Theories of Imaging Electrons in Nanostructures

A thesis presented
by

Jian Huang

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
May 2006

Thesis advisor

Eric J. Heller

Author

Jian Huang

Theories of Imaging Electrons in Nanostructures

Abstract

Experiments on nanostructures have yielded a very rich body of data. Very often the experimental temperature is low enough to make S-wave scattering applicable, and the coherence length is significant enough such that phase factors play an important role. In this thesis, multiple scattering theory (MST) is applied to studying first an acoustic experiment and then 2-dimensional electron gases (2DEGs) in nanoscale systems; a finite difference method (FDM) is employed to solve the Schrodinger equation for a single-electron quantum dot with various boundary conditions; numerical techniques are developed to study cusps in 2DEGs and their tip scan images. In total, five closely related projects are described in this thesis.

In the first project, starting with the usual Lippman-Schwinger formulation of MST, a perturbative MST is developed to express in terms of the original field the changes induced by removing, adding, shifting, and changing a scatterer. The formalism is intuitive and easy to implement numerically.

In the second project, time-reversal focusing in sound scattering is studied in terms of MST. Many important features of the experimental system are revealed for the first time. An interpretation is provided to explain some recent experimental observations.

In the third project, MST is applied to 2DEGs in the S-wave scattering limit. Many novel conductance fringe patterns are revealed and a general formulation is given for these patterns. Also provided are estimation formulas of thermal width for the circular, elliptical, and hyperbolic fringes, which are found to be consistent with numerical results.

In the fourth project, first FDM is employed to solve for ground state energy of a single-electron quantum dot, which leads to its conductance peak patterns under tip scan. Singular value decomposition (SVD) is then utilized to invert matrix to extract electron wavefunction out of the experimental conductance data. The stability and sensitivity to boundary conditions of FDM and SVD are also studied.

In the fifth project, numerical techniques for generating classical and quantum cusps in 2DEGs are developed. Their tip scan images are analyzed to show interesting dynamics in imaging process. The classical folding of quantum cusp is demonstrated at short wavelength. More complicated, overlapping cusps are shown as well. Most importantly, experimental cusps are matched nicely by numerical results from the Kirchhoff method.

Contents

List of Figures

Citations to Previously Published Work

The majority of the content of the thesis has appeared in the following papers (in chronological order):

"A perturbative multiple scattering theory", J. Huang, *Physics Letters A* **322**, 10-18 (2004);

"Imaging a single-electron quantum dot", P. Fallahi, A.C. Bleszynski, R.M. Westervelt, J. Huang, J.D. Walls, E.J. Heller, M. Hanson, A.C. Gossard, *Nano Letters* **5**, 223 (2005);

"Imaging electrons in a single-electron quantum dot", P. Fallahi, A.C. Bleszynski, R.M. Westervelt, J. Huang, J.D. Walls, E.J. Heller, M. Hanson, A.C. Gossard, *Physics of Semiconductors 2004: Proceedings of the 27th International Conference on the Physics of Semiconductors (ICPS)*, Flagstaff, Arizona, IOP publishing, 2005;

"Multiple scattering theory for two-dimensional electron gases in the presence of spin-orbit coupling ", J.D. Walls, J. Huang, R.M. Westervelt, E.J. Heller, *Physical Review B* **73(3)**, 035325 (2006);

and manuscripts:

"Local perturbation evolution in random wave field and a time-reversal diagnostic", J. Huang, S. Tomsovic, and E.J. Heller, in preparation;

"Conductance fringe patterns in two dimensional electron gases", J. Huang, J.D. Walls, and E.J. Heller, to be incorporated in a review paper;

"Imaging cusps in two dimensional electron gases", J. Huang, E.J. Heller, and R.M. Westervelt, in preparation.

Acknowledgments

Many mixed feelings arise when I look back at the past years in graduate school. It's a rewarding, yet challenging experience. First and foremost, I must thank Prof. Eric J. Heller for being such a wonderful advisor. Without his kindness and forgiveness, it's difficult to imagine me writing the thesis now. Prof. Heller shows patience when a student struggles, shows forgiveness when one makes mistakes, shows diligence when students and postdocs fill up his appointment schedule everyday. The more time I spend in the group, the more impressed I am by his depth and breadth of knowledge. His penetrating insight can quickly locate any tiny flaw in my calculation; his unique, talented way of looking at a complicated problem from its most simplified case has greatly influenced the way I think and research. The research projects have been fun and educational. I couldn't ask for a better advisor to guide my graduate research.

Prof. Michael Tinkham and Prof. Eugene Demler have been so helpful and understanding throughout my graduate school years. It's a real honor to have had these two great physicists on my thesis committee.

My early research career has been greatly influenced by two brilliant former group members, Gregory A. Fiete, a former graduate student in the group and then a postdoc in the department, and Jamie D. Walls, a postdoc in the group. Despite his busy schedule, in the half year when we shared office, Greg never hesitated to spend hours educating me on condensed matter theories. The few long conversations I had with him essentially laid foundation for my first two research projects. I have always been amazed by Greg's ability to explain complicated theories in the simplest, clearest way possible. In many ways, Greg was a mentor to me in the early days. Jamie joined

the group soon after Greg left. It has been real fortunate and pleasant to collaborate with and learn from Jamie in the past two years. Patient, nice, and yet talented, Jamie has impressed me with his expertise in nuclear magnetic resonance and spintronics. The many conversations we had over the past years have been fruitful, demonstrated by our several joint papers, published or to be published.

I am thankful to Prof. Steven Tomsovic at Washington State University for hosting me for one semester when I first joined the group while Prof. Heller was away on sabbatical.

Sheila Ferguson, Bonnie Currier, and David Norcross have lent me tremendous administrative support over the years, for which I'm forever grateful.

Coming to the United States as an average Chinese student, I have gone through a path that's not the easiest that one can imagine. I have learned much and grown up. I deeply appreciate the few friends who lent me a hand in my most difficult times. I must thank my college advisors, Prof. Guangtian Zou and Prof. Zhiren Zheng, for having so much faith in me and my college roommates for moral support.

I thank my family for always sticking with me without complaints ever since I left home for high school at 13.

Dedicated to my mother Zhi-Ju Luo,
my father Shi-You Huang,
my brother Hua Huang,
and those who lent me a helping hand.

Chapter 1

Introduction and Outline of the Thesis

1.1 Introduction

In recent years, experiments on nanostructures have yielded a very rich body of data. Technological advancements have allowed physicists to probe ever smaller and more quantum mechanical systems, and physics processes never seen before have been shown to happen on these small scales. Quantum corrals, quantum mirages, and branched electron flow through a quantum point contact are some recent examples.

Very often in these nanoscale experiments, quantum effect is large and objects behave more like waves than particles. Interactions among objects and between objects and external potentials are best described in terms of wave scattering. An arbitrary scattering among many objects can be very difficult to analyze because of high order partial wave scatterings and the large number of scatterers involved. So the most

general form of multiple scattering theory is often employed numerically and in the statistical sense. However, many of the above difficulties don't exist in most nanoscale experiments where measurements are done in mostly clean samples and low experimental temperature makes S-wave scattering, the simplest scattering, the dominant one. In these nanoscale systems, coherence length is very significant such that phase factors play an important role and coherent interference becomes abundant. Multiple scattering theory (MST) can be applied in the S-wave limit, which allows for easy and compact formulation, to explain many such phenomena observed in two-dimensional electron gases (2DEGs) in recent experiments.

Largely motivated by several previously unsolved mysteries in the electron flow imaging experiments done in the Westervelt group at Harvard years ago, I apply multiple scattering theory (MST) in the S-wave limit to give a full account of all possible conductance fringe patterns in 2DEGs and utilize the Kirchhoff method to reveal the formation of quantum cusps in 2DEGs and their tip scan images, in addition to other topics. Most of the research revolves around phenomena observed experimentally in 2DEGs in nanostructures, and MST is the most used technique, though not the only one.

1.2 Outline of the Thesis

A broad range of problems are researched in this thesis. A perturbed version of multiple scattering theory (MST) is first developed in a general matrix formalism; a classical acoustic scattering experiment is then analyzed in great detail to reveal the experimental system's internal dynamics; MST is applied to scattering in

two-dimensional electron gases (2DEGs) to explain conductance fringe patterns; a finite difference method (FDM) is employed to solve the Schrodinger equation for a single-electron quantum dot with various boundary conditions and singular value decomposition (SVD) is used to retrieve the electron wavefunction; numerical techniques are developed to study cusps in 2DEGs and their tip scan images. In total, five closely related projects are presented.

In the first project on perturbation in scattering, starting with the usual Lippman-Schwinger formulation of MST, a perturbative MST is developed to express in terms of the original field the changes induced by removing, adding, shifting, and changing a scatterer. The formalism is intuitive and easy to implement numerically. An intuitive diagrammatic representation of MST is also given, and the possible occurrence or suppression of resonance is also discussed.

In the second project on acoustic scattering, time-reversal focusing in sound scattering is studied in terms of MST. Many important features of the experimental system are revealed for the first time. An interpretation is provided to explain some recent experimental observations in similar systems.

In the third project on conductance fringe patterns in 2DEGs, MST is applied to 2DEGs in the S-wave scattering limit. Many novel conductance fringe patterns are revealed and a general formulation is given for these patterns. Also provided are estimation formulas of thermal width for the circular, elliptical, and hyperbolic fringes, which are found to be consistent with numerical results. Occurrence of resonance and signatures of impurity properties are also addressed.

In the fourth project on a single-electron quantum dot, FDM is first employed to

solve for ground state energy of a single-electron quantum dot which, upon resonant tunneling, leads to circular conductance peak patterns under tip scan. Singular value decomposition (SVD) is then utilized to invert matrix to extract electron wavefunction out of the experimental conductance data. The stability and sensitivity to boundary conditions, along with applicability, of FDM and SVD are also studied.

In the fifth project, numerical techniques for generating classical and quantum cusps in 2DEGs are developed. Their tip scan images are analyzed to show interesting dynamics in imaging process. Phase space dynamics have been particularly interesting and illuminating. The corresponding classical structure of quantum cusp is demonstrated at short wavelength. More complicated, overlapping cusps are shown as well. Lastly and most importantly, experimental cusps are matched nicely by numerical results from the Kirchhoff method.

Chapter 2

A Perturbative Multiple Scattering Theory

2.1 Introduction

Scattering theory has been well studied in the past century [62]. Initially developed to help study atomic and nuclear structures, scattering theory was soon also applied to condensed matter physics, and then went beyond quantum scattering on microscopic and mesoscopic scales to find applications in classical, macroscopic systems in optics [68, 82], and more recently in acoustic scattering experiments [14, 27]. As Maynard [52] pointed out, though quantum scattering is derived from the Schrodinger equation and classical scattering is based on a classical wave equation, many analogies between them exist under appropriate conditions. However, it remains to be seen what impact a small perturbation has on a scattering system. A perturbation theory can give important insight into the system's internal dynamics, which are often too difficult

to analyze directly.

Our work is motivated by the recent experimental and theoretical studies on perturbations in acoustic time reversal focusing [78, 70]. As the first part of our ongoing project on perturbed time reversal focusing, this chapter gives an analytical formulation of perturbations in multiple scattering, applicable to both quantum and classical systems [52]. The following chapter [38] will explore perturbations in acoustic time reversal focusing in great detail, not only providing an analytical model to reproduce the observations [78], but demonstrating analogous features in quantum scattering systems as well. By developing a perturbative multiple scattering theory in a general form in this chapter, we expect its future broad applications to both theoretical study of stability of scattering systems and experimental research in atomic and mesoscopic physics. For example, studies on photon interaction with atoms, atomic collision in Bose-Einstein condensates, and electronic transport in mesoscopic/nanoscopic systems can all benefit from an understanding of the influence of small changes on scatterers (atoms or impurities) due to perturbations like temperature fluctuations, etc. The knowledge of perturbation effects could also be used to design apparatus to monitor gradual changes over long time in various systems, such as the proposed industrial/medical imaging and wireless communication schemes based on time reversal focusing [78].

The chapter is organized as follows. First we review the conventional formulation of multiple scattering theory. Then the change in the overall incoming wave on an arbitrary scatterer is derived in terms of the unperturbed wave field for removing, adding, shifting a scatterer, and replacing a scatterer with a different one, respectively.

In the end, the collective perturbation due to changes on multiple scatterers is given in a compact form. Physical interpretation is given along with mathematical derivation. All formulations are given for a general multiple scattering system. For simplicity, only s-wave scattering is considered, since it's the most commonly explored case in multiple scattering research and can be easily obtained in mesoscopic systems by lowering temperature to achieve wavelength comparable to scatterer size.

2.2 The conventional *s*-wave multiple scattering theory

Consider an open, linear, scalar-wave system consisting of a source and a number of s-wave scatterers whose scattering strength is uniform and given by $\epsilon(\omega)$. Denote the source's and N scatterers' positions respectively as $\vec{x}_s$ and $\{\vec{x}_i; i = 1, N\}$. From an s-wave multiple scattering theory [62], the total resulting wave field $\phi^N(\vec{x};\omega)$ at an arbitrary point $\vec{x}$ can be expressed in position-frequency space using the free propagation Green function $G(\vec{x}, \vec{x}';\omega)$ as

$$\begin{aligned} \phi^N(\vec{x};\omega) &= \phi_s(\vec{x};\omega) + \epsilon(\omega)\sum_{i=1}^{N} G(\vec{x}, \vec{x}_i;\omega)\phi^N(\vec{x}_i;\omega) \\ &= \phi_s(\vec{x};\omega) + \epsilon(\omega)\sum_{i=1}^{N} G_{\vec{x},i}\phi_i^N, \end{aligned} \tag{2.1}$$

where $\phi_i^N = \phi^N(\vec{x}_i;\omega)$, $G_{\vec{x},i} = G(\vec{x}, \vec{x}_i;\omega)$, $\phi_s(\vec{x};\omega)$ is the unscattered signal directly from the source, $\phi^N(\vec{x}_i;\omega)$ is the overall incoming wave at the i^{th} scatterer, and the superscript N denotes the total number of scatterers present.

Setting $\vec{x} = \vec{x}_i$ successively in Eq. (2.1) for all i leads to a series of equations that

can be expressed formally in the matrix representation

$$M^N \cdot \phi^N \equiv \begin{pmatrix} m_{11} & \dots & m_{1N} \\ . & \dots & . \\ . & \dots & . \\ m_{N1} & \dots & m_{NN} \end{pmatrix} \begin{pmatrix} \phi_1^N \\ . \\ . \\ \phi_N^N \end{pmatrix} = \begin{pmatrix} \phi_{1,s} \\ . \\ . \\ \phi_{N,s} \end{pmatrix}, \tag{2.2}$$

where $\phi_{i,s} = \phi_s(\vec{x}_i;\omega)$, $m_{ii} = 1$, and $m_{ij} = -\epsilon(\omega)G_{i,j}$ for $i \neq j$. The standard method of solving heterogeneous matrix equations gives

$$\left|M^N\right| \phi_i^N = \begin{vmatrix} m_{11} & . & m_{1(i-1)} & \phi_{1,s} & m_{1(i+1)} & . & m_{1N} \\ m_{21} & . & m_{2(i-1)} & \phi_{2,s} & m_{2(i+1)} & . & m_{2N} \\ . & . & . & . & . & . & . \\ m_{N1} & . & m_{N(i-1)} & \phi_{N,s} & m_{N(i+1)} & . & m_{NN} \end{vmatrix}. \tag{2.3}$$

The determinant on the right hand side is obtained by replacing the i^{th} column with the vector $\{\phi_{i,s}\}$. From the standard matrix theory [29], the determinant $\left|M^N\right|$ can be written as $\sum_p \pm[m_{p(1)1}m_{p(2)2}m_{p(3)3}\dots m_{p(N)N}]$, where the sum is extended over all permutations p of the integers $1, 2, ..., N$ and a $+$ or $-$ sign is affixed to each product according to whether p is even or odd for $p(1), p(2), ..., p(N)$. Given that, for $i \neq j$, m_{ij} describes free wave propagation from the j^{th} scatterer toward the i^{th} scatterer, any term $m_{p(1)1}m_{p(2)2}m_{p(3)3}\dots m_{p(N)N}$ can be reordered to have indices linked as $m_{jl}m_{lh}m_{hk}m_{ks}m_{st}\dots m_{pj}$ after removing factors like m_{hh} ($= 1$, which means the h^{th} scatterer isn't involved in scattering). Each such term corresponds to a unique scattering path that visits every scatterer at most once and that starts and ends at the j^{th} scatterer, i.e. a closed loop as the one in Fig. 2.1(c). So $\left|M^N\right|$ describes all kinds of possible loops in scattering paths. Similarly, the right hand

side determinant of Eq. (2.3) is the sum of terms each of which can be reordered into the form $m_{ij}m_{jh}m_{hk}m_{ks}m_{st}\ldots m_{pq}\phi_{q,s}$, i.e. the index begins at i and finishes at s on the right hand side. Each term corresponds to a unique scattering path that visits every scatterer at most once, starting from the source and ending at the i^{th} scatterer as the one in Fig. 2.1(b). Now a scattering path interpretation to Eq. (2.3) is manifest. Since ϕ_i^N describes the overall incoming wave on the i^{th} scatterer from source, it contains contributions from the various scattering paths from source to the scatterer. Any of the various paths, such as the one in Fig. 2.1(a), can be decomposed into one basic path that visits every scatterer at most once, as the one in Fig. 2.1(b), and a possible second part that consists of one or more loops each of which starts and ends at an intermediate scatterer on the basic path, as the one in Fig. 2.1(c). Interestingly, the right and left hand sides of Eq. (2.3) describe all possible choices of basic paths and loops that may be involved in scattering, thereby offering a convenient and intuitive way to explain later results.

One should notice that if $\left|M^N\right|$ is small, ϕ_i^N will be large for the i^{th} scatterer with a nontrivial right hand side determinant in Eq. (2.3), which usually corresponds to certain resonance of the system. The form in Eq. (2.2) is particularly useful for generating approximations to the changes in the wave field due to removing, adding, shifting, or changing the strengths of some of the scatterers. For most applications, the natural next step is to invert the $\{m_{ij}\}$ matrix in Eq. (2.2) or calculate the left and right hand side determinants in Eq. (2.3) to obtain ϕ_i^N, almost always done numerically. But here we stay with the matrix form to develop analytical formulas for the various perturbations in the following sections.

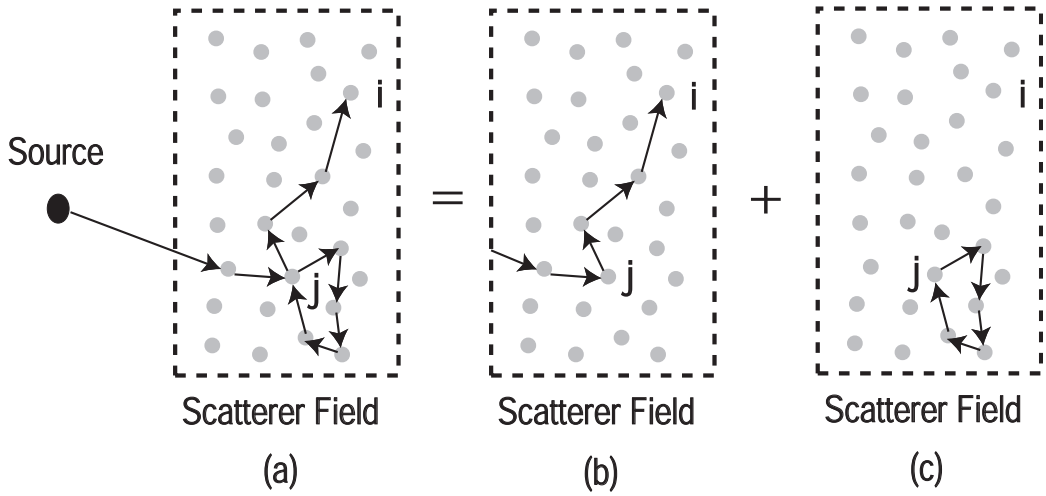

Figure 2.1: (a) Diagram of a typical scattering path from the source to the i^{th} scatterer, which visits the j^{th} scatterer twice. The entire path can be decomposed into two parts described in (b) and (c) respectively. (b) The basic component of the scattering path which visits every scatterer no more than once. (c) The loop component of the scattering path that begins and ends at the j^{th} scatterer. There can be none or multiple loops in an arbitrary scattering path.

2.3 Removing an *s*-wave scatterer

As a starting point, let us focus on the changes that will be introduced by removing an arbitrary scatterer labeled by the index N. According to Eq. (2.1), the change in the wave field is

$$\delta\phi^N(\vec{x};\omega) \equiv \phi^{N-1}(\vec{x};\omega) - \phi^N(\vec{x};\omega) = \epsilon(\omega)\left(\sum_{i=1}^{N-1} G_{\vec{x},i}\delta\phi_i^N - G_{\vec{x},N}\phi_N^N\right), \qquad (2.4)$$

where $\delta\phi_i^N = \phi_i^{N-1} - \phi_i^N$ for $1 \le i \le N-1$ is the change in the overall incoming wave on the remaining i^{th} scatterer. For later convenience, here we define $\delta\phi_N^N = \phi_{N,s} - \phi_N^N$. Eq. (2.4) expresses the change in the wave field as the sum of the propagated changes from the remaining scatterers minus the propagated field from the absent scatterer. An expression for the $\{\delta\phi_i^N\}$ is sought in terms of the unperturbed or N-scatterer wave field quantities.

For the wave field generated with just the first $N-1$ scatterers, a similar matrix representation exists as the N-scatterer case except with one fewer row and column. Thus,

$$M^{N-1}\cdot\phi^{N-1} \equiv \begin{pmatrix} m_{11} & \dots & m_{1(N-1)} \\ . & \dots & . \\ . & \dots & . \\ m_{(N-1)1} & \dots & m_{(N-1)(N-1)} \end{pmatrix} \begin{pmatrix} \phi_1^{N-1} \\ . \\ . \\ \phi_{N-1}^{N-1} \end{pmatrix} = \begin{pmatrix} \phi_{1,s,} \\ . \\ . \\ \phi_{N-1,s} \end{pmatrix}. \quad (2.5)$$

Substituting the left hand side of Eq. (2.5) into the first $N-1$ elements of the vector $\{\phi_{i,s}\}$ on the right hand side of Eq. (2.2) and subtracting $\begin{pmatrix} M^{N-1} & \mathbf{0} \\ \mathbf{0} & 1 \end{pmatrix}\cdot\phi^N$ from both sides leads to a solvable series of equations

$$\begin{aligned} \begin{pmatrix} M^{N-1} & \mathbf{0} \\ \mathbf{0} & 1 \end{pmatrix} \begin{pmatrix} \delta\phi_1^N \\ . \\ . \\ \delta\phi_N^N \end{pmatrix} &= \begin{pmatrix} 0 & ... & 0 & m_{1N} \\ . & ... & . & . \\ 0 & ... & 0 & m_{(N-1)N} \\ m_{N1} & ... & m_{N(N-1)} & 0 \end{pmatrix} \begin{pmatrix} \phi_1^N \\ . \\ . \\ \phi_N^N \end{pmatrix} \\ &= \begin{pmatrix} m_{1N} \\ . \\ . \\ m_{(N-1)N} \\ m'_{NN} \end{pmatrix} \phi_N^N, \end{aligned} \quad (2.6)$$

where $m'_{NN} = \frac{1}{\phi_N^N}\sum_{i=1}^{N-1} m_{Ni}\phi_i^N$. The changes in the wave field due to removing a single scatterer now appear only on the left hand side of the equation, whereas the right hand side contains only the original wave field quantities. Using the same

method of solution of matrix equations as before gives for $1 \leq i \leq N-1$

$$\left|M^{N-1}\right| \delta\phi_i^N = \begin{vmatrix} m_{11} & . & m_{1(i-1)} & m_{1N} & m_{1(i+1)} & . & m_{1(N-1)} \\ m_{21} & . & m_{2(i-1)} & m_{2N} & m_{2(i+1)} & . & m_{2(N-1)} \\ . & . & . & . & . & . & . \\ m_{(N-1)1} & . & m_{(N-1)(i-1)} & m_{(N-1)N} & m_{(N-1)(i+1)} & . & m_{(N-1)(N-1)} \end{vmatrix} \phi_N^N, \tag{2.7}$$

where both determinants are of $(N-1) \times (N-1)$ matrices. The right hand side is obtained by first replacing the i^{th} column with the N^{th} column, and next removing the N^{th} row and column from M^N. Now the right hand side terms are like $m_{ij}m_{jh}m_{hk}m_{ks}m_{st}\ldots m_{pN}\phi_N^N$, i.e. the index begins at i and finishes at N. Given the ordering sequence of the terms in the expanded determinants has a path interpretation, one sees that, after the N^{th} scatterer is removed, the information of ϕ_N^N's absence can reach the i^{th} scatterer through paths described on the right hand side, with possible loops at intermediate scatterers given by the left hand side terms as shown in Fig. 2.2. To obtain $\delta\phi_i^N$, one can numerically solve either Eq. (2.6) or Eq. (2.7). The sum of the various terms in either of the two determinants of Eq. (2.7) strongly depends on the convergence of high order scattering terms.

Before proceeding to a different perturbation, let us examine the implications of Eq. (2.7). The left hand side $\left|M^{N-1}\right|$ is usually nonzero for real frequencies in a non-absorbing medium. The i-dependent right hand side determinant must be nonzero for at least part of the remaining scatterers. Therefore when $\left|M^{N-1}\right|$ is very small, $\delta\phi_i^N$ will be large for the i^{th} remaining scatterer with a nontrivial right hand side

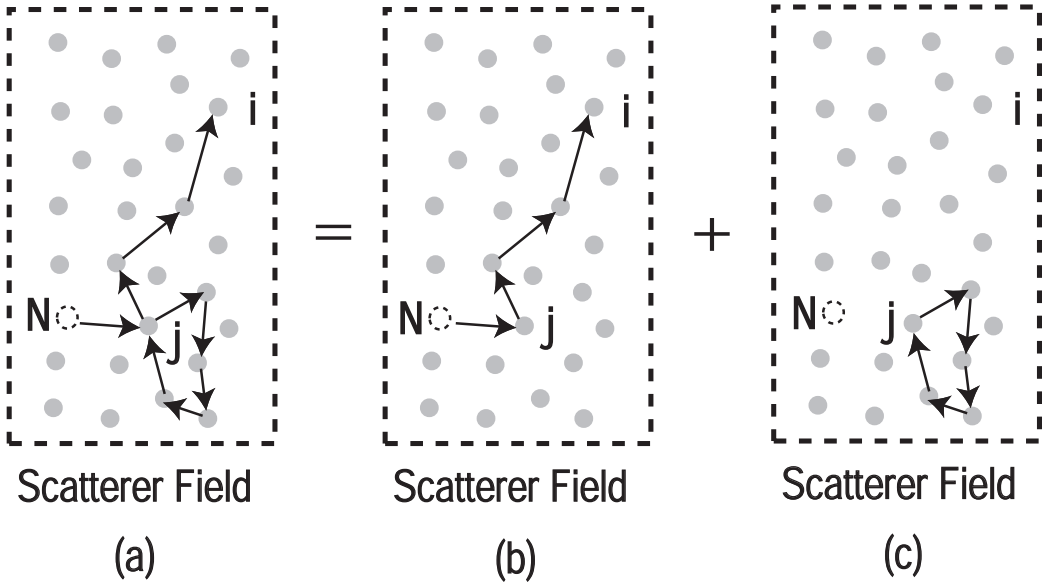

Figure 2.2: (a) Diagram of a typical path for the absence of the N^{th} scatterer to reach the i^{th} scatterer, which visits the j^{th} scatterer twice. The entire path can be decomposed into two parts described in (b) and (c) respectively. (b) The basic component of the path which visits every scatterer no more than once. (c) The loop component of the path that begins and ends at the j^{th} scatterer. There can be none or multiple loops in an arbitrary path.

determinant. In such situation, removing the N^{th} scatterer could either induce a resonance in a scattering system or suppress an existing resonance if $\left|M^N\right|$ is also small in Eq. (2.3).

Since each m_{ij} $(i \neq j)$ contains one $\epsilon(\omega)$, in the weak scattering limit in a ballistic transport system (such as a clean quantum wire), it's possible to obtain an approximate analytical solution to $\delta\phi_i^N$ using only the few lowest order terms in Eq. (2.7). For example, keeping up to the first order terms in $\epsilon(\omega)$ in Eq. (2.7) leads to $\delta\phi_i^N \approx m_{iN}\phi_N^N = -\epsilon(\omega)G(\vec{x}_i, \vec{x}_N; \omega)\phi_N^N$. Thus, among the remaining scatterers, the amplitude of $\delta\phi_i^N$ varies in proportion to that of $G(\vec{x}_i, \vec{x}_N; \omega)$, which is, for $k|\vec{x}_i - \vec{x}_N| \gg 1$, $\frac{2\pi}{k}$ in 1D; $\sqrt{\frac{2\pi}{k|\vec{x}_i - \vec{x}_N|}}$ in 2D; and $1/|\vec{x}_i - \vec{x}_N|$ in 3D ($k = 2\pi/\lambda$ and λ is the wavelength) [55]. This implies that in the weak scattering limit the amplitude of $\delta\phi_i^N$ is not only small, but non-increasing in distance $|\vec{x}_i - \vec{x}_N|$ from the perturbed scatterer. But in the strong scattering limit, such a simple relationship no longer holds and the amplitude of $\delta\phi_i^N$ could be larger in far field than in near field given appropriate conditions.

2.4 Adding an *s*-wave scatterer

Similarly, we can derive the changes that will be introduced by adding an arbitrary scatterer labeled by the index $N+1$. According to Eq. (2.1), the change in the wave field is

$$\delta\phi^{N+}(\vec{x};\omega) \equiv \phi^{N+1}(\vec{x};\omega) - \phi^N(\vec{x};\omega) = \epsilon(\omega)\left(\sum_{i=1}^{N+1} G_{\vec{x},i}\delta\phi_i^{N+} + G_{\vec{x},N+1}\phi_{N+1,s}\right), \quad (2.8)$$

where $\delta\phi_i^{N+} = \phi_i^{N+1} - \phi_i^N$ for $1 \leq i \leq N$ is the change in the overall incoming wave on the original i^{th} scatterer. We also define $\delta\phi_{N+1}^{N+} = \phi_{N+1}^{N+1} - \phi_{N+1,s}$. Eq. (2.8) expresses the change in the wave field as the sum of the propagated changes from the original scatterers plus the propagated field from the newly added scatterer. An expression for the $\{\delta\phi_i^{N+}\}$ is sought in terms of the unperturbed or N-scatterer wave field quantities.

For the wave field generated with the total $N+1$ scatterers, a similar matrix representation exists as the N-scatterer case except with one additional row and column. Thus,

$$M^{N+1} \cdot \phi^{N+1} \equiv \begin{pmatrix} m_{11} & \dots & m_{1(N+1)} \\ . & \dots & . \\ . & \dots & . \\ m_{(N+1)1} & \dots & m_{(N+1)(N+1)} \end{pmatrix} \begin{pmatrix} \phi_1^{N+1} \\ . \\ . \\ \phi_{N+1}^{N+1} \end{pmatrix} = \begin{pmatrix} \phi_{1,s,} \\ . \\ . \\ \phi_{N+1,s} \end{pmatrix}. \quad (2.9)$$

Substituting the left hand side of Eq. (2.2) into the first N elements of the vector $\{\phi_{i,s}\}$ on the right hand side of Eq. (2.9) and subtracting $M^{N+1} \cdot \begin{pmatrix} \phi_1^N \\ . \\ \phi_N^N \\ \phi_{N+1,s} \end{pmatrix}$ from both sides leads to a solvable series of equations

$$\left(M^{N+1}\right) \begin{pmatrix} \delta\phi_1^{N+} \\ . \\ . \\ \delta\phi_{N+1}^{N+} \end{pmatrix} = \begin{pmatrix} 0 & ... & 0 & -m_{1(N+1)} \\ . & ... & . & . \\ 0 & ... & 0 & -m_{N(N+1)} \\ -m_{(N+1)1} & ... & -m_{(N+1)N} & 0 \end{pmatrix} \begin{pmatrix} \phi_1^N \\ . \\ \phi_N^N \\ \phi_{N+1,s} \end{pmatrix}$$

$$= -\begin{pmatrix} m_{1(N+1)} \\ . \\ . \\ m_{N(N+1)} \\ m'_{(N+1)(N+1)} \end{pmatrix} \phi_{N+1,s}, \tag{2.10}$$

where $m'_{(N+1)(N+1)} = \frac{1}{\phi_{N+1,s}} \sum_{i=1}^{N} m_{(N+1)i}\phi_i^N$. The changes in the wave field due to adding a single scatterer now appear only on the left hand side of the equation, whereas the right hand side contains only the original wave field quantities. Using the same method of solution of matrix equations as before gives for $1 \leq i \leq N+1$

$$\left|M^{N+1}\right| \delta\phi_i^{N+} =$$

$$- \begin{vmatrix} m_{11} & . & m_{1(i-1)} & m_{1(N+1)} & m_{1(i+1)} & . & m_{1(N+1)} \\ m_{21} & . & m_{2(i-1)} & m_{2(N+1)} & m_{2(i+1)} & . & m_{2(N+1)} \\ . & . & . & . & . & . & . \\ m_{(N+1)1} & . & m_{(N+1)(i-1)} & m'_{(N+1)(N+1)} & m_{(N+1)(i+1)} & . & m_{(N+1)(N+1)} \end{vmatrix} \phi_{N+1,s}, \tag{2.11}$$

which for $1 \leq i \leq N$ simplifies to

$$\begin{vmatrix} m_{11} & . & m_{1(i-1)} & m_{1(N+1)} & m_{1(i+1)} & . & 0 \\ m_{21} & . & m_{2(i-1)} & m_{2(N+1)} & m_{2(i+1)} & . & 0 \\ . & . & . & . & . & . & . \\ m_{(N+1)1} & . & m_{(N+1)(i-1)} & m'_{(N+1)(N+1)} & m_{(N+1)(i+1)} & . & m'_{(N+1)(N+1)} - 1 \end{vmatrix} \phi_{N+1,s}$$

$$
\begin{aligned}
&= \begin{vmatrix} m_{11} & . & m_{1(i-1)} & m_{1(N+1)} & m_{1(i+1)} & . & m_{1N} \\ m_{21} & . & m_{2(i-1)} & m_{2(N+1)} & m_{2(i+1)} & . & m_{2N} \\ . & . & . & . & . & . & . \\ m_{N1} & . & m_{N(i-1)} & m_{N(N+1)} & m_{N(i+1)} & . & m_{NN} \end{vmatrix} (m'_{(N+1)(N+1)} - 1)\phi_{N+1,s} \\
&= \begin{vmatrix} m_{11} & . & m_{1(i-1)} & m_{1(N+1)} & m_{1(i+1)} & . & m_{1N} \\ m_{21} & . & m_{2(i-1)} & m_{2(N+1)} & m_{2(i+1)} & . & m_{2N} \\ . & . & . & . & . & . & . \\ m_{N1} & . & m_{N(i-1)} & m_{N(N+1)} & m_{N(i+1)} & . & m_{NN} \end{vmatrix} (\sum_{i=1}^{N} m_{(N+1)i}\phi_i^N - \phi_{N+1,s}),
\end{aligned}
\tag{2.12}
$$

where determinants are of $(N+1)\times(N+1)$ matrix on the left and $(N+1)\times(N+1)$ matrix for $i = N+1$ and $N \times N$ matrix for $i = N$ on the right. The right hand side is obtained by replacing the i^{th} column with $-\begin{pmatrix} m_{1(N+1)} \\ . \\ . \\ m_{N(N+1)} \\ m'_{(N+1)(N+1)} \end{pmatrix} \phi_{N+1,s}$. Now for $1 \leq i \leq N$ the right hand side terms are like $m_{ij}m_{jh}m_{hk}m_{ks}m_{st}\ldots m_{p(N+1)}(\sum_{i=1}^{N} m_{(N+1)i}\phi_i^N - \phi_{N+1,s})$, i.e. the index begins at i and finishes at $N+1$. Given that the ordering sequence of the terms in the expanded determinants has a path interpretation, one sees that, after the $(N+1)^{th}$ scatterer is added, the additional wave of $(\sum_{i=1}^{N} m_{(N+1)i}\phi_i^N - \phi_{N+1,s})$ can reach the i^{th} scatterer through paths described on the right hand side, with possible loops at intermediate scatterers given by the left hand side terms. To obtain $\delta\phi_i^{N+}$, one can numerically solve either Eq. (2.10) or Eq. (2.11). The sum of

the various terms in either of the two determinants of Eq. (2.11) strongly depends on the convergence of high order scattering terms.

The left hand side $\left|M^{N+1}\right|$ is usually nonzero for real frequencies in a non-absorbing medium. The i-dependent right hand side determinant must be nonzero for at least part of the remaining scatterers. Therefore when $\left|M^{N+1}\right|$ is very small, $\delta\phi_i^{N+}$ will be large for the i^{th} remaining scatterer with a nontrivial right hand side determinant. In these cases, adding the $(N+1)^{th}$ scatterer could either induce a resonance in a scattering system or suppress an existing resonance if $\left|M^{N}\right|$ is also small in Eq. (2.3).

2.5 Shifting a scatterer

Analogous equations can be developed to account for the effect of displacing a scatterer. The whole process can be viewed as a two-step operation. First, remove the N^{th} scatterer from its original position, and second, place it in its new location. The first step can be described by the same set of equations developed above for removing a scatterer. The reverse process of the second step is to remove the scatterer from its new location, which is also described by the same set of equations except that ϕ_N^N , $\delta\phi_i^N$, and m_{iN} in Eq. (2.7) are replaced by primed versions for the new location of the N^{th} scatterer. The total effect follows by subtraction of the two sets of equations. Letting $\delta\phi_i^N$ $(i < N)$ denote the full change of the compound process $\delta\phi_i^N = (\phi_i^N)' - \phi_i^N$ gives

$$\left|M^{N-1}\right| \delta\phi_i^N =$$

$$\begin{vmatrix} m_{11} & . & m_{1(i-1)} & m_{1N} & m_{1(i+1)} & . & m_{1(N-1)} \\ m_{21} & . & m_{2(i-1)} & m_{2N} & m_{2(i+1)} & . & m_{2(N-1)} \\ . & . & . & . & . & . & . \\ m_{(N-1)1} & . & m_{(N-1)(i-1)} & m_{(N-1)N} & m_{(N-1)(i+1)} & . & m_{(N-1)(N-1)} \end{vmatrix} \phi_N^N$$

$$- \begin{vmatrix} m_{11} & . & m_{1(i-1)} & m_{1N'} & m_{1(i+1)} & . & m_{1(N-1)} \\ m_{21} & . & m_{2(i-1)} & m_{2N'} & m_{2(i+1)} & . & m_{2(N-1)} \\ . & . & . & . & . & . & . \\ m_{(N-1)1} & . & m_{(N-1)(i-1)} & m_{(N-1)N'} & m_{(N-1)(i+1)} & . & m_{(N-1)(N-1)} \end{vmatrix} \phi_{N'}^N, \tag{2.13}$$

where $\phi_{N'}^N$ is the overall incoming wave at the N^{th} scatterer when placed at its new location. Note that the two determinants on the right hand side are not identical because of their dependence on the N^{th} scatterer's locations.

But Eq. (2.13) isn't the best way to express the changes since it involves $\phi_{N'}^N$, which isn't present in the original wave field. A better solution is to first write an equation similar to Eq. (2.2) for the new wave field

$$\begin{aligned} (M^N)' \cdot (\phi^N)' &\equiv \begin{pmatrix} m_{11} & \dots & m_{1(N-1)} & m_{1N'} \\ . & \dots & . & . \\ m_{(N-1)1} & \dots & m_{(N-1)(N-1)} & m_{(N-1)N'} \\ m_{N'1} & \dots & m_{N'(N-1)} & m_{N'N'} \end{pmatrix} \begin{pmatrix} (\phi_1^N)' \\ . \\ . \\ (\phi_N^N)' \end{pmatrix} \\ &= \begin{pmatrix} \phi_{1,s} \\ . \\ . \\ \phi_{N',s} \end{pmatrix}, \end{aligned} \tag{2.14}$$

which, when compared with Eq. (2.2), gives

$$(M^N)' \cdot (\phi^N)' = \begin{pmatrix} 0 \\ . \\ 0 \\ \phi_{N',s} - \phi_{N,s} \end{pmatrix} + M^N \cdot \phi^N. \tag{2.15}$$

Defining $\delta\phi_i^N = (\phi_i^N)' - \phi_i^N$ for all i and subtracting $(M^N)' \cdot \phi^N$ from both sides leads to

$$\begin{pmatrix} m_{11} & \dots & m_{1(N-1)} & m_{1N'} \\ . & \dots & . & . \\ m_{(N-1)1} & \dots & m_{(N-1)(N-1)} & m_{(N-1)N'} \\ m_{N'1} & \dots & m_{N'(N-1)} & m_{N'N'} \end{pmatrix} \begin{pmatrix} \delta\phi_1^N \\ . \\ . \\ \delta\phi_N^N \end{pmatrix} = \begin{pmatrix} m_{1N} - m_{1N'} \\ . \\ m_{(N-1)N} - m_{(N-1)N'} \\ \Delta m_{NN} \end{pmatrix} \phi_N^N, \tag{2.16}$$

where $\Delta m_{NN} = [(\phi_{N',s} - \phi_{N,s}) + \sum_{i=1}^{N-1}(m_{Ni} - m_{N'i})\phi_i^N]/\phi_N^N$. The equations can be solved to give for all i

$$\begin{vmatrix} m_{11} & \dots & m_{1(N-1)} & m_{1N'} \\ . & \dots & . & . \\ m_{(N-1)1} & \dots & m_{(N-1)(N-1)} & m_{(N-1)N'} \\ m_{N'1} & \dots & m_{N'(N-1)} & m_{N'N'} \end{vmatrix} \delta\phi_i^N =$$

$$\begin{vmatrix} m_{11} & . & m_{1(i-1)} & m_{1N} - m_{1N'} & m_{1(i+1)} & . & m_{1N'} \\ . & . & . & . & . & . & . \\ m_{(N-1)1} & . & m_{(N-1)(i-1)} & m_{(N-1)N} - m_{(N-1)N'} & m_{(N-1)(i+1)} & . & m_{(N-1)N'} \\ m_{N'1} & . & m_{N'(i-1)} & \Delta m_{NN} & m_{N'(i+1)} & . & m_{N'N'} \end{vmatrix} \phi_N^N. \tag{2.17}$$

Now only the original wave field quantities and $m_{iN'}/m_{N'i}$ (related to the new location of the N^{th} scatterer and thus necessarily included) are used to express the changes in the overall incoming wave on the i^{th} scatterer. Moreover, Eq. (2.17) gives $\delta\phi_i^N$ for all i, including the shifted N^{th} scatterer itself.

