging Methods

5.3.1 Measuring shift of the conductance peak

The first of our proposed imaging methods is analogous to the method used by Sridhar. We tune the Westervelt gate to one of its conductance resonances, scan the AFM tip over the inside of the gate, and measure the shift of the conductance peak on a dense, two-dimensional grid of points. However, it is not clear whether the measured shift is indeed proportional to $|\psi(\mathbf{r}_{AFM})|^2$, since the proof cited by Sridhar holds only for closed devices. In order to show that this proportionality continues to hold in the open case, we offer an alternative to the proof given in [35] that can be extended to open devices. Our proof makes use of the "dressed" T-matrix formalism developed in [51].

Let us consider how best to describe the scattering from the AFM tip in the presence of the gate. Conventional scattering theory gives us the full Green's function $G(E)$ for the combined system in terms of the background Green's function $G_0(E)$ and the T-matrix $T(E)$ that describes the scattering from both the gate and the tip.

$$G(E) = G_0(E) + G_0(E)T(E)G_0(E). \tag{5.1}$$

However, we would like to assume that the scattering from the AFM tip is only a weak perturbation to the strong scattering from the gate. Since the latter is responsible for defining the eigenstates of the cavity, and thus the overall analytic structure of $G(E)$, it would be physically more appealing if this split were made explicit in our equations. With this in mind, we define the dressed T-matrix for the tip, $\mathcal{T}_{afm}(E)$, implicitly with the equation

$$G(E) = G_{gate}(E) + G_{gate}(E)\mathcal{T}_{afm}(E)G_{gate}(E), \tag{5.2}$$

where $G_{gate}(E)$ is the Green's function in the presence of the gate. As shown in [51], $\mathcal{T}_{afm}(E)$ can be expressed in terms of the ordinary T-matrix, $t_{afm}(E)$, for the tip in free

space as

$$\mathcal{T}_{afm}(E) = \frac{1}{1 - t_{afm}(E)[G_{gate}(E) - G_0(E)]}\, t_{afm}(E). \tag{5.3}$$

All of the information concerning the quasistates of the gate is now encoded in $G_{gate}(E)$, while the scattering from the tip is described by $\mathcal{T}_{afm}$.

Let us assume for the moment that $G_{gate}(E)$ describes the Green's function of a closed billiard instead of our open gate. As usual the eigenenergies of the billiard are given by the poles of $G_{gate}(E)$, so if we want to understand how these energies are shifted when we introduce an AFM tip into the billiard, we simply have to find the poles of the full $G(E)$. From Eq. (5.2) we see that the analytic structure of $G(E)$ coincides with that of $\mathcal{T}_{afm}(E)$. Assuming that $t_{afm}(E)$ has no resonant structure of its own, Eq. (5.3) tells us that the poles of $\mathcal{T}_{afm}(E)$ are given by

$$t_{afm}(E)\,[G_{gate}(E) - G_0(E)] = 1. \tag{5.4}$$

In the limit of a weak perturbation, $t_{afm}(E)$ will be very small. This equation will thus only be satisfied if we are near a pole of $G_{gate}(E)$. Since we have assumed that the billiard is closed, and so describes a Hermitean system, we can expand $G_{gate}(E) - G_0(E)$ near the n-th eigenstate of the billiard as

$$G_{gate}(\mathbf{r}', \mathbf{r}; E) - G_0(\mathbf{r}', \mathbf{r}; E) = \frac{\psi_n^\dagger(\mathbf{r}')\,\psi_n(\mathbf{r})}{E - E_n} + O(1), \tag{5.5}$$

where we have restored the explicit spatial dependence. Note that we have also kept the background term because we are interested in evaluating this expression along the diagonal, and only the difference of the two Green's functions is finite there. If we assume that the AFM tip can be treated as a point scatterer, then combining Eq. (5.4) with the expansion in Eq. (5.5), we find $E - E_n \simeq t_{afm}(E_n)|\psi(\mathbf{r}_{afm})|^2$. Now let us open up the billiard just enough to attach two external leads. The resonances in the transmission spectrum between these two leads will coincide with the eigenenergies of the billiard, up to some small correction induced by the local deformation of the eigenstates near the leads; and we thus have the desired result that the AFM tip shifts the conductance peak by an amount proportional to the absolute square of the associated eigenstate at the tip.

Now let us imagine that our closed billiard is actually the Westervelt gate with the mirror continued up to the walls, and ask what happens as we deform the mirror back to its original angular size. Eqs. (5.2) and (5.3) continue to hold, but the expansion in

Eq. (5.5) changes dramatically. The existence of such an expansion depends entirely on the Hermitean nature of the closed billiard. If we think of the Westervelt gate as being simply a scattering system in the infinite plane, then it is certainly Hermitean and admits an expansion of this form. However, such an expansion ignores the fact that the cavity has a resonant structure. Since this structure is the physical reason for the conductance peaks, it should be explicit in our equations. If we restrict our attention to the gate, our system becomes nonhermitean, and we must replace the orthogonal expansion in Eq. (5.5) with a biorthogonal expansion in terms of the now different left and right eigenstates of the Hamiltonian. As in the closed case, a single term will dominate the expansion near a resonance, but we will now be left with a shift proportional to $\psi_L^\dagger(\mathbf{r}_{afm})\psi_R(\mathbf{r}_{afm})$, where the subscripts indicate the left and right eigenstates. We would like to argue that for the states that give rise to the peaks, this can be approximated by $|\psi(\mathbf{r}_{afm})|^2$, where ψ is the eigenstate of the closed gate. Since the Westervelt gate is classically stable and the QPC is smaller than a wavelength, the only way for amplitude to leave the cavity is to tunnel either out through the caustics that bound the classical motion or back through the QPC. Semiclassically these tunneling corrections, which take us from ψ to $\psi_{L,R}$, will be smaller by a factor of $\hbar$ than the contributions from the classical orbits responsible for ψ itself. Thus in the limit of small $\hbar$, the shift should again be proportional to the absolute square of the eigenstate at the tip.

5.3.2 Measuring the change in conductance

As in the previous section, we begin by splitting up our description of the scattering. Let t_{gate} and t_{afm} be the T-matrices for the Westervelt gate and the AFM tip, respectively, in eachothers absence. For convenience we will drop the explicit energy dependence. When the AFM is placed inside the gate, the new T-matrices, T_{gate} and T_{afm}, are given by the pair of coupled equations

$$T_{gate} = t_{gate} + t_{gate}G_0T_{afm}, \tag{5.6}$$

$$T_{afm} = t_{afm} + t_{afm}G_0T_{gate}. \tag{5.7}$$

If we assume that scattering from the AFM tip is weak, then we can expand these two equations, taking into account only a single scattering event from the tip. We find

$$T_{gate} = t_{gate} + t_{gate}G_0t_{afm} + t_{gate}G_0t_{afm}G_0t_{gate}, \tag{5.8}$$

$$T_{afm} = t_{afm} + t_{afm} G_0 t_{gate}. \tag{5.9}$$

The interpretation of these equations should be obvious. The first says an electron can do one of the following: scatter off the old structure; hit the AFM tip once, and then scatter off the old; or hit the old, scatter off the tip and then hit the old again. The interpretation of the second is similar.

The wave function at any point in the plane is given in terms of the T-matrices as

$$\psi = \psi_0 + G_0(T_{gate} + T_{afm})\psi_0, \tag{5.10}$$

and the current can be found from the usual expression $\mathbf{j} = (\hbar/m)\Im(\psi^\dagger \nabla \psi)$. If we combine all of the above expressions, we find that the shift in the current due to the AFM tip is given by

$$\begin{aligned} \delta\mathbf{j} = \frac{\hbar}{m}\Im\Big\{ & (G_0(1 + t_{gate}G_0)t_{afm}(1 + G_0 t_{gate})\psi_0)^\dagger \nabla(\psi_0 + G_0 t_{gate}\psi_0) \\ & + (\psi_0 + G_0 t_{gate}\psi_0)^\dagger \nabla(G_0(1 + t_{gate}G_0)t_{afm}(1 + G_0 t_{gate})\psi_0)\Big\} \end{aligned} \tag{5.11}$$

Let us now evaluate the shift in the normal component of the current through the QPC. The subexpression $(1 + G_0 t_{gate})\psi_0$ is just the original wave function evaluated at the AFM tip; and $\psi_0 + G_0 t_{gate}\psi_0$ is the original wave function at the QPC. Thus we can rewrite the shift as

$$\begin{aligned} \delta j_x = \frac{\hbar}{m}\Im\Big\{ & (G_0(1 + t_{gate}G_0)t_{afm}\psi(\mathbf{r}_{afm}))^\dagger \partial_x \psi(\mathbf{r}_{QPC}) + \\ & \psi(\mathbf{r}_{QPC})^\dagger \partial_x (G_0(1 + t_{gate}G_0)t_{afm}\psi(\mathbf{r}_{afm}))\Big\}. \end{aligned} \tag{5.12}$$

We immediately see that the shift is, up to derivatives, the interference between the original wave function at the QPC and an outgoing s-wave from the AFM tip that reflects in all possible ways from the gate before entering the QPC.