Similar to the preceding section, if $|(M^N)'|$ is small, $\delta\phi_i^N$ will be large for the i^{th} scatterer with a nontrivial right hand side determinant in Eq. (2.17). In such situation, shifting the N^{th} scatterer could either induce a resonance in a scattering system, or suppress an existing resonance if $\left|M^N\right|$ is small in Eq. (2.3).

2.6 Changing a scatterer's strength

It is also straightforward to apply a similar method to the case in which the N^{th} scatterer changes its scattering strength by a factor of an arbitrary constant γ. This can be viewed as a special case of Eq. (2.13) by replacing $m_{iN'}$ with γm_{iN} in the second determinant on the right hand side. One should note that, for $i \neq N$, it's m_{iN} instead of m_{Ni} that is multiplied by γ (m_{Ni} describes a scattering by the i^{th} scatterer toward the N^{th} scatterer), and that $\phi_{N'}^N \neq \phi_N^N$ (despite the same position for the scatterer) because the change in scattering strength of one scatterer modifies

the wave field at every point. However, as in the preceding section, there is a better expression for the changes in wave field not involving $\phi^N_{N'}$. Similar to Eq. (2.2), the new wave field has

$$
\begin{aligned}
(M^N)'' \cdot (\phi^N)'' &\equiv \begin{pmatrix} m_{11} & \dots & m_{1(N-1)} & \gamma m_{1N} \\ \cdot & \dots & \cdot & \cdot \\ m_{(N-1)1} & \dots & m_{(N-1)(N-1)} & \gamma m_{(N-1)N} \\ m_{N1} & \dots & m_{N(N-1)} & m_{NN} \end{pmatrix} \begin{pmatrix} (\phi^N_1)'' \\ \cdot \\ \cdot \\ (\phi^N_N)'' \end{pmatrix} \\
&= \begin{pmatrix} \phi_{1,s} \\ \cdot \\ \cdot \\ \phi_{N,s} \end{pmatrix}.
\end{aligned}
\tag{2.18}
$$

Compared with Eq. (2.2), Eq. (2.18) gives

$$
\begin{aligned}
(M^N)'' \cdot (\phi^N)'' &\equiv \begin{pmatrix} m_{11} & \dots & m_{1(N-1)} & \gamma m_{1N} \\ \cdot & \dots & \cdot & \cdot \\ m_{(N-1)1} & \dots & m_{(N-1)(N-1)} & \gamma m_{(N-1)N} \\ m_{N1} & \dots & m_{N(N-1)} & m_{NN} \end{pmatrix} \begin{pmatrix} (\phi^N_1)'' \\ \cdot \\ \cdot \\ (\phi^N_N)'' \end{pmatrix} \\
&= \begin{pmatrix} m_{11} & \dots & m_{1N} \\ \cdot & \dots & \cdot \\ \cdot & \dots & \cdot \\ m_{N1} & \dots & m_{NN} \end{pmatrix} \begin{pmatrix} \phi^N_1 \\ \cdot \\ \cdot \\ \phi^N_N \end{pmatrix} \equiv M^N \cdot \phi^N.
\end{aligned}
\tag{2.19}
$$

Setting $\delta\phi^N_i = (\phi^N_i)'' - \phi^N_i$ for all i and subtracting $(M^N)'' \cdot \phi^N$ from both sides

gives

$$\begin{pmatrix} m_{11} & \dots & m_{1(N-1)} & \gamma m_{1N} \\ . & \dots & . & . \\ m_{(N-1)1} & \dots & m_{(N-1)(N-1)} & \gamma m_{(N-1)N} \\ m_{N1} & \dots & m_{N(N-1)} & m_{NN} \end{pmatrix} \begin{pmatrix} \delta\phi_1^N \\ . \\ . \\ \delta\phi_N^N \end{pmatrix} = (1-\gamma) \begin{pmatrix} m_{1N} \\ . \\ m_{(N-1)N} \\ 0 \end{pmatrix} \phi_N^N. \tag{2.20}$$

Solving Eq. (2.20) gives

$$\begin{vmatrix} m_{11} & \dots & m_{1(N-1)} & \gamma m_{1N} \\ . & \dots & . & . \\ m_{(N-1)1} & \dots & m_{(N-1)(N-1)} & \gamma m_{(N-1)N} \\ m_{N1} & \dots & m_{N(N-1)} & m_{NN} \end{vmatrix} \delta\phi_i^N =$$

$$(1-\gamma) \begin{vmatrix} m_{11} & . & m_{1(i-1)} & m_{1N} & m_{1(i+1)} & . & m_{1(N-1)} \\ m_{21} & . & m_{2(i-1)} & m_{2N} & m_{2(i+1)} & . & m_{2(N-1)} \\ . & . & . & . & . & . & . \\ m_{(N-1)1} & . & m_{(N-1)(i-1)} & m_{(N-1)N} & m_{(N-1)(i+1)} & . & m_{(N-1)(N-1)} \end{vmatrix} \phi_N^N, \tag{2.21}$$

or

$$\left(\gamma \left|M^N\right| + (1-\gamma)\left|M^{N-1}\right|\right) \delta\phi_i^N =$$

$$(1-\gamma) \begin{vmatrix} m_{11} & . & m_{1(i-1)} & m_{1N} & m_{1(i+1)} & . & m_{1(N-1)} \\ m_{21} & . & m_{2(i-1)} & m_{2N} & m_{2(i+1)} & . & m_{2(N-1)} \\ . & . & . & . & . & . & . \\ m_{(N-1)1} & . & m_{(N-1)(i-1)} & m_{(N-1)N} & m_{(N-1)(i+1)} & . & m_{(N-1)(N-1)} \end{vmatrix} \phi_N^N. \tag{2.22}$$

Once again, the perturbation $\delta\phi_i^N$ on every scatterer, including the N^{th} scatterer, is represented using only the original wave field quantities. Eq. (2.22) has a rather surprising implication. If one could freely control γ (which needs further careful examination in certain cases) and $\left|M^N\right| \neq \left|M^{N-1}\right|$ (which is usually true), one could always induce resonance in a non-resonant scattering system by setting $\gamma = 1/(1 - \left|M^N\right| / \left|M^{N-1}\right|)$. Since resonance often corresponds to persisting bouncing of wave between scatterers, it's possible to trap wave inside a scatterer field for a significant amount of time by tuning the scattering strength of an arbitrary scatterer, according to Eq. (2.22). This could be an alternative to the recent experimental work of freezing photons in an atomic cloud [4].

2.7 Perturbing multiple scatterers

By repeated use of Eq. (2.4), the removal of a total of m scatterers can be viewed as the sequential removal of a single scatterer at a time, i.e.

$$\begin{aligned}
\phi^{N-m}(\vec{x};\omega) - \phi^N(\vec{x};\omega) &= \sum_{j=1}^{m} \left(\phi^{N-j}(\vec{x};\omega) - \phi^{N-j+1}(\vec{x};\omega)\right) \\
&= \sum_{j=1}^{m} \delta\phi^{N-j+1}(\vec{x};\omega) \\
&= \epsilon(\omega) \left\{ \sum_{j=1}^{m} \left(\sum_{i=1}^{N-j} G_{\vec{x},i} \delta\phi_i^{N-j+1} - G_{\vec{x},N-j+1} \phi_{N-j+1}^{N-j+1} \right) \right\} \\
&= \epsilon(\omega) \left(\sum_{i=1}^{N-m} G_{\vec{x},i} \sum_{j=1}^{m} \delta\phi_i^{N-j+1} - \sum_{i=N-m+1}^{N} G_{\vec{x},i} \phi_i^N \right), \quad (2.23)
\end{aligned}$$

where $\delta\phi_i^{N-j+1}$ is given by equations similar to Eq. (2.7). In the last form given, all the canceling terms have been removed. Interestingly, the differential contributions coming from the remaining scatterers are summed over the m systems having j scat-

terers removed, whereas the missing contributions from the removed scatterers rely only on the N-scatterer system. It's the consequence of the linearity of the system, which makes the overall change on any remaining scatterer equal the linear sum of the changes induced in each removal of the m scatterers.

Similarly, the addition of a total of m scatterers can be viewed as the sequential addition of a single scatterer at a time, which leads to

$$\phi^{N+m}(\vec{x};\omega) - \phi^{N}(\vec{x};\omega) \;=\; \epsilon(\omega)\left(\sum_{i=1}^{N+m} G_{\vec{x},i}\sum_{j=1}^{m}\delta\phi_i^{N+,j} + \sum_{i=N+1}^{N+m} G_{\vec{x},i}\phi_{i,s}\right), \quad (2.24)$$

where $\delta\phi_i^{N+,j}$, change due to the j^{th} addition, is given by equations similar to Eq. (2.11).

By linearity, the collective change in wave field due to shifting m scatterers is conveniently given as

$$\Big(\phi^{N}(\vec{x};\omega)\Big)' - \phi^{N}(\vec{x};\omega) = \epsilon(\omega)\sum_{i=1}^{N} G_{\vec{x},i}\sum_{j=1}^{m}(\delta\phi_i^N)_j, \quad (2.25)$$

where $(\delta\phi_i^N)_j$ is the change $\delta\phi_i^N$ in the incoming wave on the i^{th} scatterer due to the j^{th} shifting, given by Eq. (2.17).

Similarly, when m scatterers change their scattering strength, the collective change in wave field is

$$\Big(\phi^{N}(\vec{x};\omega)\Big)'' - \phi^{N}(\vec{x};\omega) = \epsilon(\omega)\sum_{i=1}^{N} G_{\vec{x},i}\sum_{k=1}^{m}(\delta\phi_i^N)_k, \quad (2.26)$$

where $(\delta\phi_i^N)_k$ is the change $\delta\phi_i^N$ in the incoming wave on the i^{th} scatterer due to the k^{th} scatterer's change in scattering strength, given by Eq. (2.21). The m scatterers can change their scattering strength by different factors γ. That is, γ in Eq. (2.21) can be replaced with a different γ_k each time calculating $(\delta\phi_i^N)_k$.

Above are the collective perturbations due to modifying multiple scatterers. They are analytical and strict. At certain weak scattering limit, the various $\delta\phi_i^N$ can be approximated using low scattering terms, therefore offering a convenient analytical study of perturbation effects.

2.8 Conclusion

In the preceding sections, we have expressed the changes in the overall incoming wave on an arbitrary scatterer in terms of the unperturbed wave field for removing, adding, shifting a scatterer, and replacing a scatterer with a different one respectively, and then given the overall perturbation due to changing multiple scatterers in compact forms. Most of the results can be intuitively understood in the scattering path picture. The equations can be easily solved numerically to compare the changes on different scatterers, which can help reveal the correlation strength between a remaining scatterer and the removed one(s). Also at the weak scattering limit in a ballistic transport system, the few lowest order scattering terms may approximate the change on a scatterer's overall incoming wave field analytically. The formulas are simple and intuitive, should be very helpful to research on perturbations in atomic, mesoscopic and even classical scattering systems.

Upon completion of this research project, it was recently brought to my attention that a former graduate student, Stella Chan, in the Heller group had independently worked out some similar problems in her thesis in a different formalism [12], which complements our study in this chapter very well.

Chapter 3

Local Perturbation Evolution in A Random Wave Field and A Time Reversal Diagnostic

In the previous chapter, we reviewed the conventional multiple scattering theory and its variations in the different cases of perturbation. In this chapter, we apply multiple scattering theory to a classical scattering system to treat perturbations and try to generalize it to quantum scattering as well. The next chapter will apply MST to treat quantum scattering in 2DEGs.

We investigate the effects of local perturbations on wave propagation in an open system comprised of a number of randomly placed identical scatterers. The work is motivated by the recent acoustic experiments on time reversal focusing by A. Tourin, *et al.* [*Phys. Rev. Lett.* **87**, 274301 (2001)]. A general scattering theory for time reversal focusing is first presented. Numerical simulations based on the theory demonstrate

that both the enhancement due to time reversal focusing and the transition in the deterioration pattern from a linear to square root dependence on the number of altered scatterers exist in quantum scattering systems as well. A following detailed analysis of the key features in the experimental setup is made to fully reproduce major observations on local perturbation with good accuracy. Several previously unknown experimental facts are also easily understood.

3.1 Introduction

In this chapter we consider the effects of local perturbations on wave field evolution for systems invariant under time reversal. If the dynamics of the system are chaotic (or disordered), then even the simplest, coherent initial state will rapidly evolve into a highly complicated wave field. It may be rather difficult to measure accurately the entire wave field in order to check for deviations due to small perturbations. However, perturbing the system and reversing the dynamics properly can help remove phase factors and simplify a study of the system. More specifically, the idea is to propagate a coherent initial state for a time t, alter the system to some varying degree, reverse the dynamics for the same length time, and finally check the extent to which the wave field returns to its original state. The match between the original state and the propagated/reversed one for a given perturbation indicates the system's robustness against the perturbation. Two well known experimental realizations of such a situation are spin polarization echoes in nuclear magnetic resonance [58, 90], and acoustic time reversal mirrors (TRM) [27, 78].

The experiments of Tourin *et al.* [78] can be briefly described as follows. There

are a source rod and a series of transducer receivers (120) on opposite sides of a rectangular area of 3000 randomly-positioned-with-uniform-density, long (120 mm) steel rods (scatterers) aligned parallel to the z-axis. A short pulse signal is generated by the source rod, which propagates through the scatterer field. In this arrangement, each transducer receiver records a long signal with a noisy appearance. Then a certain number of the rods are removed from the rectangular area. Various intervals of the receivers' signals are time inverted and retransmitted simultaneously by the receivers. Depending on the number of removed rods, the retransmitted waves recombine so as to reconstruct a signal similar to the initial pulse at the source. The more rods removed, the less the reconstructed signal resembles the initial pulse. Three main experimental observations are of interest to us. First, it is observed that by using a roughly $5\mu s$ interval for the time-reversal process, under certain circumstances the resulting signal at the source has a deterioration that has a nearly linear dependence on the number of removed rods, whereas other circumstances lead to a nearly square root dependence. The second result is that the half-peak deterioration of the resulting signal is near inversely proportional to the end point of the time interval used; they termed these measurements, 'dynamic time-reversal experiments'. Lastly, reversing the entire receiver signal produces a deterioration pattern approximately proportional to the square root of the number of removed rods; these measurements are dubbed here, 'static time-reversal experiments'.

There have been a number of previous theoretical works on TRM [5, 57, 9, 6, 70]. Many of them addressed the general principles of TRM, but did not concentrate on any particular experiment or its particular features. Snieder and Scales provided

original insight on the stability of time reversal focusing in a general scattering system using waves and particles [70]. However, their work does not directly explain many of the later experimental observations [78]. Therefore, our focus is to show how characteristics of the experimental arrangement lead to the observations not well understood so far and, most importantly, to provide a further analysis of the perturbation sensitivity of TRM. The experimental observations [78] can be understood simply, and our work generalizes to quantum scattering systems as well. This study has broad potential applications to problems of classical physics such as wireless communication based on TRM and industrial/medical ultrasound imaging [78], and also to quantum problems such as impurity scattering in mesoscopic/nanoscopic systems with time-reversal symmetry.

The chapter is organized as follows. In Section 3.2, a general scattering theory formalism is introduced for time reversal focusing, with or without perturbations to scatterers. Numerical experiments based on the theory are conducted in a 2D quantum scattering system. The results confirm the validity of time reversal focusing and the linear-m to $\sqrt{m}$ transition in deterioration pattern (m is the number of altered local scatterers). Some delicate issues regarding removing/adding rods and the order of rod numbers in the two stages of TRM experiments are addressed and would make good subjects for further experimental investigations. In the latter half of the chapter, we apply the theory in detail to explain the experimental observations by A. Tourin, *et al.* [78]. In Section 3.3, we introduce the differential cross section of a single ultrasound scattering on a steel rod, and derive the Green functions for direct transport and multiple scattering. With these results, many common features

of the TRM experiments, such as the slowly-varying local mean amplitude and the essential information contained in the receiver signal, appear naturally. In Section 3.4, to address the perturbation sensitivity of TRM, a diffusion model is proposed. Most experimental observations are nicely reproduced by our theory.

3.2 A General Scattering Theory of Perturbed Time Reversal Focusing

Many perturbation effects in TRM experiments are exhibited in ways that are system dependent. However, certain features are general and should apply to many other multiple scattering systems. To examine this prediction, a general scattering theory and its numerical results based on a 2D s-wave quantum scattering system in ballistic transport limit are presented here.

The 2D scattering system is shown in Fig. 3.1. The central wave frequency is $3MHz$, and the wave speed is $1.5Km/s$. So the central wavelength is $0.5mm$, much smaller than the average spacing between scatterers but much bigger than the scatterer size (here we use 40 identical zero-range-interaction scatterers), which makes an s-wave scattering treatment valid. The scattering strength is uniformly $\epsilon = 2i(e^{4i} - 1)$. The source to receiver forward propagation Green function can be easily calculated in frequency space as follows. Denote the source's, receiver's, and N scatterers' positions respectively as $\vec{x}_s$, $\vec{x}_r$, and $\{\vec{x}_i; i = 1, N\}$. From a zero-range multiple scattering theory [62], the final Green function $G^r(\vec{x}, \vec{x}_s; \omega)$ can be expressed in terms of the free propagation Green function $G_0^r(\vec{x}_s, \vec{x}_i; \omega) = -i(J_0(k|\vec{x}_s -$

$\vec{x}_i|) + iY_0(k|\vec{x}_s - \vec{x}_i|))$ ($J_0(Y_0)$ is the Bessel function of the first (second) kind and the superscript r denotes the Green function is obtained using the retarded free propagation Green function) as

$$\begin{aligned} G^r(\vec{x}, \vec{x}_s; \omega) &= G_0^r(\vec{x}, \vec{x}_s; \omega) + \sum_{i=1}^{N} \epsilon G_0^r(\vec{x}, \vec{x}_i; \omega) G^r(\vec{x}_i, \vec{x}_s; \omega) \\ &= G_{\vec{x},s}^0 + \sum_{i=1}^{N} \epsilon G_{\vec{x},i}^0 G_{i,s}, \end{aligned} \tag{3.1}$$

where $G_{i,s} = G^r(\vec{x}_i, \vec{x}_s; \omega)$ is the final Green function at the i^{th} scatterer. Setting $\vec{x} = \vec{x}_j$ successively in Eq. (3.1) for all j leads to a series of equations that can be expressed formally in the matrix representation

$$\begin{pmatrix} m_{11} & \dots & m_{1N} \\ . & \dots & . \\ . & \dots & . \\ m_{N1} & \dots & m_{NN} \end{pmatrix} \begin{pmatrix} G_{1,s} \\ . \\ . \\ G_{N,s} \end{pmatrix} = \begin{pmatrix} G_{1,s}^0 \\ . \\ . \\ G_{N,s}^0 \end{pmatrix}, \tag{3.2}$$

where $m_{ii} = 1$, and $m_{ij} = -\epsilon G_{i,j}^0$ for $i \neq j$. Inverting the matrix gives $G_{i,s}$, which can be plugged into Eq. (3.1) to solve for $G^r(\vec{x}_r, \vec{x}_s; \omega)$. Amazingly, the Green function $G^a(\vec{x}_r, \vec{x}_s; \omega)$ of the time reversed propagation can be obtained in exactly the same way using the advanced free propagation Green function $G_0^a(\vec{x}_s, \vec{x}_i; \omega) = i(J_0(k|\vec{x}_s - \vec{x}_i|) - iY_0(k|\vec{x}_s - \vec{x}_i|)) = G_0^r(\vec{x}_s, \vec{x}_i; \omega)^*$, which leads to $G^a(\vec{x}_r, \vec{x}_s; \omega) = G^r(\vec{x}_r, \vec{x}_s; \omega)^*$.

First of all, in principle time reversal refocusing should be applicable to any multiple scattering system with time reversal symmetry. As shown in Fig. 3.2 (a), a Gaussian pulse $f(t)$ is generated by the source with a frequency spectrum $F(\omega)$. It propagates through the scatterer field to record a chaotic signal on the receiver. The frequency spectrum is

$$R_N(\omega) = G_N^r(\omega) F(\omega) \tag{3.3}$$

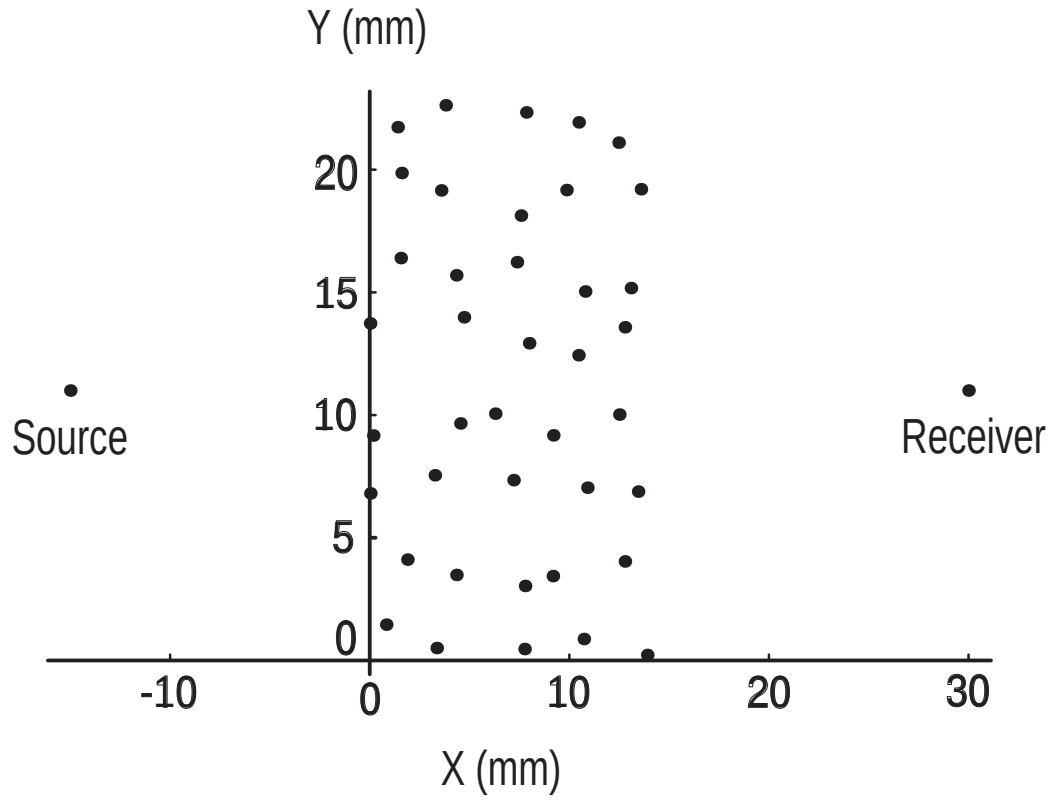

Figure 3.1: The 2D quantum system of 40 zero-range-interaction scatterers field, source, and receiver.

for the receiver signal and

$$S_{N,N}(\omega) = G^a_N(\omega)G^r_N(\omega)F(\omega) = |G^r_N(\omega)|^2F(\omega) \tag{3.4}$$

for the refocused signal at source. Here a subscript N is added to denote the number of rods involved in each propagation. Projected back into time domain, the receiver and refocused signals are shown in Fig. 3.2(b) and (c) respectively for $N = 40$. The refocused signal is obviously much cleaner than the receiver signal. For comparison, in Fig. 3.2 (d) we also draw the time reversed signal in homogeneous medium without scatterers in between source and receiver. It's easy to find the presence of the scatterer field does enhance the refocusing effect.

It is then easy to find the refocused signal when m scatterers are removed before the time reversal operation

$$S_{N-m,N}(\omega) = G^a_{N-m}(\omega)G^r_N(\omega)F(\omega). \tag{3.5}$$

or when m scatterers are added before the time reversal operation

$$S_{N+m,N}(\omega) = G^a_{N+m}(\omega)G^r_N(\omega)F(\omega). \tag{3.6}$$

We call the first stage of forward propagation toward the receiver the 'scattering stage'; the second stage of backward propagation toward the source the 'recovery stage'. When there are more (or less) scatterers present in the scattering stage compared with the recovery stage, the waves passing through the extra scatterers (or depleted region) at the scattering stage will experience the change and follow a different scattering path at the recovery stage and therefore are not focused back at the source. We should realize that, during most of the recorded duration, each of the m

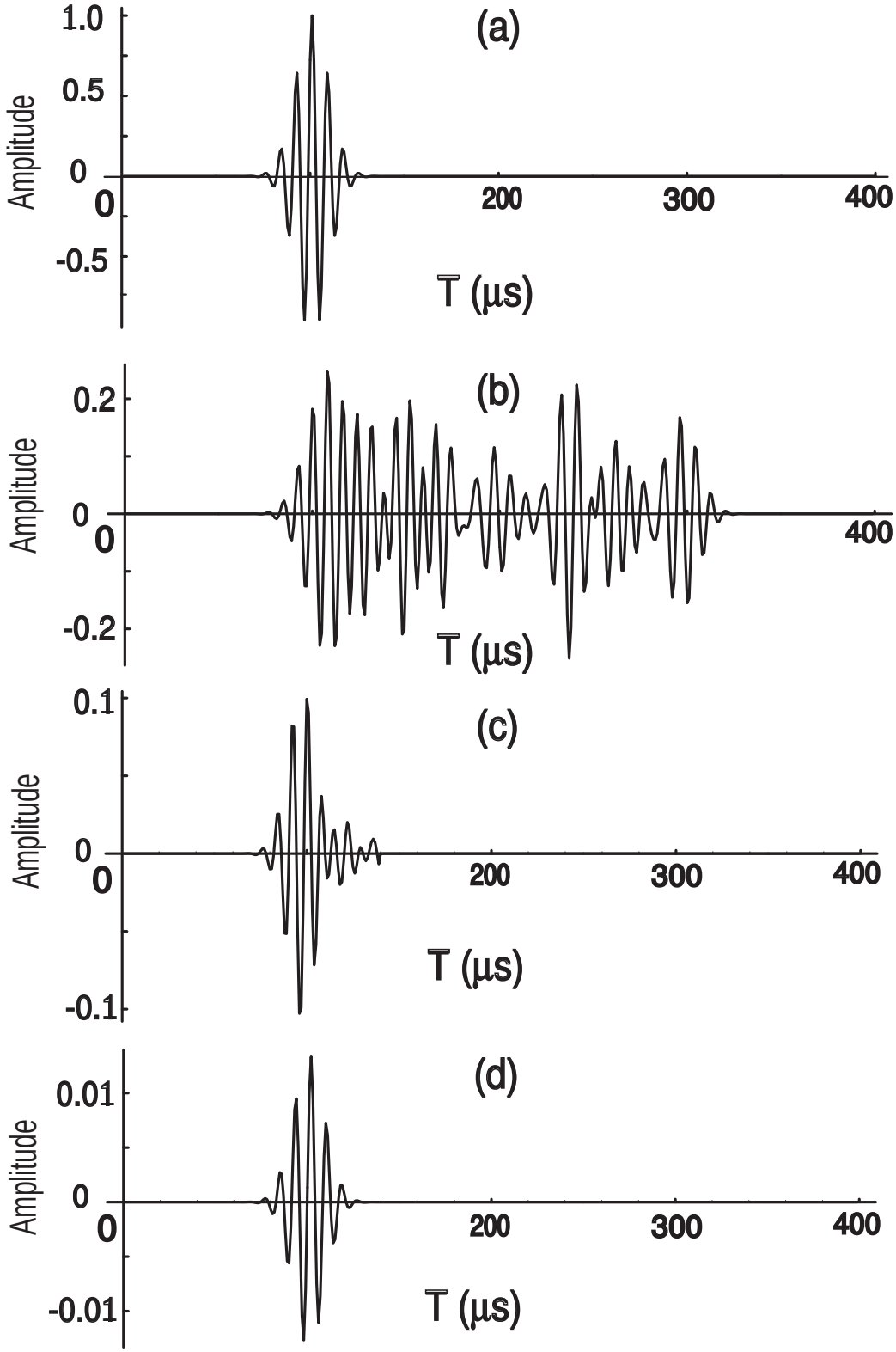

Figure 3.2: (a) source signal; (b) receiver signal; (c) signal back at source at time zero after a perfect time reversal operation; (d) refocused signal without scatterers present, i.e. in a homogeneous medium.

altered scatterers contributes a dephased component with a similar amplitude but a relatively random phase. Dephased wave has a coherent ballistic front in the early part of the receiver signal where m dephased components go through similar paths close to the direct path from the source to the receiver. The m contributions are coherent in phase and the overall deterioration is proportional to m in early time. But in later times, due to more and more scatterings contributions from the m dephasing effects start to take on relatively random phases. Since the ensemble average of a sum of m signals with the same amplitude but random phase has an amplitude that is $\sqrt{m}$ times the individual amplitude, the overall deterioration is proportional to $\sqrt{m}$ in later time. A qualitative understanding for such m to $\sqrt{m}$ transition is that, while each of the m altered scatterers dephases only unperturbed wave in early time, later on they are more and more likely to dephase waves that are already dephased by other scatterers, resulting in redundant dephasing. Since the coherent front is rather short in time, when the m full receiver signals are refocused, the overall deterioration is also proportional to $\sqrt{m}$.

Numerical experiments are carried out to demonstrate the various deterioration patterns. For $N = 40$ and various m, we repeatedly perform the static perturbed time reversal operation for 5 times and average the results to obtain the deterioration curve in Fig. 3.3 (a). A fitting curve $1 - 0.132\sqrt{m}$ is also plotted to demonstrate the $\sqrt{m}$ deterioration pattern. When various m rods are added before time reversal operation, there is also a deterioration pattern which is shown in Fig. 3.3 (b) fitting the curve $1 - 0.17\sqrt{m}$ very well. Since we remove/add the m scatterers completely randomly, not just around the field center, the $\sqrt{m}$ pattern supports our interpretation

of random phase addition among different perturbation contributions. One should also notice that, when up to 20 out of 40 scatterers, i.e. 50% of the total scatterers, are removed/added, the deterioration is only about 60%/80%, much less dramatic than the acoustic TRM experiment [78]. This suggests specific system features play important roles in deterioration rate, as we will fully explore later in the case of acoustic TRM [78].

One might notice that, for the same m, adding rods causes a slightly larger perturbation. This is understandable. Since the unperturbed or perturbed Green function is always obtained from Eq. (3.2), removing m rods leaves a perturbed matrix with $2Nm - m^2$ less terms compared with the unperturbed $N \times N$ matrix; but adding m rods brings in $2Nm + m^2$ irrelevant terms, resulting in a larger percentage of perturbation terms and thus a larger deterioration. Such an asymmetry can also be found by observing $G^a_{N+m}(\omega) \neq G^a_{N-m}(\omega)$ in Eqs. (3.5, 3.6). One should also notice the order of rod numbers in the two stages of time reversal operation makes a difference, though probably small, by observing that Eqs. (3.5, 3.6) imply

$$G^a_{N-m}(\omega)G^r_N(\omega) = (G^a_N(\omega)G^r_{N-m}(\omega))^* \neq G^a_N(\omega)G^r_{N-m}(\omega), \tag{3.7}$$

which leads to

$$S_{N-m,N}(\omega) \neq S_{N,N-m}(\omega). \tag{3.8}$$

The last test is on the linear m to $\sqrt{m}$ transition in deterioration pattern in dynamic time reversal experiments. We remove different numbers of rods at and near the field center after the forward propagation and then time reverse two $1.6\mu s$ time windows with end time $t_e = 1.6\mu s,\ 9.2\mu s$ respectively. The recovered peaks are plotted in Fig. 3.3 (c) for $t_e = 1.6\mu s$ and (d) for $t_e = 9.2\mu s$. Clearly, there is the

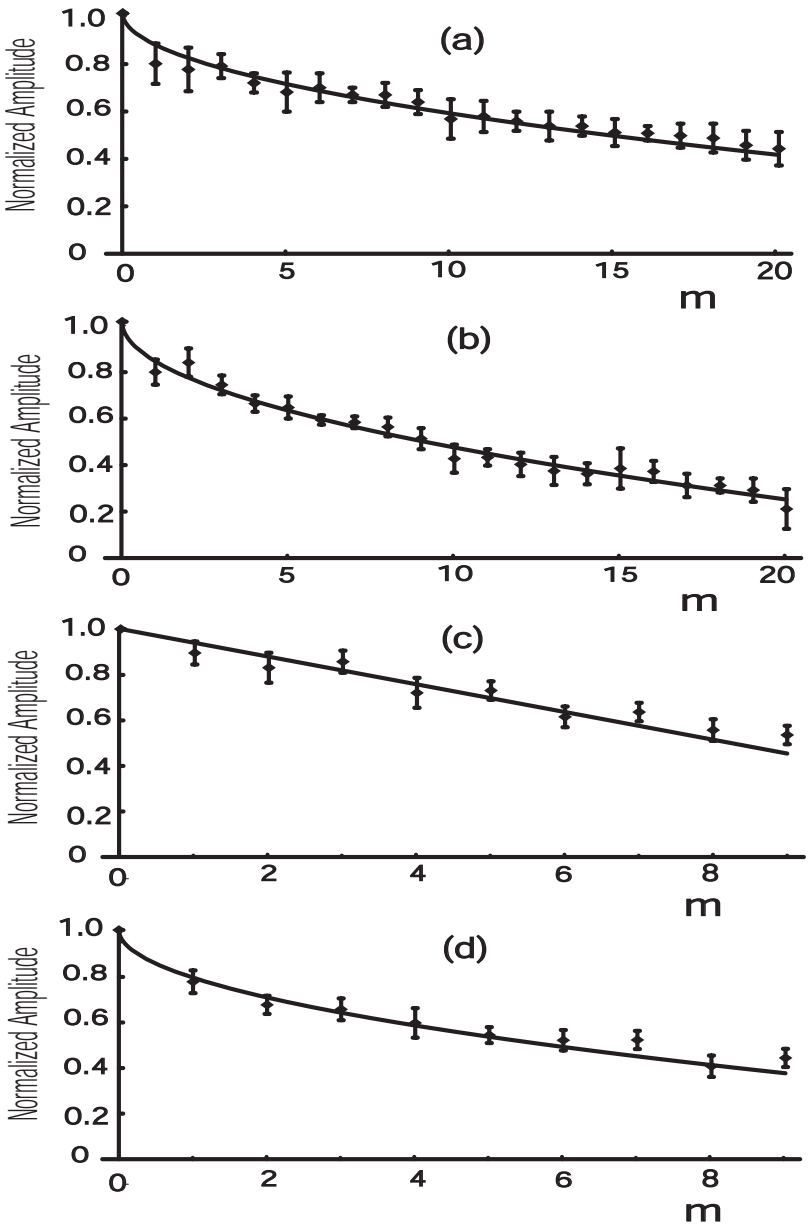

Figure 3.3: (a) Normalized perturbed peak due to removing rods, with a fitting curve $1-0.132\sqrt{m}$; (b) normalized perturbed peak due to adding rods, with a fitting curve $1-0.17\sqrt{m}$; (c) normalized perturbed peak by a $1.6\mu s$ time windows with end time $t_e = 1.6\mu s$, with a fitting curve $1-0.06m$; (d) curve of normalized perturbed peak by a $1.6\mu s$ time windows with end time $t_e = 9.2\mu s$, with a fitting curve $1-0.21\sqrt{m}$.

linear to square root transition in the m-dependence of recovered peaks.

The above theory demonstrates the general shapes of deterioration patterns. However, various systems often deteriorate at very different rates, i.e. with different coefficients in front of m or $\sqrt{m}$. The deterioration rate is a critical indication of a system's stability against perturbation and is intrinsically related to the structure of the system. In the following sections we'll study the acoustic time reversal experiments as an example and, by fully exploring its experimental setup, show how its deterioration rate is determined by its physical parameters.

3.3 General Properties of Time-Reversal Mirror Experiments

Acoustic scattering by a long rod has a rich history. Notably, there are the contributions by Philip Morse for rigid, immovable cylindrical rods [54], and by J. J. Faran, Jr. for elastic rods which incorporate the material features of the rods [24]. Faran's work is consistent with experimental data, still widely cited today, and quite relevant to the rods used in the TRM experiments of Tourin *et al.* [78]. This gives a good starting point from which to move toward the eventual purpose of understanding the time domain Green functions and multiple scattering.

3.3.1 Differential cross-section for a steel rod

The central frequency of 3.2 MHz used in their TRM experiments gives an acoustic wavelength of 0.47 mm, whereas the scattering rod's diameter is 0.8 mm.

This does not quite place the experiments within an s-wave scattering regime, as would typically be considered in the theory of disordered systems. Therefore, as a first step, we need to investigate the angular distribution of the flux of a singly scattered wave to see how well dispersed it is before proceeding to the multiple scattering regime.

For the steel rods of Tourin *et al.* the angular distribution of scattered flux in a single scattering event according to Faran's formula [24] is plotted in Fig. 3.4. It illustrates a couple of key points. First, not too surprisingly, the differential cross-

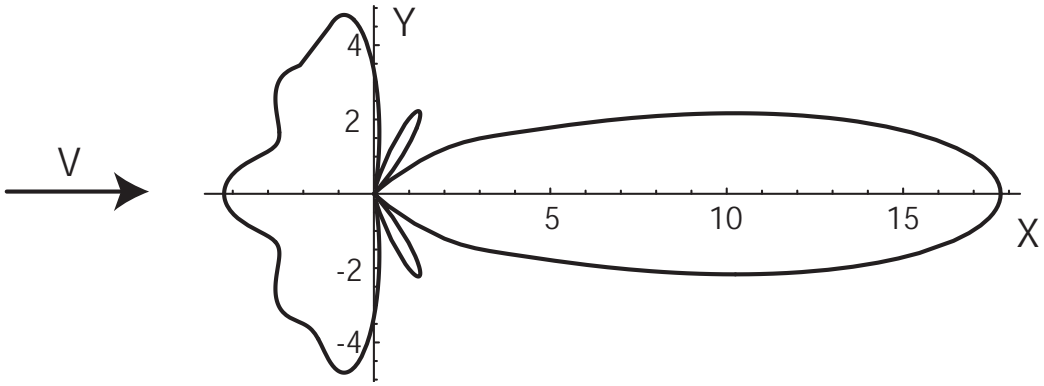

Figure 3.4: Diagram of the angular distribution of scattered sound flux by a steel rod, i.e. the differential cross section of scattering. The incident wave arrives from left and the rod sits at the origin. The rod radius is $0.4\,mm$, compressional velocity in steel $5.9\,mm/\mu s$, shear velocity in steel $3.2\,mm/\mu s$, sound velocity in water $1.5\,mm/\mu s$, density of steel rod $\rho_1 = 7.85\,g/mm^3$, and Poisson's ratio of the rod is $\sigma = 0.3$ [80].

section is greatest in the forward direction, more than three times larger than any other direction. Nevertheless, the angularly integrated forward-scattered waves have nearly equal total flux to the back-scattered waves. Furthermore, the back-scattered flux is fairly uniformly spread in angle. In a sense, if a component of the wave field is back-scattered more than a couple of times, its flux is roughly uniformly distributed in angle; i.e. once a component of the wave field has scattered roughly four or more times,

there is no preferred direction. So even though the scattering is considerably different than s-wave scattering, nevertheless there is a basis for a diffusion-like process taking place for the highly multiply scattered wave components. Second, there is an overall forward momentum associated with each scattered wave component. The reflection symmetry about the incident direction restricts the scattered wave's momentum to being aligned in the incident direction. Because roughly half of the flux is forward and half backward, the more narrowly distributed forward-scattered cross-section leads to a total forward momentum in excess of the total backward momentum. This is in the forward radial direction for the first scattering. However, the effect continues in an averaged sense for multiple scattering, i.e. some momentum in the forward radial direction, but with a decreasing amplitude factor for each successive scattering.

These observations partly explain the coherent wave that appears after averaging receiver signals over the possible configurations of scatterers [19, 80]. From symmetry, the angular flux is equally clockwise and counterclockwise at all times and cancels out after disorder averaging even including multiple scattering effects. On the other hand, there has to be an overall, positive radial momentum, coherent in all configurations, for the scattered waves. Although, that positive momentum should decrease once the wave field is primarily composed of highly, multiply scattered components.

3.3.2 Green function for a source rod

Consider first that the source rod is infinitely long. The time-dependent free-space Green function for a source can be expressed as an integral of a 3D Green function

over the z coordinate aligned along the rods:

$$\begin{aligned} G(\vec{x},\vec{x}';t) &= \int_{-\infty}^{+\infty} \mathrm{d}z'\; G_{3D}(\vec{r},\vec{r}';t) = \int_{-\infty}^{+\infty} \mathrm{d}z' \frac{\delta\left(t-\frac{1}{c}\sqrt{(\vec{x}-\vec{x}')^2+(z-z')^2}\;\right)}{4\pi\sqrt{(\vec{x}-\vec{x}')^2+(z-z')^2}} \\ &= \frac{\theta(t-\frac{1}{c}|\vec{x}-\vec{x}'|)}{2\pi\sqrt{t^2-\frac{1}{c^2}(\vec{x}-\vec{x}')^2}}, \end{aligned} \tag{3.9}$$

where c is the speed of sound and assumed constant, $\vec{x}$ lies in the plane transverse to the rods. Equation (3.9) is consistent with the formulation of 2-dimensional waves by Morse and Feshbach [55]. Note that the δ-function form of the 3D Green function is replaced by a step function with a square root singularity, and a slow time decay $\propto t^{-1}$.