Because of the complicated dependence of δj_x on $\psi(\mathbf{r}_{afm})$, the experimentally measured shifts do not form an image of the wave function inside the cavity. However, the locus of zero shifts does map the nodal lines, and we should be able to approximately reconstruct the wave function from these lines. This is done as follows. If the gate were actually closed, the nodal lines would cross, and the cavity would be effectively partitioned into a patchwork of smaller billiards, with the wave function subject to Dirichlet boundary conditions along each boundary. The wave function in each one of these billiards could be found, up to a multiplicative phase, using the boundary wall method [13, 19]. The global

wave function could then be assembled, again up to an overall phase, by matching the normal derivatives of the local wave functions across each boundary. This global matching would most likely have to be done in a least squares sense that minimizes any inconsistencies when matching around closed loops. This method is not directly applicable to the Westervelt gate because in an open billiard the nodal lines do not cross. However, because the states of interest are only coupled to the continuum via tunneling, we should be able to replace the noncrossings with crossings, up to local errors in the reconstructed wave function over regions whose sizes scale as $\hbar$.

5.4 Numerical simulations

We restrict ourselves here to the second method and compute the absolute shift of the conductance as a function of tip position. [1] All of our numerical simulations were performed using the wavelet-based boundary wall method described in [19]. The potentials defining the gate were approximated as one-dimensional Dirichlet boundaries, and the perturbation to the 2DEG induced by the AFM tip was approximated using circular Dirichlet boundaries of various radii. In Fig. 5.4 we plot the result of a numerical AFM scan taken along the axis of symmetry with the gate tuned to the 0_a peak of Fig. 5.2. As expected, the wave function itself is not well approximated by the scans, but the zeros of the scans performed with the $\lambda/16$ tip agree quite well with the positions of the nodes.

5.5 Conclusions

If we model the AFM tip as a circular Dirichlet boundary, we find that we need a $\lambda/8$ radius tip in order to map the nodes of the wave function. Current technology limits the tip radius to at best $\lambda/2$, so at the moment our method cannot be implemented experimentally. However, there is a question of whether this is a reasonable model of the tip. A real AFM tip does not carve an impenetrable hole out of the 2DEG, but rather only induces a local depression. Perhaps the need for such a small tip is to compensate for the effect of a zero boundary condition. A recent improvement to the original boundary wall

[1] When these two imaging methods were first proposed, it was thought that only the second would be experimentally feasible. However, it is now believed that the first method may, in fact, be more stable. Numerical simulations using this method are currently being performed, and if these results are found to better than those presented here, they will be used in the published version of this chapter.

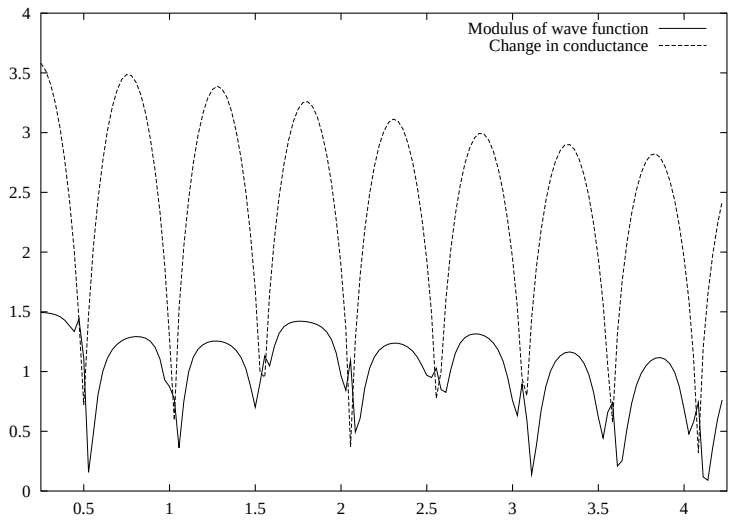

Figure 5.4: Transverse AFM scan of the 0_a peak taken along the axis of symmetry. Since the nodes of the wave function should coincide with the zero conductance shifts, we plot the shift in conductance, $C_{max} - C_{afm}(\mathbf{r})$, along with $|\psi|$. The scan was performed using a $\lambda/16$ diameter tip.

method allows us to extend our calculations to scattering from finite potential steps [51]. We are currently computing both types of scans using tips modeled as finite potentials [20].

Even if we cannot measure the details of the wave function, there is still a possible experimental use for a $\lambda/2$ tip. Another recent paper concerning the behavior of the Westervelt gate [16] predicts that when the circular mirror is pulled out from the QPC beyond its radius of curvature, the gate should support a sequence of resonances due to diffractive scattering off the ends of the mirror. The wave functions associated with these resonances are quite unique: they consist of V-shaped bands of radial nodes extending from the QPC out to the edges of the mirror. It would be interesting to perform a transverse scan of a gate tuned to one of these resonances simply to prove the existence of these states.

Appendix A

Bracket Symmetries of the Classical N=1 String

We show that it is possible to extend Moore's analysis of the classical scattering amplitudes of the bosonic string to those of the N=1 superstring. Using the bracket relations we are able to show that all possible amplitudes involving both bosonic and fermionic string states at arbitrary mass levels can be expressed in terms of amplitudes involving only massless states. A slight generalization of Moore's original definition of the bracket also allows us to determine the 4-point massless amplitudes themselves using only the bracket relations and the usual assumptions of analyticity. We suggest that this should be possible for the higher point massless amplitudes as well.

A.1 Introduction

In a recent pair of papers [53, 52], Moore showed that the classical S-matrix for the flat, twenty-six dimensional bosonic string is uniquely determined up to the string coupling constant. Using a natural algebraic structure on the space of physical string states, he first wrote down an infinite set of linear relations among the exact classical scattering amplitudes for all values of the kinematical invariants. He then used these relations to show that an arbitrary n-point amplitude could be written as a linear combination of n-point tachyon amplitudes at different values of the kinematical invariants. Finally he showed that when coupled with a number of assumptions about the analytic behavior of the scattering amplitudes, these relations could be used to determine the n-point tachyon amplitude itself, thus fixing all n-point amplitudes up to a choice of multiplicative constant

c_n. Since factorization further determines all of these potentially different c_n in terms of the string coupling, Moore actually succeeded in uniquely fixing the full S-matrix up to a choice of the string coupling. The algebraic structure used by Moore was a particular case of a more general one called the Gerstenhaber bracket, which exists for the full BRST cohomology of an arbitrary chiral operator algebra. This particular case is itself called the bracket, and the relations it generates are called bracket relations.

In this paper we will show that Moore's techniques can be extended to the flat, 10-dimensional N=1 superstring. Because of the existence of different ghost pictures, we must begin by showing that bracket is actually well defined. With this done we go on to generate a set of relations among n-point amplitudes involving only massless states. Since the four possible 4-point massless amplitudes can be calculated, these relations can be explicitly checked for $n = 4$, and they indeed hold. We next show that any n-point amplitude involving massive states can be expressed in terms of one of a finite number of massless amplitudes. Using the bracket relations generated by the massless fermionic operator—the generalization of the supersymmetry operator to nonzero momentum—the number of such amplitudes can be reduced to a maximal independent set. For $n = 4$ there are two such amplitudes, and we show that a generalization of Moore's original definition of the bracket allows us to fix the value of both of these amplitudes up to a constant. We suggest that this should be possible for the independent higher point massless amplitudes as well, and thus that the full S-matrix for the N=1 string should be uniquely determined up to a choice of coupling constant. Finally we show that the construction used by Moore in [52] to lift a restriction on n encountered in [53] can also be carried over to the $N = 1$ string.

A.2 Review of Moore's formalism

We begin with a review of the formalism introduced by Moore [53]. For the purpose of this review we will restrict ourselves to the open bosonic string. The extension to the closed string can be found in [53]. It is known that the BRST cohomology of a general chiral operator algebra admits an operation called the Gerstenhaber bracket [49] that maps $\{*,*\} : H^{g_1} \times H^{g_2} \to H^{g_1+g_2-1}$. We are interested in this bracket for the case of $g_1 = g_2 = 1$, since H^1 is just the space of physical string states. We have the following explicit contour

integral representation of the bracket:

$$\{\mathcal{O}_1, \mathcal{O}_2\}(z) = \oint dw\, (b_{-1}\mathcal{O}_1)(w)\mathcal{O}_2(z). \tag{A.1}$$

Note that up to the factor of c implicitly contained in $\mathcal{O}_2$ [1], this is nothing more than the commutator of two dimension-one operators. We also note that since all of the operators for the flat $N = 1$ string contain a factor of the form $e^{ip\cdot X}$, and $e^{iq\cdot X(w)}e^{ip\cdot X(z)} \sim (w - z)^{q\cdot p}e^{i(q+p)\cdot X(z)} + \cdots$, this bracket is only defined for operators whose momenta satisfy $q \cdot p \in \mathbb{Z}$.