Although the rods are very long compared to other scales in the experiment, their finite length does cut off the slow decay of the Green function beyond some time. This depends on the planar distance $|\vec{x}-\vec{x}'|$, the rod length L (120 mm), and the value of z. The finite-rod Green function is given by $\int_{-L/2}^{L/2}\mathrm{d}z'\; G_{3D}(\vec{r},\vec{r}';t)$, and in general, it involves two cut-off times $t_1=\sqrt{(\vec{x}-\vec{x}')^2+(L/2-z)^2}/c$ and $t_2=\sqrt{(\vec{x}-\vec{x}')^2+(L/2+z)^2}/c$ which become identical for the points in the mid-plane ($z=0$). The Green functions "turn on" at $t=|\vec{x}-\vec{x}'|/c$ and behave as the infinite-rod model until $t=Min[t_1,t_2]$; deviation from the infinite model appears in the window $Min[t_1,t_2]<t<Max[t_1,t_2]$ and it vanishes beyond $t=Max[t_1,t_2]$. For example, within the zone of the scattering rods and in the mid-plane, for the mean free planar path, 4 mm, the Green function "turns on" at2.7 μs, and shuts off at 40.1 μs, with a non-zero duration of 37.4 μs.

The time cutoffs are more pronounced from the source rod due to the distance to the scattering field, $|\vec{x}-\vec{x}'|=110\ mm$. The Green function has a nonzero duration of

9 μs in the rod middle plane beginning at 73 μs. In addition, the waves propagating into the scatterer field arrive within the angular range of roughly $[-\frac{\pi}{4}, \frac{\pi}{4}]$ from the horizontal plane as shown in Fig. 3.5. Ignoring the effects of tilting cylinders (i.e. ignoring paths which alter their angle with respect to the horizontal), signals scattering from one rod to another within the scatterer field effectively see only a very short segment of their neighboring rods. This drastically reduces the duration of Green functions inside the scatterer field to $\frac{0.41|\vec{x}-\vec{x}'|}{c}$. For the mean free path of 4 mm, the duration now becomes 1.1 μs, down from the above 37.4 μs by considering the finite length of rods. We show later that such short durations in the Green function can provide a measure of the number of scattering events that a wave signal experiences. Note that there are two confining plates supporting the scatterers at either of their ends. To a certain extent, they could serve to effectively lengthen the scatterers by reflecting the wave directed at the plates back into the field. Such a virtual extension in the z direction by reflection does not significantly extend the signal since the source rod is left open in space and large θ wave signals have already escaped from the scatterer field.

For convenience ahead, we shall use a relative time $t_r = t - \frac{1}{c}|\vec{x}-\vec{x}'|$ and introduce an approximation to the infinite-rod Green function

$$G(\vec{x}, \vec{x}'; t) = \frac{\theta(t - \frac{1}{c}|\vec{x} - \vec{x}'|)}{2\pi\sqrt{t^2 - \frac{1}{c^2}(\vec{x} - \vec{x}')^2}} = \frac{\theta(t_r)}{2\pi\sqrt{t_r(t_r + a)}} \approx \frac{\theta(t_r)}{2\pi\sqrt{t_r a}}, \tag{3.10}$$

where $a = \frac{2}{c}|\vec{x} - \vec{x}'|$ is twice the free transport time, and in the last step $(a + t_r)$ is replaced with a in the denominator since the Green function is dominated by $\sqrt{t_r}$ at small t_r. For the mean free path in the experiment which corresponds to $a = 5.4\mu s$, both the exact and approximate infinite-rod Green functions are plotted in Fig. 3.6(a)

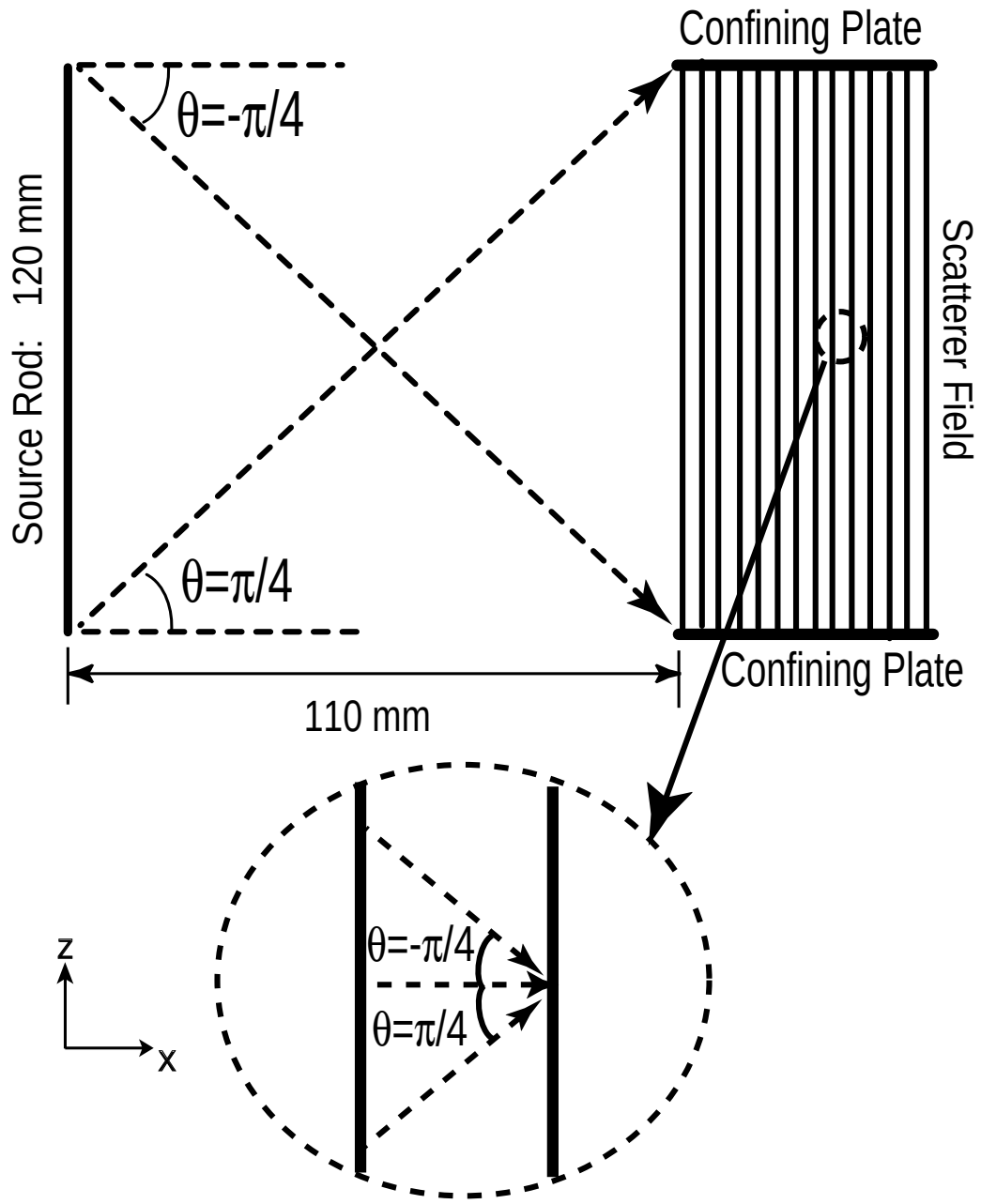

Figure 3.5: Diagram of the angular range of wave trapped inside the scatterer field. 110 *mm* is the closest distance between the source rod and the scatterer field, which corresponds to a maximum incoming angle range of roughly $[-\frac{\pi}{4}, \frac{\pi}{4}]$. The limited angle range further reduces the Green function duration inside the field, as illustrated in the inset taken at the middle of two rods.

for comparison. The approximation does not introduce noticeable deviation until late in the long time tail, long after the average duration 1.1 μs for Green functions inside the scatterer field when the limited angular range of incoming wave is considered.

3.3.3 Green functions of a single scattering and beyond

The next step is to consider one or more scattering events. The first quest is to determine the signal received at some point from a singly scattered wave without considering the finite rod length. Our interest is in its time dependence, not the overall amplitude. Thus, we neglect the amplitude prefactors from the source strength and the angular dependence of the differential cross-section. The wave component traveling directly from the first rod to the third rod is ignored here, but its amplitude is covered by the results in the previous section. For simplicity, the scattering strength $\epsilon(\omega)$ is taken as 1. The single scattering signal can be written analytically in real time as

$$\begin{aligned} G_1 \quad & (\vec{x}-\vec{x}_i, \vec{x}_i-\vec{x}'; t) \\ = \quad & \int_0^t \mathrm{d}t' \; G(\vec{x}, \vec{x}_i; t-t') G(\vec{x}_i, \vec{x}'; t') \\ = \quad & \frac{1}{4\pi^2} \int_0^t \mathrm{d}t' \; \frac{\theta(t-t'-\frac{1}{c}|\vec{x}-\vec{x}_i|)}{\sqrt{(t-t')^2-\frac{1}{c^2}(\vec{x}-\vec{x}_i)^2}} \frac{\theta(t'-\frac{1}{c}|\vec{x}_i-\vec{x}'|)}{\sqrt{t'^2-\frac{1}{c^2}(\vec{x}_i-\vec{x}')^2}} \\ = \quad & \frac{\Theta\left(t-\frac{1}{c}\left\{|\vec{x}-\vec{x}_i|+|\vec{x}_i-\vec{x}'|\right\}\right)}{2\pi^2\sqrt{t^2-\frac{1}{c^2}\left(|\vec{x}-\vec{x}_i|-|\vec{x}_i-\vec{x}'|\right)^2}} K\left(\sqrt{\frac{t^2-\frac{1}{c^2}\left(|\vec{x}-\vec{x}_i|+|\vec{x}_i-\vec{x}'|\right)^2}{t^2-\frac{1}{c^2}\left(|\vec{x}-\vec{x}_i|-|\vec{x}_i-\vec{x}'|\right)^2}}\right), \end{aligned} \tag{3.11}$$

where $K(z)$ is the complete elliptic integral of the first kind, and $\Theta(z)$ is the Heavyside step function as before. G_1 first jumps to the constant $c/(8\pi\sqrt{|\vec{x}-\vec{x}_i||\vec{x}_i-\vec{x}'|})$, and decays very slowly thereafter. As in the preceding subsection, the single scattering

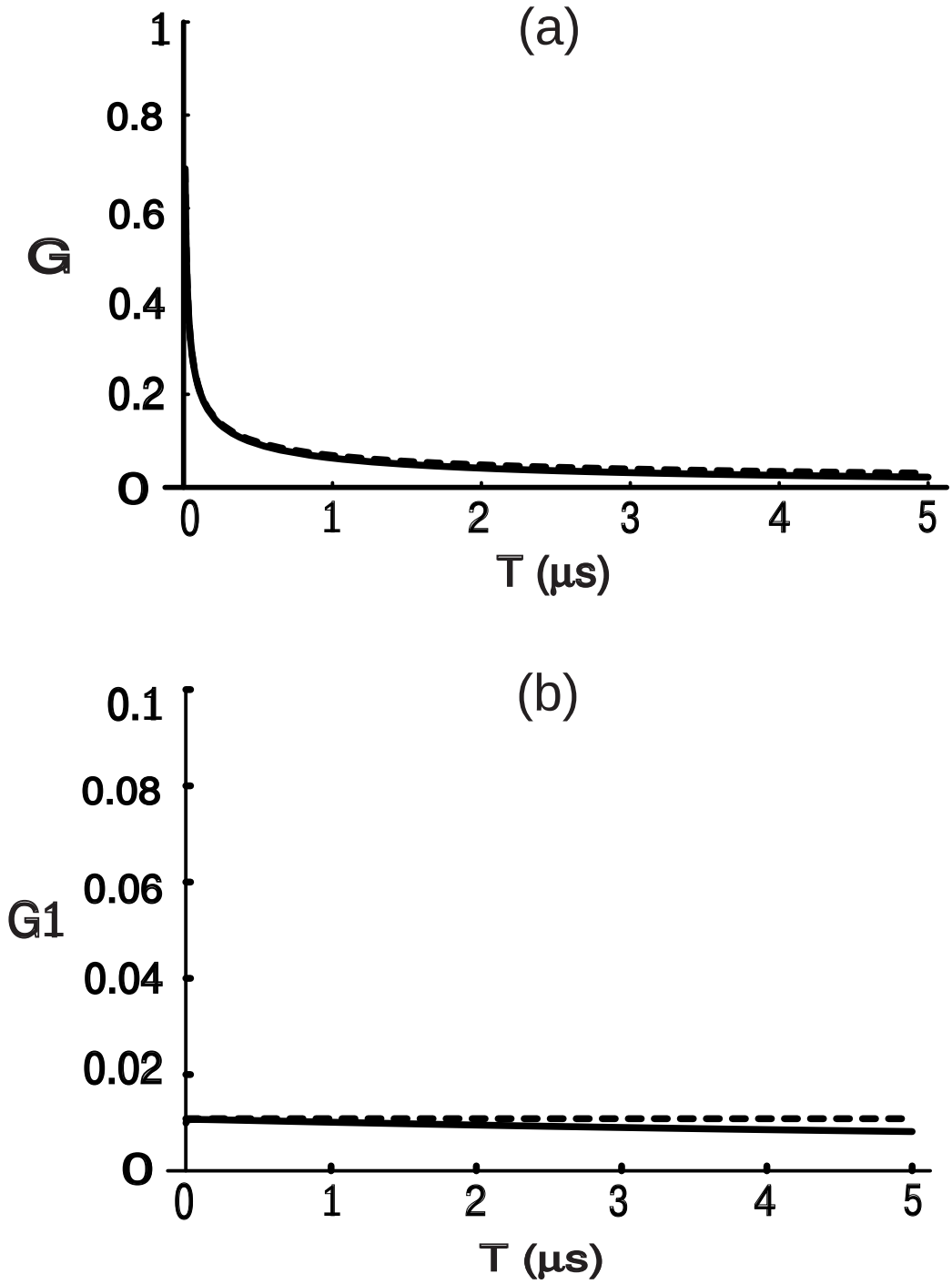

Figure 3.6: (a) The Green function (solid line) and its approximation (dashed line) for the mean free path, 4 mm, in the time-reversal experiment. (b) The incoming wave on a third rod (solid line) after a δ wave from the first rod is scattered once by the second rod, i.e. the Green' function of a single scattering event, for $t_{\hat{x},i} = 2.7\mu s,\ t_{i,\hat{x}'} = 5\mu s$ in the text. The approximate form (dashed constant line) $(8\pi\sqrt{t_{\hat{x},i}t_{i,\hat{x}'}})^{-1}$ is also drawn for comparison. The difference between the two curves gets smaller for longer $t_{\hat{x},i},\ t_{i,\hat{x}'}$.

Green function can also be formulated in terms of relative times at short time as

$$G_1(t_{\hat{x},i}, t_{i,\hat{x}'}; t_r) = \frac{\Theta(t_r)}{8\pi\sqrt{t_{\hat{x},i}t_{i,\hat{x}'}}}, \tag{3.12}$$

where $t_{\hat{x},i},\ t_{i,\hat{x}'}$ are free travel times between rods (or, from source to rod and rod to receiver), and t_r is generalized to include the sum over all segments of a path. An example of both the exact and approximate forms of the single scattering Green function is drawn in Fig. 3.6(b) for comparison.

Quite interestingly, a single scattering event completely removes the singularity and flattens out the entire wave signal over a long time range. The spreading feature is enhanced with successive scatterings. For higher order G_n, the short, relative time expression becomes

$$G_n(t_{\hat{x},i} \ldots t_{j,\hat{x}'}; t_r) = \frac{t_r^{(n-1)/2}}{\pi^{1/2}(8\pi)^{(n+1)/2}\sqrt{t_{\hat{x},i} \ldots t_{j,\hat{x}'}}} \prod_{i=1}^{n} \frac{\Gamma\left(\frac{i}{2}\right)}{\Gamma\left(i\right)}, \tag{3.13}$$

where $\Gamma(z)$ is the Gamma function. This expression illustrates two important physical features. First, each successive scattering leads to an additional $t_r^{1/2}$ factor in the short, relative time expression. Second, each successive scattering is reduced by a factor $\Gamma(n/2)/2\sqrt{2\pi}\Gamma(n)\sqrt{t_{ij}}$, which falls off extremely rapidly for large-n (there is also the extra scattering strength $\epsilon(\omega)$ factor). While the number of distinct scattering paths increases exponentially with n, using Stirling's formula to give the leading asymptotic behavior of the factorials suggests that the exponential number of paths, $\exp(an)$, cannot compensate for the resultant $\exp(-n\ln n/2)$ behavior found in the decreasing prefactors.

The expressions derived above have been based on an infinite-rod model. One has to be alert to the deviation between the infinite-rod model and the experimental

system, which grows quickly with scattering order as the infinite-rod model predicts a divergence in amplitude for multiple scattering signals which always vary slowly in experiments.

3.3.4 Amplitude flattening and information carriage of multiply scattered signals

In understanding the slowly-varying amplitude of receiver signals in TRM experiments, cut-off times in the Green functions due to the finite rod length play a critical role and can be easily shown in one example. For an infinite-rod model, a signal, say, t is expected to be proportional to $t^{3/2}$ after time convolution with an infinite-rod Green function. However, a cut-off time t_c of the Green function means the signal after time convolution is still linear in t for $t > t_c$. In addition, if the incoming signal itself has a cut-off time t_i, the convoluted signal only grows as the theory up to $t_1 = Min[t_c, t_i]$, then decreases instead of growing as t, and reaches zero at $t_2 = (t_c + t_i)$. Inside the scatterer field, a typical cut-off time is as small as $1.1\mu s$ so the deviation is striking. The deviation is further amplified in multiple scattering. In the infinite-rod model, one can show the fully summed n-scattering term on a rod/receiver diverges much quicker than $t^{(1+n)/2}$ for $n \geq 2$; but experimentally it always drops to zero at a finite time.

A simplified analysis can explain the experimental signal. A single n-scattering contribution involves $(n+1)$ cutoff times $t_1, t_2, ..., t_{n+1}$. Only the very first portion of length $Min[t_1, t_2, ..., t_{n+1}]$, which is smaller than $1.1\mu s$, is consistent with the infinite-rod derivations. The physical n-scattering signal evolves slowly but in a complicated

way thereafter and goes to zero at $t = \sum_{i=1}^{n+1} t_i$. An $n(\geq 2)$-scattering component in a finite rod's signal has a tiny front portion, even smaller than $t^{(1+n)/2}$, due to a higher power of t and a prefactor decreasing with increasing n. The peak of the real $n(\geq 2)$-scattering signal appears later with increasing n, as the peak of high order scattering signal is generated by the peak of its preceding order. With increasing order, the peak amplitude decreases to zero due to the limited overall signal spread into longer time. For large n, the real signal eventually becomes a very flat, near-zero, background signal in a range $[t = 0, \sum_{i=1}^{n+1} t_i]$. On the receiver, a signal with small scattering order n arrives early with a relatively large amplitude. On the other hand, the number of distinct scattering paths increases exponentially with n and they arrive late at the receiver. So there are more contributions, despite each being relatively weaker, to later part of the receiver signal. When all factors are taken into account, the receiver signals in experiments become quite flat over a long range of time, followed by a long fading tail [78, 20, 42, 21]. Even though a short pulse is used as source signal in the experiments which may facilitate the formation of the slowly-varying receiver signal, the spreading-and-flattening mechanism due to finite rod length should exist for any other source signal with finite duration.

The spread in time can give us some estimate about the maximum number of nontrivial scatterings a wave experiences before reaching a receiver. For example, the experimental receiver signal has a nonnegligible range of 250 μs [18]. Taking out the two cutoff times between the scattering field and the source rod/receiver, the 231 μs remaining is 210 times the cutoff period 1.1 μs for the mean free path 4 mm, which corresponds to 210 nontrivial scatterings within the scattering field. For

a small perturbation originating from the field center, an average of 105 nontrivial scatterings is expected before the perturbed wave reaches a receiver. This is consistent with a diffusive approach to the transport of both the unperturbed and perturbed wave signals to be discussed later.

Given that the acoustic signal arriving at a later time has on average scattered more times and given causality, the latter part of the receiver signal, in a sense, collects more information about the source and scatterers than the earlier part. There is enhancement of refocusing at the source after time reversal because the field collects waves directed toward its edges, which would otherwise be missed by the receiver array in a homogeneous medium [28]. Those otherwise missed waves are redirected to the receiver array through multiple scattering. Since the process of multiple scattering spreads waves toward a later signal time, the time reversal of the latter part of the receiver signal can better recover the source pulse; i.e. it contains waves from a wider range of space as observed by Derode *et al.* [20]. This more-information-at-later-time feature also offers a qualitative explanation for the dynamic time reversal experimental observations in [78], in which the recovered source pulse using the latter part of the receiver signal has greater sensitivity to the perturbation than the front part. Remarkably, every point in the same x-y plane on the receiver end bears the same feature and similar information due to the redirection of waves through scatterings. Thus, even one transducer can be used for time reversal [20].

As a final remark, note that the minimum spacing 1.5 mm between rods [18] is far larger than the central wavelength 0.47 mm, thus a persistent bouncing of wave amplitude between two closely placed rods, which can cause divergences in certain

systems [11] cannot exist in these experiments.

3.4 Perturbed Time-Reversal Mirror Experiments

From the above considerations, a small perturbation originating from the center of the scattering region would scatter from a few to a couple hundred times before the perturbed wave reaches a receiver. This motivates a diffusive approach to the transport of both the unperturbed and perturbed signals.

3.4.1 Perturbation evolution in a diffusive scattering field

The basic picture of the perturbed TRM experiments can be depicted as follows. In a strongly scattering medium where diffusive transport of wave signal is dominant, any local perturbation in the scattering medium is imprinted onto wave signal passing the perturbed area by changing the signal's scattering path orientation, length, and therefore phase. The perturbed wave signal subsequently transports diffusively inside the medium as well. As time increases, more and more of the original wave signal diffuses past the altered scatterers giving rise to greater perturbed wave signal. According to Snieder and Scales [70], both ballistic and coda waves have critical length scales of stability against scatterer displacement that are much smaller than the wavelength λ and shrink with increasing number of scatterers N as $\frac{1}{\sqrt{N}}$. Here $\lambda = 0.47\ mm$ and scatterer diameter is $0.8\ mm$. So even if a single scatterer is removed or added, any wave signal diffusing into the perturbed region is completely dephased, and will not be refocused back to the source after the time-reversal operation.

For simplicity of presentation, our analysis focuses on the central receiver. How-

ever, the arguments are general and applicable to any other receiver. From the various experiments [78, 20, 19, 80, 79, 42, 21, 28] and the arguments of the preceding section, strongly multiply scattered wave signals do not vary a great deal in their mean intensity within a long time window, except for an initial transient from a ballistic wave front. Let the constant A denote the average diffusive wave intensity in the scatterer field. In the following calculation, we consider only the intensity of the unperturbed and perturbed scattered wave signals because recovered signal at the source after the time reversal operation in the unperturbed case is a constructive addition of all contributions.

Consider the perturbation of a single scatterer. The perturbed TRM experiment is viewed in the following way: there are N rods in forward propagation and $(N-1)$ rods in backward refocusing. The perturbation is caused by the scattering of the extra rod in the forward propagation stage whose perturbed, resultant wave field does not refocus at the source in the second stage. The 2-dimensional diffusive transport of the perturbed wave intensity has a Green function [55]

$$G_{diff}(r,t) = \frac{1}{t} e^{-r^2/4Dt} \theta(t), \tag{3.14}$$

where $t = 0$ denotes the time when the perturbed signal first arrives without any scattering and r is the distance between the rod and observation point. D is the diffusion constant, which is measured to be 3.2 $mm^2/\mu s$, and shown to be consistent with theoretical predictions [79]. It is appropriate to use the 2D diffusion Green function in short time because the various effects due to the differential scattering cross-section for a single rod and the limited rod length are contained in the experimentally measured diffusion constant D. Note that the receiver signal arrives directly from wave

scattered by rods in a region near the right edge of the rectangular scattering zone with a thickness of the mean free path 4 mm as shown in Fig. 3.7. This is termed "emission zone" in the following. A dephased diffusing wave has to reach a rod in emission zone first before being recorded by the receiver. The perturbed scatterer field is sketched in Fig. 3.7. The time to travel the distance L_0 from the field center to its upper/lower edges is $t_0 = 73\mu s$ for free waves [18], and much longer than $73\mu s$ for diffusing waves due to the delay by scattering.

With the assumed roughly constant intensity A for the overall signal near the center, the relative intensity of the perturbed waves on a rod a distance r away becomes

$$\begin{aligned} RA(r,t) &= \frac{A'(r,t)}{A} = \frac{\int_0^t A \cdot G_{diff}(r,t')\ \mathrm{d}t'}{A} = \int_0^t \frac{1}{t'} \mathrm{e}^{-r^2/4Dt'}\ \mathrm{d}t' \\ &\approx \frac{4Dt}{r^2}\mathrm{e}^{-r^2/4Dt} \qquad t < \frac{r^2}{4D} \\ &\approx \frac{r^2}{4Dt} - \ln\frac{r^2}{4Dt} - \gamma \qquad t > \frac{r^2}{4D}. \end{aligned} \tag{3.15}$$

The last two forms follow from the asymptotic and series results for the exponential integral function, respectively, which describes A'/A. For $r = 20\ mm$, the relative amplitude A'/A of the dephased wave on a rod at the emission zone center is plotted in Fig. 3.8(a), which is reasonably approximated by a linear curve $(7.74 \times 10^{-3})t$ in the time range $[0\mu s,\ 200\mu s]$ of interest to us. Besides the diffusing portion, dephased wave also has a short, small ballistic front. When both parts are considered, the linear curve offers a fairly good approximation for the overall perturbed intensity in both short and intermediate times.

The perturbed waves recorded by the receiver are mostly from rods at or near the center of the emission zone, according to the angular distribution of scattered

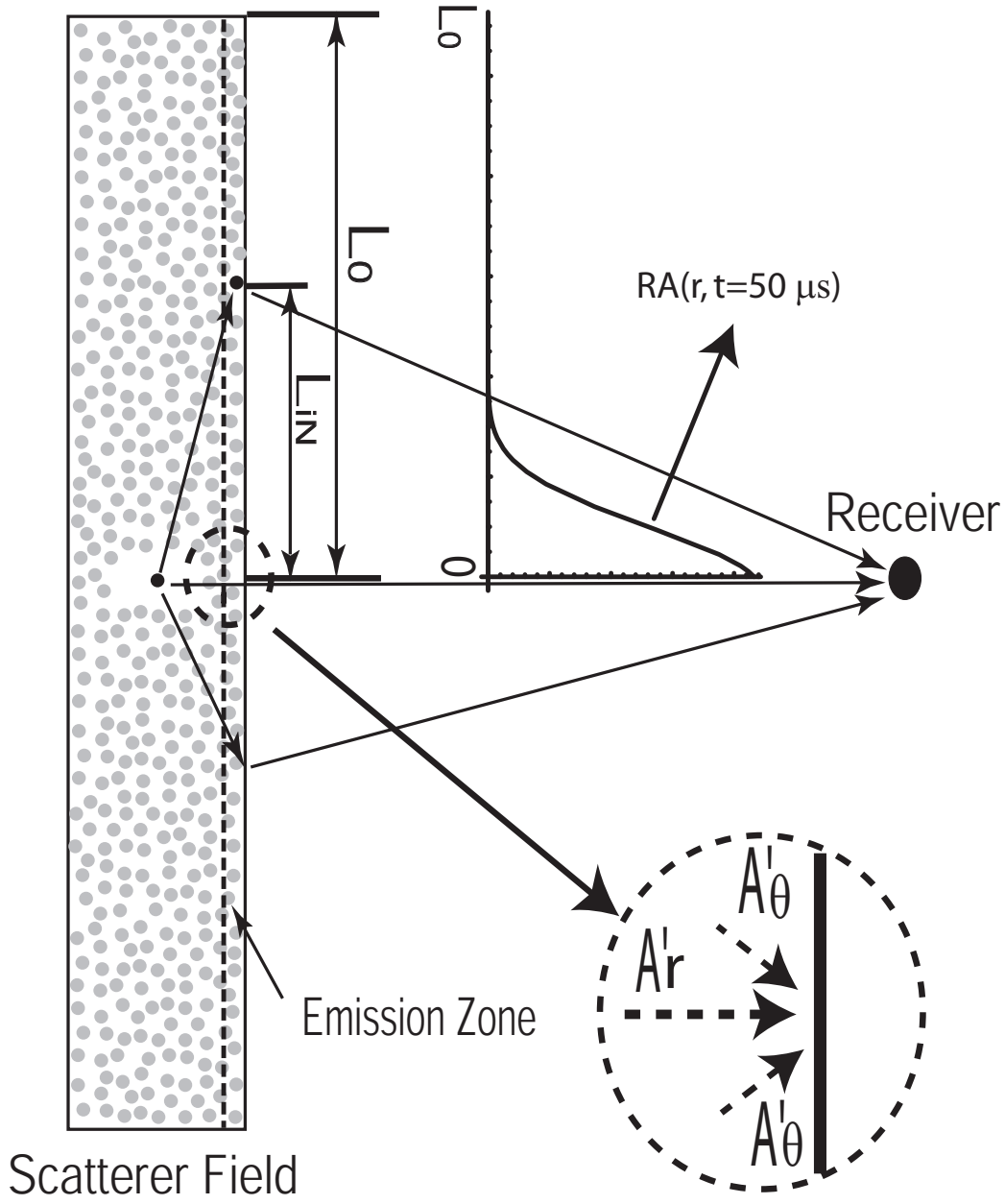

Figure 3.7: The top view of the scatterer field after m rods are removed from its center. The dotted area consists of a uniform, yet random, distribution of scatterers. The central blank area is where m rods are removed. The emission zone is defined as between the dashed line and the right edge of the field. Perturbed waves have to diffuse into the emission zone first before being scattered by a rod in the zone toward the receiver. The $RA(r, t = 50\mu s)$ curve is the relative intensity of perturbed wave along the emission zone $50\mu s$ after it first reaches the center of the emission zone. The curve takes into account the fact that $t = 0\mu s$, the moment of the first arrival of perturbed wave, is defined differently along the emission zone.

flux in Fig.3.4. Perturbed waves reaching a rod far from the center experience many more scatterings, and thus smaller flux as can be seen from the r-dependence in Eq. (3.14). These waves also arrive at the receiver much later since they diffuse longer distances. The above calculated dephased incident wave on an emission zone rod can be separated into a radial part A'_r and an angular part A'_θ as shown in Fig. 3.7. The two parts have different fractions transmitted to the receiver due to their orientation. For the radial part, to the rods near the center of the emission zone that contribute significant dephased waves to the receiver, the receiver sits at different forward-scattering angles. When all contributions are summed up, it equals $\frac{A}{2}$, half of the total incident wave on a single emission zone rod. For the angular part, to all the rods in the emission zone, the receiver sits at either their forward-scattering angles with very small flux or the backscattering angles of those rods far away (thus small flux and late arrival on receiver). So the angular part only gets a small fraction transmitted. Overall, the transmitted wave is a portion more than 50% of but still less than the total dephased wave incident on a rod at the emission zone center, which is denoted for now with a coefficient η. This limitation does not apply to the overall incident waves on the receiver, which can be viewed as from many diffusive sources inside the field instead of the only source at the field center for dephased waves. So the transmitted dephased waves toward the receiver is $\eta A'(t)$ and the overall incident wave is roughly A, which gives the lost portion of amplitude of coherent waves on the receiver due to removing one rod as ($t \leq 200\mu s$)

$$f(t) = 1 - \sqrt{1 - \eta A'(t)/A} \approx 1 - \sqrt{1 - \eta(7.74 \times 10^{-3})t}. \tag{3.16}$$

Here we return the discussion to amplitude for the convenience of considering the

relative phase among waves perturbed by different rods.

Now consider the overall dephasing effect by $m(m \geq 1)$ rods instead of 1 rod. The previous general theory section shows the m dephasing effects coherently add up together in early time. We obtain

$$f(m,t) \approx m(1 - \sqrt{1 - \eta(7.74 \times 10^{-3})t}). \tag{3.17}$$

One should take some caution when employing $f(t)$ and $f(m,t)$, the percentage of dephased wave in the overall receiver signal. They are rather loosely defined for long time. While the near constant amplitude A for receiver signal is taken from experiments with finite rods, $A'(t)$ calculates the dephased wave as if in an infinite-rod system, despite the diffusion constant D makes the short time prediction fairly precise. Therefore $\eta(7.74 \times 10^{-3})t$ grows beyond 1 at long time, clearly overcounting dephased waves. However, attempts to redefine it within the range of $[0, 1]$ will not succeed without making corrections to Eq. (3.15) to better account for finite rod length, which will not be easy and also unnecessary. As shown later, using Eqs. (3.16) and (3.17) with caution can still yield good reproduction of experimental observations.

3.4.2 The dynamic and static time-reversal experiments

Once the dynamic time-reversal operation is performed, all unperturbed receiver signal will be in phase at the source location at time zero; and the dephased part of $f(m,t)$ is obviously out of phase and becomes background noise. The normalized deteriorated source signal reconstructed by a short $5\mu s$ time window with end time t_e becomes

$$P_d(m,t_e) = 1 - f(m,t_e) \approx 1 - m(1 - \sqrt{1 - \eta(7.74 \times 10^{-3})t_e}). \tag{3.18}$$

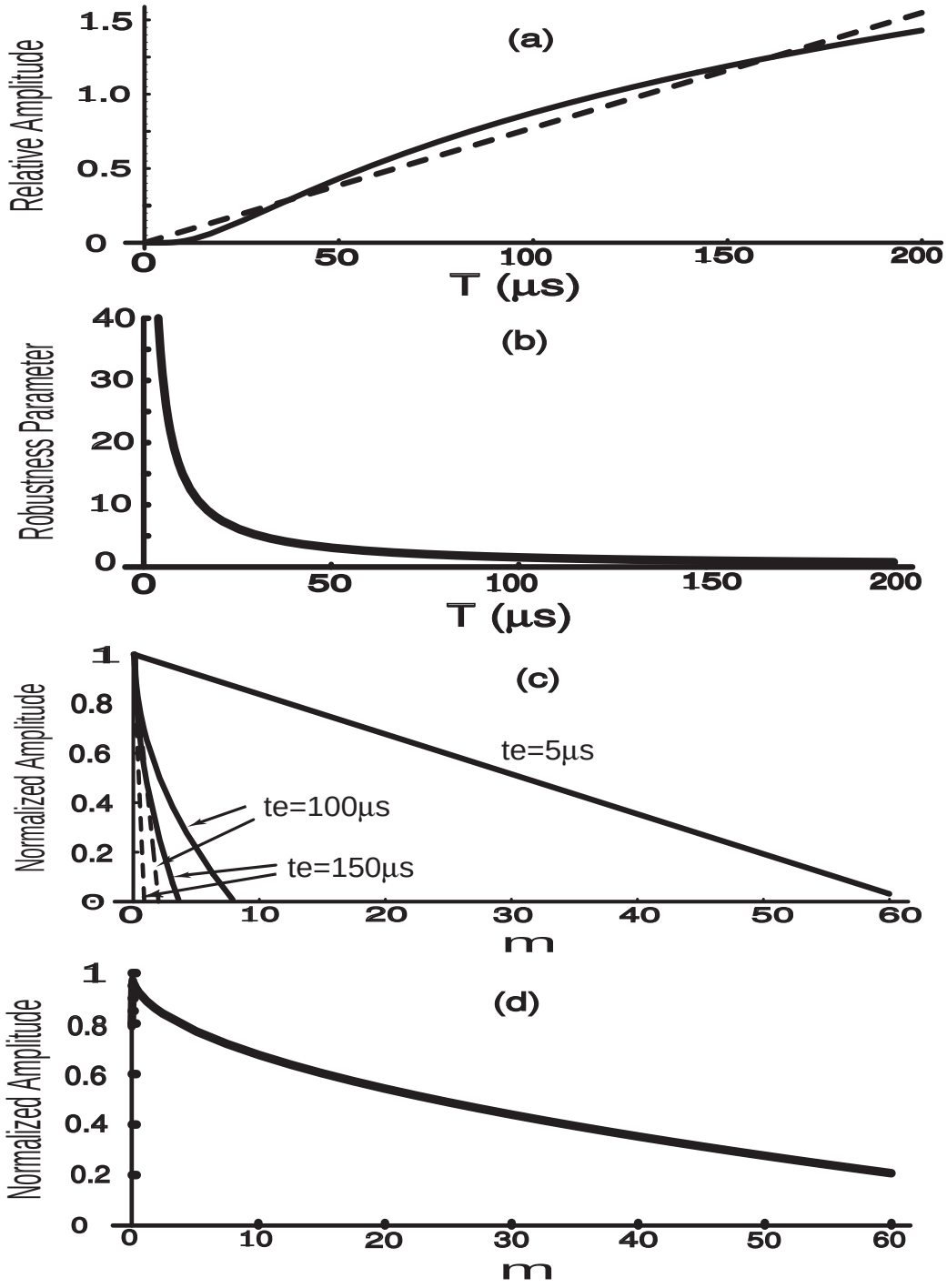

Figure 3.8: (a) The relative intensity of perturbed wave at $r = 20\ mm$. The solid line represents the relative intensity. The dashed line is a fitting curve $(7.74 \times 10^{-3}) * t$. $[0\mu s,\ 200\mu s]$ is the time range in which the vast majority of experimental signal was recorded. (b) The half-peak deterioration curve. The robustness parameter is the number of removed rods. (c) Normalized peak amplitude versus the number of removed rods for different t_e of the end of the time-reversal window. $t_e = 5, 100, 150\mu s$ for the straight lines from right to left. For $t_e = 100, 150\mu s$, the $\sqrt{m}$ correction is taken into account and the deterioration curves are redrawn in solid. (d) Curve of the recovered source signal (normalized) in the static time-reversal experiment.

Setting $P_d(m, t_e) = 1/2$, with a choice of $\eta = 0.8$, Eq. (3.18) gives the robustness parameter as half peak deterioration

$$m = \frac{1}{2(1 - \sqrt{1 - \eta(7.74 \times 10^{-3})t_e})}. \tag{3.19}$$

The $m - t_e$ relation is drawn in Fig. 3.8(b), which is the same as the experimental observation in [78]. It's worth noting that η is between 1/2 and 1, just as predicted above. Also Eq. (3.18) shows that for a fixed t_e its recovered peak decreases linearly with the number of removed rods, and for different t_e the decrease rate increases linearly with t_e, which are shown in Fig. 3.8(c) and fit the experimental results [78] well except some tail deviation to be explained later.

According to the general theory section, we should realize that during most of the $250\mu s$ duration each of the m perturbed rods contributes a dephasing effect with the same magnitude but a randomly distributed phase. So the m in the deterioration term of Eq. (3.18) should be replaced with $\sqrt{m}$ for large t_e to account for the relative phases. The two curves $t_e = 100, 150\mu s$ are redrawn in Fig. 3.8(c). The fitting is improved. But such a $\sqrt{m}$ correction doesn't affect Fig. 3.8(b) because, as t_e grows bigger, even one added rod is enough to create a half-peak deterioration and the random phase effect can't be observed.

To fully reproduce the static time-reversal experimental observation, Eq. (3.18) should be integrated with the $\sqrt{m}$ correction over the entire $250\mu s$ duration

$$\begin{aligned} P(m) &\approx 1 - \frac{\int_0^{250} f(m, t, \gamma)\, dt}{250} \\ &= 1 - \sqrt{m}\frac{\int_0^{250}(1 - \sqrt{1 - 0.8\gamma(7.74 \times 10^{-3})t})\, dt}{250}, \end{aligned} \tag{3.20}$$

where a factor $0 < \gamma < 1$ is added in to counterbalance the overcounting of perturba-

tion at long time as discussed at the end of the preceding subsection. With $\gamma = 0.24$, a reasonable choice for the experimental signal [78], the theoretical curve is given in Fig. 3.8(d). The fitting between our curve and the experimental observation, which is also an ensemble average of 5 measurements, is excellent. As mentioned before, the $\sqrt{m}$ pattern in the static time-reversal experiment is essentially an ensemble average feature.

3.5 Conclusion

In conclusion, we have developed a general scattering theory for perturbed time reversal focusing, which is then used to diagnose the evolution of a local perturbation in a random wave field in two cases, one ballistic and quantum and the other diffusive and classical. We have shown that dephasing is the reason for deterioration in reconstructed source signal. Exhibition of perturbation effects in measurement can be affected by the characteristics of a specific system. But a number of features are shared by many, if not all, multiple scattering systems. That is, the linear-m to $\sqrt{m}$ transition in the dynamic deterioration pattern and the $\sqrt{m}$ pattern in the static deterioration are general features, though the specific coefficients in front of m or $\sqrt{m}$, which determine the deterioration rates, can be system dependent. The detailed analysis of the time reversal mirror experiments provides a number of insights into many experimental observations not previously well understood, such as the slowly-varying amplitude of the receiver signal, different focusing qualities by the different parts of the receiver signal, and the coherent wave appearing after ensemble averaging the receiver signals. Studying perturbation effects, especially in diffusive

media, can reveal a great deal about the internal dynamics of a system since the perturbed wave travels inside the medium before being recorded. The basic theory presented here can be modified to treat perturbation problems in a number of areas involving discrete scatterers, such as impurity scattering in solids, etc.

Chapter 4

Conductance Fringe Patterns in Two Dimensional Electron Gases

In this chapter, we employ scattering theory to investigate conductance fringes often observed in two dimensional electron gases (2DEGs) at low temperature. Typical fringe patterns at weak scattering limit are analyzed. A general, convenient formula for evaluating thermal averaging of an arbitrary physical signal is given and then used to study electron flux through an open QPC. The appearance of resonances is shown to be likely only in very limited ranges. The impact on fringe patterns by different impurities is also discussed.

4.1 Introduction

In the past few years, with fast advancements in experimental techniques, physicists have been able to obtain many fascinating images in nanostructures. In partic-

ular, scanning probe microscope (SPM) has been used to investigate various features of systems in either open or closed configuration [76, 49, 48, 22, 15, 16, 53, 91, 30, 51, 41, 89, 33]. However, the current SPM technique is not refined enough so that its presence introduces negligible change to the state of the system under measurement. Frequently the perturbation introduced by an SPM tip potential is comparable with and sometimes even dominates over the system's own dynamical factors. Thus measurements obtained through an SPM reflects strongly perturbed states of the system, instead of the original state intended to measure. One example is using an SPM tip to probe a single-electron quantum dot [23], where the shrinkage in conductance rings accompanying the changing SPM tip voltage applied to the same quantum dot clearly testifies the strong impact of measurement. Therefore one important task facing the SPM imaging research is to understand the change made in a system by an SPM tip, and thereby connect measurement data to the original unperturbed state of the system. Significant progress has been made toward this goal. In particular, Fiete and Heller [26] showed how scattering theory can be applied to understanding quantum corrals and quantum mirages. Our work follows their initiative by applying scattering theory to the recent experiments in open Quantum Point Contact (QPC) [76, 49]. Several theoretical works have been dedicated to interesting phenomena in QPC like branched flow [66, 84] and interferometry [36]. This chapter will take a unique, yet simple approach to the problem by first examining scattering patterns due to one or two impurities and a tip, and then increasing the number of impurities to see the evolution of conductance fringe patterns.

Besides identifying general features of conductance fringes, the chapter also de-

velops a useful formula to estimate the impact of thermal averaging on the fringes, which as a signature of coherence are supposed to survive in much reduced regions at non-zero temperature. The possible presence of resonances is investigated as well, because when a resonance develops with the tip away from impurities the sudden change in conductance can very likely be misread as the tip sitting on an impurity. A brief discussion of differences in fringe pattern caused by different impurities is presented near the end of the chapter.

4.2 Conductance Fringe Patterns

In this section, scattering theory is applied to an open QPC system probed by an SPM tip. The discussion is mostly limited to low scattering orders, which is usually the case for the clean samples used in many of the imaging experiments [76, 49]. Analytical expressions and numerical simulations of the conductance as a function of SPM tip position are first derived for the single impurity case. Extensions to multiple impurities are also provided, and numerical simulations of two and five impurities confirm qualitatively predictions on fringe patterns and show some signatures observed previously in experiments [76, 49].

4.2.1 Back-Scattered Flux through an Open QPC

A sketch of the experimental QPC system [76, 49] is shown in Fig. 4.1(a), which consists of an open QPC, an SPM tip, and a set of random impurities. The incoming electron wave, ϕ_s, propagates from the QPC towards the tip and impurities and gets partially reflected back through the QPC. When the SPM tip is moved, the flux

through the QPC also changes. For a fixed potential bias between the source and the drain, this change in flux manifests itself as a change in conductance, which is measured as a function of SPM tip position to give an image of the electron flow. In this section, analytical expressions are derived for the change in conductance as a function of tip position in the presence of a single impurity.

A few approximations can be made to simplify the analysis. When the impurity and the SPM tip are placed far away from the boundary walls of the QPC, wave propagation toward and amongst the scatterers can be well described by the conventional $2D$ Green's function in free space. Using an approximate proportionality between flux and wave density established in Eq. (A.2) of Appendix A, scattering theory can be used to calculate the flux as a function of SPM tip position for the system shown in Fig. 4.1(a).

Denote the total wave incident upon an impurity and the tip as Ψ_1 and Ψ_t respectively and the total reflected wave evaluated at the QPC located at $\vec{r}_s$ as Ψ_s. Ψ_1, Ψ_t and Ψ_s are given by

$$\begin{aligned} \Psi_1 &= G_{1,s}\phi_s + \epsilon_t G_{1,t}\Psi_t \ , \\ \Psi_t &= G_{t,s}\phi_s + \epsilon_1 G_{t,1}\Psi_1 \ , \\ \Psi_s &= \epsilon_1 G_{s,1}\Psi_1 + \epsilon_t G_{s,t}\Psi_t \ , \end{aligned} \tag{4.1}$$

where $G_{i,j}$ is the $2D$ Green's function between points i and j given by $G_{i,j} = -i\frac{m_e}{2\hbar^2}H_0^{(1)}\left(k\left|\vec{r}_i - \vec{r}_j\right|\right)$, where $H_0^{(1)}$ is the Hankel function of the first kind, $\hbar k = \sqrt{2m_e E}$, ϕ_s is the source wave emanating from the QPC at $\vec{r}_s$. The scattering amplitude for scatterer j is given by $\epsilon_j = \frac{-8\hbar^2}{m_e}\sin\left(\delta_j(E)\right)e^{i\delta_j(E)}$, where m_e is effective electron mass and $\delta_j(E)$ is the induced phase shift of the scattered wave. Solving for

Ψ_s in Eq. (4.1) gives

$$\Psi_s = \frac{\left(\epsilon_1 G_{1,s}^2 + \epsilon_t G_{t,s}^2 + 2\epsilon_1\epsilon_t G_{1,s}G_{s,t}G_{t,1}\right)\phi_s}{1-\epsilon_1\epsilon_t G_{1,t}^2}\ . \tag{4.2}$$

The first two terms in the numerator correspond to a single scattering trajectories off the fixed impurity and the SPM tip respectively, whereas the third term in the numerator represents double scattering trajectories between the impurity and the tip. The denominator can be expanded in a power series of $\epsilon_1\epsilon_t G_{1,t}^2$, where each successive term corresponds to an additional bouncing trajectory between the impurity and the SPM tip. In the weak scattering limit considered here ($|\epsilon_1|$, $|\epsilon_t| \ll 1$), only the lowest scattering terms are significant. Keeping the terms up to double scattering, the reflected wave and density back at the QPC are

$$\Psi_s \approx \phi_s\left(\epsilon_1 G_{1,s}^2 + \epsilon_t G_{t,s}^2 + 2\epsilon_1\epsilon_t G_{1,s}G_{s,t}G_{t,1}\right), \tag{4.3}$$

$$\begin{aligned}|\Psi_s|^2 \approx\ & |\phi_s|^2(|\epsilon_1|^2|G_{1,s}|^4 + |\epsilon_t|^2|G_{t,s}|^4 + 4|\epsilon_1|^2|\epsilon_t|^2|G_{1,s}|^2|G_{s,t}|^2|G_{t,1}|^2 \\ & +2\ Re\left[\epsilon_1\epsilon_t^* G_{1,s}^2 G_{t,s}^{2\,*}\right] + 4\ Re\left[\left(\epsilon_1^* G_{1,s}^{2\,*} + \epsilon_t^* G_{t,s}^{2\,*}\right)\epsilon_1\epsilon_t G_{1,s}G_{s,t}G_{t,1}\right]).\end{aligned} \tag{4.4}$$

When the SPM tip is moved around, the observed conductance through the QPC is modulated due to changes in the reflected electron flux. For $kr >> 1$, $G_{i,j} \approx -\mathrm{i}\frac{m_e}{2\hbar^2}\sqrt{\frac{2}{\pi k r_{i,j}}}e^{\mathrm{i}(kr_{i,j}-\frac{\pi}{4})}$, where $r_{i,j} = |\vec{r}_i - \vec{r}_j|$. The flux as a function of SPM tip position can then be written as

$$\begin{aligned}J\ & (\vec{r}_t, E) \\ \approx\ & \hbar k|\Psi_s(\vec{r}_t)|^2 \\ \approx\ & \frac{16 m_e|\phi_s|^2\cos\theta_0}{\pi^2\hbar^3 k}\Big(\frac{\sin^2(\delta_1(E))}{r_{1,s}^2} + \frac{\sin^2(\delta_t(E))}{r_{t,s}^2} + \frac{128\sin^2(\delta_1(E))\sin^2(\delta_t(E))}{\pi k r_{1,s}r_{s,t}r_{t,1}}\end{aligned}$$

$$+\frac{2\sin\left(\delta_1(E)\right)\sin\left(\delta_t(E)\right)}{r_{1,s}r_{t,s}}\ \cos\left[2k\left(r_{1,s}-r_{t,s}\right)+\delta_1(E)-\delta_t(E)\right]$$
$$-\frac{16\sqrt{2}}{\sqrt{\pi k r_{1,s}r_{s,t}r_{t,1}}}\left(\frac{\sin^2\left(\delta_1(E)\right)\sin\left(\delta_t(E)\right)\sin\left[k\left(r_{1,t}+r_{t,s}-r_{1,s}\right)+\delta_t(E)-\frac{\pi}{4}\right]}{r_{1,s}}\right.$$
$$\left.+\frac{\sin^2\left(\delta_t(E)\right)\sin\left(\delta_1(E)\right)\sin\left[k\left(r_{1,t}+r_{1,s}-r_{t,s}\right)+\delta_1(E)-\frac{\pi}{4}\right]}{r_{t,s}}\right), \tag{4.5}$$

where Eq. (A.2) in Appendix A has been used to relate $|\Psi_s|^2$ to $J(\vec{r}_t, E)$.

4.2.2 Conductance Fringe Patterns for a Single Impurity

The first term of $J(\vec{r}_t, E)$ in Eq. (4.5) does not depend upon tip position at all, and thus can be neglected in the following discussion. The magnitude of the other terms do depend upon the SPM tip position and scale roughly as $(r_{t,s})^{-2}$. On top of this decay with tip position, tip-dependent conductance oscillations are evident in Eq. (4.5) in the fourth, fifth, and sixth terms, which leads to three types of coherent fringes which can be observed in the flux measurements.

The fourth term in Eq. (4.5), which represents the interference of single-scattering trajectories off of the SPM tip and the fixed impurity, leads to circular fringes centered about the QPC, defined by the relation

$$\phi_c(\vec{r}_t) = 2kr_{t,s} - 2kr_{1,s} + \delta_1(E) - \delta_t(E) = constant, \tag{4.6}$$

which are maximal when $\phi_c = 2n\pi$, for integer n.

The fifth term in Eq. (4.5), which represents the interference of the double scattering trajectory between the tip and impurity with the single scattering trajectory from the fixed impurity, leads to elliptical fringes with the foci given by the QPC and

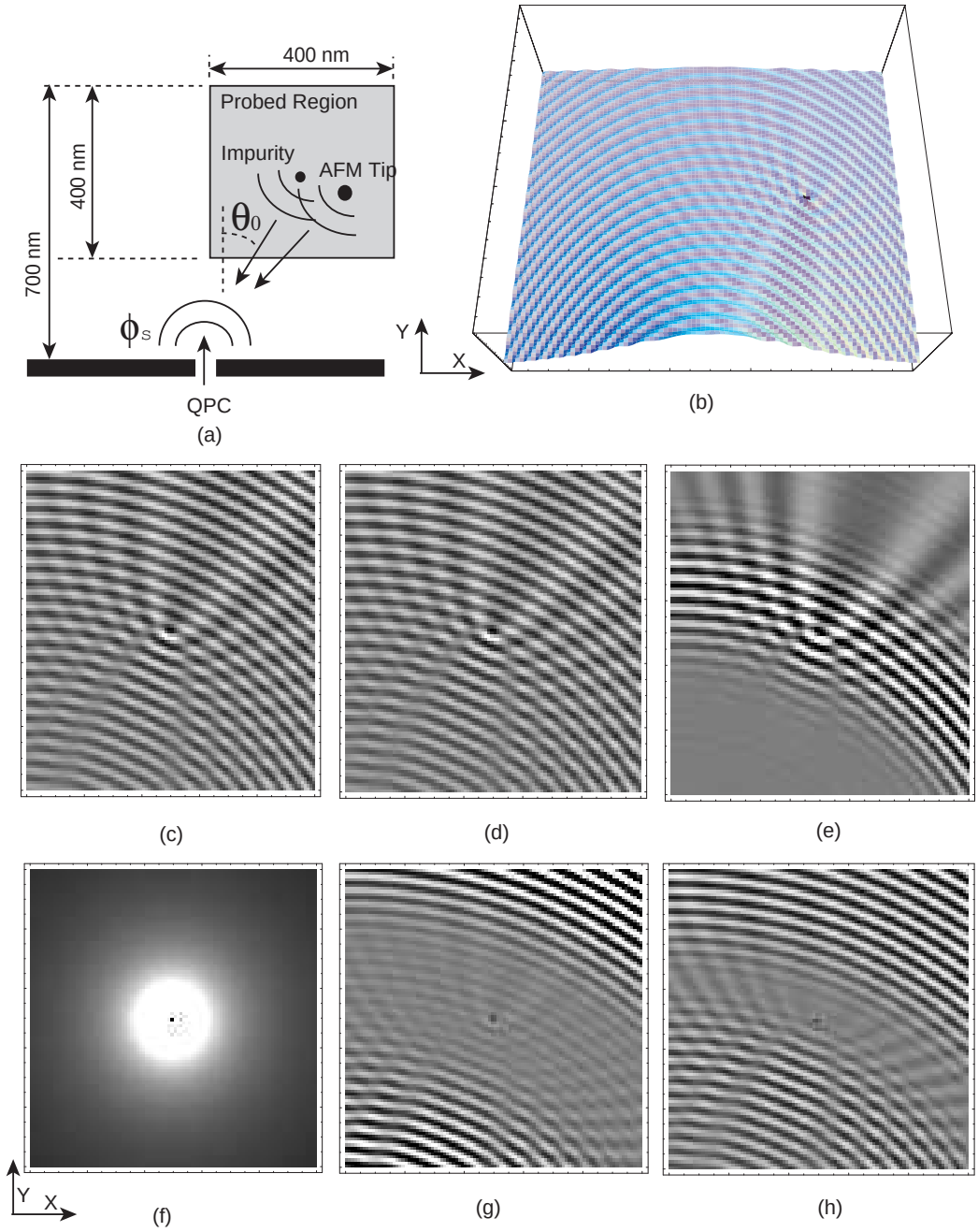

Figure 4.1: (a) Open QPC system diagram. The thick black lines next to QPC are its boundary walls; (b) A $3D$ view of the reflected flux in the probed region in (a) and an equal-sized region to its left; (c-e) Reflected flux by an impurity and an SPM tip in the probed region at $T = 0\ K$, $1.7\ K$, and $17\ K$. All are obtained by multiplying the total flux by $\frac{y}{r^3}$ to get only the oscillations; (f) Flux distribution without the SPM tip; (g) Thermal contribution at $1.7\ K$, obtained by subtracting (c) from (d); (h) Thermal contribution at $17\ K$, obtained by subtracting (c) from (e).

the fixed impurity. The elliptical fringes are defined by the relation

$$\phi_e(\vec{r}_t) = k(r_{1,t} + r_{t,s}) - kr_{1,s} + \delta_t(E) - \frac{\pi}{4} = constant, \tag{4.7}$$

which are maximal when $\phi_e = (4n+1)\pi/2$ for integer n.

Finally, the sixth term in Eq. (4.5), which represents the interference of the double scattering trajectory between the tip and impurity with the single scattering trajectory from the SPM tip, leads to hyperbolic fringes with the foci again given by the QPC and the fixed impurity. The hyperbolic fringes are defined by the relation

$$\phi_h = k(r_{1,t} - r_{t,s}) + kr_{1,s} + \delta_1(E) - \frac{\pi}{4} = constant, \tag{4.8}$$

which are maximal when $\phi_h = (4n+1)\pi/2$ for integer n.

All three fringe patterns are evident in Figs. 4.1(c-e), which plot the exact numerical flux calculations. The numerical calculation shows near perfect patterns, which validates the approximations made above in analytical derivation. For fixed energy E, the circular fringe contours, ϕ_c, have a period and a physical spacing of $\lambda/2$, where $k\lambda = 2\pi$. Similarly, the elliptical and the hyperbolic fringes both have a period of λ, although the fringe spacing is not constant, as is seen in Fig. 4.1 (c-e). Such a difference in spatial period doesn't seem significant in the single scatterer case but does play a critical role in the case of multiple impurities as discussed in the next section.

The above results are for a fixed energy, typically chosen to be at the Fermi energy, E_f, which corresponds to the zero temperature case. The conductance measurements, however, are always performed at nonzero temperature, which means that the results must be thermally averaged. If the current injected from the QPC is a result of a

small potential bias between the source and the drain, the above conductance must be averaged over the function

$$-f^{'}(E,T) \;=\; \frac{1}{k_BT}\frac{\exp\left(\frac{E-E_F}{k_BT}\right)}{1+\exp\left(\frac{2[E-E_F]}{k_BT}\right)} \tag{4.9}$$

Thermal averaging reduces the fringe heights and therefore is important when considering the observation of these effects. This will be discussed in more detail later in the chapter.

4.2.3 Conductance Fringe Patterns for Multiple Impurities

Up to double scattering, Eq. (4.3) can be generalized in the case of $N \geq 2$ impurities and a single SPM tip as follows:

$$\Psi_s \approx \left(\epsilon_t G_{t,s}^2 + \sum_{i=1}^{N}\left[\epsilon_i G_{i,s}^2 + +2\epsilon_i\epsilon_t G_{i,s}G_{s,t}G_{t,i} + 2\sum_{n<i}\epsilon_i\epsilon_n G_{i,s}G_{s,n}G_{n,i}\right]\right)\phi_s. \tag{4.10}$$

The four terms in Eq. (4.10) have a clear physical interpretation. The first and second terms represent the single scattering trajectory from the SPM tip and fixed impurities respectively. The third term represents all all double scattering trajectories involving an impurity and the SPM tip, whereas the last term represents all double scattering trajectories involving two different impurities. Despite the fact that the reflected wave now includes many different electron trajectories, the overall conductance fringe pattern still consists of circular, elliptical, and hyperbolic fringes as in the single scatterer case.

The circular fringes now arise from the interference between the single scattering trajectory from the tip, $(\epsilon_t G_{t,s}^2)$, and either the single scattering trajectories from the fixed impurities $(\epsilon_i G_{i,s}^2)$ or the double scattering trajectories involving any two fixed

impurities, $(\epsilon_i \epsilon_j G_{i,s} G_{i,j} G_{j,s})$. Although the amplitude for the interference with the double scattering trajectories are smaller by a factor of ϵ_i relative to the amplitude of the interference with the single scattering trajectories, the number of such trajectories scales as N^2 so their contribution to the circular fringes can not be ignored relative to the single scattering trajectories, which are only linear in the number of scatterers. Although the above fringes are all circular, the phase of each individual fringe is related to impurity location (see Eq. (4.6) for the interference between the single scattering trajectories) . When these circular fringe patterns are superimposed, the overall circular fringe pattern may get enhanced or damped, largely depending on impurity configuration.

The elliptical fringes are now produced by the interference between the tip and impurity double scattering trajectory $(\epsilon_j \epsilon_t G_{j,s} G_{s,t} G_{t,j})$ with either the impurity single scattering trajectory $(\epsilon_i G_{i,s}^2)$ or the two impurity double scattering trajectory $(\epsilon_m \epsilon_n G_{m,s} G_{s,n} G_{n,m})$. The foci of the elliptical fringes are located at the QPC and at the location of the impurity involved in the tip and impurity double scattering trajectory. As in the circular case, different elliptical fringes possess a different phase relation to each other [as seen in ϕ_e in Eq. (4.7)]. Many of the elliptical fringes have a focus on different impurities and are readily distinguishable. However, some of them do share the same two foci. For example, the two elliptical fringes from the interference between $(\epsilon_j \epsilon_t G_{j,s} G_{s,t} G_{t,j})$ and $(\epsilon_j G_{j,s}^2)$ or $(\epsilon_j \epsilon_n G_{j,s} G_{s,n} G_{n,j})$ both have one focus at the QPC and the other at impurity j. However, the superposition of these two elliptical fringes is strongly dependent on the location of impurity n since the phase relation of the two fringes are different. Thus the height of the fringes will again

depend upon the impurity configuration.

As before, one set of hyperbolic fringes originate from the interference between the SPM tip single scattering trajectory ($\epsilon_t G_{t,s}^2$) with the tip and impurity double scattering trajectory ($\epsilon_j \epsilon_t G_{j,s} G_{s,t} G_{t,j}$). Hyperbolic fringes formed in this way have the QPC as a common focus, with the other focus located at the location of the relevant impurity [thus a reduction of fringe heights is again expected due to the difference in phase relations between the various hyperbolic fringes, given by ϕ_h in Eq. (4.8)]. Another set of hyperbolic fringes can arise which involve the interference between two double scattering trajectories involving the tip with different impurities [($\epsilon_j \epsilon_t G_{j,s} G_{s,t} G_{t,j}$) and ($\epsilon_p \epsilon_t G_{p,s} G_{s,t} G_{t,p}$) with $j \neq p$]. For these hyperbolic fringes, the foci are located on the two relevant impurities. Although some of these hyperbolic fringes might share one common focus, none of them share both foci together. In practice, when there are only a few impurities the overall pattern already becomes highly complicated due to superposition of different fringes. Careful examination is required to extract some useful signatures out of the complex flux map.

4.2.4 Numerical Simulation of the Backscattered Flux for Multiple Impurities

Numerical simulations for the conductance as a function of SPM tip position in the presence of two and five impurities are shown in Fig. 4.2. Fig. 4.2(a), (d), (g), (j) show $|\Psi(\vec{r})|^2$ in the absence of an SPM tip, while the rest of the figures represent the flux measurements in the presence of an SPM tip. All flux patterns seem to consist of a circular fringe background along with many elliptical and hyperbolic fringes, which

is expected from the discussion in the previous section. Comparing Fig. 4.2(b), (e), and (h), the flux patterns clearly depend upon the impurity configuration. For two impurities separated by more than a few (more than five is good enough in numerical simulation) wavelengths apart [Fig. 4.2(b) and (c)], the interference between the two impurities is weak, which results in two sets of elliptical and hyperbolic fringes that are easily distinguishable. In the second and third cases in Fig. 4.2(d-i), the two impurities are close to each other, roughly $\lambda/2$ apart, but with different orientations with respect to the QPC. The two sets of elliptical and hyperbolic fringes strongly interfere with each other and the interference strongly depends on the phase difference between the initial wave incident on the two impurities. In the second case, the two impurities are $\lambda/2$ apart radially. Since the radial fringes have a period of $\lambda/2$ and the elliptical and the hyperbolic fringes have a period of λ, the $\lambda/2$ separation causes a constructive addition of radial fringes and a mostly destructive addition of elliptical and hyperbolic fringes. The relative intensity of radial/hyperbolic fringes to radial fringes is minimized. Fig. 4.2(e) and (f) show strong radial fringes and very weak elliptical/hyperbolic fringes, demonstrating the predicted change in relative intensity when compared to the single impurity case in Fig. 4.1(c-e) (each density plot has its own color scale so relative intensity among fringes is best suited to make comparison among plots). In the third case, however, the impurities are aligned horizontally. When projected to the radial direction their distance gives an equivalent separation of slightly less than $\lambda/4$, which is half the period of radial fringes and a quarter for elliptical/hyperbolic fringes. Therefore the overall pattern should show damped radial fringes and enhanced elliptical/hyperbolic fringes; and the relative intensity should

favor elliptical/hyperbolic fringes more compared to Fig. 4.1(c-e). And it does in Fig. 4.2(h) and (i).

The last case in Fig. 4.2 is for five far separated impurities. Despite large separation, many sets of fringes overlap strongly in the plot region. At low temperature, strong coherent interference among impurities leads to a highly complicated overall pattern in Fig. 4.2(k), where the artificially marked impurities are undistinguishable. This is not surprising given the many sets of fringes predicted for multiple impurities in the preceding subsection.

Even though both Figs. 4.1 and 4.2 show that the presence of an SPM tip largely changes flux pattern, there is a rough correspondence of regions of high flux between plots with and without a tip. This is most obvious in Fig. 4.2 between (g) and (h), and between (j) and (k). It is a good news because it justifies the basic principle of using an SPM tip to probe the approximate electron flux distribution. One should also notice that most elliptical and hyperbolic fringes are oriented radially outward from the QPC and create many angular nodes on circular fringes in Fig. 4.2(k). These features seem to have actually appeared in experiments. The many big or small branchings of electron flux in [76, 49] are aligned along the radial direction from the QPC, and could be signature of the hyperbolic fringes discussed here.

4.2.5 Patterns in Strong Scattering Limit

When the scattering strength increases, multiple scattering trajectories must be taken into account when calculating $J(\vec{r}_t)$. Consider first the single impurity case discussed in the previous section. There are only three relevant lengths in the system,

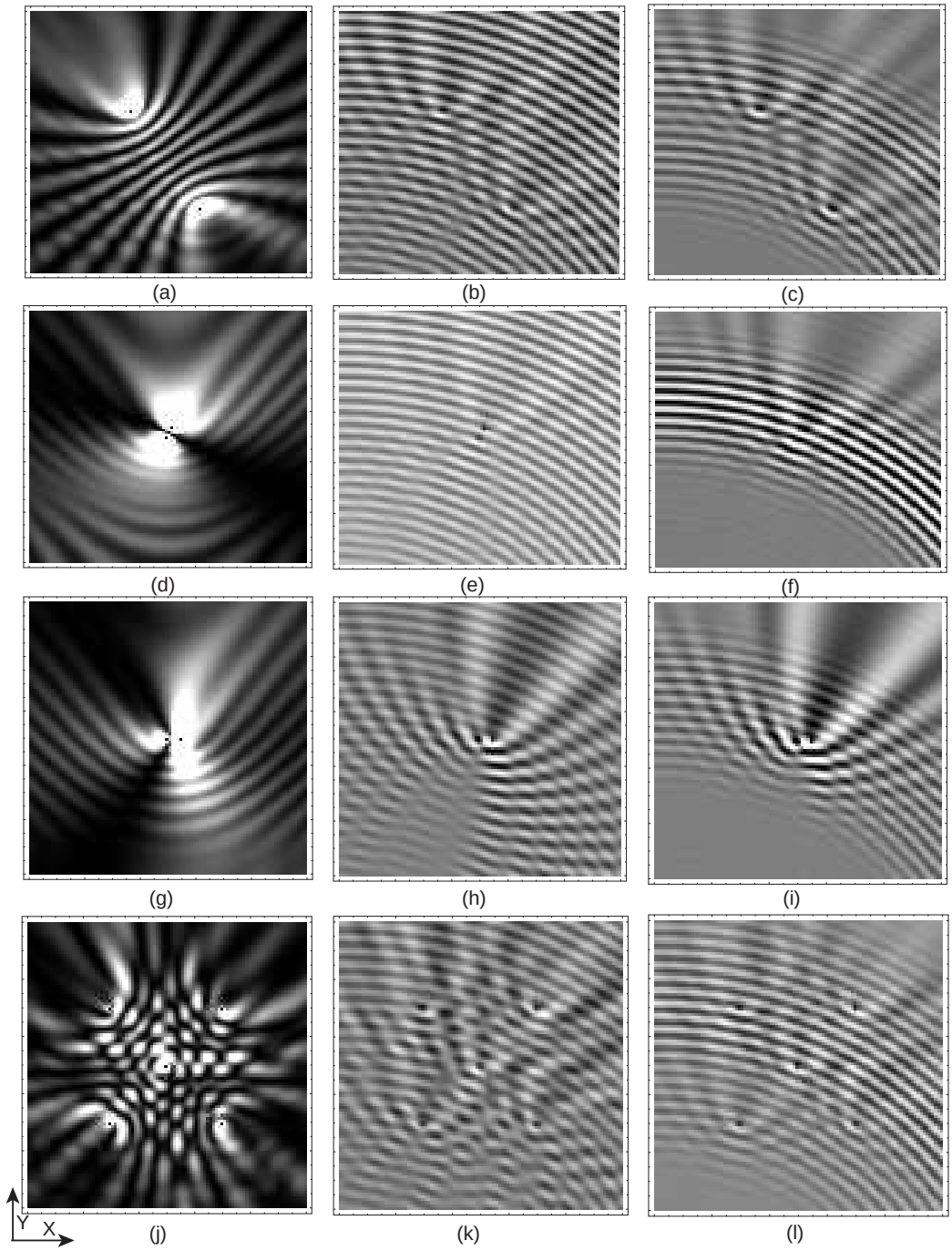

Figure 4.2: The left, middle, and right columns are flux distribution without an SPM tip, and reflected flux in the probed region (as in Fig. 4.1) at $T = 1.7\ K$ and $17\ K$. First three rows: two impurities far apart, close with a mostly destructive interference, and close with a constructive interference. Last row: five impurities regularly spaced. All impurities are marked by a black dot. More detail in the text.

$r_{1,t}$, $r_{t,s}$, and $r_{1,s}$, which can be readily seen by expanding Eq. (4.2). In the multiple scattering limit, the phase along an arbitrary scattering path starting and ending at the QPC (and neglecting the various δ scattering phase shifts) can be written as $\phi = k\left(lr_{1,t} + mr_{t,s} + nr_{1,s}\right)$, where either $(m,n) \in \{(2,0),(0,2)\}$ with $l = 0,2,4,...$ or $(m,n) = (1,1)$ with $l = 1,3,5,....$ The phase difference between two arbitrary scattering paths can be expressed as

$$\Delta\phi = k\left(lr_{1,t} + mr_{t,s} + nr_{1,s}\right) \tag{4.11}$$

where either $(m,n) \in \{(\pm 2,\mp 2),(0,0)\}$ and $l = 0,\pm 2,\pm 4,...$ or $(m,n) = (\pm 1,\mp 1)$ and $l = \pm 1,\pm 3,\pm 5,....$ Eq. (4.11) gives the general expression for fringe patterns as a function of tip position:

$$k\left(lr_{1,t} + mr_{t,s}\right) = constant \tag{4.12}$$

For $l = 0$ and $m = \pm 2$, circular fringes around the QPC defined by $\pm 2r_{t,s}$ =constant are formed. For $m = 0$ and $l \neq 0$, circular fringes around the impurity defined by $r_{1,t}$ =constant are formed. For $lm > 0$, elliptical fringes with foci located at the QPC and the impurity are formed, and for $lm < 0$, hyperbolic fringes with foci at the impurity and the QPC are formed.

The above analysis can be extended to the case of multiple impurities in the presence of an SFM tip. The phase accumulated along an arbitrary scattering trajectory starting and ending at the QPC can always be expressed as a linear sum of relevant lengths in the system, $\phi = k\left(\sum_{i=1}^{N} l_i r_{i,s} + \sum_{i=1}^{N}\sum_{i<j} m_{i,j} r_{i,j} + \sum_{i=1}^{N} n_i r_{i,t} + \eta r_{s,t}\right)$ where either $\eta = 2$, $l_i = 0$ for $i \in [1,N]$, $m_{i,j}, n_i \in \{0,1,2,3,....\}$, or $\eta = 0$, $\sum_{i=1}^{N} l_i = 2$ with $l_i \in \{0,1,2\}$ and $m_{i,j}, n_i \in \{0,1,2,3,...\}$, or finally $\eta = 1$, $\sum_{i=1}^{N} l_i = 1$ for

$l_i \in \{0, 1\}$, and $m_{i,j}, n_i \in \{0, 1, 2, 3, ...\}$. The fringe patterns generated by the movable SPM tip can be defined by the following phase relation:

$$\Delta\phi(\vec{r}_t) = k\left(\sum_{i=1}^{N} n_i r_{i,t} + \eta r_{t,s}\right) = constant \tag{4.13}$$

where $\eta \in \{0, \pm 1, \pm 2\}$ and $\sum_{i=1}^{N} n_i \in \{0, \pm 1, \pm 2, ...\}$. Let us analyze each possible scenario of (n_i, η). If $n_i = 0$ for all i and $\eta \neq 0$, circular fringes centered about the QPC and defined by $kr_{t,s}$ =constant are formed. If $\eta = 0$ but $n_i \neq 0$ for some i, fringes defined by the relation $\sum_{i=1}^{N} n_i r_{i,t}$ =constant are formed. For these fringes if only n_k is nonzero, circular fringes about impurity k are formed, whereas if only n_k and n_j are nonzero, elliptical ($n_j n_k > 0$) or hyperbolic fringes ($n_j n_k < 0$) can be formed with the foci centered on impurities j and k. For more than two n_i nonzero, even more complicated patterns can be formed. Finally, if $\eta \neq 0$ and at least one $n_i \neq 0$, there$\sum_{i=1}^{N} \frac{n_i}{\eta} r_{i,t} + r_{s,t}$ =constant. It defines fringes focused at the QPC and an impurity if there is exactly one $n_i \neq 0$ (an ellipse for $n_i = \eta$; an hyperbola for $n_i = -\eta$; a deformed ellipse for $n_i \eta > 0$; a deformed hyperbola for $n_i \eta < 0$). More complicated, often fragmented patterns appear for more than one $n_i \neq 0$. The superposition of the various fringe patterns renders the overall conductance map highly complicated. However, most SPM experiments are performed either in the few-body limit or in the weak scattering limit. Most of the above-mentioned patterns don't appear at equal strength, and given the right conditions, some of the fringe patterns may dominate over others.

A rigorous investigation of systems with many impurities and strong scattering can be very beneficial to mesoscopic physics. For example, in submicrometer disordered metallic systems at low temperatures, interesting conductance fluctuation effects arise

owing to the interference of coherent electron wave transmission through the system. Such phenomena have been studied and shown to yield novel predictions for the correlations in laser speckle patterns and have provided the possibility of studying the positions and motions of scatterers in a medium which multiply scatter the probing light [25].

4.3 Thermal Averaged Conductance Fringes

4.3.1 A General Formalism of Thermal Averaged Fermionic Signal

The goal in this part is to develop an asymptotic formula to conveniently express a Fermionic signal at finite temperature in terms of both itself at zero temperature and separated thermal terms. The Fermionic signal is denoted $\vec{J}(\vec{r}, T)$, which in this chapter refers to the reflected electron flux through an open QPC in Fig. 4.1(a). But the formula applies to a general Fermionic signal. The thermal broadening function at a finite temperature T is [17]

$$-f'(E,T) = \frac{1}{k_B T} \frac{e^{(E-E_F)/k_B T}}{(1+e^{(E-E_F)/k_B T})^2}. \tag{4.14}$$

In $2D$, the density of states is

$$a(E)\ dE = \frac{2A}{\hbar^2} 2\pi p\ dp = \frac{4\pi A m_e}{\hbar^2} dE, \tag{4.15}$$

where $dE = d(\frac{p^2}{2m_e})$, A is the area of the QPC, and a factor of 2 is added to account for the spin degree of freedom.

For a signal like electron flux $\vec{J}(\vec{r},E)$ at fixed energy E, the thermally averaged signal $\vec{J}(\vec{r},T)$ is given by

$$\begin{aligned}
&\vec{J}\ \ (\vec{r},T)\\
&= \int \vec{J}(\vec{r},E)\ (-f'(E,T))\ a(E)\ dE = \frac{4\pi A m_e}{\hbar^2 k_B T}\int_0^{+\infty} \frac{e^{(E-E_F)/k_BT}}{(1+e^{(E-E_F)/k_BT})^2}\ \vec{J}(\vec{r},E)\ dE\\
&= \frac{-4\pi A m_e}{\hbar^2}\int_{E=0}^{+\infty}\ \vec{J}(\vec{r},E)\ d\left(\frac{1}{1+e^{(E-E_F)/k_BT}}\right)\\
&= \frac{-4\pi A m_e}{\hbar^2}\left[\left(\frac{\vec{J}(\vec{r},E)}{1+e^{(E-E_F)/k_BT}}\right)_{E=0}^{+\infty} - \int_{E=0}^{+\infty}\ \frac{\partial\vec{J}(\vec{r},E)}{\partial E}\frac{{}_{\mathrm{d}}E}{1+e^{(E-E_F)/k_BT}}\right]\\
&= \frac{4\pi A m_e}{\hbar^2}\int_0^{+\infty}\frac{\partial\vec{J}(\vec{r},E)}{\partial E}\frac{{}_{\mathrm{d}}E}{1+e^{(E-E_F)/k_BT}}\\
&= \frac{4\pi A m_e}{\hbar^2}\int_0^{+\infty}\frac{\partial\vec{J}(\vec{r},x)}{\partial x}\frac{{}_{\mathrm{d}}x}{1+e^{x-x_F}},
\end{aligned} \tag{4.16}$$

where $x = E/k_BT$, $x_F = E_F/k_BT$.

From [59], the generalization of Sommerfeld's lemma says that for any well-behaved function $f(x)$ at $x_F \gg 1$

$$\int_0^{+\infty}\frac{f(x)}{1+e^{x-x_F}}\ dx \approx \int_0^{x_F} f(x)\ dx + \frac{\pi^2}{6}\left(\frac{df}{dx}\right)_{x=x_F} + \frac{7\pi^4}{360}\left(\frac{d^3f}{dx^3}\right)_{x=x_F} + \cdots . \tag{4.17}$$

In the low temperature fringing experiments considered here [76, 49], $x_F = 109 \gg 1$ for $T = 1.7K$ and $E_F = 16\ meV$. Eq. (4.17) can be used to expand Eq. (4.16) into an asymptotic series

$$\begin{aligned}
&\vec{J}\ \ (\vec{r},T)\\
&\approx \frac{4\pi A m_e}{\hbar^2}\left[\int_0^{x_F}\ \vec{J}'(\vec{r},x)\ dx + \frac{\pi^2}{6}\left(\frac{\partial^2\vec{J}}{\partial x^2}\right)_{x=x_F} + \frac{7\pi^4}{360}\left(\frac{\partial^4\vec{J}}{\partial x^4}\right)_{x=x_F} + ...\right]\\
&= \frac{4\pi A m_e}{\hbar^2}\left[\ \vec{J}(\vec{r},x_F) - \vec{J}(\vec{r},0)\ + \frac{\pi^2}{6}\left(\frac{\partial^2\vec{J}}{\partial x^2}\right)_{x=x_F} + \frac{7\pi^4}{360}\left(\frac{\partial^4\vec{J}}{\partial x^4}\right)_{x=x_F} + ...\right]\\
&= \frac{4\pi A m_e}{\hbar^2}\left[\ \vec{J}(\vec{r},x_F) + \frac{\pi^2}{6}\left(\frac{\partial^2\vec{J}}{\partial x^2}\right)_{x=x_F} + \frac{7\pi^4}{360}\left(\frac{\partial^4\vec{J}}{\partial x^4}\right)_{x=x_F} + ...\right]
\end{aligned}$$

$$= \frac{4\pi A m_e}{\hbar^2}\left[\vec{J}(\vec{r},E_F)+\frac{(\pi k_B T)^2}{6}\left(\frac{\partial^2\vec{J}}{\partial E^2}\right)_{E=E_F}+\frac{7\,(\pi k_B T)^4}{360}\left(\frac{\partial^4\vec{J}}{\partial E^4}\right)_{E=E_F}+...\right] \tag{4.18}$$

The first term in Eq. (4.18) corresponds to flux at the Fermi energy at $T = 0\ K$. The other terms in Eq. (4.18) represent contributions from thermal excitations, which arise at nonzero temperatures. At finite temperatures, these excitations will compromise the coherent fringes originating from $\vec{J}(\vec{r},E_F)$ and at higher temperatures will eventually dominate over and destroy the coherent fringes. In the following, for simplicity only the first thermal term is examined in the single impurity case shown in Fig. 4.1 in order to provide a quantitative understanding of the impact of finite temperature.

Asymptotic expansions are not new to the study of Fermionic systems. However, the unique beauty of having a stand-alone zero-temperature term in Eq. (4.18) is only obtained after combining thermal averaging with the asymptotic expansion given in Eq. (4.17). Previously, an approximate integral form has been proposed in order to describe thermal averaging in 2DEGs [67]. It provided good approximation to Eq. (4.14), but it possesses severe disadvantages in providing analytical integration for the general case of multiple scattering. Such disadvantages motivated the use of Eq. (4.18).

Compared with the original integral form in Eq. (4.16), there are three major advantages for employing the asymptotic series in Eq. (4.18). First of all, the zero-temperature and the thermal terms are separated in the series. Isolating the thermal terms helps understand their magnitude and dependence upon energy and temperature. Secondly, analytical integration can only be performed on a very limited set

of functions whereas differentiation can be performed on any well-behaved function strictly, thereby making an asymptotic series consisting of derivatives much easier to study analytically. For example, the phase shift of most scatterers have a nontrivial energy dependence, rendering analytical integration for the flux nearly impossible. As demonstrated later, when a scatterer size is large, like an SPM tip, the energy dependence of the scattering phase shift can contribute significantly to thermal terms. Finally, the thermal terms can be calculated term by term to see how they originate from the zero-temperature flux, and how thermal excitations can generate decoherent terms which tend to destroy the zero-temperature fringes. Again, though Eq. (4.18) is for electron flux through an open QPC, under the condition $x_F \gg 1$ thermal averaging of many other Fermionic signals can be expanded in similar fashion, with the aforementioned advantages.

However, strong caution must be taken in using Eq. (4.18). Because of the differentiation with respect to E, the thermal terms in Eq. (4.18) are proportional to various powers of certain lengths in the system. The higher the thermal order, the higher the power of the lengths, as shown in the next section. In the near field regime, the various length scales are small, higher power makes thermal terms appear as small perturbations to the Fermi term, thus the expansion in Eq. (4.18) converges. In the far field regime of large length scales, these thermal terms diverge, even though they have different phase factors and can mutually cancel each other to still satisfy the condition for the expansion in Eq. (4.17). Fortunately, the remaining fringes at non-zero temperatures survive only in near field regime, where thermal terms aren't significant enough to destroy the Fermi term. So Eq. (4.18) can be safely employed,

as shown next.