The bracket allows us to find relations among scattering amplitudes as follows. We begin by choosing $n + 1$ physical state operators $V_i, i = 1, \ldots, n$, and J, with momenta p_i and q, respectively. We assume that the momenta satisfy $q + \sum_i p_i = 0$ and $q \cdot p_i \in Z$, so that J is mutually local with respect to each of the V_i. A different notation is used for the last operator since it will be used to generate the relations among (derivatives of) the other operators. Now consider the correlation function $\langle 0|V_1(z_1)V_2(z_2)\ldots V_n(z_n)\oint J|0\rangle$, where the contour is taken around a small circle that does not enclose any other operators. Since we are dealing with the open string, this contour should really be restricted to the upper-half plane, but for the moment let us imagine that J can be analytically continued so that the correlator is well defined for a general contour. Since the contour does not enclose another operator, the correlator vanishes. However, if we deform the contour back around infinity, we pick up a contribution from each of the operators equal to just the bracket of that operator with J. Finally if we fix the positions of the first, second, and n-th operators at the points ∞,1, and 0, respectively; and integrate a fixed ordering of the positions of the remaining $n - 3$ operators over the interval $[0, 1]$, we find

$$\sum_{i=1}^{n}(-1)^{q\cdot(p_i+\cdots+p_n)}\mathcal{A}(V_1, \cdots, \{J, V_i\}, \cdots, V_n) = 0, \tag{A.2}$$

where $\{J, V_i\}$ is just the bracket we defined above. These are Moore's bracket relations or finite difference relations, so called because the amplitudes are evalutated at different values of the kinematic invariants.

We must stop here to note that the above procedure can only be carried out for scattering amplitudes involving at most twenty-six V_i. This restriction is easily seen as

[1] Since the physical operators are elements of the level one BRST cohomology, each operator is actually an equivalence class of operators. For the theory at hand we can always choose a representative of the form cV, where V is a dimension one operator.

follows. Our ability to write down the bracket relations rests entirely upon our ability to choose momenta such that their sum is zero, all of the operators are on-shell, and J is mutually local with respect to each of the V_i. If d is the number of noncompact target space dimensions, then for $n \leq d$ the number of independent conditions on the momenta is equal to the number of independent kinematical invariants available to us. Thus we can always find the necessary momenta. However, because of the linear relations among $d+1$ vectors in a d-dimensional space, when $n > d$ the number of conditions is larger than the number of invariants by $n - d$. Thus we are no longer assured of being able to choose the necessary momenta. In his second paper [52] Moore was able to overcome this restriction by embedding the ordinary twenty-six dimensional string in a string theory with $26 + 2m$ target space dimensions. We shall see later that this same trick works for the $N = 1$ string.

A.3 The N=1 String

A.3.1 The spectrum

We begin by recalling the operator content of the $N = 1$ critical superstring [28]. The space $\mathcal{H}$ of physical operators is a direct sum of two subspaces, $\mathcal{H} = \mathcal{H}_b \oplus \mathcal{H}_f$, whose elements have either bosonic or fermionic target space statistics, respectively. Each of these subspaces carries both a discrete and a continuous grading, and may be written as

$$\mathcal{H}_* = \bigoplus_{n \in Z_+} \int_{p \in R^{1,9}} dp\, \mathcal{H}_*[p, n], \tag{A.3}$$

where $\mathcal{H}_*[p, n]$ is the space of operators of momentum p at mass level n. These spaces are null unless $p^2 = -2n$; and when this condition is satisfied, dim $\mathcal{H}_b[p, n] =$ dim $\mathcal{H}_f[p, n] = p_8(n)$, since the spectrum of the string is supersymmetric.

Before we write down any examples of physical operators, let us recall that the spectrum of the $N = 1$ string admits an infinite number of inequivalent, irreducible representations [24]. These representations are labeled by their charge with respect to the field ϕ, where ϕ enters into the bosonization of the (β, γ) system. An operator with charge n is said to be in the n-th ghost picture, or to carry ghost charge n. Bosonic operators always carry integral ghost charge, and fermionic operators half-odd integral charge. We may pass between the different representations using the picture changing operation: $\mathcal{O}_{q+1} = [Q_{BRST}, 2\xi\mathcal{O}_q]$. Since we will be only be interested in computing scattering am-

plitudes, it is only necessary to keep the part of Q_{BRST} that conserves the (b,c) ghost charge, namely $Q_{BRST} = \oint dz\, \frac{1}{2}\psi \cdot \partial X e^{\phi}\eta(z)$. The (-1)- and $(-1/2)$-pictures are called the canonical pictures. Operators take on their simplest form when written in one of these pictures.

The massless spectrum consists of a 10-dimensional target space vector and a 10-dimensional target space spinor. These operators have representatives [24] in the canonical pictures of the form

$$V_{-1}(\zeta,p) = c\zeta\cdot\psi_{(-1)}e^{ip\cdot X} = c\zeta\cdot\psi e^{-\phi}e^{ip\cdot X}, \tag{A.4}$$

$$V_{-1/2}(u,p) = c\bar{u}S_{(-1/2)}e^{ip\cdot X} = c\bar{u}Se^{-\phi/2}e^{ip\cdot X}. \tag{A.5}$$

BRST invariance requires that $p^2 = 0$ for both operators, $\zeta\cdot p = 0$ and $\bar{u}p\cdot\gamma = 0$; and the GSO projection requires that u be a chiral spinor. We will also need the form of the bosonic vertex in the 0-ghost picture. This is given by

$$V_0(\zeta,p) = -c\zeta\cdot(\partial X + ip\cdot\psi\psi)e^{ip\cdot X}. \tag{A.6}$$

The general form of the physical operators at the first massive level has been worked out by Koh *et al.* in [43]. We will not list their results here, but rather simply make use of them as the need arises.

A.3.2 The bracket

We now want to extend the bracket to the $N = 1$ string. Here the existence of an infinite number of inequivalent ghost pictures immediately raises the following question: does the bracket depend on which picture we use? Consider the derivation of the bracket relations, in particular the correlator $\langle 0|V_1(z_1)V_2(z_2)\ldots V_n(z_n)\oint J|0\rangle$. It is always possible to redistribute the ghost charges inside a correlation function, so let us imagine shifting one unit of ghost charge from some V_i to a different V_j, both before and after we perform the contour deformation. Since the resulting relations should be the same, this suggests that taking the bracket of an operator $\mathcal{O}_1$ with the picture changed version of another operator $\mathcal{O}_2$ should be the same as the picture changed version of $\{\mathcal{O}_1,\mathcal{O}_2\}$. For a more direct proof of this, consider the expression $[Q_{BRST}, 2\xi\{\mathcal{O}_1,\mathcal{O}_2\}]$. Since neither $\mathcal{O}_1$ nor $\mathcal{O}_2$ contains η, we can move 2ξ inside the bracket next to $\mathcal{O}_2$. Furthermore since $\mathcal{O}_1$ is a always assumed to be a physical operator, Q_{BRST} sees only $2\xi\mathcal{O}_2$. Thus we can move Q_{BRST} inside the bracket,

leaving us with the desired expression $\{\mathcal{O}_1, [Q_{BRST}, 2\xi\mathcal{O}_2]\}$. Using the relation $\{\mathcal{O}_1, \mathcal{O}_2\} = -(-1)^{q\cdot p}\{\mathcal{O}_2, \mathcal{O}_1\}$, we can also write this as $\{[Q_{BRST}, 2\xi\mathcal{O}_1], \mathcal{O}_2\}$. Thus the picture changing operation commutes with the bracket. Along these same lines, the argument that says we can redistribute the ghost charges inside a correlation function [24] also tells us that we can redistribute the ghost charge inside the bracket. These two properties tell us that the bracket is independent of the ghost pictures of the operators involved. We will see below that this is indeed borne out by example.