4.3.2 Circular, Elliptical, and Hyperbolic Thermal Widths

With the aid of Eq. (4.18), it is easy to understand the difference among Fig. 4.1(c), (d), and (e). The first term in Eq. (4.18), $J(\vec{r}_s, T = 0K) = J(\vec{r}_s, E_F)$, is plotted in Fig. 4.1(c) plots. The coherent fringes are clearly evident. In order to examine the effects of thermal averaging, only $\frac{(\pi k_B T)^2}{6}\left(\frac{\partial^2 J}{\partial E^2}\right)_{E=E_F}$, the second right-hand term in Eq. (4.18), is evaluated, which turns out to capture most of the essential features when compared to the exact numerical integration of Eq. (4.16). Of the many k-dependent terms in Eq. (4.5), the last three terms oscillate with changing $(r_{1,s} - r_{t,s})$, $(r_{1,t} + r_{t,s} - r_{1,s})$, and $(r_{1,t} + r_{1,s} - r_{t,s})$ as an SPM tip moves around. These terms produce the conductance fringe patterns discussed in the previous section. It turns out that each kind of fringe possesses a different thermal width at finite temperatures.

The first thermal term in Eq. (4.18), $\frac{(\pi k_B T)^2}{6}\left(\frac{\partial^2 J}{\partial E^2}\right)_{E=E_F}$, can be rewritten as $\frac{(\pi m_e k_B T)^2}{6\hbar^4 k_F^2}\left(\frac{\partial^2 J}{\partial k^2} - \frac{1}{k}\frac{\partial J}{\partial k}\right)_{k=k_F}$, which generates a number of terms. When either double or single differentiation is performed on the prefactor $\frac{1}{k}$ in Eq. (4.5), coherent terms exactly in phase with $J(\vec{r}_s, E_F)$ are produced; when double differentiation in k is performed consecutively on the same *sin* or *cos* functions in Eq. (4.5), decoherent terms exactly opposite in phase with $J(\vec{r}_s, E_F)$ are produced; all other thermal terms are at relatively random phase with $J(\vec{r}_s, E_F)$ and of less interest in this study (these random terms superimpose, not cancel, the coherent and decoherent terms, much like the way elliptical/hyperbolic fringes superimpose circular fringes). When the tip moves away from the impurity or QPC, the decoherent terms can grow very large and

exactly offset the coherent fringes. Such cancellation leads to our first-order estimate of each fringe's thermal width, which is a measure of remaining coherence range.

For circular fringes, at large $|r_{1,s} - r_{t,s}|$, $\frac{\partial^2 J}{\partial E^2}$ will be dominated by terms of $\frac{\partial^2 \cos[2k(r_{1,s}-r_{t,s})+\delta_1(E)-\delta_t(E)]}{\partial k^2}$, which include a decoherent term with a prefactor $\left(2\left(r_{1,s} - r_{t,s}\right) + \frac{\partial \delta_1(E)}{\partial k} - \frac{\partial \delta_t(E)}{\partial k}\right)^2$. When this dominant term grows comparable to the coherent circular terms in $J(\vec{r}_s, E_F)$ and $\frac{(\pi k_B T)^2}{6}\left(\frac{\partial^2 J}{\partial E^2}\right)_{E=E_F}$, the circular fringes are destroyed, to the first order approximation. The cancellation of coherent and decoherent terms of $\cos\left[2k\left(r_{1,s} - r_{t,s}\right) + \delta_1(E) - \delta_t(E)\right]$ can be expressed in their prefactors as

$$\frac{(\pi m_e k_B T)^2}{6\hbar^4 k_F^2}\left[\frac{3}{k_F^2} - (\delta_1'(E)^2 + \delta_t'(E)^2)_{k=k_F} - \left(2\left(r_{1,s} - r_{t,s}\right) + (\delta_1'(E) - \delta_t'(E))_{k=k_F}\right)^2\right] +1 \approx 0,$$

$$r_{t,s} \approx r_{1,s} + \frac{1}{2}\left((\delta_1'(E) - \delta_t'(E))_{k=k_F} \pm \sqrt{\frac{6\hbar^4 k_F^2}{(\pi m_e k_B T)^2} + \frac{3}{k_F^2} - (\delta_1'(E)^2 + \delta_t'(E)^2)_{k=k_F}}\right). \tag{4.19}$$

where $\delta'(E) = \partial\delta(E)/\partial k$. Eq. (4.19) shows that at finite temperature, coherent fringes survive thermal averaging only when the tip is within a small radial range centered about a distance equal to the distance between QPC and the impurity. The range is centered at $r_{1,s} + \frac{1}{2}\left(\delta_1'(E) - \delta_t'(E)\right)_{k=k_F}$ with a half width of

$$\begin{aligned} l_c &= \frac{1}{2}\sqrt{\frac{6\hbar^4 k_F^2}{(\pi m_e k_B T)^2} + \frac{3}{k_F^2} - (\delta_1'(E)^2 + \delta_t'(E)^2)_{k=k_F}} \\ &\geq \frac{1}{2}\sqrt{\frac{6\sqrt{2}\hbar^2}{\pi m_e k_B T} - (\delta_1'(E)^2 + \delta_t'(E)^2)_{k=k_F}}\ , \end{aligned} \tag{4.20}$$

which decreases with temperature and whose less k_F-dependent lower limit is derived using $A^2 + B^2 \geq 2AB$.

In order to get a feel for what the thermal width is for physical samples, choose an effective electron mass, m_e, to be 0.067 times the electron mass, which corresponds to the observed m_e found in the $2DEG$ formed in GaAs/AlGaAs heterostructures[76, 49]. Let $\delta_1(E)$ and $\delta_t(E)$ take forms in Eq. (B.3) in Appendix B with $a_1 = 1\ nm$ and $a_t = 30\ nm$. At $T = 1.7K$ and $k_F = \frac{2\pi}{37.5nm}$, the half width is 507.08 nm and the lower limit is 70.84 nm. At $T = 17K$, the half width becomes 48.7 nm and the lower limit becomes 17.3 nm. The half width prediction of Eq. (4.20) matches Fig. 4.1(d), and (e) pretty well and can be further improved by considering higher order thermal terms in Eq. (4.18). The independence of the lower limit upon k_F in Eq. (4.20) comes first as a surprise but is easily understood since thermal averaging is due to excitation in a range of width k_BT around E_F, regardless of what the Fermi energy is. The tip/impurity-induced dependence on k_F in the width formula has much more to do with their scattering properties than thermal excitation. The role of scatterers' properties will be discussed in the final section. l_c will be referred to as the circular thermal half width and indicates the spatial range where circular fringes will exist. Also interesting is that Eq. (4.19) shows that the survived coherent fringes are centered at $\frac{1}{2}\left(\delta_1'(E) - \delta_t'(E)\right)_{k=k_F}$ away from the impurity location, due to again the different scattering properties of the impurity and the tip. This displacement is roughly 15.1 nm using the above parameters, close to the period of circular fringes, $\lambda_F/2 \approx 18.7\ nm$. Note that Fig. 4.1(e) and (h) do show a displacement close to one full fringe. This spatial displacement depends on the energy derivative of scattering

phase shifts, and is the time delay effect first discovered by Wigner [85] and later extensively researched by others [69, 1, 31, 13].

Thermal widths for elliptical and hyperbolic fringes can also be derived in similar ways. Consider the last two terms in Eq. (4.5) that depend on $(r_{1,t} + r_{t,s} - r_{1,s})$ and $(r_{1,t} + r_{1,s} - r_{t,s})$. At finite temperatures, cancellation of $(r_{1,t} + r_{t,s} - r_{1,s})$-dependent coherent and decoherent terms in Eq. (4.18) leads to

$$\frac{(\pi m_e k_B T)^2}{6\hbar^4 k_F^2}\left[\frac{5.25}{k_F^2} - (2\delta_1'(E)^2 + \delta_t'(E)^2)_{k=k_F} - \Big((r_{1,t} + r_{t,s} - r_{1,s}) + (\delta_t'(E))_{k=k_F}\Big)^2\right] +1 \approx 0,$$

$$r_{t,s} + r_{1,t} \approx r_{1,s} - (\delta_t'(E))_{k=k_F} \pm\sqrt{\frac{6\hbar^4 k_F^2}{(\pi m_e k_B T)^2} + \frac{5.25}{k_F^2} - (2\delta_1'(E)^2 + \delta_t'(E)^2)_{k=k_F}}\ . \tag{4.21}$$

which defines an elliptical thermal half width of

$$\begin{aligned} l_e &= \sqrt{\frac{6\hbar^4 k_F^2}{(\pi m_e k_B T)^2} + \frac{5.25}{k_F^2} - (2\delta_1'(E)^2 + \delta_t'(E)^2)_{k=k_F}} \\ &\geq \sqrt{\frac{3\sqrt{14}\hbar^2}{\pi m_e k_B T} - (2\delta_1'(E)^2 + \delta_t'(E)^2)_{k=k_F}}\ . \end{aligned} \tag{4.22}$$

Cancellation of $(r_{1,t} + r_{1,s} - r_{t,s})$-dependent coherent and decoherent terms leads to

$$\frac{(\pi m_e k_B T)^2}{6\hbar^4 k_F^2}\left[\frac{5.25}{k_F^2} - (\delta_1'(E)^2 + 2\delta_t'(E)^2)_{k=k_F} - \Big((r_{1,t} + r_{1,s} - r_{t,s}) + (\delta_1'(E))_{k=k_F}\Big)^2\right] +1 \approx 0,$$

$$r_{t,s} - r_{1,t} \approx r_{1,s} + (\delta_1'(E))_{k=k_F} \pm\sqrt{\frac{6\hbar^4 k_F^2}{(\pi m_e k_B T)^2} + \frac{5.25}{k_F^2} - (\delta_1'(E)^2 + 2\delta_t'(E)^2)_{k=k_F}}\ . \tag{4.23}$$

which defines an hyperbolic thermal half width of

$$\begin{aligned} l_h &= \sqrt{\frac{6\hbar^4 k_F^2}{(\pi m_e k_B T)^2} + \frac{5.25}{k_F^2} - (\delta_1'(E)^2 + 2\delta_t'(E)^2)_{k=k_F}} \\ &\geq \sqrt{\frac{3\sqrt{14}\hbar^2}{\pi m_e k_B T} - (\delta_1'(E)^2 + 2\delta_t'(E)^2)_{k=k_F}} \ . \end{aligned} \tag{4.24}$$

l_e and l_h are very similar and slightly more than twice l_c as expected. At $T = 1.7K$ and $k_F = \frac{2\pi}{37.5nm}$, l_e (l_h) is 1014.2 nm (1013.8 nm) and its lower limit is 163.9 nm (161.1 nm) . At $T = 17K$, l_e (l_h) becomes 97.9 nm (93.2 nm) and the lower limit becomes 43.3 nm (31.2 nm). However, l_e and l_h are not as directly reflected in Fig. 4.1(e) as l_c because they describe the sum or difference of two distances $r_{t,s}$ and $r_{1,t}$. At the special direction defined by connecting the QPC and impurity, $r_{t,s}$ and $r_{1,t}$ do align with each other. Along this direction, an overall increase of l_e/l_h in sum/difference is split between $r_{t,s}$ and $r_{1,t}$, each changing by $l_e/2 \approx l_h/2 \approx l_c$. Elliptical and hyperbolic fringes along this direction have the same fringe spacing as the circular fringes, i.e. $\lambda/2$. Therefore the number of survived elliptical/hyperbolic fringes should equal that of circular ones, though all three sets of fringes surround the impurity in different manners and are displaced from each other. Fig. 4.1(e) does confirm this point, despite the fact that the elliptical fringes are difficult to identify in the plot. Eqs. (4.21) and (4.23) also show a displacement $-(\delta_t'(E))_{k=k_F}$ / $(\delta_1'(E))_{k=k_F}$ in the sum/difference due to the scattering properties of the tip and the impurity, which will be further studied later.

In Fig. 4.1, the three various types of coherent fringes are plotted for various temperatures, along with thermal contributions. The figures are calculated from the strict integral form in Eq. (4.16) with the exact $2D$ Green's functions and exhibit

many of the features predicted using just the first thermal term in Eq. (4.18) alone. One should notice that elliptical and hyperbolic fringes always appear in pair and distribute on the opposite sides of the impurity, which helps to locate the impurity, whereas circular fringes can be used only to estimate the radial distance of an impurity. Comparison of the plots in Fig. 4.1 and Fig. 4.2 shows one benefit of thermal averaging. When temperature rises, all fringes shrink closer towards the scattering impurities, allowing better identification of impurities. Fig. 4.2(k) and (l) best illustrate this unexpected advantage. Survival of coherent fringes far away from the QPC at finite temperatures is essentially due to interferometry between different paths, which has been demonstrated in the recent electron interferometer experiment in 2DEG [48]. Additionally, note that Fig. 4.1(f) is drastically different from Fig. 4.1(c-e), even though they both depict the same s-wave scattering system. The elliptical and hyperbolic fringes look as if they were due to higher order partial wave scatterings. Caution must be taken in interpreting measured data.

4.4 Resonances in Open QPC Imaging Experiments

In the preceding sections, it has become clear that conductance map obtained through SPM probing can differ largely from the original wave field distribution due to additional scattering paths associated with the presence of the SPM tip. There is also the possibility that misidentification of impurity location can occur, due to the possibility of resonances occurring in the scattering. Let us illustrate this point in the single impurity case. Despite the complex flux map, it is assumed that as the tip gets closer to the impurity, the reflected flux increases, reaching a local maximum when

the tip is right atop of the impurity[74]. However, it could happen that resonance occurs when the SPM tip is placed at some special location that's not right on top of the impurity. Resonance shows up in sudden increase in reflected flux and can be misinterpreted as the tip hitting an impurity. Thus a thorough study of conditions when resonances can occur will prove to be very important in correct reading of measurement data. In the following, possibility of resonances is analyzed for the single impurity case in the open QPC imaging experiments in the s-wave scattering limit.

Resonances are quasi-bound states with positive energy (potential at infinity is defined as zero), and correspond to poles in the Green's function of system [62]. When resonance energy drops to below zero, a bound state is developed. In the open QPC experiments considered here, all electrons passing through the QPC possess positive energy. So bound state is unlikely to form and resonance is a much more likely concern. For resonances to occur in the single impurity case, poles of Eq. (4.2) must exist with a positive real part, which defines resonance energy, and a small imaginary part, which defines resonance width. Considering only the real part of the denominator, on resonance it has

$$\begin{aligned} Re\left(1-\epsilon_1\epsilon_t G_{1,t}^2\right) &= 0\ , \\ Re\left(\epsilon_1\epsilon_t G_{1,t}^2\right) &= 1\ . \end{aligned} \tag{4.25}$$

The left hand side of Eq. (4.25) depends on both k, the wave number, and $r_{1,t}$, the distance between the SPM tip and the impurity. The equation can be solved numerically to find for a specific k the possible tip location that can give rise to a resonance. For our example system, in the range $k \in [0.9k_F, 1.1k_F]$ or $E \in [0.81E_F, 1.21E_F]$ no

positive values of $r_{1,t}$ can be found. Since $k_BT\,/\,E_F = 1/109$ here and electrons can almost never be excited to more than $0.2E_F$ away from the Fermi surface, possibility of resonances can be ruled out at the s-wave scattering limit. Analytical proof of absence of resonances can be provided at the far and near field limits for a general one-impurity, s-scattering system as follows.

In the far field limit of $k_F r_{1,t} \gg 1$, $G_{1,t} = -i\frac{m_e}{2\hbar^2}H_0^{(1)}\left(kr_{1,t}\right) \approx -i\frac{m_e}{2\hbar^2}\sqrt{\frac{2}{\pi kr_{1,t}}}e^{i(kr_{1,t}-\frac{\pi}{4})}$. Eq. (4.25) can be simplified into

$$\begin{aligned} \pi kr_{1,t} &= -32\sin\delta_1\sin\delta_t\cos\left(\delta_1+\delta_t+2kr_{1,t}-\frac{\pi}{2}\right) \le 32\ , \\ r_{1,t} &\le 32/\pi k = 16\lambda/\pi^2 \approx 1.62\,\lambda \approx 1.62\,\lambda_F\ . \end{aligned} \tag{4.26}$$

Eq. (4.26) however contradicts the far field assumption. So no resonance can occur in far field. This conclusion is intuitively easy because two s-wave scatterers far apart can't keep wave for long before it leaks into empty space.

In the near field limit of $k_F r_{1,t} \ll 1$, $G_{1,t} = -i\frac{m_e}{2\hbar^2}H_0^{(1)}\left(kr_{1,t}\right) \approx \frac{m_e}{\pi\hbar^2}\lg(kr_{1,t})$. Eq. (4.25) reduces to

$$\frac{64}{\pi^2}\sin\delta_1\sin\delta_t\cos\left(\delta_1+\delta_t\right)\left(\lg(kr_{1,t})\right)^2 = 1\ . \tag{4.27}$$

Under the condition $\sin\delta_1\sin\delta_t\cos\left(\delta_1+\delta_t\right) > 0$, Eq. (4.27) can be solved to give

$$r_{1,t} = \frac{1}{k}\exp\left(-\frac{\pi}{8\sqrt{\sin\delta_1\sin\delta_t\cos\left(\delta_1+\delta_t\right)}}\right)\ . \tag{4.28}$$

However, resonance also requires the imaginary part of the denominator be small, i.e. $\left|Im\left(1-\epsilon_1\epsilon_t G_{1,t}^2\right)\right| \ll 1$, which defines resonance width. In near field the requirement translates to

$$\frac{64}{\pi^2}\left|\sin\delta_1\sin\delta_t\sin\left(\delta_1+\delta_t\right)\right|\left(\lg(kr_{1,t})\right)^2 \ll 1\ . \tag{4.29}$$

Comparison of Eqs. (4.27) and (4.29) leads to

$$\sin(\delta_1 + \delta_t) \approx 0;\ |\cos(\delta_1 + \delta_t)| \approx 1\ . \tag{4.30}$$

Eq. (4.30) can be analyzed to show two sets of solutions

$$\begin{aligned}
\cos(\delta_1 + \delta_t) &\approx 1\ and\ \sin\delta_t \approx -\sin\delta_1\ for\ (\delta_1 + \delta_t) \approx 2n\pi\ ; \\
\cos(\delta_1 + \delta_t) &\approx -1\ and\ \sin\delta_t \approx \sin\delta_1\ for\ (\delta_1 + \delta_t) \approx (2n+1)\pi\ ; \\
&(n = 0, \pm1, \pm2, ...)\ .
\end{aligned} \tag{4.31}$$

However, Eq. (4.31) contradicts the condition $\sin\delta_1 \sin\delta_t \cos(\delta_1 + \delta_t) > 0$ in Eq. (4.28). Thus no s-wave resonance exists in near field either.

The above proof of non-resonance in far and near fields applies to a general s-scattering system consisting of two scatterers. In the intermediate range of $r_{1,t}$ comparable to λ_F, s-scattering resonance could exist given appropriate E_F, $\delta_1(E)$, and $\delta_t(E)$. Of course, considering p-wave and even higher order partial wave scatterings increases the possibility and range of resonances in a system, so does the number of scatterers, as shown in Fig. 4.2(h) and (i). Detailed studies of proximity resonance can be found in [35, 50]. Given that scattering property of impurities is important to occurrence of resonances, what other differences can be made if one impurity is replaced by another? The last section is dedicated to some preliminary discussion on this topic.

4.5 Signatures of Impurity Properties in Conductance Fringe Patterns

The goal of SPM imaging is not only to shed light upon the configuration of impurities in mesoscopic/nanoscopic structures but also to help elucidate the scattering properties of the impurities. Much of the discussion in the preceding sections has focused on scattering path, which strongly relates to impurity placement. However, if the configuration is kept unchanged but the impurities are replaced by different ones, can there be signatures in the conductance map? Let us illustrate the impact again in the single impurity case. The difference will certainly be significant if an impurity changes its dimensions drastically. Scattering in the system will change accordingly from s-wave to higher order partial wave as the scattering length increases. What if the dimensions are fixed but the impurity changes its potential and thus scattering property, for example, if a potential barrier turns into a potential well of the same size? Can we tell the difference from a conductance map?

Fringe patterns depend mostly on scattering path, but, as noted earlier, they can be displaced by different amounts for different impurities. Eqs. (4.19) and (4.23) show that the displacement of both circular fringes and hyperbolic fringes depends on $(\delta_1'(E))_{k=k_F}$, which is small for either potential barrier or potential well in s-wave scattering limit, as shown in Appendix B. This displacement appears in circular and hyperbolic fringes because they involve exactly one scattering off the impurity. Can it accumulate in multiple scattering? It is possible but likely won't yield anything significant. As shown in Section 4.2.4, interference among multiple scattering paths

leads to many kinds of regular, deformed, or even discontinuous patterns. The overall flux field will become so complex that it's difficult to tell different patterns apart, save their displacements. Similarly, when there are more impurities, such displacement of patterns becomes impossible to identify even at low scattering limit, as shown in Fig. 4.2(k).

Thermal widths have dependence on the scattering property of impurity as well. Eqs. (4.20), (4.22), and (4.24) all contain $(\delta_1'(E))_{k=k_F}$, a small factor in s-wave scattering limit. Numerical results confirm that all three thermal widths change little without this term.

Resonance has been proven to be impossible for a single impurity in s-scattering except in intermediate range, regardless of the nature of impurity potential. Addition of more impurities certainly increases the chance for resonance. But then resonance is induced mostly by the configuration of impurities, not so much by the property of an individual impurity.

In conclusion, in s-wave scattering limit changes in an impurity's potential can hardly show up in any non-trivial way. This is consistent with the notion of s-wave scattering because relatively long wavelength can't tell the internal structure of a small impurity. When the relative size of impurity gets larger, high order partial wave scatterings appear and are capable of conveying information of scattering property of an impurity. Then changes in impurity potential are easy to detect. Of course, impurities of significantly different sizes can have distinct signatures as well, which in turn relates to high order partial wave scattering.

4.6 Conclusion

In summary, we have carefully examined conductance fringe patterns in 2DEG from several important perspectives and gained a great deal of valuable insight in the mechanism of SPM imaging. Identifying general patterns in conductance fringes reveals existence of interferometry in probing and prevents us from mistaking these patterns as signatures of high order partial wave scatterings. The analysis is compared to numerical simulations and extended to cases of multiple impurities and multiple scatterings. Then a useful general formula is developed to estimate thermal averaging of fringes in terms of separate thermal terms. A thermal width is given for each fringe pattern to indicate the range of survived coherent fringes at finite temperature. Since resonance could be misread as an SPM tip hitting an impurity, the possible occurrence of resonances is studied as well. In s-wave scattering limit, the possibility is shown to be low and can exist for only intermediate separation. Lastly, a brief examination of the impact on fringe pattern by properties of impurity shows minimal changes in s-wave scattering limit. Significant change in impurity size or potential change in large impurities can both lead to detectable differences, thus allowing for distinguishing impurities by SPM probing.

Chapter 5

Imaging A Single-Electron Quantum Dot

In this chapter, we employ finite difference method (FDM) to solve for ground state energy of a single-electron quantum dot, which gives rise to a circular conductance peak pattern under SPM probing. SPM tip introduces strong perturbation and is shown to be critical to the formation of the observed conductance ring and its variation with tip voltage. The ultimate goal is to obtain information about a single electron's wavefunction. For this purpose, singular value decomposition (SVD) is utilized to invert matrix to extract electron wavefunction out of the experimental conductance data. The stability and sensitivity to boundary conditions of FDM and SVD are also studied.

5.1 Introduction to Experimental Setup

Few-electron quantum dots have many exciting properties, including discrete energy levels, atom-like shell filling, and many-body effects, made visible by the Coulomb blockade. Single-electron quantum dots in particular are promising candidates for spin qubits in quantum information processing. The electron spin in each dot acts as a qubit, and tunneling entangles spins on adjacent dots. Scanning probe microscopy provides a powerful tool to probe the behavior of a single electron in quantum dots and dot circuits. Westervelt group at Harvard has used a liquid He cooled scanning probe microscope (SPM) with a charged tip to image a single-electron quantum dot in the Coulomb blockade regime. Fig. 5.1 illustrates the imaging technique.

The quantum dot, shown in Fig. 5.1(b), was formed in a GaAs/$Al_{0.3}Ga_{0.7}As$ heterostructure containing a 2DEG by Cr surface gates. The 2DEG is 52 nm below the surface, with measured density 3.8×10^{11} cm^{-2} and mobility $470,000$ cm^{2}V^{-1}s^{-1} at $4.2K$. The heterostructure was grown by molecular beam epitaxy with the following layers: 5nm GaAs cap layer, 25 nm $Al_{0.3}Ga_{0.7}As$, Si delta-doping layer, 22 nm $Al_{0.3}Ga_{0.7}As$, 20 nm GaAs, 100 nm $Al_{0.3}Ga_{0.7}As$, a 200 period GaAs/$Al_{0.3}Ga_{0.7}As$ superlattice, 300 nm GaAs buffer and a semi-insulating GaAs substrate. The 2DEG is formed in a 20 nm wide GaAs square well between two $Al_{0.3}Ga_{0.7}As$ barriers. The sample was mounted in a liquid-He cooled SPM [46, 47, 72] and cooled to $T = 1.7K$.

A conducting SPM tip scanned above the surface can change the induced charge in a quantum dot, as shown in Fig. 5.1(a), and change the number of electrons. An image is obtained by recording the dot conductance G as the tip is scanned across the sample. The tip voltage V_{tip} perturbs the bathtub-like background potential that

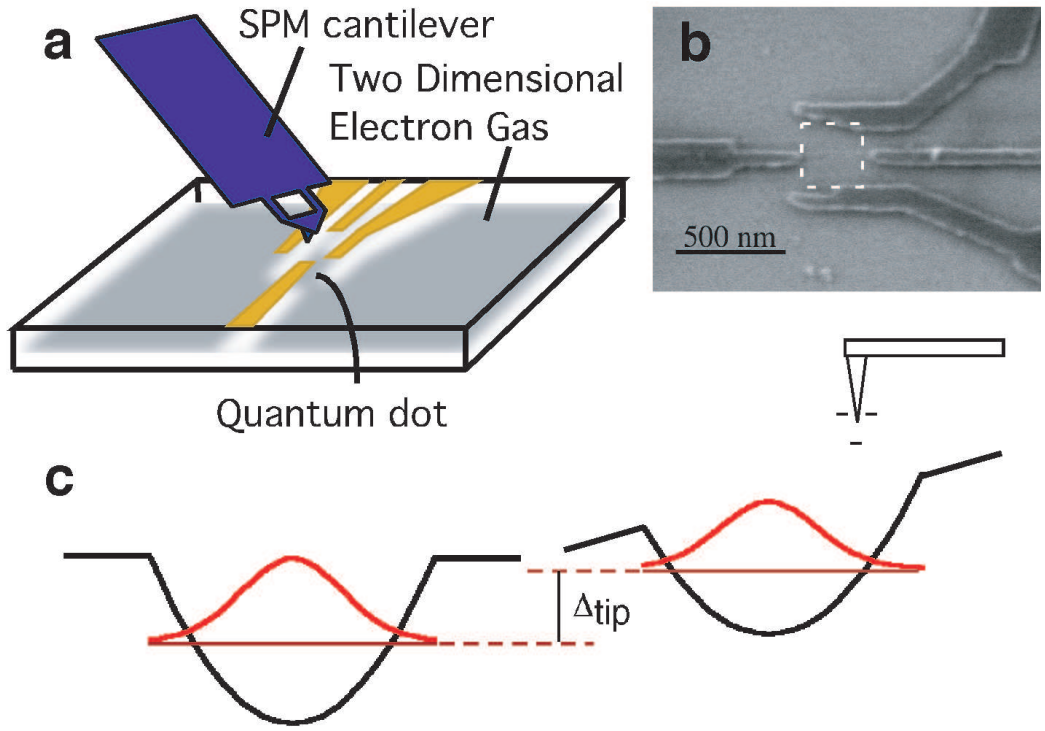

Figure 5.1: (a) Schematic diagram of the experimental set-up used to image electrons in a single-electron quantum dot. A charged scanning probe microscope (SPM) tip is scanned at a fixed height above the surface of the GaAs/AlGaAs heterostructure containing the dot. Images are obtained by recording the Coulomb blockade conductance G vs. tip position. (b) A scanning electron micrograph of the quantum dot. The dashed line indicates the area covered by the conductance images. (c) Schematic diagram that shows how the potential holding an electron in the dot is affected by the charged tip. The tip induced shift Δ_{tip} in the energy of the electron state changes the Coulomb blockade conductance of the dot.

holds electrons in the dot as in Fig. 5.1(c). When the tip is far away, V_{tip} moves the bathtub up and down without changing its shape, in a manner similar to the side-gate voltage. When the tip is close, V_{tip} can change the shape of the bathtub and the energy of the electron state above the bottom. The overall shift Δ_{tip} in the ground state energy, indicated in Fig. 5.1(c), moves the gate-voltage position of the Coulomb-blockade conductance peak.

Without the tip present, the quantum dot could be tuned to contain 0 or 1 electrons in the Coulomb blockade regime, as shown in Fig. 5.2. It has been determined that the one-electron charging energy is $4.2meV$ and the ground-state to first-excited-state energy spacing is $3.1meV$. Images of the single-electron quantum dot were obtained at $T = 1.7K$ by recording the Coulomb blockade conductance for fixed tip and gate voltages, V_{tip} and V_G, with source-to-drain voltage $V_{SD} = 0$ V, while the tip was spatially scanned over the quantum dot, 100 nm above the surface. The charged tip shifts the lowest energy level in the dot and creates a ring in the image.

A series of images are shown in Fig. 5.6(a-d) for tip voltages 40 mV, 50 mV, 60 mV, and 80 mV, respectively. In each image, a ring-shaped feature is observed, centered on the middle of the dot. The ring represents a contour of constant tip to dot coupling at which the Coulomb blockade conductance is on a peak. This peak corresponds to resonant tunneling through the lowest energy level of the dot. The dot contains one electron when the tip is outside the ring and zero when the tip is inside the ring. As V_{tip} is increased in a series of steps from Fig. 5.6(a) to (d), the radius of the ring shrinks to a small value. The line shape of the ring provides a window through which one can extract information about the dot. The probing window can

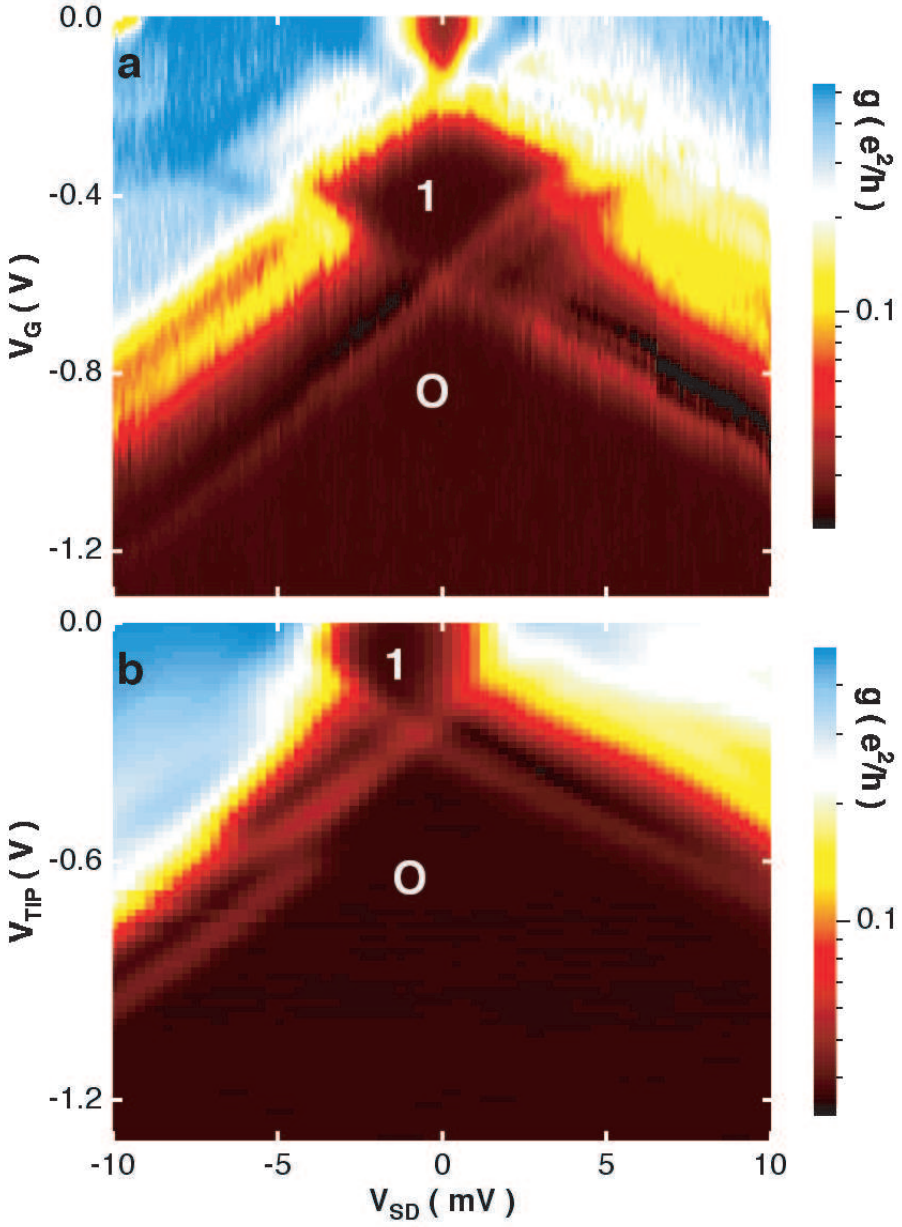

Figure 5.2: (a) Plot of differential conductance $g = dI/dV_{SD}$ as a function of side-gate voltage V_G and source-to-drain voltage V_{SD} at $T = 1.7K$, showing Coulomb blockade diamonds for 0 and 1 electrons, and resonant tunneling through the ground and first excited energy levels separated by 3.1 meV. (b) Plot of differential conductance $g = dI/dV_{SD}$ vs. SPM tip voltage V_{tip} and V_{SD} at $T = 1.7K$ for a fixed tip position, showing that the tip acts as a movable gate.

be moved to any desired location with respect to the dot by changing the ring radius. Numerically fitting the line shape of the ring can help determine the tip-induced shift of electron energy in the dot, as shown below.

5.2 Finite Difference Method (FDM) for Partial Differential Equation (PDE)

The central issue is to look for the ground state energy of an electron confined in a finite region with arbitrary background potential. Solving Schrodinger's equation numerically becomes a must. This section introduces the most popular numerical technique for solving partial differential equations (PDE) like Schrodinger's equation, i.e. the finite difference method (FDM). There are several tricky aspects to pay special attention to in applying FDM. First, choosing appropriate boundary conditions for a specific problem can be critical to obtaining the correct solution, especially for one in a closed geometry as our quantum dot problem. Second, if FDM is used to solve PDE describing time evolution of a system, stability of the solution is critical. If a small amount of initial error due to approximation gets amplified over time, there is no converging solution when approximation must be made at initial or intermediate time. Given the fact that almost all practical problems involve approximation to certain extent, stability analysis is a must. In the following subsections common boundary conditions are compared, and different iterations of FDM over time and their stability are discussed.

5.2.1 Hardwall, Periodic, and Damped Boundary Conditions

In a closed geometry, a system normally sits at one or a combination of eigenstates. Boundary shape and magnitude set the eigenstates to a large extent. Typical approximations of boundary condition for a confined system include hardwall, periodic, and damped boundary conditions.

Hardwall boundary assumes an infinitely high step function along the boundary:

$$\begin{aligned} V &= 0, \; if \; inside \; the \; boundary; \\ &= +\infty, \; if \; outside \; the \; boundary. \end{aligned}$$

Hardwall approximation is often used in cases where a system is tightly confined in a small region comparable to the predominant wavelength and reflection is not important, as in our quantum dot problem.

Periodic boundary assumes that the entire space is periodic and repeats at the scale of the system's length/width. In one dimensional case the approximation is expressed as:

$$\begin{aligned} V &= V(x), \; for \; x \in [0, \; L]; \\ &= V(mod(x, L)), \; for \; x \notin [0, \; L], \end{aligned}$$

where the system is assumed to lie in the range of $[0, \; L]$ and $mod(x, L)$ indicates the remainder from dividing x by L. Periodic approximation best suits cases where the system is confined in a region much larger than its wavelength and where reflection is important, such as reflection of atoms in a quantum wire/waveguide [81].

Damped boundary is a modification of hardwall boundary. Instead of having a potential going from zero abruptly to infinity at the boundary, the confining potential

can be set to rise gradually and smoothly, exponential as an example,

$$\begin{aligned} V &= 0, \; if \; inside \; the \; boundary; \\ &= e^x - 1, \; if \; outside \; the \; boundary, \end{aligned}$$

where x denotes distance from the boundary. When wavepacket technique [36] is used in confined geometry, damped boundary condition can be implemented such that, when a wavepacket hits boundary, the wavepacket is exponentially absorbed by multiplying it with a damping function exponential in distance outside the boundary.

When modeling tip scattering, a tip can be approximated as a hard disk or more realistically as a soft tip whose potential rises quickly but smoothly.

5.2.2 Explicit, Implicit, and Crank-Nicholson FDM

Another important application of FDM is to study the time evolution of a semi-open system, for example, electron flow in 2DEGs [75, 76] and heat transfer in solid. The important issue here is the treatment of iteration over time. Fig. 5.3 illustrates three typical ways of projecting a PDE onto a grid. Let's compare the differences with one dimensional heat equation as an example. The 1-D heat equation is:

$$\frac{\partial^2 \Psi}{\partial x^2} = c \frac{\partial \Psi}{\partial t}, \tag{5.1}$$

where c is a constant that is system specific. In FDM the equation is projected into a 2-D space of $X - T$, as shown in Fig. 5.3. On the grid $\frac{\partial \Psi}{\partial t}$ is always discretized as $(\Psi(x_i, \; t) - \Psi(x_i, \; t-1))/\Delta t$, where $\Delta t = 1$ according to the discretization in time. The discretization of $\frac{\partial^2 \Psi}{\partial x^2}$ is more complicated and is where the three FDMs differ.

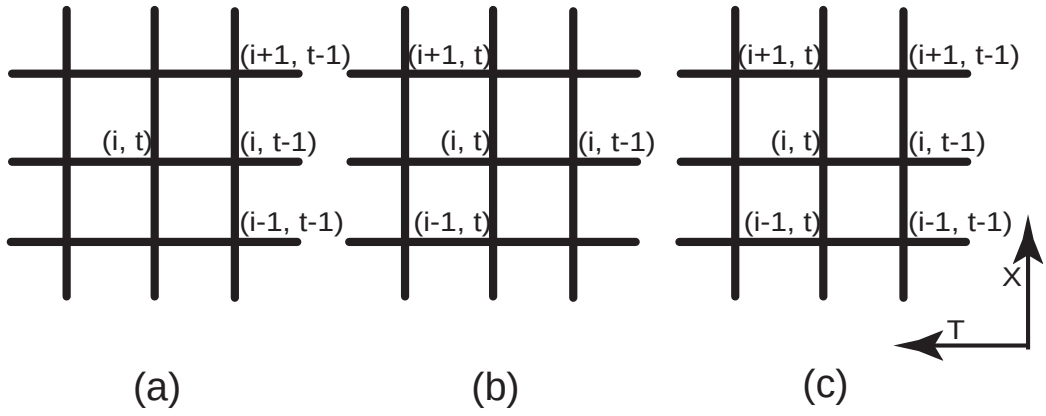

Figure 5.3: (a) Explicit FDM, (b) Implicit FDM, (c) Crank-Nicholson FDM.

Explicit FDM projects $\frac{\partial^2\Psi}{\partial x^2}$ onto the X-axis at time $t-1$. Since $\frac{\partial\Psi}{\partial x}(x_i)$ can be discretized as $(\Psi(x_{i+1},\ t-1)-\Psi(x_i,\ t-1))/\Delta x$, $\frac{\partial^2\Psi}{\partial x^2}(x_i)$ can be expressed as $(\frac{\partial\Psi}{\partial x}(x_i,\ t-1)-\frac{\partial\Psi}{\partial x}(x_{i-1},\ t-1))/\Delta x = (\Psi(x_{i+1},\ t-1)+\Psi(x_{i-1},\ t-1)-2\Psi(x_i,\ t-1))/\Delta x^2$. The explicit FDM formulation of Eq.(5.1) is

$$\frac{\Psi(x_{i+1},\ t-1)+\Psi(x_{i-1},\ t-1)-2\Psi(x_i,\ t-1)}{\Delta x^2} = c\frac{\Psi(x_i,\ t)-\Psi(x_i,\ t-1)}{\Delta t}, \quad (5.2)$$

which relates three variables, $\Psi(x_{i+1},\ t-1)$, $\Psi(x_{i-1},\ t-1)$, and $\Psi(x_i,\ t-1)$, known at time $t-1$ to one unknown variable $\Psi(x_i,\ t)$ at time t. Fig. 5.3(a) illustrates how explicit FDM discretization relates points on the $X-T$ grid. The advantage of Eq. (5.2) is that each variable $\Psi(x_i,\ t)$ at time t can be directly solved for in one equation. Such an advantage actually comes at a price, with the stability of explicit FDM being compromised as shown in the following subsection.

Implicit FDM projects $\frac{\partial^2\Psi}{\partial x^2}$ onto the X-axis at time t. With $\frac{\partial\Psi}{\partial x}(x_i)$ discretized as $(\Psi(x_{i+1},\ t)-\Psi(x_i,\ t))/\Delta x$, $\frac{\partial^2\Psi}{\partial x^2}(x_i)$ can be expressed as $(\frac{\partial\Psi}{\partial x}(x_i,\ t)-\frac{\partial\Psi}{\partial x}(x_{i-1},\ t))/\Delta x = (\Psi(x_{i+1},\ t)+\Psi(x_{i-1},\ t)-2\Psi(x_i,\ t))/\Delta x^2$. The implicit FDM formulation of Eq. (5.1)

is

$$\frac{\Psi(x_{i+1},\ t) + \Psi(x_{i-1},\ t) - 2\Psi(x_i,\ t)}{\Delta x^2} = c\frac{\Psi(x_i,\ t) - \Psi(x_i,\ t-1)}{\Delta t}, \tag{5.3}$$

which relates three variables, $\Psi(x_{i+1},\ t)$, $\Psi(x_{i-1},\ t)$, and $\Psi(x_i,\ t)$, unknown at time t to one variable $\Psi(x_i,\ t-1)$ known at time $t-1$. Fig. 5.3(b) illustrates how implicit FDM discretization relates points on the $X-T$ grid. The disadvantage of Eq. (5.3) is that each variable $\Psi(x_i,\ t)$ at time t can't be directly solved for in one equation. One has to apply Eq. (5.3) to each $\Psi(x_i,\ t)$ at time t and obtain a full series of equations relating $\Psi(x_i,\ t)$ to $\Psi(x_i,\ t-1)$ and solve for all $\Psi(x_i,\ t)$ simultaneously. Despite this disadvantage, implicit FDM is often preferred to explicit FDM because it is unconditionally stable as shown later.