Having shown that the bracket is well defined, we want to compute the brackets needed to write down the relations among scattering amplitudes involving only massless operators. Because the bracket is picture independent, we will choose the pictures such that the brackets take on a simple form. The bracket of two massless bosonic operators, with one in the (-1)-picture and the other in the 0-picture, whose momenta q and p satisfy $q \cdot p = 0$, is given by

$$\begin{aligned}
\{V_{-1}(\zeta, q), V_0(\zeta', p)\} &= -\oint dw\, \zeta \cdot \psi_{(-1)} e^{iq\cdot X(w)} c\zeta' \cdot (\partial X + ip \cdot \psi\psi) e^{ip\cdot(z)} \\
&= ic(\zeta \cdot p\zeta' - \zeta' \cdot q\zeta - \zeta \cdot \zeta' p) \cdot \psi_{(-1)} e^{i(q+p)\cdot X(z)} \\
&= i\zeta \cdot pV_{-1}(\zeta', q+p) - i\zeta' \cdot qV_{-1}(\zeta, q+p) \\
&\qquad -i\zeta \cdot \zeta' V_{-1}(p, q+p). \qquad \text{(A.7)}
\end{aligned}$$

To verify that this result is indeed picture independent, let us compute the bracket with both operators in the (-1)-picture.

$$\begin{aligned}
\{V_{-1}(\zeta, q), V_{-1}(\zeta', p)\}(z) &= \oint dw\, \zeta \cdot \psi_{(-1)} e^{iq\cdot X(w)} c\zeta \cdot \psi_{(-1)} e^{iq\cdot X(z)} \qquad \text{(A.8)} \\
&= c(\zeta \cdot \psi\zeta' \cdot \psi - i\zeta \cdot \zeta' q \cdot \partial X) e^{-2\phi} e^{i(q+p)\cdot X(z)}. \qquad \text{(A.9)}
\end{aligned}$$

It is easy to check that the picture changed version of this result is the same as above. Because of this independence we will write

$$\{V_b(\zeta, q), V_b(\zeta', p)\} = i\zeta \cdot pV_b(\zeta', q+p) - i\zeta' \cdot qV_b(\zeta, q+p) - i\zeta \cdot \zeta' V_b(p, q+p), \qquad \text{(A.10)}$$

with the understanding that the ghost picture will be chosen so that the ghost charges balance.

The mixed case of the bracket between a massless boson and a massless fermion of momenta q and p, again with $q \cdot p = 0$, is

$$\{V_b(\zeta, q), V_f(u, p)\} = -\frac{i}{2} V_f((\not{\zeta}\not{q} + \not{p})\not{\zeta} u, q+p). \qquad \text{(A.11)}$$

It is easy to check that both $\{V_0(\zeta,q),V_{-1/2}(u,p)\}$ and $\{V_{-1}(\zeta,q),V_{1/2}(u,p)\}$, and the picture changed version of $\{V_{-1}(\zeta,q),V_{-1/2}(u,p)\}$ all give this answer. Finally the bracket between two massless fermionic operators, again with $q\cdot p=0$, is given by

$$\{V_f(v,q),V_f(u,p)\}=-\frac{1}{\sqrt{2}}V_b(\bar{v}\gamma u,q+p). \tag{A.12}$$

A.3.3 The massless relations

Now let us use these brackets to derive relations among scattering amplitudes involving only massless operators. We begin with the relation generated by $n+1$ massless bosonic operators $\{V_b(\zeta_1,p_1),\ldots,V_b(\zeta_n,p_n);V_b(\zeta,q)\}$, whose momenta satisfy $q\cdot p_i=0$ for all p_i. Again since the bracket is picture independent, we will always specify the operators by giving the spinors, polarization tensors, *etc.* needed to characterize them in the canonical picture, with the understanding that the ghost charges will be chosen so that they sum to -2. The relation associated to these operators is

$$\sum_{i=1}^{n}\zeta\cdot p_i\mathcal{A}_{b\ldots b}\begin{pmatrix}\cdots & \zeta_i & \cdots\\ \cdots & q+p_i & \cdots\end{pmatrix}\Big|s_{ij}\Big)+\sum_{i=1}^{n}\zeta_i\cdot q\mathcal{A}_{b\ldots b}\begin{pmatrix}\cdots & \zeta & \cdots\\ \cdots & q+p_i & \cdots\end{pmatrix}\Big|s_{ij}\Big)$$
$$+\sum_{i=1}^{n}\zeta\cdot\zeta_i\mathcal{A}_{b\ldots b}\begin{pmatrix}\cdots & p_i & \cdots\\ \cdots & q+p_i & \cdots\end{pmatrix}\Big|s_{ij}\Big)=0. \tag{A.13}$$

The finite difference relation among scattering amplitudes involving only massless fermions can be derived from $\{V_f(u_1,p_1),\ldots,V_f(u_n,p_n);V_b(\zeta,q)\}$, where again $q\cdot p_i=0$.

$$\sum_{i=1}^{n}\mathcal{A}_{f\ldots f}\begin{pmatrix}\cdots & (\not{q}+\not{p}_i)\not{\zeta}u_i & \cdots\\ \cdots & q+p_i & \cdots\end{pmatrix}\Big|s_{ij}\Big)=0. \tag{A.14}$$

Finally we can find relations among scattering amplitudes involving both massless bosons and fermions by choosing the generating operator to be fermionic. For example the relation associated with $\{V_f(u_1,p_1),V_f(u_2,p_2),V_f(u_3,p_3),V_b(\zeta_4,p_4);V_f(v,q)\}$ is

$$\mathcal{A}_{ffff}\begin{pmatrix}u_1 & u_2 & u_3 & (\not{q}+\not{p}_4)\not{\zeta}_4 v\\ p_1 & p_2 & p_3 & q+p_4\end{pmatrix}\Big|s,t\Big)+\mathcal{A}_{ffbb}\begin{pmatrix}u_1 & u_2 & \bar{v}\gamma u_3 & \zeta_4\\ p_1 & p_2 & q+p_3 & p_4\end{pmatrix}\Big|s,t\Big)+$$
$$\mathcal{A}_{fbfb}\begin{pmatrix}u_1 & \bar{v}\gamma u_2 & u_3 & \zeta_4\\ p_1 & q+p_2 & p_3 & p_4\end{pmatrix}\Big|s,t\Big)+\mathcal{A}_{bffb}\begin{pmatrix}\bar{v}\gamma u_1 & u_2 & u_3 & \zeta_4\\ q+p_1 & p_2 & p_3 & p_4\end{pmatrix}\Big|s,t\Big)=0. \tag{A.15}$$

Since V_f is the supersymmetry operator at nonvanishing momentum, this should be thought of as a generalized supersymmetry relation. For general n, there are $[(n+1)/2]$ such relations. Since the four possible 4-point scattering amplitudes have all been computed [28], it is possible to check these relations for $n=4$, and indeed they hold. This provides a nontrivial check on our formalism.

A.3.4 The Massive Relations

Having shown that Moore's ideas can be extended to the $N = 1$ string, we want to prove that one of his nontrivial assertions, namely that we can express any n-point amplitude in terms of the n-point tachyon amplitude, has a counterpart in the superstring. Consider the set of n operators $V_b^l = V_b(\zeta, lq), l = 1, \ldots, n$, where $\zeta \cdot q = q^2 = 0$. Combining the bracket with these operators, we have the maps

$$\{V_b^l, *\} : \mathcal{H}_*[p - (n-l)q, n-l] \to \mathcal{H}_*[p - nq, n], \tag{A.16}$$

where we have assumed $q \cdot p = 0$. If we then note that these maps are equivalent to the action of the bosonic DDF operators $\zeta \cdot A_l$, then the no-ghost theorem implies that the map had by summing over l is onto [28]. This means that given any m-point bosonic scattering amplitude involving string states at levels $n_1, \ldots, n_m$, we can always find an l such that the relation generated by $\{V_b(n_1 - l), \ldots, V_b(n_m); V_b^l\}$ expresses our original amplitude in terms of amplitudes involving at least one string state at a strictly lower mass level. Repeating this process we can express any bosonic m-point function in terms of the m-point amplitude for the massless bosonic states. This argument easily generalizes to amplitudes involving an even number of fermionic states. Thus we can say that an arbitrary amplitude can be expressed in terms of one of a finite number of massless amplitudes. For the first few cases $n = 4, 5, 6, 7$ there are naïvely $4, 4, 8, 10$ amplitudes. Using the generalized supersymmetry relations, we can reduce this to $2, 2, 5, 7$ independent amplitudes.

A.4 Recursion Relations

A.4.1 An initial attempt

We would next like to derive a set of recursion relations for the massless scattering amplitudes that would allow us to evaluate them without ever having to go through the usual conformal field theory calculations. We will restrict ourselves to the simplest case of 4-point amplitudes. Following the pattern of Moore's original work, we are tempted to proceed as follows. We choose the generating current to be $V_b(\zeta, q)$. This guarantees that we will not change the number of either bosons or fermions. We then choose the momenta such that one of the amplitudes involves a massive string state. To see why this is necessary, note that in the massless relations derived above, all of the amplitudes are

evaluated at the same values of s and t. These sort of relations are useless if we want to have a recursive method for finding the dependence of the scattering amplitudes on the kinematical invariants. Choosing the momenta such that $q \cdot p_i \neq 0$ for some p_i guarantees that we will have different values of s and t. Finally we try to choose the polarization vectors and spinors such that the massive state is BRST trival, leaving us again with a relation among only massless amplitudes.

We will begin with $\{V_f(u_1, p_1), \ldots, V_f(u_4, p_4); V_b(\zeta, q)\}$, with the momenta chosen such that $p_1 \cdot q = 1$, $p_2 \cdot q = p_3 \cdot q = 0$ and $p_4 \cdot q = -1$. This requires computing $\{V_{-1}(\zeta, q), V_{-1/2}(u, p)\}$ for $q \cdot p = -1$. This is given by

$$\frac{1}{\sqrt{2}}(Y \cdot \partial X + R \cdot \psi \not{\psi}) S_{(-3/2)} e^{i(q+p)\cdot X(z)}, \tag{A.17}$$

where we have defined $Y = iq\bar{u}\not{\zeta}$ and $R = \frac{1}{4}\bar{u}\zeta - \frac{1}{72}\bar{u}\not{\zeta}\gamma$. Note that this is the vertex operator for a *massive* string state. Now we want to choose ζ and possibly a number of the u_i such that the massive vertex is BRST trivial. One of the two possible sets of conditions that render the vertex trivial is given by the pair of equations [43]

$$8R + Y(\not{q} + \not{p}) + \not{R}\gamma = 0, \tag{A.18}$$

$$Y - \not{R}(q + p) - 4R(\not{q} + \not{p}) - \frac{1}{9}\not{Y}\gamma = 0. \tag{A.19}$$

For our particular Y and R, the first of these simplifies to $2\bar{u}\zeta + q\bar{u}\not{\zeta}(\not{q} + \not{p}) = 0$. The only solution of this equation is $\zeta = q$, and in this case all of the other states generated by $V_b(\zeta, q)$ vanish identically. Thus we fail to generate a nontrivial relation. There is another set of conditions on Y and R that also yields BRST trivial states, but these require that $Y \propto q + p$, so we can never write our state as one of these.

Let us try again with $\{V_b(\zeta_1, p_1), \ldots, V_b(\zeta_4, p_4); V_b(\zeta, q)\}$, $p_1 \cdot q = 1$, $p_2 \cdot q = p_3 \cdot q = 0$, and $p_4 \cdot q = -1$. We begin by computing $\{V_{-1}(\zeta, q), V_{-1}(\zeta', p)\}$ for $q \cdot p = -1$:

$$\left\{\zeta \cdot \psi \zeta' \cdot \psi iq\partial X - \frac{1}{2}\zeta \cdot \zeta' \left(iq \cdot \partial^2 X + (iq \cdot \partial X)^2\right) + \zeta \cdot \partial\psi \zeta' \cdot \psi\right\} e^{-2\phi} e^{i(q+p)\cdot X}. \tag{A.20}$$

If we apply the picture changing operator to this we find

$$(\alpha_{\mu\nu\rho}\psi^\mu\psi^\nu\psi^\rho + (\sigma_{\mu\nu} + \alpha_{\mu\nu})\partial X^\mu \psi^\nu + \sigma_\mu \partial\psi^\mu) e^{-\phi} e^{i(q+p)\cdot X}, \tag{A.21}$$

where

$$\alpha_{\mu\nu\rho} = -\frac{i}{6} q_{[\mu}\zeta_\nu \zeta'_{\rho]}, \tag{A.22}$$

$$\sigma_{\mu\nu} + \alpha_{\mu\nu} = -\zeta \cdot p\zeta'_\mu q_\nu + \zeta \cdot q\zeta'_\mu q_\nu - \zeta \cdot \zeta' q_\mu q_\nu - \zeta'_\mu \zeta_\nu, \tag{A.23}$$

$$\sigma_\mu = i\zeta \cdot \zeta' q_\mu - i\zeta' \cdot q\zeta_\mu. \tag{A.24}$$

Here $\sigma_{\mu\nu}$ denotes the symmetric piece and $\alpha_{\mu\nu}$ the antisymmetric. If we choose $\zeta = q$, then this vertex operator vanishes. However, as in the case of four fermions, all of the verticies generated from $V_b(\zeta, q)$ in this case will be BRST trivial. Since we cannot set $\zeta = q$, we must have a nontrivial σ_μ. From [43] we know that σ_μ contributes only to the BRST trivial part of a vertex. Thus we should separate out the contribution from σ_μ and check whether there is any remaining physical piece that cannot be removed. First of all σ_μ must satisfy $\sigma \cdot (p + q) = 0$. This condition can be written as $\zeta \cdot \zeta' + \zeta' \cdot q\zeta \cdot p = 0$. The BRST trivial piece of $\sigma_{\mu\nu}$ is $\sigma^{trivial}_{\mu\nu} = (q + p)_{(\mu}\sigma_{\nu)}$. Subtracting this from the above expression for $\sigma_{\mu\nu}$, we find

$$\sigma^{phys}_{\mu\nu} = -\frac{1}{2}\zeta \cdot p(\zeta'_\mu q_\nu + q_\mu \zeta'_\nu) - \frac{1}{2}(\zeta'_\mu \zeta_\nu + \zeta_\mu \zeta'_\nu) \tag{A.25}$$

$$+\frac{1}{2}\zeta \cdot \zeta'(p_\mu q_\nu + q_\mu p_\nu) - \frac{1}{2}\zeta' \cdot q(p_\mu \zeta_\nu + \zeta_\mu p_\nu). \tag{A.26}$$

With the help of the the above condition we can show that $\sigma^{phys}_{\mu\nu}$ satisfies $(q + p)^\mu \sigma^{phys}_{\mu\nu} = 0$ and $\eta^{\mu\nu}\sigma^{phys}_{\mu\nu} = 0$, so that $\sigma^{phys}_{\mu\nu}$ is indeed a phyical state [43]. The result of this is that we cannot arrange things such that the massive vertex is BRST trivial, and we again fail to find a recursion relation.

A.4.2 A Generalized Bracket

From what we have seen above, it would appear impossible to determine any of the 4-point amplitudes using only the bracket relations. What we want to show now is that we have been too narrow in our thinking. Since the bracket was originally conceived as a special case of the Gerstenhaber bracket, we assumed that it could only be applied to the physical operators of the $N = 1$ string. However there is no reason we cannot consider $\{\mathcal{O}_1, \mathcal{O}_2\}$, where $\mathcal{O}_2$ is a physical operator but $\mathcal{O}_1$ is just a dimension one chiral operator constructed from the fields of our theory, as long as the result is again a physical operator. $\mathcal{O}_1$ must still have dimension one, since otherwise our contour deformation arguments would not go through. To show that there are indeed examples of such operators, let us compute the bracket of a massless fermion and what can be thought of as an on-shell, bosonic string

tachyon:

$$\{V_T(q), V_{-1/2}(u,p)\} = \oint dw\, e^{iq\cdot X(w)} c\bar{u} S_{(-1/2)} e^{ip\cdot X(z)} = V_{-1/2}(u, q+p), \tag{A.27}$$

where we have assumed that their momenta satisfy $q \cdot p = -1$. If we replace the fermion by a boson, we find the analagous result

$$\{V_T(q), V_{-1}(\zeta, p)\} = V_{-1}(\zeta, q+p), \tag{A.28}$$

where again $q \cdot p = -1$.

Since the effect of the tachyon is to shift the momentum of the fermion, this would appear to be just what we need to generate recursion relations. Thus let us consider the relation associated with $\{V_{-1/2}(u_1,p_1), V_{-1/2}(u_2,p_2), V_{-1/2}(u_3,p_3), V_{-1/2}(u_4,p_4); V_T(q)\}$, where $q\cdot p_1 = q\cdot p_2 = q\cdot p_3 = -q\cdot p_4 = 1$. We find

$$\mathcal{A}_{ffff}(s-1, t-1) = \mathcal{A}_{ffff}(s-1,t) + \mathcal{A}_{ffff}(s, t-1). \tag{A.29}$$

This is just the kind of relation found by Moore for the 4-point tachyon amplitude for the bosonic string [53]. As before it is easy to show that this relation is satisfied. Note that in all of the above computations, we never used the picture independent notation of Sec. A.3, but rather made it clear that we were always working in the canonical ghost pictures. This was done for two reasons. First of all our arguments for picture independence do not go through if both operators in the bracket are not physical. Thus we had to restrict ourselves to some definite picture. Second of all, had we worked in say the $(+1/2)$- and 0-pictures, we would have found that the tachyon did *not* map physical states into physical states. The tachyon can be used in the canonical pictures because there the massless states are the product of a tachyon and something that the tachyon does not see.