Crank-Nicholson FDM appears like the average of explicit and implicit FDMs. It projects $\frac{\partial^2\Psi}{\partial x^2}$ onto the X-axis at both time t and time $t-1$. Now $\frac{\partial^2\Psi}{\partial x^2}(x_i)$ can be expressed as $(1/2)(\Psi(x_{i+1},\ t) + \Psi(x_{i-1},\ t) - 2\Psi(x_i,\ t) + \Psi(x_{i+1},\ t-1) + \Psi(x_{i-1},\ t-1) - 2\Psi(x_i,\ t-1))/\Delta x^2$. The Crank-Nicholson FDM formulation of Eq.(5.1) is

$$\begin{aligned} & \frac{1}{2\Delta x^2}(\Psi(x_{i+1},\ t) + \Psi(x_{i-1},\ t) - 2\Psi(x_i,\ t) \\ & +\Psi(x_{i+1},\ t-1) + \Psi(x_{i-1},\ t-1) - 2\Psi(x_i,\ t-1)) \\ = \ & c\frac{\Psi(x_i,\ t) - \Psi(x_i,\ t-1)}{\Delta t}, \end{aligned} \tag{5.4}$$

which relates three variables, $\Psi(x_{i+1},\ t)$, $\Psi(x_{i-1},\ t)$, and $\Psi(x_i,\ t)$, unknown at time t to three variables, $\Psi(x_{i+1},\ t-1)$, $\Psi(x_{i-1},\ t-1)$, and $\Psi(x_i,\ t-1)$, known at time $t-1$. Fig. 5.3(c) illustrates how Crank-Nicholson FDM discretization relates points on the $X-T$ grid. The disadvantage of Eq. (5.4) is again that each variable $\Psi(x_i,\ t)$ at time t can't be directly solved for in one equation. One has to apply Eq. (5.4)

repeatedly to $\Psi(x_i,\ t)$ at time t and obtain a full series of equations relating $\Psi(x_i,\ t)$ to $\Psi(x_i,\ t-1)$ and solve for all $\Psi(x_i,\ t)$ simultaneously. Despite this disadvantage, like implicit FDM, Crank-Nicholson FDM is unconditionally stable as shown next.

Other forms of FDM exist. For example, one can apply explicit FDM to solve for half of $\Psi(x_i,\ t)$, alternating along the X-axis, at time t directly and then apply implicit FDM to directly solve for the other half of $\Psi(x_i,\ t)$. These techniques are much more advanced and can be found in books on numerical methods [61].

5.2.3 Stability of FDM when Iterating Over Time

The preceding subsection presents three typical forms of FDM. They differ mainly in the discretization of $\frac{\partial^2\Psi}{\partial x^2}$, which turns out to have a profound impact on the stability of solution. A brief analysis is given as follows.

For the three kinds of FDM, Eqs. (5.2, 5.3, 5.4) can all be cast into a general form

$$\begin{aligned} A\cdot\Psi^t \ &\equiv \ \begin{pmatrix} A_{11} & \dots & A_{1N} \\ . & \dots & . \\ . & \dots & . \\ A_{N1} & \dots & A_{NN} \end{pmatrix} \begin{pmatrix} \Psi_1^t \\ . \\ . \\ \Psi_N^t \end{pmatrix} \\ &= \ \begin{pmatrix} B_{11} & \dots & B_{1N} \\ . & \dots & . \\ . & \dots & . \\ B_{N1} & \dots & B_{NN} \end{pmatrix} \begin{pmatrix} \Psi_1^{t-1} \\ . \\ . \\ \Psi_N^{t-1} \end{pmatrix} \equiv B\cdot\Psi^{t-1}, \end{aligned} \tag{5.5}$$

where $\Psi_i^t = \Psi(x_i, t)$ and A and B are coefficient matrices. Eq. (5.5) can be extended

to between t and $t+1$ or $t+N$

$$\Psi^{t+1} = A^{-1}B \cdot \Psi^t, \tag{5.6}$$

$$\Psi^{t+N} = (A^{-1}B)^N \cdot \Psi^t, \tag{5.7}$$

A stable solution requires that a small error introduced at time t will go to zero at a distant future time $t+N$, which in turn requires

$$(A^{-1}B)^N \rightarrow 0 \; as \; N \rightarrow +\infty. \tag{5.8}$$

It has been shown [65] that for a 1-D heat equation with $c = 1$, explicit FDM is stable only if

$$0 < \Delta t/\Delta x^2 \leq 1/2. \tag{5.9}$$

while implicit and Crank-Nicholson FDM both are stable for any positive value of the ratio $\Delta t/\Delta x^2$.

5.2.4 Comparison with Finite Element Method (Eigenstate Decomposition Method)

Another method of solving for the ground state energy of a confined electron is finite element method or, more commonly called by physicists, eigenstate decomposition method [61]. The idea is to assume that the ground state of a confined electron in the presence of an SPM tip is a linear combination of the lowest few energy states of the electron without the tip. For our single-electron quantum dot configuration, this decomposition has some disadvantages. First, the assumption of the perturbed ground state composed of the lowest few unperturbed states needs careful examination, particularly when the tip perturbation is strong, and the appropriate number

of unperturbed states to use needs to be chosen carefully. Second, the unperturbed states have to be calculated with the proper boundary condition. The lowest energy states of a simple unconstrained harmonic oscillator can't be used directly because they don't reflect the presence of boundary walls. Lastly, given the first two complications, applying eigenstate decomposition to our quantum dot problem would incur significant amount of extra computation to ensure that important features of the system are correctly accounted for. It is easier to directly project the ground state wavefunction onto a two-dimensional grid in coordinate space and solve the Schrodinger's equation. Time evolution is not considered here since we look at only stable eigenstates.

To solve Schrodinger's equation on a coordinate space grid, the potential term in Schrodinger's equation has to be specified properly. We assume a smooth harmonic background potential generated by the confining gates. The other part of the overall potential is due to the SPM tip, which needs to be analyzed carefully in the next section.

5.3 SPM Tip Modeling and Its Impact on Imaging

A scanning probe microscope (SPM) is a valuable tool for studying mesoscopic systems. An SPM can image and manipulate single particles at low temperatures. Single molecules have been positioned on a surface to perform logic operations through molecule cascades [34], and magnetic-resonance-force microscopy can detect a single electron spin through electron-spin resonance (ESR)[64]. Capacitively coupled SPM tips have been used to image the flow of electron waves in a two-dimensional electron

gas (2DEG) [77, 75, 76], and the motion of electrons in the quantum-Hall regime [30, 88, 91, 2, 40]. Coulomb charge oscillations were observed in SPM images of quantum dots formed in carbon nanotubes [87].

However, an SPM tip often introduces a strong perturbation to the original state of the system that is desired. Part of the reason is that the current SPM technique hasn't reached the capability of producing a very localized tip potential compared with the system being probed. In our case the tip potential has a half width of about 140 nm, while the entire quantum dot has a width of about 220 nm. The other part of the reason arises from electronic noise in measurement system and impurity potential in a physical system itself. To make the desired signal distinguishable from electronic noise and impurity perturbation, often SPM tip potential has to be strong enough to produce detectable signal, as was the case in imaging electron flow through an open QPC [75, 76]. So it is not recommended to employ perturbation theory to solve for the perturbed ground state energy of electron in the single-electron quantum dot. However, perturbation theory is convenient for extracting electron wavefunction out of measured conductance signal and is utilized later in combination with singular value decomposition to obtain some primitive, approximate information on electron wavefunction, mainly for future reference.

A Gaussian or Lorentzian potential is often used as an approximation for SPM tip potential, as in electron flow imaging [75, 76]. However, such rough approximation is not suitable for the small quantum dot under consideration here, as the error in approximation could shift energy levels and modify level spacings significantly. A better analytical model of the tip potential consistent with boundary conditions on

the voltage gates is desired. There is no direct, strict analytical model for SPM tip in existing literatures. However, the design and working mechanism of an SPM tip is very close to those of a field-emission tip. The existing rich volume of field-emission tip modeling can help us achieve a better modeling of the SPM tip.

A typical field emitter is described as a cone and a sphere at the apex [63]. One model that gives a set of equipotentials near the surface of the tip shows the spatial distribution of potential in cylindrical coordinate as [63]

$$V(r,\theta) = (V_0/a^{\lambda})(r^{\lambda} - r_0^{2\lambda+1} r^{-\lambda-1}) P_{\lambda}(\cos\theta), \tag{5.10}$$

where V_0 is the potential difference between the tip and a constant-voltage anode a distance a in front of it, r_0 is the tip radius, $P_{\lambda}(\cos\theta)$ is the Legendre function, θ_0 is the cone angle, and λ is chosen so that $P_{\lambda}(\cos\theta_0) = 0$.

For our experiment, the constant-voltage anode can be replaced by the constant-voltage gates and r_0 set to zero. λ is set to equal 0.22 given that $\theta_0 = 12$ degrees. So the analytical model modified for SPM tip becomes

$$V(r,\theta) = V_0((r/a)^{0.22} P_{0.22}(\cos\theta) - (r_g/a)^{0.22}), \tag{5.11}$$

where r_g indicates the location of voltage gates and the constant term $-(r_g/a)^{0.22}$ is added to shift the potential distribution so that the gates are at $V = 0$, a boundary condition assumed in solving Schrodinger's equation in the next section.

Numerical plots show that Eq. (5.11) gives good fit to the boundary condition of equipotentials at both tip surface and voltage gates. More importantly, Eq. (5.11) builds the dependence on tip voltage V_0 into the potential generated by the tip inside the quantum dot. Experimental observation shows a continuous shrinkage of conductance ring with increasing tip voltage V_0. Theoretical simulation in the next section

based on Eq. (5.11) successfully shows the same amount of shrinkage using the same set of V_0 while keeping all other parameters fixed! This is a strong verification of the validity of Eq. (5.11).

5.4 Conductance Rings for a Symmetric Square Quantum Dot

With a fairly good model for tip potential established in the preceding section, the remaining part of the overall potential is the background potential. The exact shape of the background potential decides electron wavefunction distribution and therefore the shape of conductance ring under SPM probing. In this section a symmetric, harmonic potential is used to demonstrate the main experimental observations, i.e. the width and shrinking speed with tip voltage of the conductance ring. The distortion in the rings is explored in the next section when asymmetric background potential is considered.

5.4.1 Energy Spectrum of a Confined Electron

When the background potential is harmonic and 1-dimensional, electrons in it should reside in various degenerate energy levels given as [32]

$$E_n = (n + 1/2)\hbar\omega, \tag{5.12}$$

where n =0, 1, 2, ... , with wavefunctions given as

$$\psi_n(x) = A_n(a_+)^n e^{-\frac{m\omega}{2\hbar}x^2}, \tag{5.13}$$

where A_n is the normalization factor, a_+ is the raising operator, and $V(x) = \frac{1}{2}m\omega^2 x^2$.

When the background potential is harmonic, symmetric in 2-dimensions, as assumed in our case, the energy levels are

$$E_n = E_{x,n} + E_{y,n} = (n_x + n_y + 1)\hbar\omega = (n+1)\hbar\omega, \tag{5.14}$$

where n_x =0, 1, 2, ... , n_y =0, 1, 2, ... , and $n = n_x + n_y$ =0, 1, 2, ... Here for an energy level n, it's $(n+1)$-fold degenerate.

However, once an SPM tip is added into the overall potential, the symmetry is broken and the degeneracy is lifted. In Fig. 5.4, we plot both the lowest 55 energy levels and the ground state electron wavefunction, when the SPM tip is placed close to the upper left corner of the dot as shown in Fig. 5.4(b). The electron is obviously pushed towards the lower right corner. The lowest 55 energy levels that previously made up the lowest 10 discrete levels now become 55 non-degenerate levels due to the presence of the SPM tip, as shown in Fig. 5.4(a). The lowest few levels are still closely grouped according to their previous degeneracy, while the highest plotted levels have a distribution that shows almost no sign of previous degeneracy.

What is relevant to us is the ground state of electron in the quantum dot. The ground state energy fluctuates as the SPM tip moves around. When it matches the energy required for electron to tunnel out of the dot, the conductance through the quantum dot drastically increases and produces a peak in the conductance map. Given the symmetry of the background potential, a circular ring of high conductance is expected in the conductance map for a given tip voltage. However, the half width of the ring is yet to be decided.

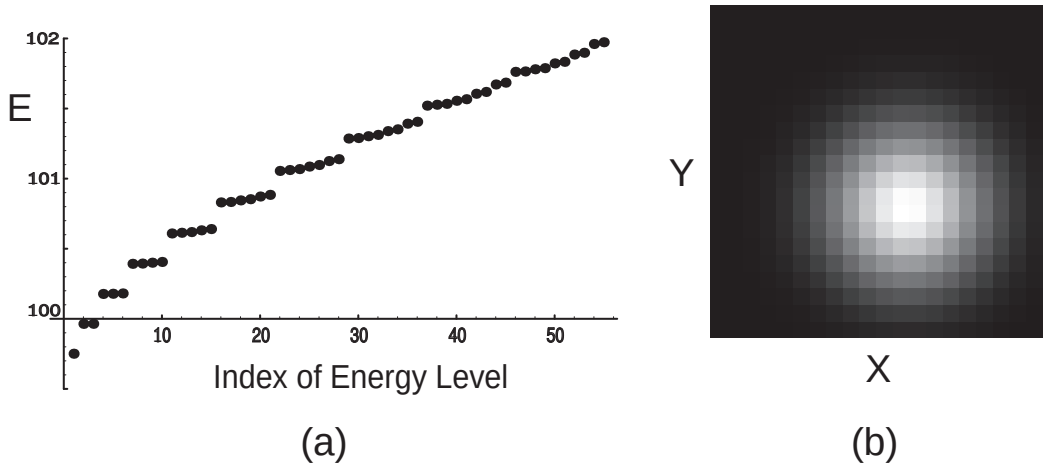

Figure 5.4: (a) The energy levels of the lowest 55 eigenstates, corresponding to the lowest 10 degenerate energy levels without tip perturbation, (b) The ground state electron wavefunction when an SPM tip is placed near the upper left corner of the plot.

5.4.2 Line Shape of Resonant Tunneling in Ground State

Beenakker [7] has shown that, for resonant tunneling, the line shape of conductance is given by

$$G = G_{\max}[\cosh(\Delta/2k_BT)]^{-2}, \tag{5.15}$$

where Δ is the energy difference between the lowest energy level in the dot and the Fermi energy in the leads. The energy difference is zero at resonance and deviates from zero as the tip moves away from its resonant position. The dot conductance at resonance is [7]

$$G_{\max}(\vec{r}) = (e^2/4k_BT)\Gamma(\vec{r}), \tag{5.16}$$

where $\Gamma(\vec{r})$ is the tunneling rate.

When we assume $\Gamma(\vec{r})$ to be constant, $G_{\max}$ becomes constant and the line shape of G is decided by $[\cosh(\Delta/2k_BT)]^{-2}$. For fixed T and Fermi energy in the leads, varying the tip voltage leads to different maps of Δ in the dot and therefore different

rings, as shown in Fig. 5.5.

For $T = 1.7K$ in the experiment, the Fermi energy in the leads is found by matching the theoretical ring position with the experimental one for a specific tip voltage, say, 40 mV as shown in Fig. 5.6(a) and (e). If our model is correct, the same Fermi energy in the leads should give rise to a series of shrinking rings as in the experiment when the tip voltage increases. Fig. 5.6(a-h) do show the same rate of shrinkage with tip voltage between experimental data and theoretical simulation. Even more, the width of the rings is matched reasonably well, which comes out naturally without any additional adjustment of parameters. This serves as another justification of the validity of our model.

In the experiment under consideration, $\Gamma(\vec{r})$ alters as the SPM tip is scanned above the dot, due to changes in the coupling between the tip and the point contacts. It results in variations in $G_{\max}$ along the ring as shown in Fig. 5.6(a-d). The exact mechanism underlying such variations would be an interesting topic to pursue in the future.

5.5 Conductance Rings for a Deformed Quantum Dot

By assuming a symmetric harmonic potential in the preceding section, we have produced a set of conductance rings that have replicated many important features observed in the experiment. However, it is obvious that the distortion in the experimental rings is largely due to asymmetry in the background potential. Therefore

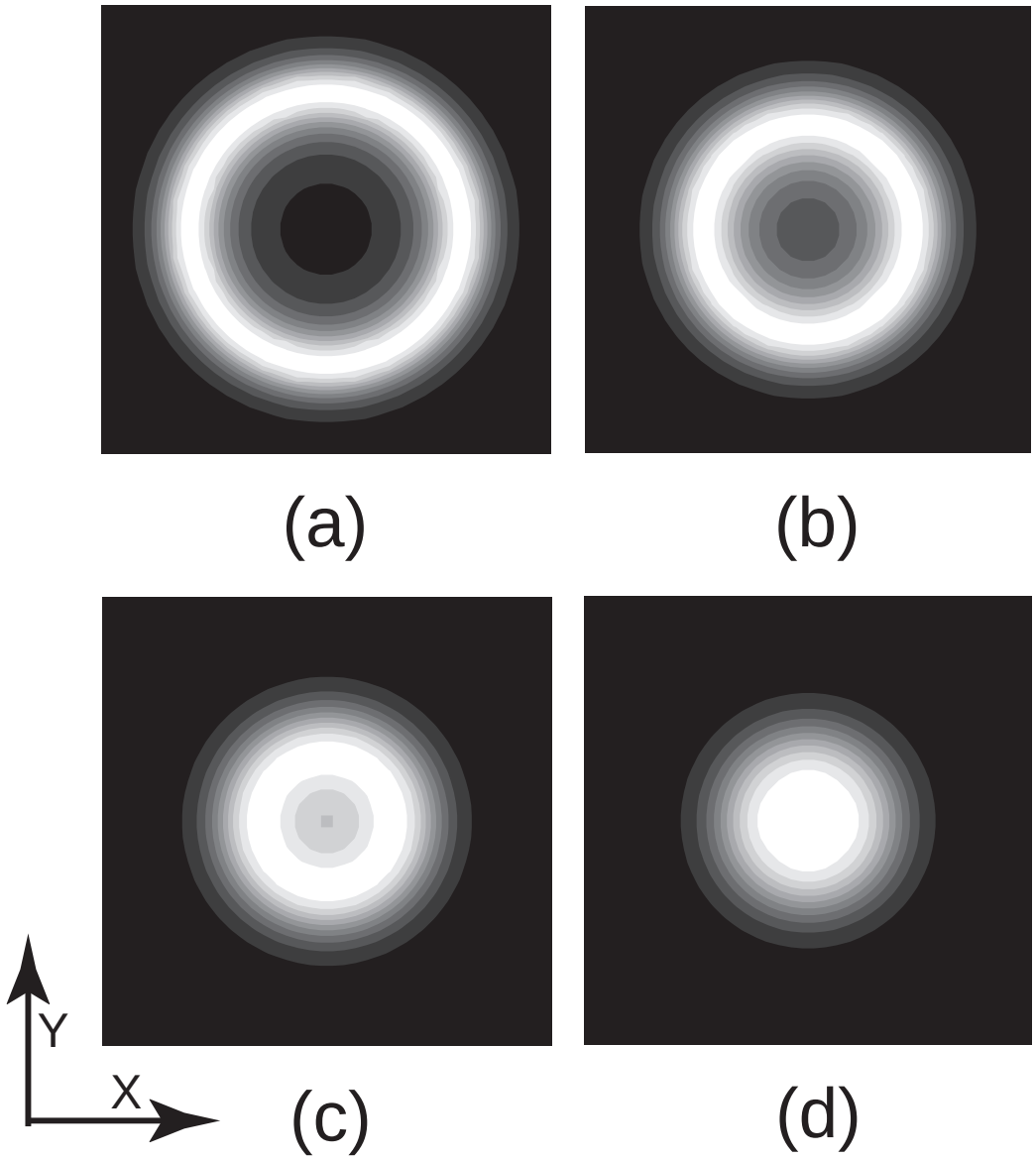

Figure 5.5: Theoretical simulation of conductance rings for a symmetric quantum dot. (a) $V_{tip} = 40mV$, (b) $V_{tip} = 50mV$, (c) $V_{tip} = 60mV$, and (d) $V_{tip} = 80mV$.

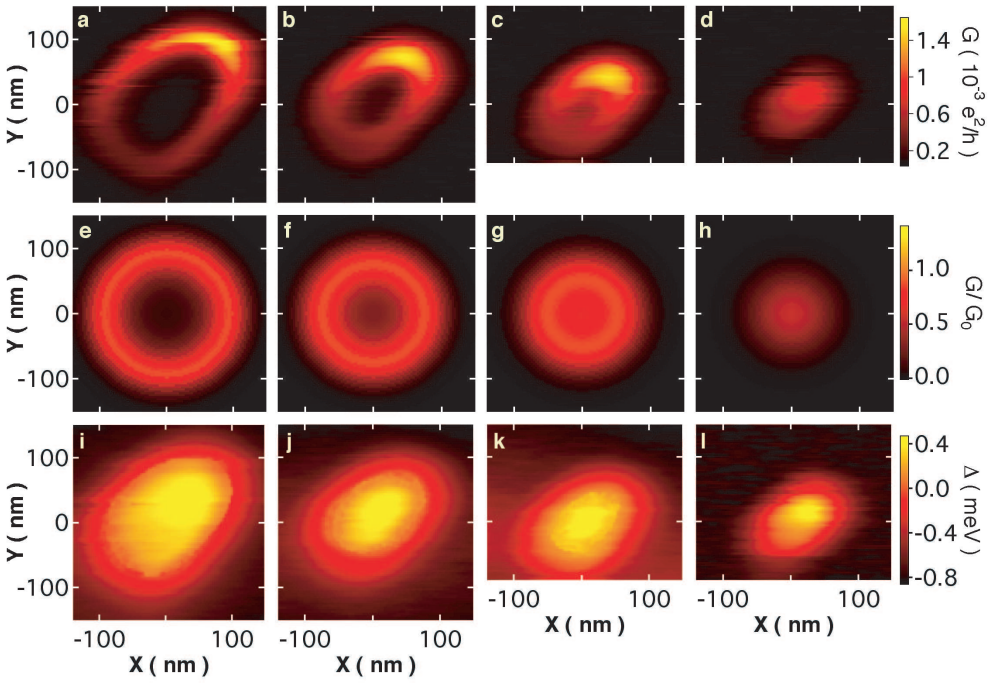

Figure 5.6: (a-d) Coulomb blockade images of a single-electron quantum dot at $T = 1.7K$, showing the dot conductance G vs. tip position. The ring of high conductance around the center of the dot is formed by the Coulomb blockade peak between 0 and 1 electron in the dot. The tip voltages V_{tip} for a-d are $40mV$, $50mV$, $60mV$ and $80mV$ respectively. (e-h) Theoretical simulations of the images in a-d for a dot formed by a parabolic potential with energy spacing $3.1meV$ (energy of first excited state from Fig. 5.2a) for the same tip voltages as a-d. (i-l) Experimental maps of the energy shift Δ of electrons in the dot vs. tip position, extracted from the measured line shape of the Coulomb blockade conductance peak forming the rings in images a-d.

some simulations with a deformed two-dimensional harmonic potential have been performed. Asymmetry is introduced by setting two different oscillation frequencies, ω_1 and ω_2, along the two diagonal directions for the dot potential. When the two oscillation frequencies differ further and further, the ground state wavefunction of electron also develops similar asymmetry accordingly. The conductance rings are found to show the same kind of asymmetry as well, but to a much lesser extent.

The reduced sensitivity to asymmetry in conductance rings can be traced back to insensitivity to asymmetry in ground state energy. When there are two oscillation frequencies defined for two orthogonal directions, the ground state energy for a two-dimensional oscillator is given by

$$E_0 = E_{1,0} + E_{2,0} = 1/2\hbar(\omega_1 + \omega_2) \; . \tag{5.17}$$

Suppose we keep ω_1 unchanged, but lower ω_2 gradually starting with $\omega_2 = \omega_1$. The component of the confining potential corresponding to ω_2 will flatten out gradually and the ground state wavefunction deforms accordingly. But the ground state energy E_0 only goes down from $1/2\hbar(\omega_1 + \omega_2) = \hbar\omega_1$ to at least $1/2\hbar\omega_1$, that is, reduction of less than a factor of 2. Therefore conductance rings corresponding to resonant tunneling don't deform as much as the background potential in the dot. The most deformation in the rings that has been obtained is shown in Fig. 5.7 for the same set of tip voltages as in Fig. 5.5.

It would be interesting to study whether other kinds of asymmetry than the simple asymmetric harmonic one here can produce more deformation in the rings. One thing that will remain true is that the amount of asymmetry induced in ground state wavefunction will be different from that in ground state energy and conductance ring.

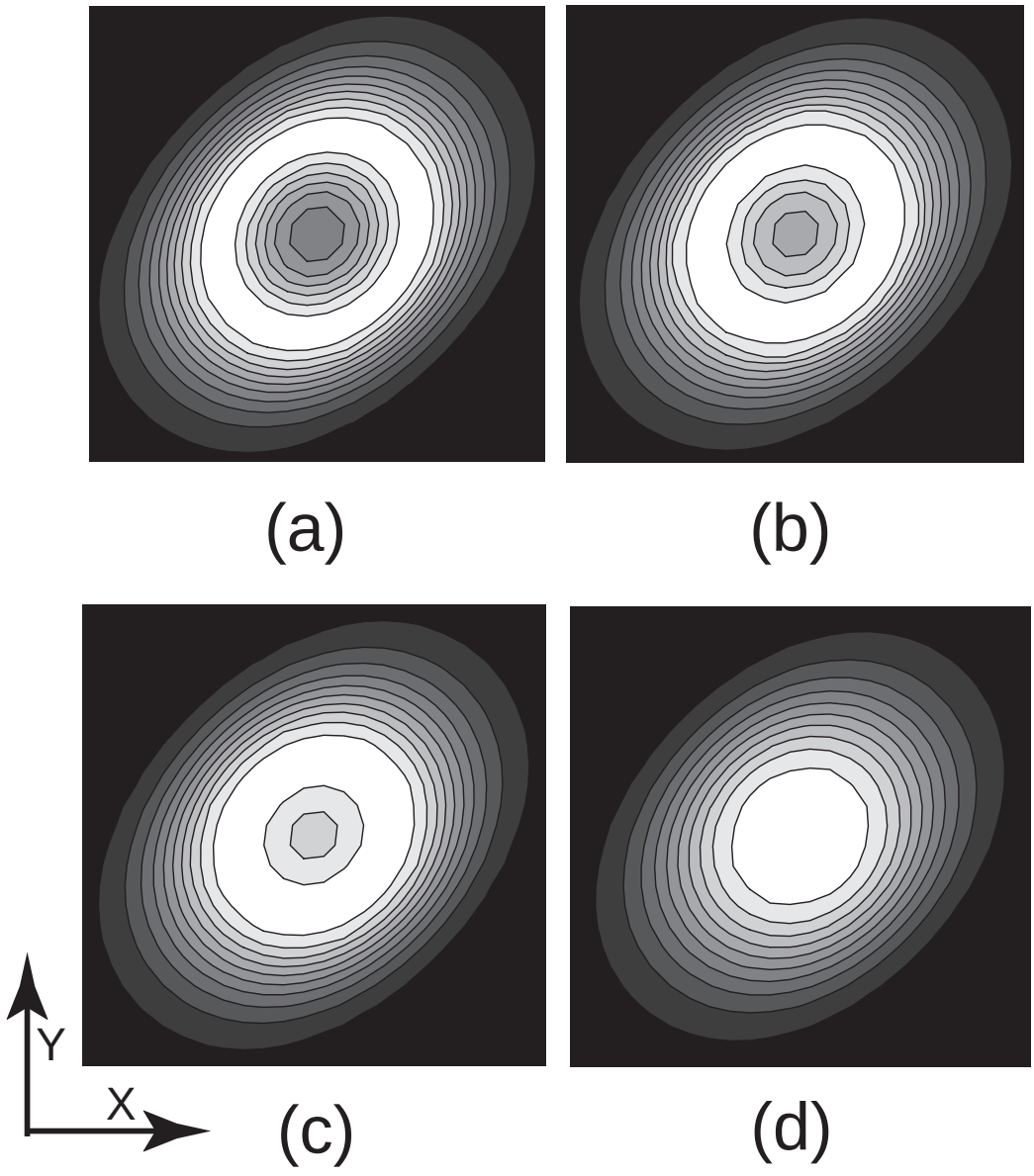

Figure 5.7: Theoretical simulation of conductance rings for a deformed quantum dot. (a) $V_{tip} = 40mV$, (b) $V_{tip} = 50mV$, (c) $V_{tip} = 60mV$, and (d) $V_{tip} = 80mV$.

5.6 Signal Processing Using Singular Value Decomposition (SVD)

In the preceding sections we have assumed certain kinds of background potential for the quantum dot and then solved for the corresponding electron ground state wavefunction, energy, and tunneling conductance map under SPM probing to match the experimental data. However, as shown already, to reproduce the deformation of and variations along the ring requires many numerical trials with different dot potentials and, most troublesomely, coupling between tip and QPCs. It seems a lot easier if the problem is approached from a totally different direction, i.e. using digital signal processing to extract true conductance map and then electron wavefunction. This and the next section will focus on this method.

Digital signal processing is a field of studying ways to filter noise out of data collected using digital electronic devices. It has been and still is an actively researched subject. There have established many popular algorithms like Kalman filter, tracking filter, and inverse filter, *etc.* [43]. The one that is most relevant to our needs is singular value decomposition (SVD). Singular value decomposition has been employed previously in the study of time-reversal focusing of ultrasound [71], and is also the method of choice for solving most linear least-squares problems, in particular the ones singular or very close to singular [61].

SVD is based on a linear algebra theorem that any $M \times N$ matrix A whose number of rows M is greater than or equal to its number of columns N, can be written as the product of an $M \times N$ column-orthogonal matrix U, an $N \times N$ diagonal matrix W

with positive or zero elements (the singular values), and the transpose of an $N \times N$ orthogonal matrix V

$$A = U \cdot W \cdot V^T \ . \tag{5.18}$$

Here column-orthogonal means that the product of any two columns is zero unless they are the same column. The diagonal elements of matrix W often correspond to intrinsic eigenvalues of the system described by the matrix A. The inverse of the matrix A is conveniently given as

$$A^{-1} = V \cdot W^{-1} \cdot U^T \ , \tag{5.19}$$

where W^{-1} is a diagonal matrix whose diagonal elements are the inverse of those of the matrix W. Problems arise when W contains some near zero elements, which often correspond to noise, and whose inverses in W^{-1} dominate over those from real signal in W. Noise filtering becomes critical.

SVD is particularly well suited for our problem. The tunneling conductance matrix A can be decomposed to show in matrix W its components from electron eigenstates and noise. Noise filtering is achieved by setting to zero certain number of lowest diagonal elements in W and their counterparts in W^{-1}, which are assumed to correspond to noise component [71]. This noise filtering has two important features that make SVD desirable. First, by eliminating only the smallest elements in matrix W, matrix A is not significantly changed. Second and most importantly, without low level noise in matrix W, its inversion and therefore the inversion of matrix A is no longer singular. Thus, noise filtering and non-singular inversion of matrix are accomplished in the same time. The challenge is to decide what number of the lowest elements in matrix W should be set to zero in noise filtering, which requires rigorous study [71].

A primitive application of SVD to the single-electron quantum dot is shown in the next section.

5.7 Extracting Electron Wavefunction Using SVD

5.7.1 Weak Perturbation Assumption

One of the constraints of applying SVD to extracting electron wavefunction is the feasibility of expressing energy level fluctuation with tip movement, ΔE, as the product of electron wave density, Ψ, and tip potential, T,

$$\Delta E = \Psi \cdot T \ . \tag{5.20}$$

Given Eq. (5.20), SVD can then be applied to invert matrix T into T_{SVD}^{-1} and obtain the approximate electron wavefunction Ψ_{SVD}

$$\Psi_{SVD} = \Delta E \cdot T_{SVD}^{-1} = \Psi \cdot (T \cdot T_{SVD}^{-1}) \ . \tag{5.21}$$

However, Eq. (5.20) is valid only if tip potential can be treated as weak perturbation to the original wavefunction [32]

$$\Delta E =< \psi | V_{tip} | \psi >= \Psi \cdot T \ . \tag{5.22}$$

As discussed previously, SPM tip potential in the experiment was strong perturbation due to technical constraints and detection needs. However, for future reference, weak perturbation theory is assumed here to demonstrate how electron wavefunction can be extracted out of energy shift map, which itself is obtained from conductance map using Eq. (5.15).

5.7.2 Stability Analysis of SVD

To apply singular value decomposition to extracting electron wavefunction, its stability must be examined because of the particular noise filtering technique of removing the lowest diagonal elements in the decomposition of the original matrix. It is equivalent to examining whether $(T \cdot T_{SVD}^{-1})$ is close to 1 in Eq. (5.21). Fig. 5.8 shows for a test wavefunction the original function, the one obtained through Eq. (5.21), and the difference between the two. During noise filtering in SVD only the highest 6 out of 1600 diagonal elements in the decomposed matrix are kept non-zero. The match between the original and the approximation using SVD is reasonably good. Taking into account that the original wavefunction is not smooth and naturally causes errors when being smoothed out in numerical processing, the precision of recovering the original wavefunction is remarkable.

5.7.3 The Recovered Electron Wavefunction

Applying SVD by keeping only the highest 6 diagonal elements during noise filtering, we obtain in Fig. 5.9 an approximate wavefunction for each of the four conductance maps in Fig. 5.6(a-d). If weak perturbation approximation is valid, the four maps should lead to the same original wavefunction since they differ only in tip voltage. The match among Fig. 5.9(a-d) is not good enough, though the tilt in wavefunction is consistently present in each of them. The mismatch is not surprising since weak perturbation approximation is not applicable to the present experiment, as explained above.

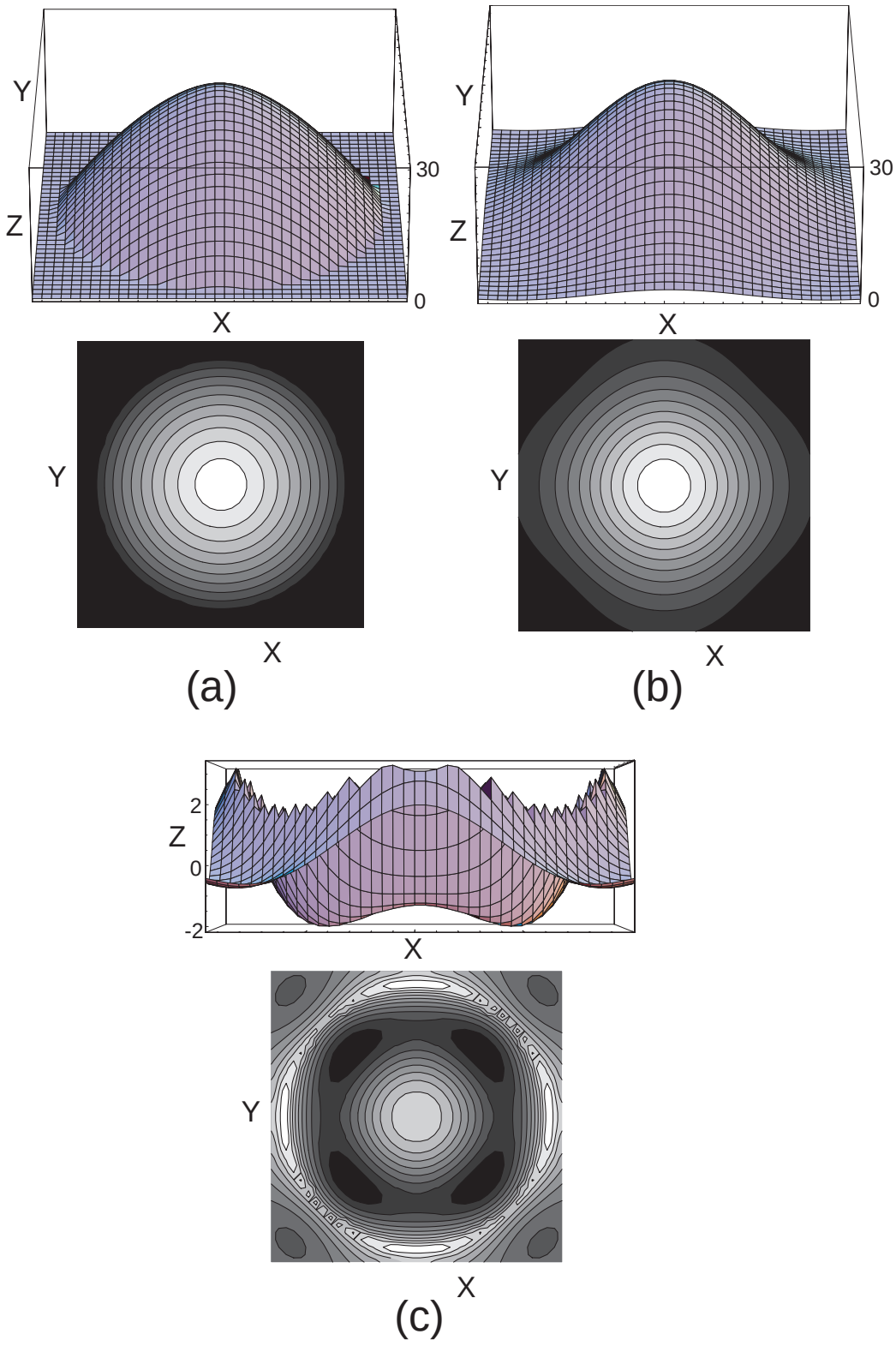

Figure 5.8: (a) The original wavefunction, (b) the wavefunction retrieved using SVD, (c) the difference between the retrieved and original wavefunctions.

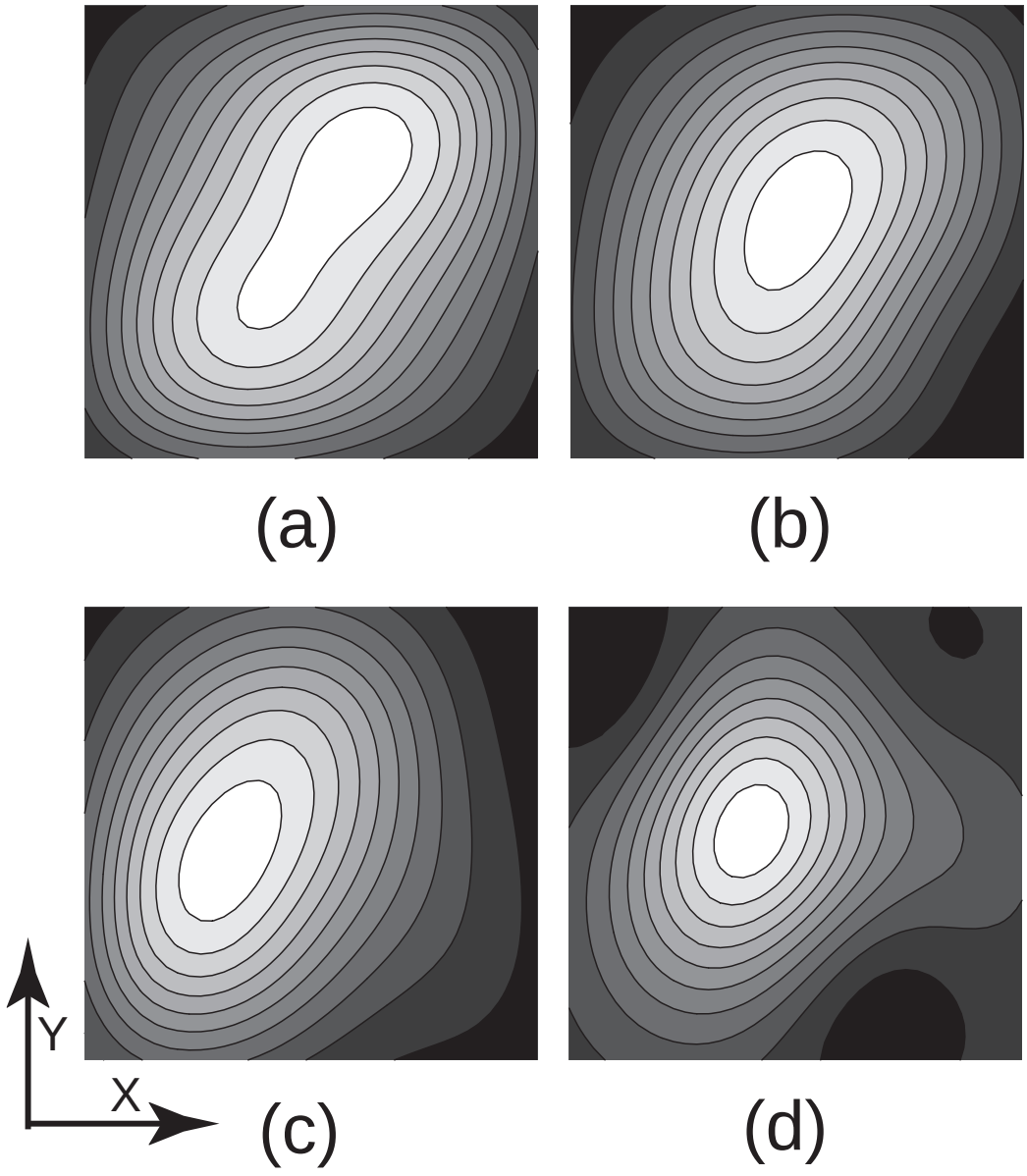

Figure 5.9: The electron wavefunctions retrieved from experimental conductance data. (a) $V_{tip} = 40mV$, (b) $V_{tip} = 50mV$, (c) $V_{tip} = 60mV$, and (d) $V_{tip} = 80mV$.