As with the tachyon amplitude, Eq. (A.29) is not enough to solve for the 4-point fermionic amplitude. Thus let us continue along the path followed by Moore, and consider the bracket of a massless fermionic operator and what would be a physical photon of the bosonic string.

$$\{V_\gamma(\zeta,q), V_{-1/2}(u,p)\} = \oint dw\, \zeta\cdot\partial X e^{iq\cdot X(w)} \bar{u} S_{-1/2} e^{ip\cdot X(z)} \tag{A.30}$$

$$= \begin{cases} -i(\zeta + (\zeta\cdot p)q)\cdot p V_{-1/2}(u, q+p) & q\cdot p = 0 \\ (\zeta + (\zeta\cdot p)q)\cdot\partial X\, \bar{u} S_{(-1/2)} e^{i(q+p)\cdot X} & q\cdot p = -1 \end{cases} \tag{A.31}$$

Since $q^2 = 0$, we can define $\zeta' = \zeta + (\zeta \cdot p)q$ and still preserve the relation $\zeta' \cdot q = 0$. This greatly simplifies the bracket relations that follow. Now consider the relation generated by $\{V_{-1/2}(u_1, p_1), V_{-1/2}(u_2, p_2), V_{-1/2}(u_3, p_3), V_{-1/2}(u_4, p_4); V_\gamma(q)\}$, where $q \cdot p_1 = q \cdot p_2 = 0$ and $-q \cdot p_3 = q \cdot p_4 = 1$:

$$\mathcal{A}_{ff\zeta f}(s,t) = -i\zeta \cdot p_1 \mathcal{A}_{ffff}(s,t) - i\zeta \cdot p_2 \mathcal{A}_{ffff}(s,t-1), \tag{A.32}$$

where the subscript ζ on the left hand side of the relation denotes the massive state $\zeta \cdot \partial X \bar{u} S_{(-1/2)} e^{i(q+p)\cdot X(z)}$. If by some convenient choice of ζ we could render this state BRST trivial, Eq. (A.32) would reduce to the recursion relation necessary to determine $\mathcal{A}_{ffff}$. But according to Koh *et al.* [43], there is no value of ζ for which this state is BRST trivial. Thus we seem to be in the same straits as before. However since the form of this operator is so much simpler than what we previously encountered for massive states, let us blindly proceed for the moment.

We begin by imposing the BRST invariance conditions for the massive state. These imply that $\zeta \cdot (p_3 + q) = \bar{u}_3 \not{q} = \bar{u}_3 \not{\zeta} = 0$. Recall that we also have the initial conditions $q \cdot p_3 = -1$, $q^2 = p_3^2 = \zeta \cdot q = \bar{u}_3 \not{p}_3 = 0$. Now $\mathcal{A}_{ffff}(s,t)$ must be of the form

$$f(s,t)\bar{u}_1 \gamma u_2 \cdot \bar{u}_3 \gamma u_4 + g(s,t)\bar{u}_2 \gamma u_3 \cdot \bar{u}_4 \gamma u_1 \tag{A.33}$$

since $\gamma_{\alpha\beta} \cdot \gamma_{\delta\epsilon}$ and $\gamma_{\beta\delta} \cdot \gamma_{\epsilon\alpha}$ are the only two independent $SO(10)$ invariants with four spinor indicies. Let us set $u_2 = u_3$. This restricts our attention to the unknown function $f(s,t)$, and also implies that $\bar{u}_3 \not{p}_2 = 0$. Using $\bar{u}_3 \not{\zeta} = 0$ along with the other conditions on u_3, we find that the most general form of ζ is $\zeta = q + \alpha p_2 + \beta p_3$. Using $\zeta \cdot (q + p_3) = 0$ we can then solve for β to find $\zeta = q + \alpha p_2 + (\alpha t - 1)p_3$. Thus we are left with a one-parameter family of massive states. Now let us recall that we want to get rid of the massive state. Since we cannot choose ζ such that the massive state is BRST trivial, perhaps we can choose α such that the amplitude involving the massive state vanishes.

The piece of the scattering amplitude involving only X and its derivatives is given by

$$\begin{aligned} &\langle e^{ip_1 \cdot X(z_1)} e^{ip_2 \cdot X(z_2)} \zeta \cdot \partial X e^{ip_3 \cdot X(z_3)} e^{ip_4 \cdot X(z_4)} \rangle \\ &= \left(\tfrac{\zeta \cdot p_1}{z_3 - z_1} + \tfrac{\zeta \cdot p_2}{z_3 - z_2} + \tfrac{\zeta \cdot p_4}{z_3 - z_4} \right) \langle e^{ip_1 \cdot X(z_1)} e^{ip_2 \cdot X(z_2)} e^{ip_3 \cdot X(z_3)} e^{ip_4 \cdot X(z_4)} \rangle. \end{aligned} \tag{A.34}$$

If we fix the $SL(2,\mathbb{C})$ symmetry as usual, then this reduces to

$$\left(-\frac{\zeta \cdot p_2}{1-x} + \frac{\zeta \cdot p_4}{x} \right) x^s (1-x)^t, \tag{A.35}$$

where we have also included the contributions of the (b, c) ghost system. If we note that the remaining contribution to the scattering amplitude does not contain any explicit factors of s or t, then we can think of the denominators x and $1-x$ as effectively shifting s and t; and requiring this amplitude to vanish implies that $f(s,t)$ must satisfy the additional relation

$$\zeta \cdot p_2 f(s, t-1) = \zeta \cdot p_4 f(s-1, t). \tag{A.36}$$

Since ζ depends on only a single parameter, with this equation we should have enough relations both to find the needed α and to solve for $f(s,t)$.

Assuming that the desired α does indeed exist, Eq. (A.32) becomes

$$\zeta \cdot p_1 f(s,t) = -\zeta \cdot p_2 f(s, t-1). \tag{A.37}$$

Combining this with Eq. (A.36) we find

$$\zeta \cdot p_1 f(s,t) = -\zeta \cdot p_4 f(s-1, t). \tag{A.38}$$

If we use these equations to write the right hand side of Eq. (A.29) in terms of $f(s-1, t-1)$, the resulting equation allows us to solve for α, with the result $\alpha = 1/(s+t)$. With this value of α Eqs. (A.37) and (A.38) become

$$f(s,t) = \frac{t}{s+t} f(s, t-1) = \frac{s-1}{s+t} f(s-1, t). \tag{A.39}$$

Using these recursion relations we find

$$f(s,t) = c_4 \frac{\Gamma(s)\Gamma(t+1)}{\Gamma(s+t+1)} \tag{A.40}$$

in agreement with the known result. Of course our relations only allow us to determine $f(s,t)$ for integer s and t, but as in [53] we can analytically continue our result to the entire complex plane by using two mild assumptions about the analyticity of the amplitudes.

In order to solve for the unknown function $g(s,t)$, we simply repeat the entire process, setting $u_1 = u_2$ rather than $u_2 = u_3$. This yields $g(s,t) = c_4' \Gamma(s)\Gamma(t+1)/\Gamma(s+t+1)$, again in agreement with the known result. Finally we can use the cyclical symmetry of the amplitude to fix $c_4' = c_4$. Our generalized bracket can also be used to compute $\mathcal{A}_{bfbf}$, and then the two relations of the form A.15 can be used to determine the two remaining amplitudes $\mathcal{A}_{bbff}$ and $\mathcal{A}_{bbbb}$. This completes the solution of the 4-point massless amplitudes using only the bracket.

A.4.3 Extension to $n > 4$

Let us begin with the $2n$-point massless fermionic amplitude for $n > 3$. The generalizations of Eq. (A.29) are generated by

$$\{V_{+1/2}(p_1), \ldots, V_{+1/2}(p_{n-3}), V_{-1/2}(p_{n-2}), \ldots, V_{-1/2}(p_{2n}); V_T(q)\}, \tag{A.41}$$

with $q \cdot p_i = 1$ for $i = 1, \ldots, n-1$ and $q \cdot p_i = -1$ for $i = n, \ldots, 2n$, and its various permutations. Note that we omit the polarization tensors for simplicity's sake. These relations should serve as the analogues of Moore's triangle relations [53], which serve to reduce the $2n$-point amplitude from a function of $n(2n-3)$ variables to one of $2n-3$ variables. The generalizations of Eq. (A.32) are the relations generated by

$$\{V_{+1/2}(p_1), \ldots, V_{+1/2}(p_{n-3}), V_{-1/2}(p_{n-2}), \ldots, V_{-1/2}(p_{2n}); V_\gamma(q)\}, \tag{A.42}$$

with $q \cdot p_i = 1$ for $i = 1, \ldots, n$ and $q \cdot p_i = -1$ for $i = n+1, \ldots, 2n$, and its permutations, along with the relations constraining the amplitudes involving massive states to vanish. These would then be used to solve for the dependence on the remaining variables. The six-point amplitude has in fact been worked out in [45] using a rather involved technique developed in [44] for computing correlators. It would be an interesting exercise to try to use the above bracket relations to reproduce their results.

Similar relations to these can also be generated for amplitudes involving either one or two bosonic operators. However, when we try to use the tachyon to generate relations for amplitudes with three or more bosons, we find that it is impossible to do so. Now as we saw in the previous subsection, one of the problems with using the generalized bracket is that it is not picture independent. For the particular operators we chose to use, V_T and V_γ, we had to restrict ourselves to the canonical picture. When we couple this restriction with the necessity of having to choose representatives of operators whose ghost charges sum to -2, we find that the two requirements are sometimes incompatible. For example in the case of three or more bosons, no matter how we choose the ghost pictures and the momenta, we end up having to compute the bracket of the tachyon with a state in a picture other than the canonical picture, which as we mentioned does not give us back a physical state. When we have four or more bosons, we can no longer even use the photon. One possible way of avoiding these restrictions would be to use higher mass states from the bosonic string to generate relations. Another possibility would be to allow the tachyon and photon to

generate higher mass states of the $N = 1$ string. While the former approach seems quite tenable, the latter is hampered by the fact that the BRST analysis for the $N = 1$ string has not yet been perfomed at even the second massive level.

A.5 Lifting the Restriction $n \leq 10$

In his original work [53] Moore found that he could only apply his bracket relations to scattering amplitudes involving at most tweny-six string states. The origin of this restriction was explained in Sec. A.2. In [52] he showed how this restriction could effectively be lifted. This was done as follows. The general form of the restriction on n is that it be no greater than the number of uncompactified target space dimensions. Thus the idea that immediately presents itself is to embed our theory in another with a larger target space. The problem with this is that doing so takes us off criticality. However we can restore the central charge to zero by tensoring our enlarged theory with a number of ghost systems.

To be more exact, let us denote by $C(M)$ the conformal field theory with target space M; e.g., $C(\mathbb{R}^{1,25})$ is the theory that underlies the ordinary critical bosonic string. Moore embedded the theory based on $C(\mathbb{R}^{1,25})$ into that based on $C(R^{1+E,25+E}) \bigotimes_{i=1}^{E} [< \xi_i, \eta_i > \cap \ker(\oint \eta_i)]$, where $< \xi_i, \eta_i >$ is a spin $(0,1)$ fermionic ghost system. This allowed him to compute amplitudes involving up to $26 + 2E$ string states, and since E is arbitrary, this effectively removed the restriction on n. Here we want to show that the same procedure also works for the $N = 1$ string if we extend the ghost fields to $N = 1$ superfields. We denote by $SC(M)$ the superconformal field theory (SCFT) with target space M. Since the $N = 1$ string is a SCFT with target space $\mathbb{R}^{1,9}$, in order to apply Moore's construction we need to extend not only the bosonic target space fields but also their fermionic partners. To cancel the contribution to the central charge from these extra fields, we will tensor our theory with an additional E copies of a spin $(1/2, 1/2)$ bosonic ghost system $< \tilde{\xi}, \tilde{\eta} >$. Note that this system does not have a $U(1)$ anomaly, so there is no need to restrict ourselves to a subspace of this theory. We can combine these two ghost systems into a single superghost system described by $\Xi = \xi + \theta\tilde{\xi}$ and $H = \tilde{\eta} + \theta\eta$, and thus our full theory may be written as $SC(\mathbb{R}^{1+E,9+E}) \bigotimes_{i=1}^{E} [< \Xi_i, H_i > \cap \ker(\oint \int d\theta H_i)]$.

Once we have solved for the amplitudes in this extended theory, how do we pull out the amplitudes for our original theory? We begin by restricting ourselves to amplitudes involving states built only from the ordinary (supersymmetric) matter fields. Since the

bracket is closed with respect to such states, we need not worry about states involving ghost fields propagating in any of the intermediate channels; after all the coeffecients of the bracket relations are nothing more than the three-point functions. Having computed these amplitudes, which live in a $10+2E$ dimensional target space, we simply continue our results to momenta whose last $2E$ components vanish.

A.6 Discussion

In this paper we have succeeded in extending most of Moore's original analysis to the the $N = 1$ string, thus doing the second of his things that "we should do." This required a slight extension of the notion of the bracket to include the bracket between a physical operator and an ordinary dimension one chiral field. Another way of thinking about this extension is to say that the $N = 1$ string admits an algebra of external operators, or that the states of the $N = 1$ string (in the canonical picture) form a module over this external algebra. The most glaring shortcoming of the present work is a lack of a general proof that the massless amplitudes are computable entirely in terms of the bracket. However, being that the 4-point amplitude is computable, we have every confidence that the general amplitude should be so as well.

Having shown that the bracket can be extended to the $N = 1$ string, we can now begin to ask questions about more interesting theories such as the type II superstring and the heterotic string. Furthermore the bracket should be applicable not only to the flat backgrounds studied up until now, but also to more general conformal field theories as long as they contain some noncompact sector. In particular we could consider the above-mentioned theories, but compactified on some nontrivial, internal conformal field theory; or we could even consider noncritical theories. In any of these examples it would be interesting to see to what extent the bracket determines the S-matrix.

Bibliography

[1] V. I. Arnol'd. *Catastrophe Theory*. Springer-Verlag, New York, 1984.

[2] Harold U. Baranger and A. Douglas Stone. Electrical linear-response theory in an arbitrary magnetic field: A new Fermi-surface formulation. *Phys. Rev. B*, 40(12):8169–8193, 1989.

[3] Alex Barnett, Areez Mody, and E. J. Heller. To be submitted.

[4] M. V. Berry and C. Upstill. Catastrophe optics: Morphologies of caustics and their diffraction patterns. *Progress in Optics*, XVIII:257–346, 1980.

[5] G. Beylkin, R. Coifman, and V. Rokhlin. Fast wavelet transforms and numerical algorithms I. *Comm. Pure Appl. Math.*, 44:141–183, 1991.

[6] J. H. Bramble, Z. Leyk, and J. E. Pasciak. The analysis of multigrid algorithms for pseudo-differential operators of order minus one.

[7] Henrik Bruus and Niall D. Whelan. Edge diffraction, trace formulae and the cardiod billiard. *Nonlinearity*, 9:1023–47, 1996.

[8] M. Büttiker. Four-terminal phase-coherent conductance. *Phys. Rev. Lett.*, 57:1761–4, 1986.

[9] A. Cohen, I. Daubechies, and J.-C. Feauveau. Biorthogonal bases of compactly supported wavelets. *Comm. Pure Appl. Math.*, XIV:485–560, 1992.

[10] I. Daubechies and W. Sweldens. Factoring wavelet transforms into lifting steps. Technical report, Bell Laboratories, Lucent Technologies, 1996.

[11] Ingrid Daubechies. *Ten Lectures on Wavelets*. SIAM, Philadelphia, 1992.

[12] Geoffrey M. Davis. A wavelet-based analysis of fractal image compression. 1997.

[13] M. G. E. de Luz, A. S. Lupu-Sax, and E. J. Heller. Quantum scattering from arbitrary boundaries. *Phys. Rev. E*, 56(3):2496–507, 1997.

[14] Hai Deng and Hao Ling. Fast solution of electromagnetic integral equations using adaptive wavelet packet transform. 1997.

[15] David L. Donoho. *Recent Advances in Wavelet Analysis*, chapter Smooth Wavelet Decompositions with Blocky Coefficient Kernels. Academic Press, 1993.

[16] Jonathan D. Edwards, Michael Haggerty, and E. J. Heller. Diffractive resonances in open mesoscopic cavities. chapter 4 of dissertation.

[17] Jonathan D. Edwards, Michael R. Haggerty, and E. J. Heller. Work in progress.

[18] Jonathan D. Edwards and E. J. Heller. Work in progress.

[19] Jonathan D. Edwards and E. J. Heller. Wavelet analysis of two dimensional quantum scattering. chapter 3 of dissertation.

[20] Jonathan D. Edwards, Adam. S. Lupu-Sax, and E. J. Heller. Work in progress.

[21] M. A. Eriksson, R. G. Beck, M. Topinka, J. A. Katine, R. M. Westervelt, K. L. Campman, and A. C. Gossard. Cryongenic scanning probe characterization of semiconductor nanostructures. *Appl. Phys. Lett.*, 69(5):671–3, 1996.