5.8 Conclusion

In this chapter, theoretical aspects of a single-electron quantum dot experiment are investigated. Finite difference methods are introduced and utilized to reproduce conductance rings and their shrinkage with tip voltage observed experimentally. An approximate model of SPM tip potential is employed as an improvement over the common, oversimplified Gaussian or Lorentzian model. The impact of tip perturbation and asymmetry in dot potential are both analyzed. Singular value decomposition is then introduced as an alternative approach to studying the experiment, by extracting electron wavefunction out of energy level shift. Recovery of electron wavefunction using SVD is demonstrated under weak perturbation assumption. The match isn't ideal for the current experiment, but will likely improve in future experiments that will be placed well in the weak perturbation regime.

Chapter 6

Imaging Cusps in 2-Dimensional Electron Gases

In this chapter, we explore another interesting and important phenomenon observed in electron flow imaging experiments: cusp formation in two-dimensional electron gases (2DEGs). Typical patterns of classical and quantum cusps are studied. Their respective dynamics and correspondence are explained. Finally, experimentally observed cusps are well reproduced using the Kirchhoff method.

6.1 Introduction

Cusps have been thoroughly studied in classical dynamic systems. As the second simplest universality class of catastrophe theory [73, 3], a cusp describes a certain kind of singularity with an infinite number of coalescing contributions concentrated at one point in space. In momentum and coordinate phase space, a cusp singularity is

identified by a vertical segment of the curve which shows a large number of momenta accumulated at one coordinate point. However, a cusp is less singular than a total focusing point, which in phase space is represented by all momenta concentrated at a focal point.

The singularity introduced by a cusp and other catastrophe phenomena has been a particular challenge in the application of semiclassical methods. Semiclassical methods are routinely used to probe systems in regimes between classical and quantum limits, for example, mesoscopic devices where the coherence length is comparable to the size of the device and wave-like features such as interference are prominent. Semiclassical methods extend classical mechanics by attaching wave functions to the "skeleton" defined by a system's classical dynamics, and thus making the transition into the quantum regime where all objects are characterized by their respective wave functions. However, the semiclassical approximation fails in singular or near singular cases where there are multiple contributions to the wave function within the volume of a single Planck cell. Various methods have been developed to rescue semiclassical methods in cases of catastrophe [10, 8, 60], and recently in mesoscopic devices [86, 56]. Very recently, a new method called "Replacement Manifolds" has been proposed by Vanicek and Heller [83] and has been applied successfully to study coherent two-dimensional electron flow in the semiclassical setting [84].

Though for a long time a purely classical concept, with the fast advancement of experimental techniques, cusps have been observed in mesoscopic devices in quantum regime, most notably in the recent two-dimensional electron flow imaging experiments [77, 75, 76]. While branching [76, 66] and fringing [66, 67, 39, 48] in the electron flow

imaging experiments are well understood, the formation of cusps in a quantum gas of electrons remains a mystery and therefore has inspired the rigorous investigation presented in this chapter. Our study looks at a cusp singularity at the classical and quantum limits and explores the connection between electron flow in these two regimes. Most importantly, generation of quantum cusps by Kirchhoff method is demonstrated, and tip scan imaging process is analyzed at both classical and quantum limits. Quantum cusps observed in electron flow imaging experiments [77, 75, 76] are reproduced nicely by the Kirchhoff method.

This chapter is organized as follows: the first two sections present various ray patterns and tip scan images of classical cusps formed by electron flows in the classical regime. In the third section, the phase space dynamics of classical tip scan are analyzed. The fourth section builds a connection between the classical and quantum cusps generated by the Kirchhoff method, where various quantum cusps and their tip scan images are presented. In the final section, the cusps observed experimentally are compared to cusps and their tip scan images produced by Kirchhoff method. The theoretical model matches experimental results very well.

6.2 Ray Patterns of Classical Cusps

A classical cusp is best characterized by its evolution in phase space. As shown in Fig. 6.1, electrons released from the left hand side, upon passing an attractive potential, form a cusp behind it. Taken at selected cross sections before, at, and after the focal point, the $P_y - Y$ curve has its two opposite peaks rising and approaching and passing each other. Right at the focal point, the two peaks are so close to each

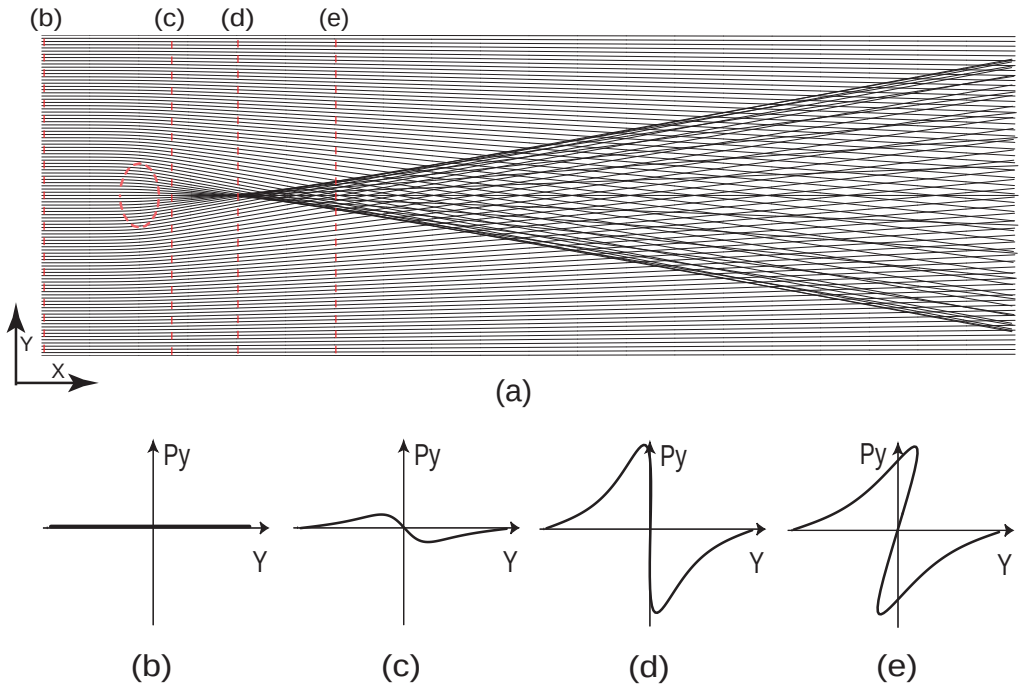

Figure 6.1: (a) Classical trajectories of electrons passing by an attractive potential with uniform velocity. The red dotted ellipse indicates the half width of the potential, which is circularly symmetric and compressed into an ellipse due to different scalings of the axes. (b-e) Phase space diagrams for four cross sections before, at, and after the focal point.

other that most of the points between the two peaks lie on a vertical section of the curve, which represents a large number of rays passing through the focal point at the same time. However, for rays far away from the potential, P_y remains close to zero throughout, and the rays are not bent in any visible way. The singularity arising from focusing of only the inner portion of incident rays is the signature of classical cusp. The entire space is divided into two parts, a one-fold part and a three-fold part.

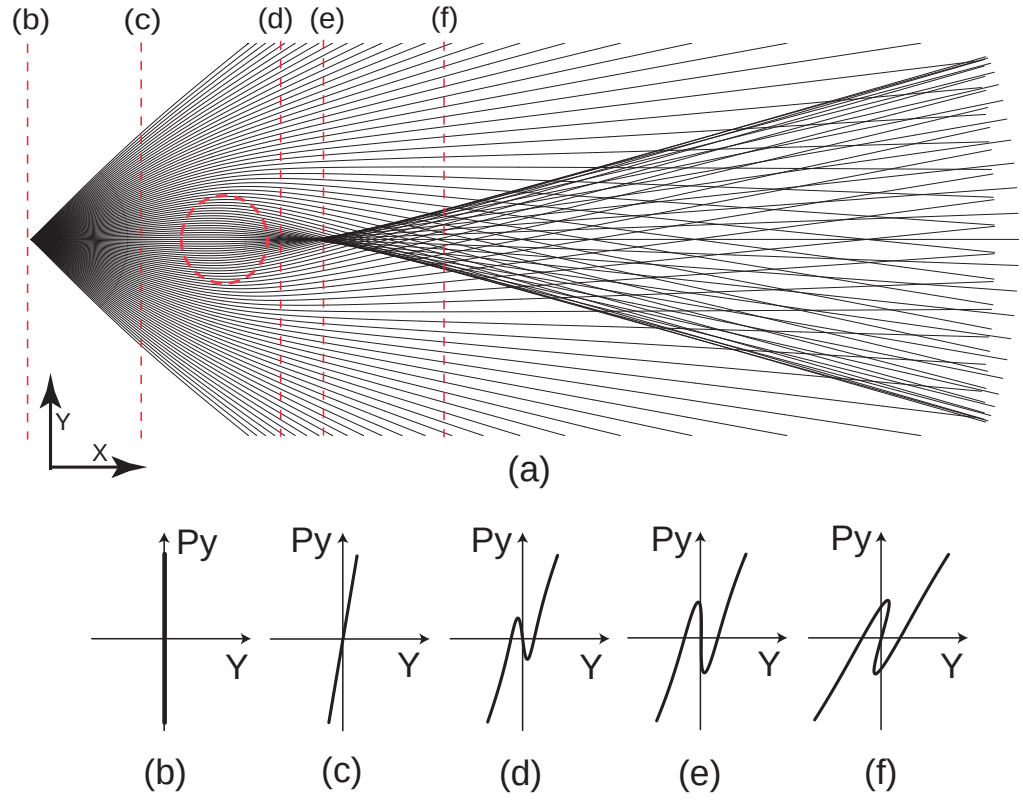

Figure 6.2: (a) Classical trajectories of electrons from a QPC passing by an attractive potential. Electron velocities are equal in amplitude. The red dotted circle indicates the half width of the symmetric potential. (b-f) Phase space diagrams for five cross sections before, at, and after the focal point.

The initial condition in Fig. 6.1 can be modified to have electrons coming out of one point, which corresponds to a quantum point contact (QPC) in later studies,

with equal speeds oriented uniformly within an angular range. Fig. 6.2 shows that a cusp is still formed behind the attractive potential. Comparing Fig. 6.1 and Fig. 6.2, it is easy to see that despite the totally opposite initial conditions in Fig. 6.1(b) and Fig. 6.2(b), a cusp is formed by partial focusing of incident rays in both cases, which is represented by a vertical middle section between two local extremes on the $P_y - Y$ curve in Fig. 6.1(d) and Fig. 6.2(e).

While Fig. 6.1 and Fig. 6.2 achieve focusing by an attractive potential which extends in space and continuously bends electron trajectories, much like a thick lens bending incident light rays continuously, the formation of a cusp is determined solely by the distribution of P_y along Y axis. If electrons are released in free space with the characteristic distribution of P_y as in Fig. 6.1(c) or Fig. 6.2(d), a cusp can still be formed despite the absence of any attractive potential. What the extended potentials in Fig. 6.1 and Fig. 6.2 have done is adjusting the P_y of incident rays to some desired distribution for a cusp to form. The role of the potentials can be replaced by releasing electrons with the proper P_y distribution at the very beginning, just like a thick lens can be replaced by a thin lens with the same focal length which can bend light rays to the desired orientations immediately.

There are many possible P_y distributions that can lead to cusp formation. One analytical distribution has been given by Vanicek and Heller [84]

$$P_y(y) = -\alpha y e^{-y^2}, \tag{6.1}$$

where α is some positive constant. Eq. (6.1) corresponds to the kind of P_y distribution in Fig. 6.1(c).

By releasing electrons with a pre-set P_y distribution similar to that in Eq. (6.1),

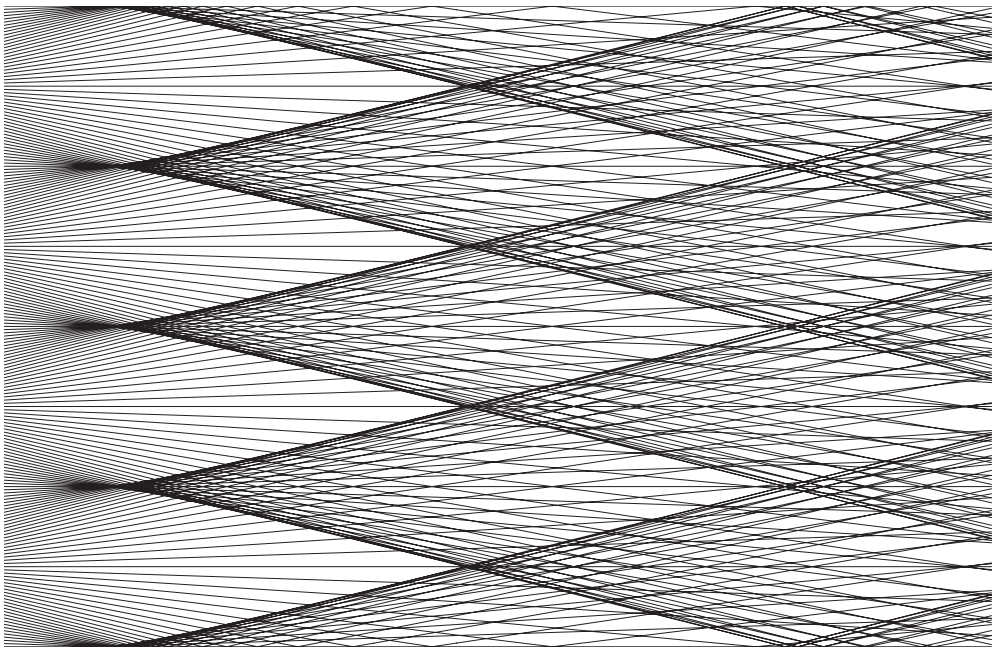

Figure 6.3: Classical trajectories of electrons released with equal velocities oriented differently.

Fig. 6.3 demonstrates the formation of multiple cusps in free space. Furthermore, the cusps overlap with each other downstream, giving rise to 5-fold and 7-fold regions in addition to the 1-fold and 3-fold regions for a single cusp. The cusp formation in Fig. 6.3 verifies the critical importance of the P_y distribution.

However, it remains to be seen how a tip scan image of a classical cusp compares with the original cusp. The introduction of a repulsive tip can greatly complicate the distribution of electron trajectories, especially when the tip is large. The following section explores tip scan images in great detail.

6.3 Tip Scan Images of Classical Cusps through a QPC

One of the major goals of this study is to understand the quantum cusps obtained in electron flow imaging experiments using a scanning probe microscope (SPM) tip. In these experiments [77, 75, 76], a repulsive SPM tip is placed downstream from a QPC to deflect electrons transported through a QPC. The conductance image taken by the SPM tip essentially measures the amount of flux reflected back through the QPC as a function of tip position. In order to develop a parallel analysis of tip scan image of classical electron flow, the electron flow in Fig. 6.2 is probed using an SPM tip modeled by a Gaussian tip (which is so steep that it's close to being a hard disk). The exact shape of the tip potential, Gaussian or Lorentzian, doesn't affect the results in any noticeable way, as long as it's steep enough. The tip scan images are taken in both single and double-QPC systems.

6.3.1 Tip Scan Images from a Single QPC

In a single-QPC system, electrons are released from a QPC, and the tip scan image counts the number of electrons reflected back through the QPC as a function of tip position. Fig. 6.4 shows for one tip position how the presence of a repulsive tip can drastically alter the electron's trajectory. Despite the large number of overall trajectories, only a small number of them are deflected by the tip, and an even smaller number of them actually return to the proximity of the QPC.

By moving the tip center over the entire region enclosed by the dotted rectangle

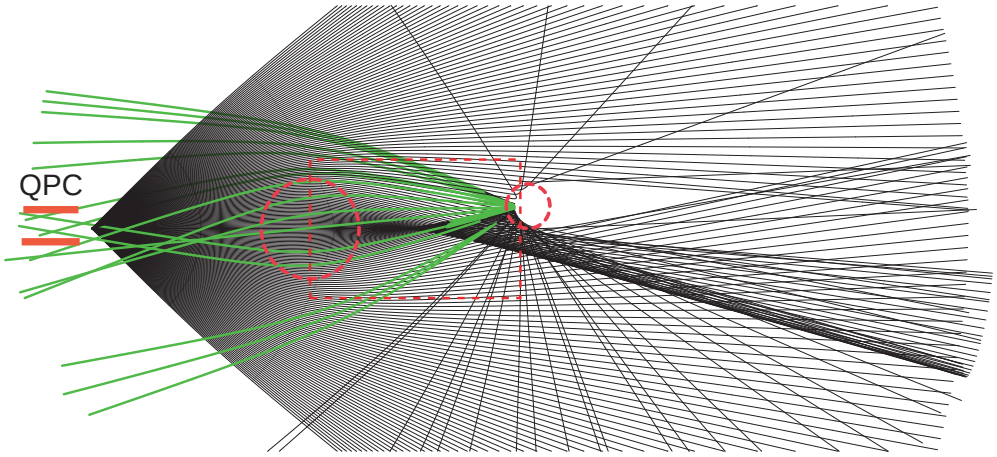

Figure 6.4: Classical trajectories of electrons from a QPC passing by an attractive potential and scattering from a repulsive tip. Electron velocities are equal in amplitude. The large and small red dotted circles indicate the half width of the potential and the 1% width of the tip.

in Fig. 6.4, and by counting the number of returning electron trajectories through the QPC for each tip position, a tip scan image is obtained. Such an image is shown in Fig. 6.5(b). The QPC width is set equal to the 1% radius of the tip (where the tip potential drops to 1% of its peak value), which is also 1/6 of the height of the scanned region. In a later simulation, the QPC width was reduced to 1/5 of the 1% radius of the tip, which gave a better match to the parameters in the actual electron flow imaging experiments. For later convenience, the 1% radius of the tip is referred to as the tip radius, though most electron trajectories penetrate a little further into the tip before turning back as shown in Fig. 6.4.

Fig. 6.5 compares the underlying ray density in the scanned region to the tip scan image. Both figures are plotted as $Log[x+1]$, where x is the actual data and Log the natural logarithm (in order to better show their structures). While Fig. 6.5(a) shows a ray density for a typical cusp structure with 1-fold and 3-fold regions separated by

bright cusp edges, Fig. 6.5(b) shows a much more complicated structure for the tip scan image.

A close examination of Fig. 6.5(b) reveals a 3-layer structure along the cusp edges, and they all lie around the cusp edges in Fig. 6.5(a). A strong middle layer sits exactly atop the cusp edges in the ray density; a weaker and thinner inner layer is shifted downstream from the middle layer by a horizontal distance of roughly one tip radius; a weak and broad outer layer extends upstream by more than one tip radius. Furthermore, the 3D plots of Fig. 6.5 (not shown here because of their large sizes) show that the peak in tip scan image actually is positioned one tip radius behind the peak position, i.e. at the focus of the cusp in ray density.

To understand the three-layer structure of the tip scan image requires a detailed analysis of electron reflection from the tip. There are two possible mechanisms for electrons to be reflected back through the QPC. One is for an electron to direct itself at the tip *center* and, upon reflection, travel its own trajectory backwards into the QPC, which is referred to as "self reflection". Self reflection exists in both 1-fold and 3-fold regions. The other mechanism is for an electron to hit the tip *surface* and, upon reflection, travel another electron's trajectory backwards into the QPC while the other electron does the exact opposite, which is referred to as "exchange reflection". Exchange reflection is possible only in 3-fold region because it requires more than one trajectory crossing at the same point allowing for an exchange coupling among the trajectories to exist. In general exchange reflection is less efficient even in the 3-fold region because exchange coupling doesn't always exist for crossing trajectories. Fig. 6.5(c) illustrates the difference between self and exchange reflections.

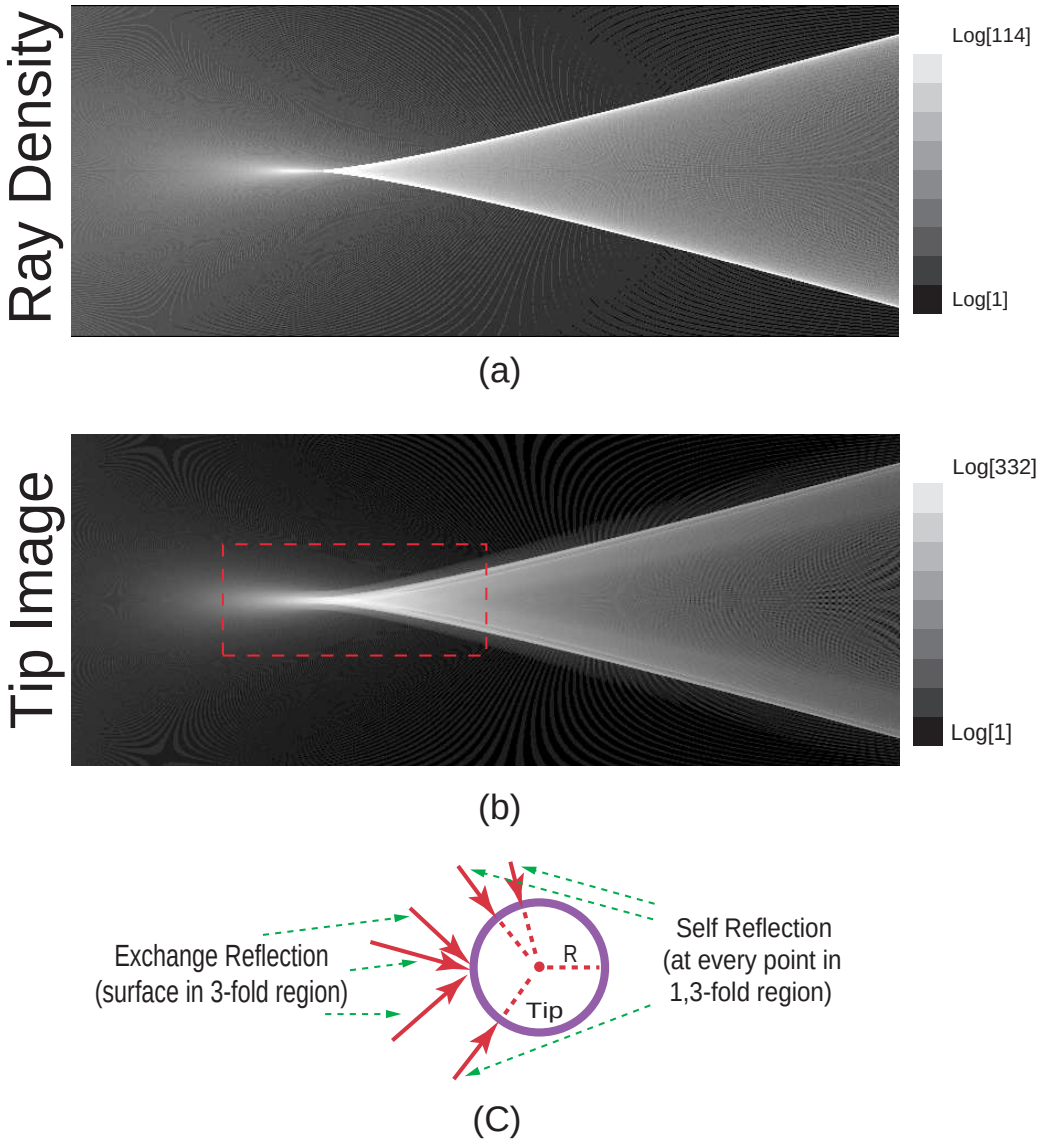

Figure 6.5: (a) Log ray density plot. (b) Log tip scan image plot based on the same ray density as in (a). (c) Illustration of self and exchange reflections on the tip.

The strictly defined self reflection returns to the QPC only 1 electron in 1-fold region and 3 electrons in 3-fold region, while the strictly defined exchange reflection returns to the QPC only in the 3-fold region either 0 electron in the absence of proper exchange coupling or 2 electrons with exchange coupling (except at focal point, both mechanisms reflect back all electrons passing the focus). The dependence on ray density fluctuations within the 1-fold or 3-fold regions would not exist in Fig. 6.5(b), nor would the three-layer structure of cusp edges if the QPC was reduced to a point with zero width and if the tip was a perfect hard point. Apparently Fig. 6.5(b) accounts for contributions from more than the strict self and exchange reflections.

The conditions for self and exchange reflections can be relaxed a bit when the QPC has a finite size and the tip is large and soft. When the QPC has a non-zero width, electrons returning to the proximity of the QPC could be counted as being reflected back through the QPC. Now electrons directed slightly off tip center seem to be able to return via "self reflection", and electron trajectories slightly off each other on the tip surface seem to send electrons back via "exchange reflection". Therefore for each tip position a range of electron trajectories can be returned to the QPC via "self" or "exchange" reflection, thus allowing the tip scan image to depend on the underlying ray density. On the other hand, when the tip is large, the high density cusp edges in Fig. 6.5(a) can return electrons either via self reflection when the tip center is right on the edges or via the less efficient exchange reflection when the left tip surface is on the edges (the tip center is now shifted one radius downstream). This tip size effect should explain the bright middle and less bright inner layers of the cusp edges in Fig. 6.5(b), which is further examined in later simulation. When the tip is soft, electrons

can penetrate further into the tip before turning back. This penetration doesn't affect self reflection, and modifies exchange reflection only in the sense that now an incoming trajectory is coupled to another trajectory after it penetrates and emerges out of the tip at a different surface point, instead of the trajectory intersecting it at the entering surface point as in the case of hard tip.

Among the three factors - QPC width, tip size, and tip softness, QPC width and tip size do introduce significant impacts on tip scan image, while tip softness doesn't show any important influence until the phase space dynamics of tip scan is studied. In the rest of this subsection, the impact of QPC width and tip size on tip scan image is analyzed. In the following section on phase space dynamics, the role of tip softness will be discussed.

Small vs. Large QPC

To better understand the effects of QPC width on tip scan images, the small region around the cusp focus enclosed by the dotted rectangle in Fig. 6.5(b) is replotted with its 3D version in Fig. 6.6(b). The same region is scanned by the same tip again after the QPC width is reduced to 1/5 its original width, and the resulting 3D and log density plots are shown in Fig. 6.6(a).

Comparing Fig. 6.6(a) and (b), the most striking differences are the absence of the weak outer layer and the thinning of both the middle and inner layer in Fig. 6.6(a). The peak in the tip scan image in Fig. 6.6(a) still sits one tip radius behind the cusp focus position in ray density plot in Fig. 6.5(a).

Clearly a wide QPC allows for more relaxed conditions for "self" and "exchange"

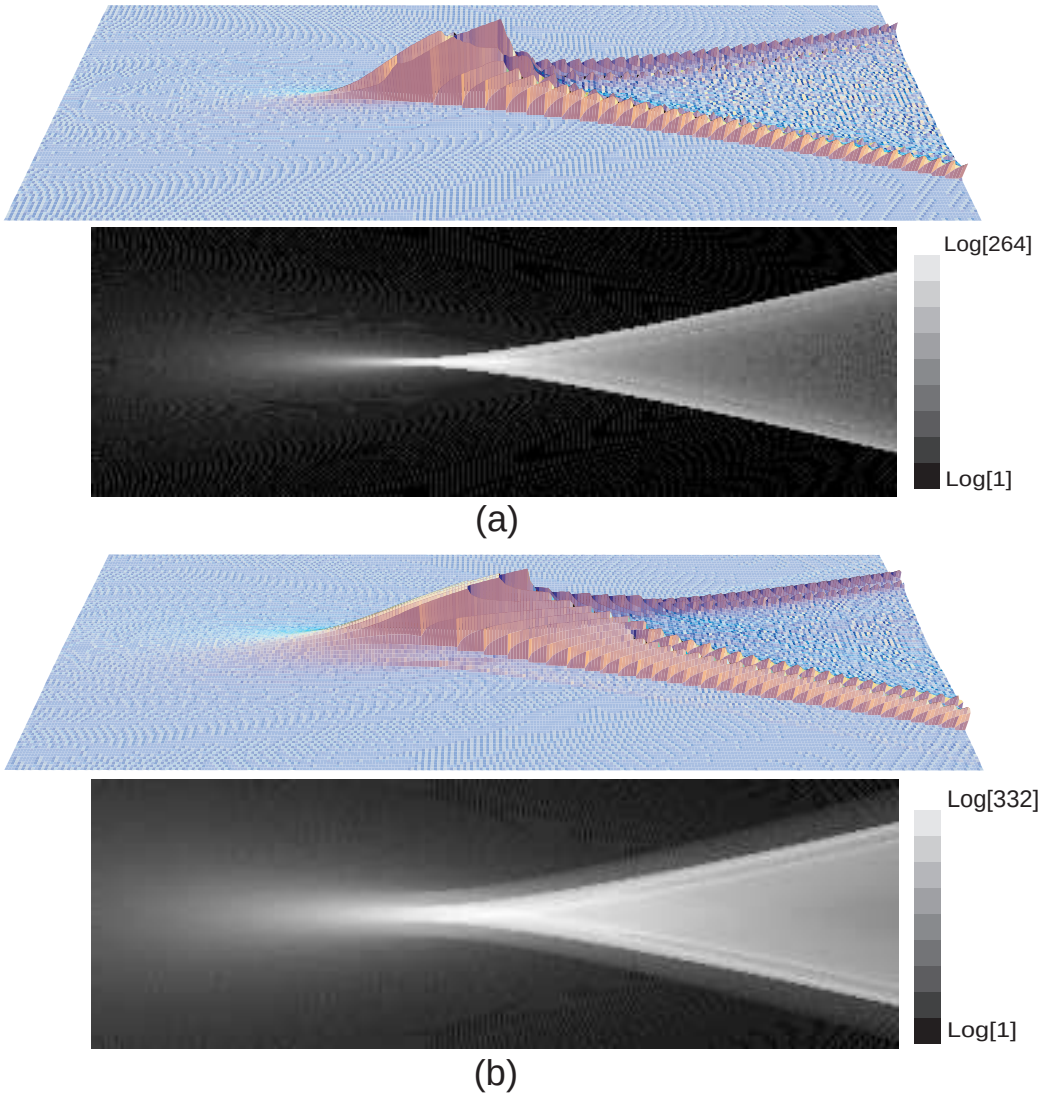

Figure 6.6: 3D and log density plots of tip scan for (a) a small QPC (width equal to one tenth of density plot height) and (b) a large QPC (width equal to half the height of density plot).

reflections, as discussed above, and thus the appearance of the weak outer layer and the broadening of middle and inner layers in Fig. 6.6(b). The fact that the peak in the tip scan image stays at the same location implies its independence on QPC width. It should also be noticed that tip radius to QPC width is 5 : 1 in Fig. 6.6(a), close to the ratio in the electron flow imaging experiments [77, 75, 76].

Small vs. Large Tip

To better understand the effects of tip size upon the tip scan images, the small region around the cusp focus enclosed by the dotted rectangle in Fig. 6.5(b) is replotted with its 3D version in Fig. 6.7(b). The same region is scanned again by a tip with a radius 1/5 that of the original tip radius, and the resulting 3D and log density plots are shown in Fig. 6.7(a).

Comparing Fig. 6.7(a) and (b), the most striking differences are the absence of the inner layer and the advancing of the peak in Fig. 6.7(a). The weak, broad outer layer still exists in Fig. 6.7(a), most visible around the cusp focus.

Clearly a large tip allows the same high density cusp edges to appear twice in tip scan via self and exchange reflections, respectively, as discussed above, and thus the appearance of the middle and inner layers in Fig. 6.7(b). When the tip size becomes small enough, the middle and inner layers merge into one which looks sharper because of its steeper rise from its background. The advancing of tip scan peak indicates a strong dependence upon tip size. The existence of the weak, broad outer layer in Fig. 6.7(a) is consistent with the preceding argument that it arises from a wide QPC.

The appearance of tip scan peak behind cusp focus in ray density can be under-

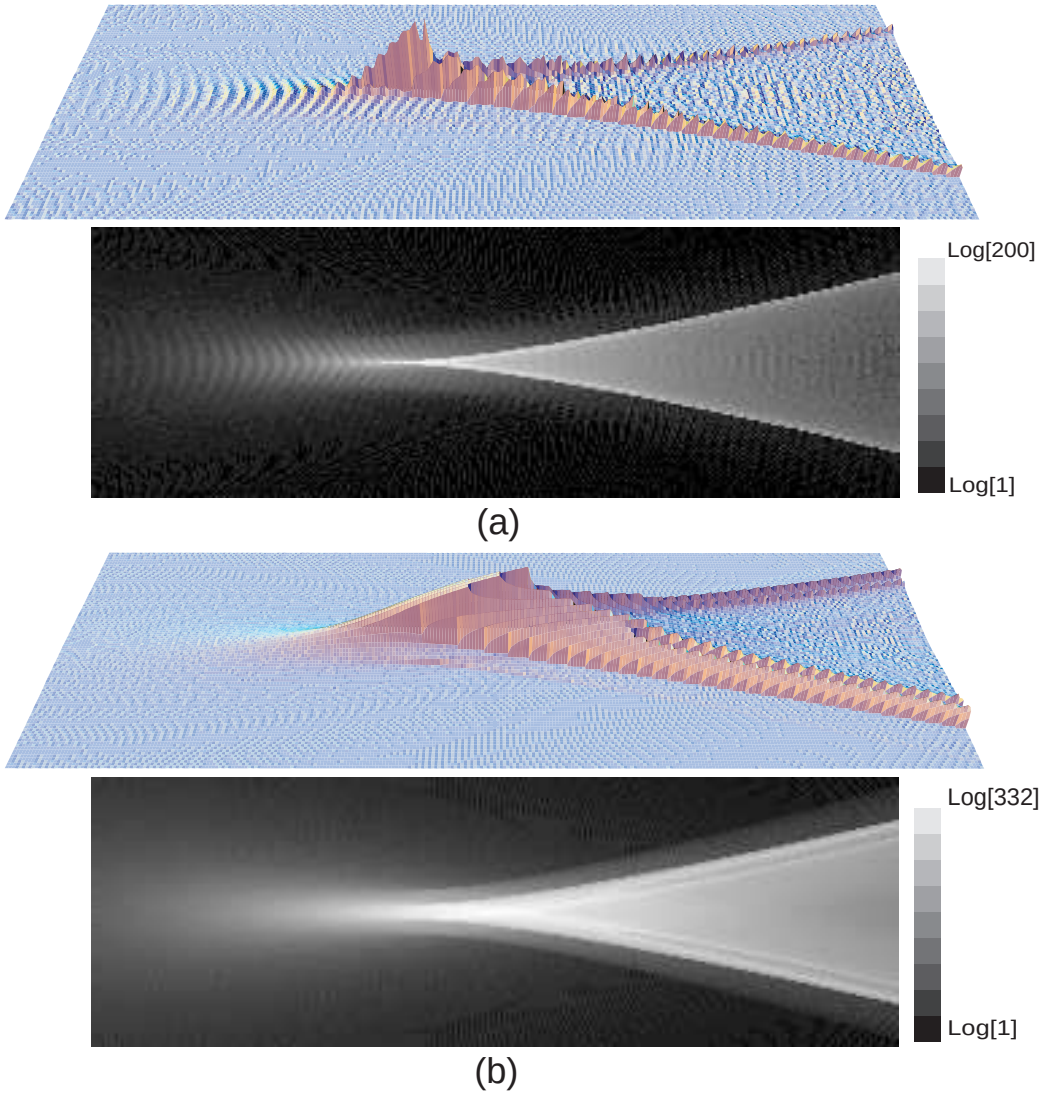

Figure 6.7: 3D and log density plots of tip scan for (a) a small tip (1% radius equal to one tenth of density plot height) and (b) a large tip (1% radius equal to half the height of density plot).

stood as follows. If the tip center is placed right at the cusp focal point, self reflection is maximized, but since the tip's left surface is now in the 1-fold region, no exchange reflection can happen. However, if the tip is shifted downstream such that cusp focus sits on the tip's left surface, exchange reflection is then maximized. Since the tip center is still in a high ray density region, self reflection is also strong. The combined self and exchange reflections exceed the sole self reflection when the tip center is at the cusp focus. Therefore tip scan peak appears one tip radius downstream from the cusp focus in ray density plot.

6.3.2 Tip Scan Images from Double QPCs

It would be interesting to consider a modified version of the experiment. What if we detect electrons deflected back toward a second QPC other than the original QPC? Fig. 6.8 illustrates the setting. A second QPC (QPC2) is placed above the original QPC (QPC1). Electrons are still released from QPC1, but only electrons deflected back into QPC2 are counted as signal. The tip scan image and the magnified focal region are plotted in log density and 3D, respectively. All parameters except the addition of a second QPC remain the same as in Fig. 6.5(b).

The most striking feature of the tip scan image is the appearance of a second, tilted, weaker, and broader cusp that overlaps with the original cusp seen in the single QPC system. The original cusp appears mostly the same except it's weaker and has only one fold except for the part overlapping with the second cusp.

It is expected that the original cusp in 1-QPC system will largely remain the same because by shifting the tip a little downward the electrons going back to QPC1 will

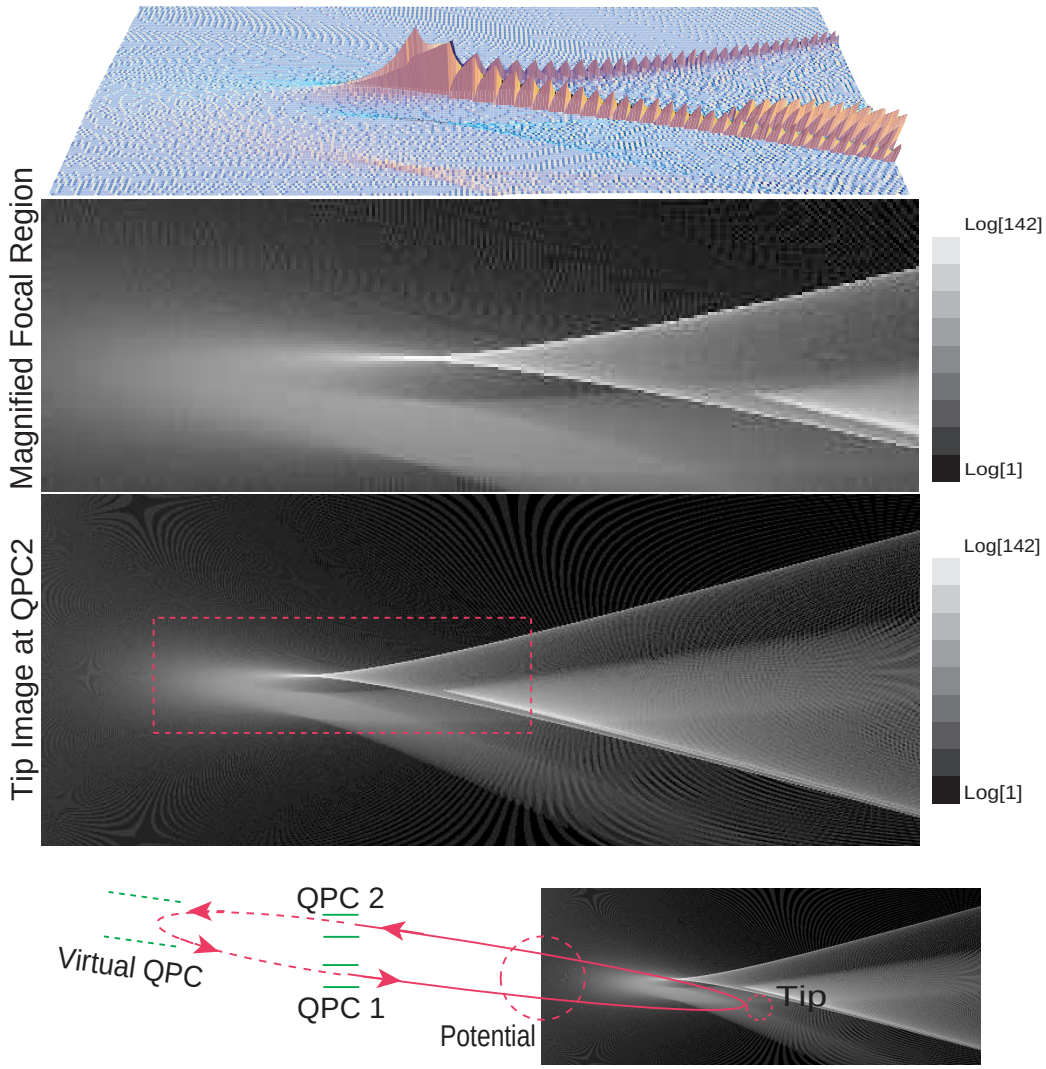

Figure 6.8: 3D and log density plots of tip scan for a double-QPC system. The bottom diagram shows the setting and illustrates a virtual QPC for the formation of a weaker, tilted second cusp.

instead reach QPC2. Still QPC2 is a less efficient detector because of its asymmetric location relative to QPC1 and the attractive potential. While prior self reflection to QPC1 can be re-directed to QPC2 by shifting the tip downward, prior exchange reflection demands a high symmetry between trajectories for exchange coupling to happen. Shifting tip downward might be able to re-direct one of the two trajectories in exchange coupling toward QPC2; the other trajectory however will most likely miss QPC2. This partially explains the disappearance in Fig. 6.8 of the less bright inner layer of cusp edges, which heavily depends on exchange reflection as explained above. Also QPC2 doesn't face returning electrons normally as QPC1 does. Projecting QPC2 onto the perpendicular direction of returning electrons' velocities gives an equivalent QPC with smaller width, which explains the disappearance of the weak outer layer, the thinning of the middle layer, and contributes to the disappearance of the inner layer of cusp edges in Fig. 6.8.