[22] G. Fernández, S. Periaswamy, and Wim Sweldens. LIFTPACK: A software package for wavelet transforms using lifting. In M. Unser, A. Aldroubi, and A. F. Laine, editors, *Wavelet Applications in Signal and Image Processing IV*, pages 396–408. Proc. SPIE 2825, 1996.

[23] Roland W. Freund and Noël M. Nachtigal. Qmrpack: A package of qmr algorithms. *ACM Transactions on Mathematical Software*, 22(1):46–77, 1996.

[24] D. Friedan, E. Martinec, and S. Shenker. Conformal invariance, supersymmetry and string theory. *Nucl. Phys. B*, 271:93, 1986.

[25] J. Gao and J.B. Delos. Quantum manifestations of bifurcations of closed orbits in the photoabsorption spectra of atoms in electric fields. *PRA*, 56(1):356–61, 1997.

[26] P. Gaspard and S. A. Rice. Exact quantization of the scattering from a classically chaotic repellor. *Journal of Chemical Physics*, 90(4):2255–62, 1989.

[27] Gene H. Golub and Charles F. Van Loan. *Matrix Computations.* Johns Hopkins University Press, Baltimore, 1996.

[28] M. Green, J. Schwartz, and E. Witten. *Superstring Theory.* Cambridge University Press, 1987.

[29] Martin C. Gutzwiller. *Chaos in Classical and Quantum Mechanics.* Springer-Verlag, New York, 1990.

[30] W. Hackbusch. *Integral Equations: Theory and Numerical Treatement.* Birkhäuser Verlag, Boston, 1995.

[31] E. J. Heller. Bound-state eigenfunctions of classically chaotic hamiltonian systems: scars of periodic orbits. *Phys. Rev. Lett.*, 53(16):1515–18, 1984.

[32] E. J. Heller and S. Tomsovic. Postmodern quantum mechanics. *Physics Today*, 46(7):38–46, 1993.

[33] Ishio Hiromu. Private communication.

[34] M. Holschneider. *Wavelets : An analysis tool.* Clarendon Press, New York, 1995.

[35] L. C. Maier Jr. and J. C. Slater. Field strength measurements in resonant cavities. *J. Appl. Phys.*, 23(1):68–77, 1952.

[36] Gerald Kaiser. *A Friendly Guide to Wavelets.* Birkhaüser, Boston, 1994.

[37] J. A. Katine, M. J. Berry, R. M. Westervelt, and A. C. Gossard. Measurement of the electronic dephasing time via the weak localization effect in a narrow channel. *Superlatt. and Microstruct.*, 16(2):211–215, 1994.

[38] J. A. Katine, M. A. Eriksson, A. S. Adourian, R. M. Westervelt, J. D. Edwards, A. Lupu-Sax, E. J. Heller, K. L. Campman, and A. C. Gossard. Point contact conductance of an open resonator. *Phys. Rev. Lett.*, 79(24):4806–9, 1997.

[39] Jordan Katine. Private communication.

[40] Jordan Asher Katine. *Electronic quantum interference in ballistic semiconductor nanostructures*. PhD thesis, Harvard University, 1996.

[41] Joseph B. Keller. Diffraction by an aperture. *J. Appl. Phys.*, 28(4):426, 1957.

[42] Joseph B. Keller. Geometrical theory of diffraction. *J. Opt. Soc. Am.*, 52(2):116–30, 1962.

[43] I. G. Koh, W. Troost, and A. Van Proyen. Covariant higher spin vertex operators in the ramond sector. *Nucl. Phys. B*, 292:201, 1987.

[44] V. A. Kostelecky, O. Lechtenfeld, W. Lerche, S. Samuel, and S. Watamura. Conformal techniques, bosonization and tree-level string amplitudes. *Nucl. Phys. B*, 288:173, 1987.

[45] V. A. Kostelecky, O. Lechtenfeld, S. Samuel, K. Verstegen, S. Watamura, and D. Sahdev. The six-fermion amplitude in the superstring. *Phys. Lett. B*, 183:299, 1987.

[46] Robert G. Kouyoumjian and Prabhkar H. Pathak. A uniform geometrical theory of diffraction for an edge in a perfectly conducting surface. *Proc. IEEE*, 62(11):1448–61, 1962.

[47] L. D. Landau and E. M. Lifshitz. *Quantum Mechanics—Non-relativistic Theory*. Pergamon Press, New York, 1977.

[48] R. Landauer. *IBM J. Res. Dev.*, 1:223, 1957.

[49] B. Lian and G. Zuckerman. New perspectives on the brst algebraic structure of string theory. *CMP*, 154:613, 1993. hep-th/9211072.

[50] Adam Lupu-Sax. Private communication.

[51] Adam S. Lupu-Sax and E. J. Heller. To be submitted.

[52] G. Moore. Addendum to: Symmetries of the bosonic string s-matrix. YCTP-P1-94, RU-94-19; hep-th/9404025.

[53] G. Moore. Symmetries of the bosonic string s-matrix. YCTP-P19-93; hep-th/9310026.

[54] Nicolas Pavloff and Charles Schmit. Diffractive orbits in quantum billiards. *Phys. Rev. Lett.*, 75(1):61–4, 1995.

[55] A. D. Peters, C. Jaffe, and J. Delos. Closed-orbit theory and the photodetachment cross section of H^- in parallel electric and magnetic fields. *PRA*, 56(1):331–44, 1997.

[56] A. D. Peters, C. Jaffe, J. Gao, and J. Delos. Quantum manifestations of bifurcations of closed orbits in the photodetachment cross section of H^- in parallel fields. *PRA*, 56(1):345–55, 1997.

[57] Tim Poston and Ian Stewart. *Catastrophe Theory and its Applications*. Pitman, San Francisco, 1978.

[58] William H. Press, Brian P. Flannery, Saul A. Teukolsky, and William T. Vetterling. *Numerical Recipes in C*. Cambridge University Press, New York, 1992.

[59] Per E. Rosenqvist, Gábor Vattay, and Andreas Wirzba. Application of the diffraction trace formula to the three-disk scattering system. *J. Stat. Phys.*, 83(1/2):243–57, 1996.

[60] J. J. Sakurai. *Advanced Quantum Mechanics*. Addison-Wesley, Reading, MA, 1993.

[61] C. D. Schwieters, J. A. Alford, and J.B. Delos. Semiclassical scattering in a circular semiconductor microstructure. *Phys. Rev. B*, 54(15):10652–10668, 1996.

[62] K. Seeger. *Semiconductor Physics*. Springer-Verlag, New York, 1991.

[63] M. Sieber. Bifurcations of periodic orbits and uniform approximations. chao-dyn/9701022.

[64] M. Sieber. Uniform approximation for period-quadrupling bifurcations. chao-dyn/9708013.

[65] Martin Sieber, Nicolas Pavloff, and Charles Schmit. Uniform approximation for diffractive contributions to the trace formula in billiard systems. *Phys. Rev. E*, 55(2):2279–99, 1997.

[66] A. Sommerfeld. *Optics*. Academic Press Inc., New York, 1954.

[67] S. Sridhar. Experimental observation of scarred eigenfunctions of chaotic cavities. *Phys. Rev. Lett.*, 67(7):785–8, 1991.

[68] Jakob J. Stamnes and Bjørn Spjelkavik. Evaluation of the field near a cusp of a caustic. *Optica Acta*, 30(9):1331–1358, 1983.

[69] W. Sweldens. The lifting scheme: A custom-design construction of biorthogonal wavelets. *Appl. Comput. Harmon. Anal.*, 3(2):186–200, 1996.

[70] W. Sweldens. The lifting scheme: A construction of second generation wavelets. *SIAM J. Math. Anal.*, 29(2), 1997.

[71] W. Sweldens and P. Schröder. Building your own wavelets at home. In *"Wavelets in Computer Graphics," ACM SIGGRAPH Course Notes*, 1996.

[72] T. J. Thornton, M. L. Roukes, A. Scherer, and B .P. van der Gaag. Boundary scattering in quantum wires. *Phys. Rev. Lett.*, 63(19):2128–31, 1989.

[73] Gábor Vattay, Andreas Wirzba, and Per E. Rosenqvist. Periodic orbit theory of diffraction. *Phys. Rev. Lett.*, 73(17):2304–7, 1994.

[74] N. D. Whelan. Geometric and diffractive orbits in the scattering from confocal hyperbolas. *Phys. Rev. E*, 51(4):3778–81, 1995.

[75] N. D. Whelan. Semiclassical quantization using diffractive orbits. *Phys. Rev. Lett.*, 76(15):2605–8, 1996.

[76] Mladen Victor Wickerhauser. *Adapted Wavelet Analysis from Theory to Software*. A. K. Peters, Wellesley, MA, 1994.

[77] A. Wirzba. Validity of the semiclassical periodic orbit approximation in the two- and three-disk problems. *Chaos*, 2(1):77–83, 1992.