The tilted, weaker, and broader second cusp can be explained in terms of a virtual QPC, as illustrated in Fig. 6.8. When an electron released from QPC1 passes the potential, scatters from the tip, and returns to QPC2, its trajectory can be viewed as part of the trajectory of another electron released from a virtual QPC behind both QPC1 and QPC2. The double-QPC system therefore reduces to a single-QPC system, in part because of the symmetry of the setup. Thus a second, tilted cusp appears. Certainly such equivalence is not perfect. As explained above, QPC2 reduces to a narrower QPC after projection onto the perpendicular direction of returning electrons' velocities. Also, to allow for electrons, with a wide spread of initial velocity orientation, originating from QPC1 to QPC2 to appear as if they came out of the

same virtual QPC, the virtual QPC needs to be wide. Thus the second cusp is very weak and has a broad focal region. Having explored tip scan images under different settings, it would be very interesting to see how the imaging process works in phase space. The next section provides findings from this unique perspective.

6.4 Phase Space Dynamics of Tip Scan of Classical Cusps

Modern dynamical systems are often analyzed in phase space to reveal some unique features that wouldn't be seen otherwise. The study of tip scan imaging of classical cusps wouldn't be complete without investigation of its phase space dynamics. Of the many choices, the Poincare-Birkhoff phase space diagram has been popularly employed in studying particle scattering on a billiard boundary, which closely resembles electron scattering from a tip surface (boundary) in our setting. Therefore it's natural to consider tip scan imaging of classical cusps in Poincare-Birkhoff phase space.

Fig. 6.9 shows the phase space diagrams for tip scattering at selected points. Since angular momentum L (only amplitude considered, positive direction defined as out of page) is conserved during scattering relative to the tip center, L is plotted against ray index of electrons in the first row, with ray 0 for an electron released at $-\pi/4$ relative to the horizontal axis and ray 2000 for an electron released at $\pi/4$ relative to the horizontal axis and the rest distributed uniformly between. Only angular momentum L of the rays scattering from the tip is registered and the rest L are set to zero. Normally a Poincare-Birkhoff diagram plots P_s against S, where S is the

length along the billiard boundary, and P_s is the component of linear momentum, P, tangent to the billiard boundary at S. Here given the circular symmetry of the tip and that the tip boundary is a circle of constant tip radius R, P_s is used to represent the angular momentum, L, for easy comparison with the first row plots of L, and the angle S/R is used as the horizontal coordinate to avoid dependence on the specific R. The resulting $P_s - S/R$ diagram is in linear proportion to the conventional Poincare-Birkhoff diagram, because angular momentum is R (a constant) times the component of linear momentum tangent to the tip surface and angle S/R is boundary length S divided by R. The $P_s - S/R$ diagrams for incoming and outgoing rays are shown in the second and third rows, respectively, and are combined in the fourth row to better demonstrate the action by tip scattering.

Comparing (a-d) in Fig. 6.9, the phase space dynamics of tip scattering are clearly revealed. First of all, in all the plots of L and P_s, there is a clear division between 1-fold and 3-fold regions, which corresponds well with the same division of space for the underlying cusp. When the tip moves, the 3-fold part on L and P_s curves moves from below the $L = 0$ or $P_s = 0$ line when tip is outside the cusp, to being tangent to the 0 line when tip is on the middle layer of cusp edge, to slightly above the 0 line when tip is on the inner layer of cusp edge, to high above and symmetric about the 0 line when tip is in the middle of the cusp.

Fig. 6.9(e) plots the phase space diagrams for the peak position in tip scan image. Comparison with Fig. 6.9(c) proves that the tip actually sits behind the focus to get optimal exchange reflection on its surface. Had the tip sit right on the focus, a straight middle segment corresponding to $L = 0$ and $P_s = 0$ would appear on

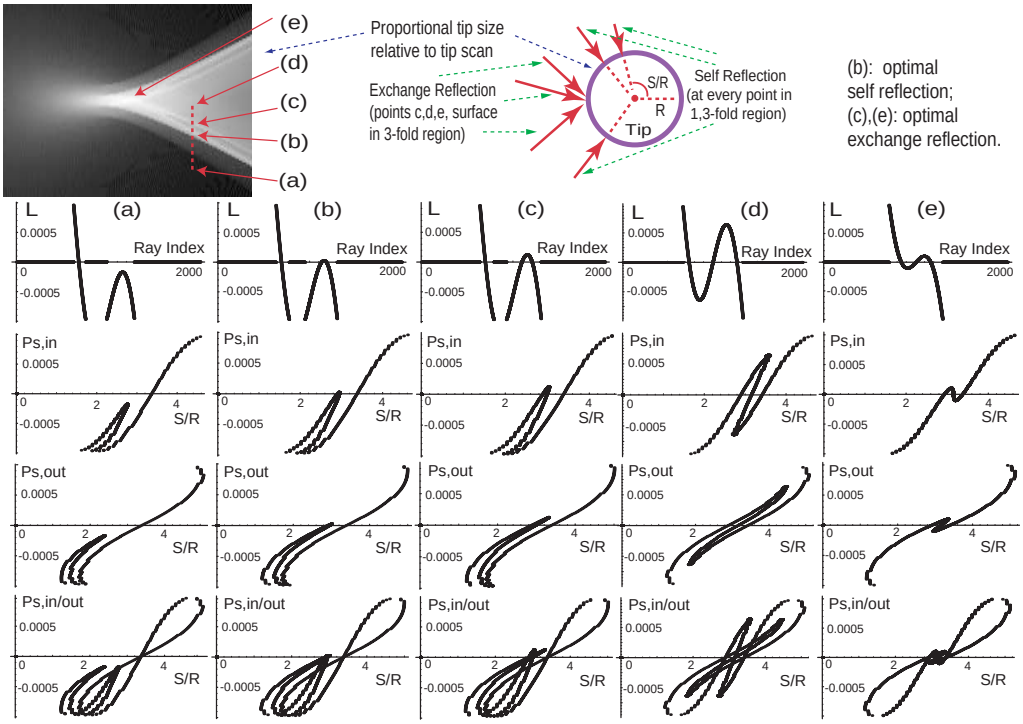

Figure 6.9: Phase space diagrams for selected points inside or outside the cusp. The first row plots angular momentum vs. ray index. The second and third rows plot incoming and outgoing P_s vs. S/R, respectively, where R is the 1% tip radius. The fourth row combines the second and third rows.

the angular momentum and Poincare-Birkhoff curves, which represents optimal self reflection instead.

The self and exchange reflections now have a phase space interpretation. Self reflection happens at the points of $L = 0$ and $P_s = 0$ on these curves. Fig. 6.9(b) corresponds to optimal self reflection and apparently is helped by the part of the L or P_s curve that is tangent to $L = 0$ or $P_s = 0$.

Exchange reflection is represented in the following way. First of all, it's easy to see from the ray diagram at the top of Fig. 6.9 that exchange reflection can happen to only rays confined between the outmost two rays of $L = 0$ and $P_s = 0$, i.e. the $L = 0$ rays with min and max ray indices or the $P_s = 0$ rays with min and max S/R. This is a further refinement of possible exchange reflection region than simply the 3-fold region. When exchange reflection does occur, viewed from the last row in Fig. 6.9, the relevant point on the incoming P_s curve is shifted horizontally (to conserve angular momentum) to the corresponding point on the outgoing P_s curve. If the point on the outgoing P_s curve can match a point on the incoming P_s curve which has the same S/R and exactly opposite P_s, then exchange reflection happens. Otherwise, no strict exchange reflection is possible except the relaxed "exchange reflection" due to a wide QPC as discussed above. Thus optimal exchange reflection requires strong overlap of positive (negative) part of the outgoing P_s curve and negative (positive) part of the incoming P_s curve between the two outmost $P_s = 0$ points. Fig. 6.9(c) and (e) represent points of optimal exchange reflection and seem to fit the requirement.

Tip softness plays an important role here. As shown in the fourth row in Fig. 6.9, apparently the incoming and outgoing P_s curves for the same tip position have

only the points of $P_s = 0$ (self reflection) in common, except for the extreme points representing electrons that barely touch the tip. The outgoing P_s curve is obtained by stretching horizontally (to maintain angular momentum conservation) the nonzero parts of the outer two folds of the incoming P_s curve away from the $P_s = 0$ point on them, that is, the positive P_s part to the right and the negative P_s part to the left. The amount of stretching depends on the softness of the tip. For a hard-disk tip, there is no stretching at all and the incoming and outgoing P_s curves are identical.

The last row in Fig. 6.9 shows that the incoming and outgoing P_s curves connect at their two ends. This is logical because the end points represent trajectories that barely scatter on the tip and thus have the same incoming and outgoing angle S/R.

The phase space discussion completes the study of classical cusps and their tip scan images. Many interesting results have been obtained. However, it remains to be seen how these classical results compare to their quantum counterparts. The remaining half of the chapter first builds the connection between classical and quantum cusps, and then explores the unique features of quantum cusps and their tip scan images.

6.5 Correspondence between Classical and Quantum Cusps

As discussed at the beginning of the chapter, what an extended potential does to the formation of a cusp in electron flow is tuning electron velocities to the proper distribution, for example, like the one in Eq. (6.1). Much like a thick lens can be replaced by a thin lens with the same focal length in optics, in forming classical cusps

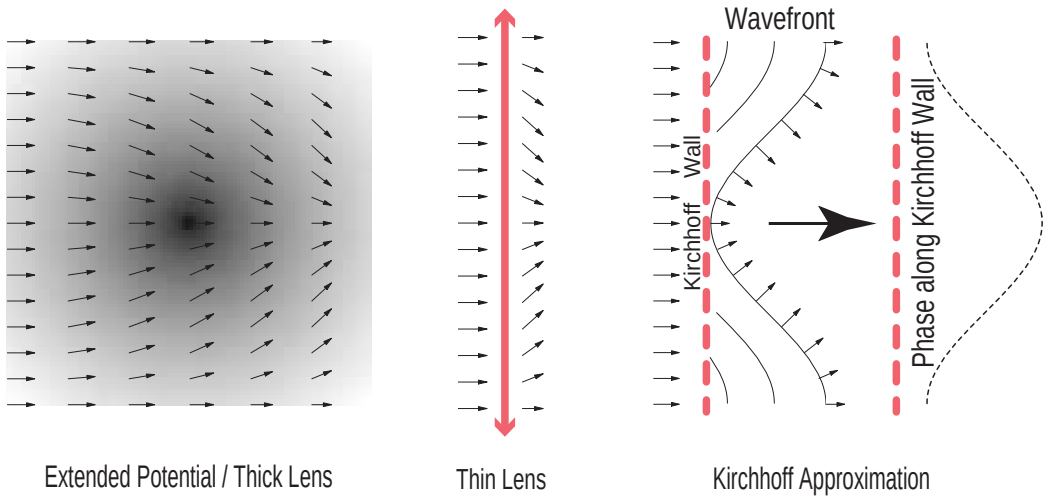

Figure 6.10: Comparison between extended potential/thick lens, thin lens, and Kirchhoff method.

an extended potential can be replaced by simply preparing electrons with the desired velocity distribution at the very beginning, as shown in Fig. 6.10. But how does the critically important classical velocity distribution translate into an equally important counterpart in forming quantum cusps? This section explores such a connection between classical and quantum cusps in detail.

There have been various studies of cusp formation in wave fields [45, 44]. According to past research in the Heller group [37], the Kirchhoff method has been able to provide a good and consistent approximation to the exact quantum wave field, and therefore becomes our method of choice in studying quantum cusps.

The Kirchhoff method is based on the idea that a wave field can be reproduced if a cross section of the field is taken and all points on the cross section act as new sources with the same phases and amplitudes as the points have had in the original wave field. The essence of the Kirchhoff method is the determination of phases and amplitudes along a chosen cross section in the wave field. To start, one must specify

what kind of wavefront is desired, which can be, as an example,

$$x = f(y) + \phi_0, \tag{6.2}$$

where $f(y)$ is a continuous function of y and ϕ_0 is an arbitrary constant. Eq. (6.2) can be rearranged into

$$\phi_0 = x - f(y), \tag{6.3}$$

which can be interpreted as a specific wavefront in the $X-Y$ space that has a constant phase. To get a wavefront for more than one phase, a further assumption about the phase dependence on $(x - f(y))$ needs be made. Typically a linear relationship is assumed, which gives the phase at every point in space:

$$\phi(x, y) = k(x - f(y)), \tag{6.4}$$

where k is a constant called wave number, which sets the phase-changing rate with spatial coordinates and is often related to wavelength λ by $k = 2\pi/\lambda$. When $f(y) = -\cos(y)$, Eq. (6.4) gives the wavefront shown in Fig. 6.10. Wavefront velocity and its y-component corresponding to Eq. (6.4) are given as

$$\begin{aligned} \vec{v} &= \vec{\nabla}\phi(x, y) = (k\hat{x} - k\frac{\partial f(y)}{\partial y}\hat{y}), \\ v_y &= \frac{\partial \phi(x, y)}{\partial y} = -k\frac{\partial f(y)}{\partial y}, \end{aligned} \tag{6.5}$$

where $\hat{x}$ and $\hat{y}$ are unit vectors.

If a cross section is taken of the wave field along the line of $x = x_0$, where x_0 is a constant, Eq. (6.4) specifies the phase along the cross section as $\phi(x_0, y) = k(x_0 - f(y))$. If the amplitude distribution in space, or at least along the cross

section, is also specified, the Kirchhoff method is equipped to produce the desired wave field.

Comparing the "thin lens" case to the Kirchhoff approximation case in Fig. 6.10, it is easy to see that the velocity of wavefront in the Kirchhoff approximation is the quantum version of the classical velocity of electron in the "thin lens" case. Releasing electrons with a desired velocity distribution corresponds to setting phases along the Kirchhoff Wall such that the wavefront velocity at every point along the Kirchhoff Wall fits the same kind of distribution of classical electron velocities. A connection between classical cusps and quantum cusps by Kirchhoff method is thus established. The following section will discuss in detail the critical parameters in generating quantum cusps by Kirchhoff method.

6.6 Quantum Cusps Generation - Dimensionless Parameters in Kirchhoff Method

The preceding section details the general procedure in generating quantum cusps by Kirchhoff method. For the rest of the chapter on quantum cusps, we use a specific case of Eq. (6.4) along the line of $x = 0$

$$\phi(0, y) = k\cos(\alpha y) = \frac{2\pi}{\lambda}\beta\cos(\alpha y), \tag{6.6}$$

where β is an arbitrary constant (added to avoid the strong coupling between phase ϕ and wavelength λ implied by the conventional wave number, i.e. λ can correspond to a phase change of $2\pi\beta$), α is a parameter that controls for fixed λ and β the valley-to-valley spacing in phase in y along the Kirchhoff Wall.

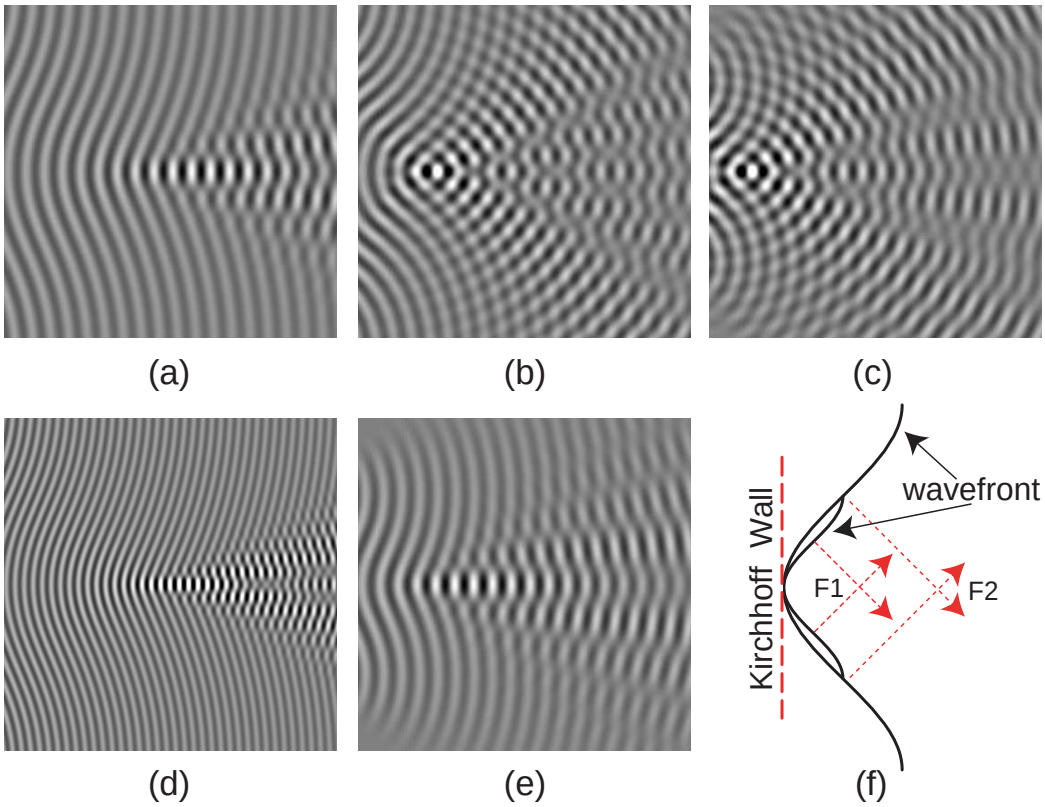

Figure 6.11: (a) Original cusp. (b) After increasing source phase peak-to-valley height by 2. (c) After reducing source phase valley-to-valley spacing by 1/2. (d) After reducing wavelength by 1/2 and increasing phase peak-to-valley height by 2. (e) After reducing source phase valley-to-valley spacing and peak-to-valley height by 1/2. (f) An illustration for (e).

If the amplitude is assumed to be constant along the Kirchhoff Wall, except for some damping on the two ends to remove any edge effect, the Kirchhoff method implemented in Eq. (6.6) takes the basic structure illustrated in Fig. 6.10 and gives rise to the various cusps studied in this and next section. Fig. 6.11(a) plots a cusp generated using Eq. (6.6) for a certain set of λ, β, and α.

Clearly Eq. (6.6) identifies three fundamental dimensionless parameters in the Kirchhoff method: the wavelength λ, the overall phase constant β (which controls peak-to-valley height in phase), and the valley-to-valley spacing parameter α. Each of the three parameters affects the resulted cusp in different manners, as shown in Fig. 6.11 and discussed next.

6.6.1 Source Phase Peak-to-Valley (PV) Height

First consider β, which controls the peak-to-valley (PV) height in source phase and therefore PV height in the generated wavefront. Increasing β leads to an increased PV height in the source phase and wavefront, and decreasing β leads to a reduced PV height in the source phase and wavefront. This proportional relationship between β and PV height in source phase is readily seen in Eq. (6.6) as β is a linear factor in source phase amplitude. Fig. 6.11(b) shows that, upon increasing β by 2, the PV height of wavefront is increased by 2 compared to the original wave field in Fig. 6.11(a). As a result of this change, the original cusp in Fig. 6.11(a) is now drawn closer to the source wall and the long string of bright nodes at cusp focus is reduced to a few bright spots, while the cusp opens wider in Fig. 6.11(b). Decreasing β in turn leads to a reduced PV height of wavefront, a cusp focus further away from the source

wall, a longer string of bright nodes at cusp focus, and a narrower cusp opening.

6.6.2 Source Phase Valley-to-Valley (VV) Spacing

Next consider α, which controls the valley-to-valley (VV) spacing in source phase and therefore the peak-to-peak (PP) spacing in the generated wavefront. Increasing α leads to a reduced VV spacing in the source phase and PP spacing in the wavefront, and decreasing α leads to an increased VV spacing in the source phase and PP spacing in the wavefront. This inversely proportional relationship between α and the VV spacing is readily seen in Eq. (6.6). Keeping αy at a constant value corresponding to a valley in Eq. (6.6), changing α one way leads to inversely proportional change in y in the opposite direction. Fig. 6.11(c) shows that, upon increasing α by 2, the PP spacing of wavefront is reduced by 1/2 compared to the original wave field in Fig. 6.11(a). As a result of this change, the original cusp in Fig. 6.11(a) is now drawn very close to the source wall, and the long string of bright nodes at a cusp focus is reduced to a few bright spots, while the cusp opens wider in Fig. 6.11(c). Decreasing α in turn leads to an increased PP spacing of wavefront, a cusp focus further away from the source wall, a longer string of bright nodes at cusp focus, and a narrower cusp opening.

6.6.3 Varying Wavelength with Fixed Wavefront

Finally consider the wavelength λ. It is expected that, keeping the wavefront (the classical "skeleton") fixed, reducing the wavelength leads to a more classical looking cusp, i.e. brighter cusp edges. However, Eq. (6.4) shows that maintaining

a fixed wavefront while adjusting the wavelength requires an inversely proportional change in the PV height of the overall source phase, which is also implied by Eq. (6.6). Intuitively one can see that filling the same space with the same wavefront shape but a smaller wavelength implies more phase fluctuation over the same spatial distance. Fig. 6.11(d) shows that, upon decreasing λ by 1/2 while maintaining a fixed wavefront, the basic wave field structure is similar to that of the original wave field in Fig. 6.11(a), except that now the cusp edges become brighter, more classical looking, which is as expected. Increasing λ with fixed wavefront in turn leads to more wave-like, less classical-looking cusps whose edges are less distinguishable from the inside.

6.6.4 Varying Source Phase VV Spacing and PV Height Jointly

It is interesting to notice from the above analysis that either α or β can adjust in the same direction the wavefront shape, cusp focus position, cusp focal nodes, and cusp opening. But do α and β have the same adjustment power? To answer this question, one can increase α and reduce β by inverse proportions to observe their interaction. From the above analysis, it is expected that if α is more powerful, the PP spacing of wavefront will decrease, the cusp focus will move toward the source wall, the string of bright focal nodes will shorten, and the cusp will open wider; if β is more powerful, the exact opposite outcomes will occur. If α and β are equally powerful, the wave field will remain the same as in Fig. 6.11(a). Fig. 6.11(e) shows the result after increasing α by 2 and reducing β by 1/2, which shows a reduced PP spacing of wavefront, the cusp focus drawn closer to the source wall, less bright focal

nodes, and a wider cusp opening. It appears α exerts more influence on the resulted wave field than β does. Fig. 6.11(f) illustrates how wavefront and cusp focus respond to increasing α by 2 and reducing β by 1/2 in Fig. 6.11(e). The cusp focus, roughly defined by the crossing of the most sloped wavefront velocities, moves from F_2 to F_1 following the adjustments. Conversely, increasing β by 2 and reducing α by 1/2 leads to an increased PP spacing of wavefront, a cusp focus further away from the source wall, more bright focal nodes, and a narrower cusp opening. A cusp focus would move from F_1 to F_2 in Fig. 6.11(f) following such adjustments.

The above analysis completes the technical study of generating quantum cusps using the Kirchhoff method. The next few sections examine how quantum cusps thus generated exemplify some basic wave physics features.

6.7 Quantum to Classical Transition at Short Wavelength

First of all, it is interesting and important to see how quantum cusps transition into classical cusps at short wavelengths, as this further verifies the connection between quantum and classical cusps established in the previous section. As already shown in Fig. 6.11(a) and (d), reducing the wavelength with fixed wavefront leads to more visible cusp edges and a more classical looking cusp. Fig. 6.12 plots a cusp with an even shorter wavelength. It can be directly compared to the classical cusp in Fig. 6.1(a). The bright cusp edges and focal region in Fig. 6.12 match the high density counterparts in Fig. 6.1(a) very well. The cusp edges in Fig. 6.12 are

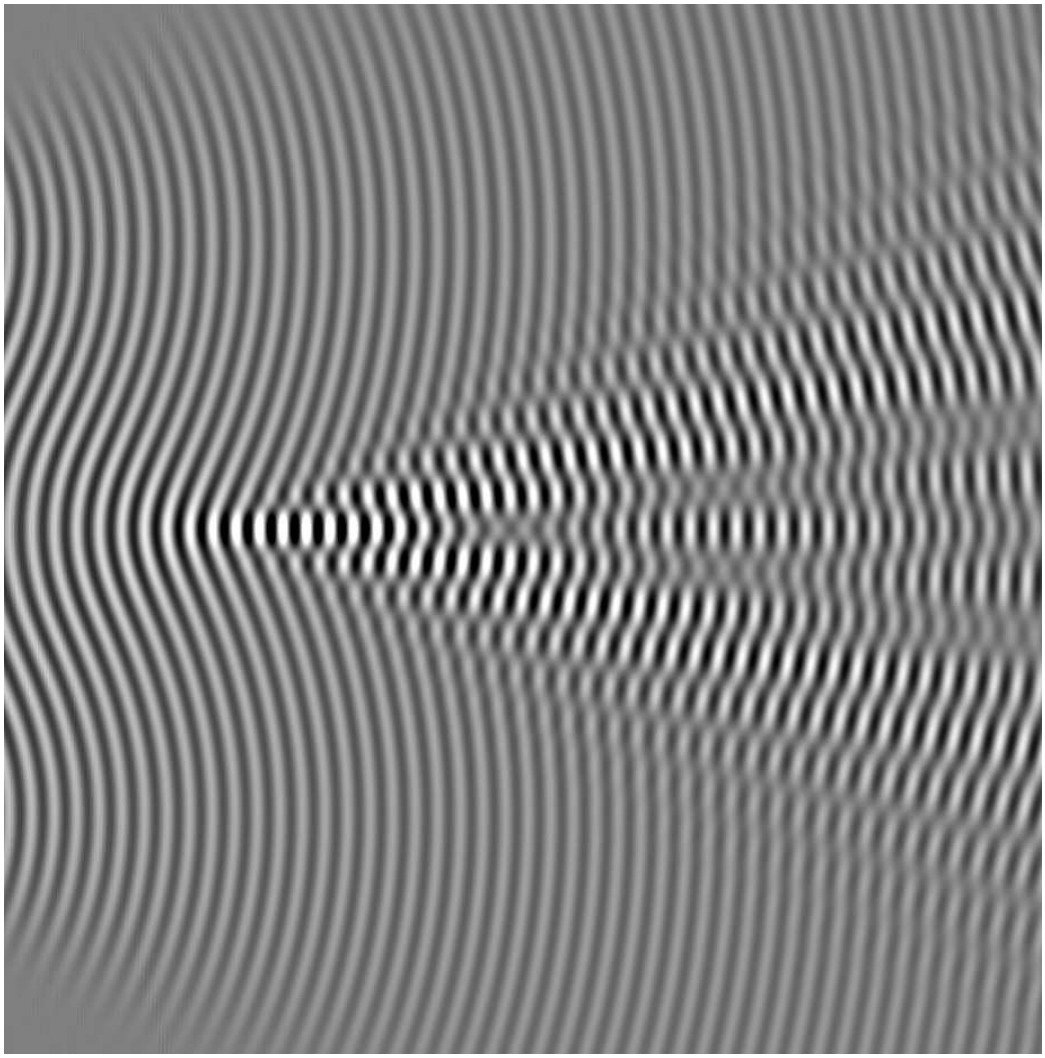

Figure 6.12: A cusp with short wavelength.

mainly constructed from many bright, nearly horizontal small lobes. The horizontal orientation of the small lobes is due to superposition of waves from different directions. Reducing the wavelength further will yield cusps that further resemble the classical cusp in Fig. 6.1(a).

6.8 Overlapping Quantum Cusps with Varying Phase Difference

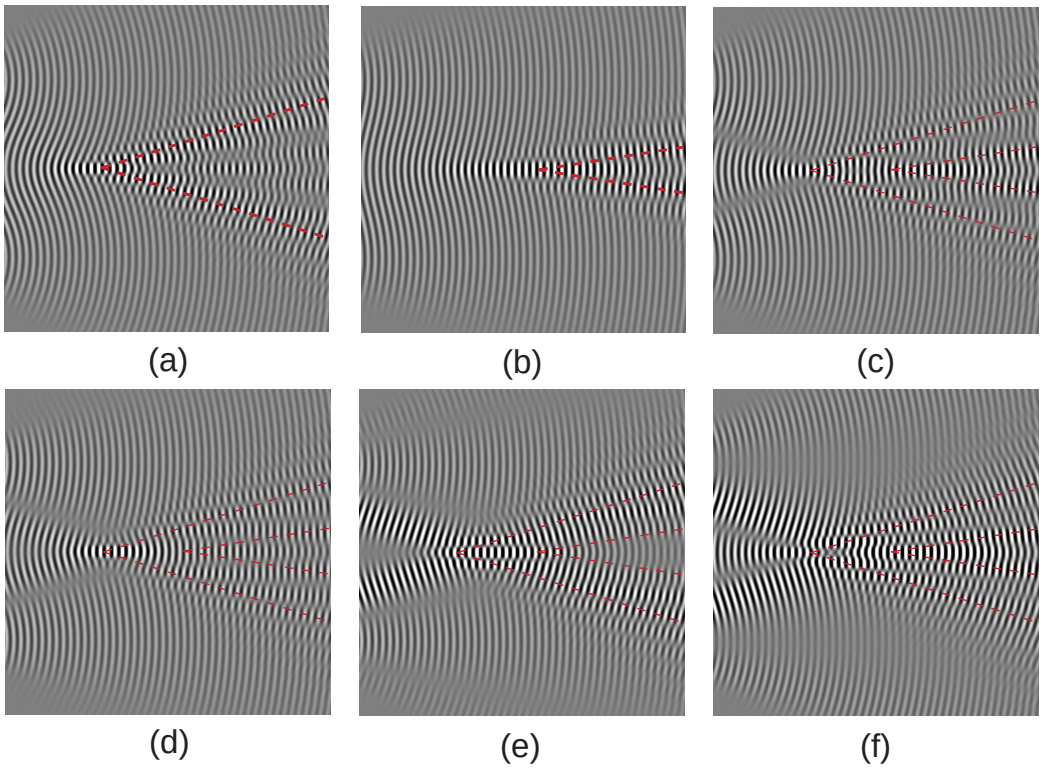

Figure 6.13: (a,b) The original two cusps. (c-f) The overall wave field after superimposing (b) to (a) with (b) lagging 0π, $1/2\pi$, π, and $3/2\pi$.

When there is more than one cusp, how do different cusps interact with each other? The simplest case is a two-cusp system. The two cusps can be aligned in

series (downstream), as investigated in this section, or in parallel, to be shown in the following section.

The series alignment of cusps has one additional advantage over the parallel alignment. It's easier to observe interference between two different cusp foci along the propagation direction of wave. In particular, when the initial phase of a cusp field is adjusted, the entire wave field adjusts its phase along the wave propagation direction, represented by the fringes shifting down or upstream in the field, making cusps in a series alignment more sensitive to changes in the initial phase.

Fig. 6.13 (a) and (b) show two individual cusps with slightly different focal lengths from the same source wall. Fig. 6.13 (c-f) show the superposition of the two cusps when the smaller cusp lags in initial phase by 0, $\pi/2$, π, and $3\pi/2$ (strictly speaking, this is the phase difference between the middle points of the source walls in Fig. 6.13 (a) and (b)), respectively.

When there is no initial phase difference, Fig. 6.13 (c) shows two largely separated cusps with separate foci. When the initial phase difference is $\pi/2$, Fig. 6.13 (d) shows that the two cusps merge together nicely with only one bright focus at the original large cusp's focal point, and the cusp edges are indistinguishable from the cusp inside. When the initial phase difference is π, Fig. 6.13 (e) shows a striking phenomenon. The small cusp disappears, along with the region in front of the large cusp's focus. The region between the edges of the two cusps is significantly enhanced, so are the two branches reaching out from the source wall to the focus of the large cusp. The drastic depletion and enhancement in different parts of the field can be attributed to the exact π difference in phase, which often leads to perfect constructive or perfect

destructive superposition. When the initial phase difference is $3\pi/2$, Fig. 6.13 (f) shows that the two cusps start to separate again, and their edges are distinguishable from their inside.

In all cases, the interference has always led to angular nodes in the common pre-focus region of the two cusps, i.e. the region between the source wall and the large cusp's focus. The two original cusp foci always correspond to some special points in the combined wave field.

6.9 Small, Medium, and Large Tip Scan Images of Quantum Cusps

Finally, tip scan images of quantum cusps must be examined. In particular, the impact of tip size on tip scan images must be studied carefully. The experimental tip is pretty large, with a radius roughly 4 times the Fermi wavelength. In this section, a single-cusp and a double-cusp system are separately probed by a small, medium, and large tip in turn. The different tip scan images are compared among themselves and with the original cusps.

Fig. 6.14 (a-c) show in turn a single-cusp wave field, its density distribution, and its total flux distribution (nearly the same as the distribution of horizontal component of flux, due to weak vertical flux). Fig. 6.14 (d-f) show the tip scan images of the cusp in Fig. 6.14 (a) by a tip of radius equal to 1/200, 1/2, and 4 wavelengths, respectively. The blank slab on the left of the tip scan images is due to offset of one tip radius from the source wall when imaging the same area using different tips. Their width

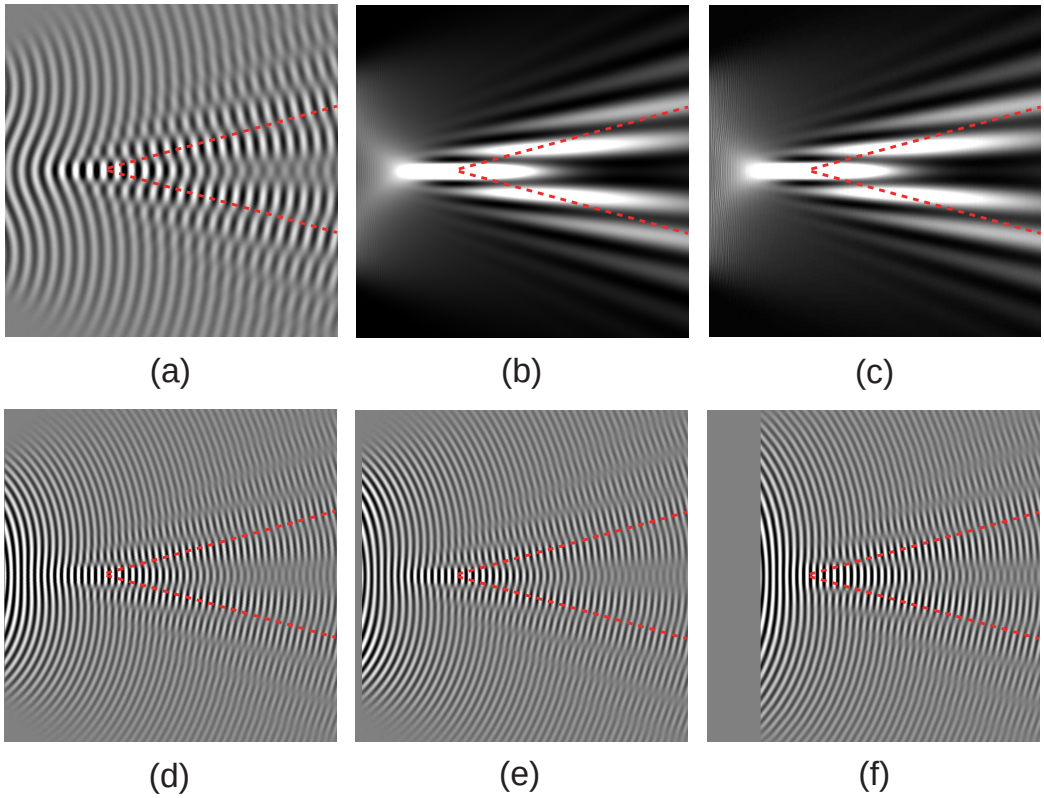

Figure 6.14: (a) A single-cusp wave field. (b) The wave field density distribution. (c) The total flux distribution. (d-f) Tip scan images with tip radius equal to 1/200, 1/2, and 4 wavelengths.

equals the corresponding tip radius.

The tip scan images are obtained by the Kirchhoff method as well. The tip surface is viewed as consisting of a number of points that receive incident waves and then send the waves out as new sources. Signal is collected at an observation point, the equivalent of a QPC in the experiments, that is 50 wavelengths to the left of the middle point on the Kirchhoff wall. Only the illuminated half of tip surface contributes to the QPC. When damping is used to discount contributions from parts of the tip that are on the top and bottom sides of the illuminated surface and therefore farther away from the QPC, the tip scan images remain mostly the same as without the damping, except that some bright angular nodes are enhanced slightly. Here we take tip scan images without any damping.

Comparing Fig. 6.14 (d-f) with (a), the most striking change is that the fringe spacing in Fig. 6.14 (d-f) is only half the fringe spacing in Fig. 6.14 (a). This is because wave travels from the source wall to the tip and then from the tip to the QPC behind the source wall. When the tip shifts a half wavelength up or downstream, wave travels one wavelength less or more to reach the QPC (this has been discussed in Chapter 4 on conductance fringe patterns in 2DEGs).

It is easy to see that the presence of tip does make a tip scan image look different from the original wave field, even when the tip is very small and close to being a point scatterer as shown in Fig. 6.14 (d). When the tip radius increases, the cusp tends to be deformed more and shifted to the right accordingly in Fig. 6.14 (e) and (f). But still, the changes are not drastic and the size, shape, and location of the cusp remain roughly the same as those of the original in Fig. 6.14 (a). This testifies to the validity

of detecting cusps using a large SPM tip in the experiments.

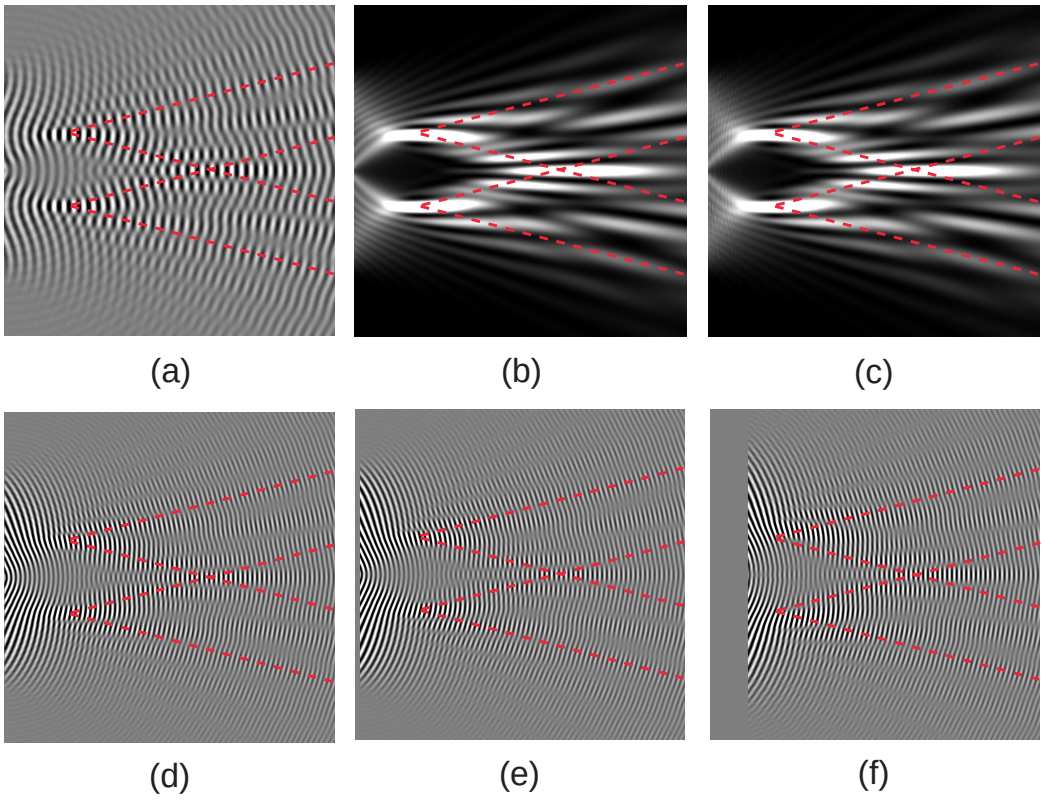

Figure 6.15: (a) A double-cusp wave field. (b) The wave field density distribution. (c) The total flux distribution. (d-f) Tip scan images with tip radius equal to 1/200, 1/2, and 4 wavelengths.

Fig. 6.15 presents results for a different system, two cusps aligned in parallel. Fig. 6.15 (a-c) show in turn the original wave field, its density distribution, and its total flux distribution. Fig. 6.15 (d-f) show the tip scan images of the cusps in Fig. 6.15 (a) by a tip of radius equal to 1/200, 1/2, and 4 wavelengths, respectively. The blank slab on the left of the tip scan images is again due to offset by tip radius. Signal is collected at the same point as in Fig. 6.14.

Compared with Fig. 6.15 (a), Fig. 6.15 (d-f) show the reduction in fringe spacing as well. The presence of tip still makes a difference even in Fig. 6.15 (d). When

tip radius increases, the cusps tend to be deformed more and shifted to the right accordingly in Fig. 6.15 (e) and (f). But, as in the single-cusp case, the changes are not drastic and the size, shape, and location of the cusps remain roughly the same as those of the originals in Fig. 6.15 (a), consistent with detecting cusps using a large SPM tip in the experiments.

6.10 Matching Quantum Cusps in 2DEGs

With both classical and quantum cusps and their corresponding tip scan images thoroughly studied above, we are ready to compare the calculated quantum cusps by the Kirchhoff method and their tip scan images to the cusps seen in the experiments. Among the many similar looking cusps observed in the electron flow imaging experiments [77, 75, 76], Fig. 6.16(a) is the clearest and most cited one.

In Fig. 6.16(b), a cusp is generated by the Kirchhoff method and compared to the experimental cusp. The width, spacing, and total number of the bright nodes are nicely reproduced. Also the angular node right under the string of bright focal nodes is reproduced in simulation. The absence of a symmetric angular node above the bright nodes in Fig. 6.16(a) is due to another cusp starting to develop above the cusp shown here and thus complicating the region above the bright nodes.

In Fig. 6.16(c) and (d), the tip scan image of the cusp in Fig. 6.16(b) is generated by Kirchhoff method and compared to the experimental cusp, which is also a tip scan image. The width, spacing, and total number of the bright nodes are matched nicely. The angular node right under the string of bright focal nodes is also reproduced in tip scan simulation.

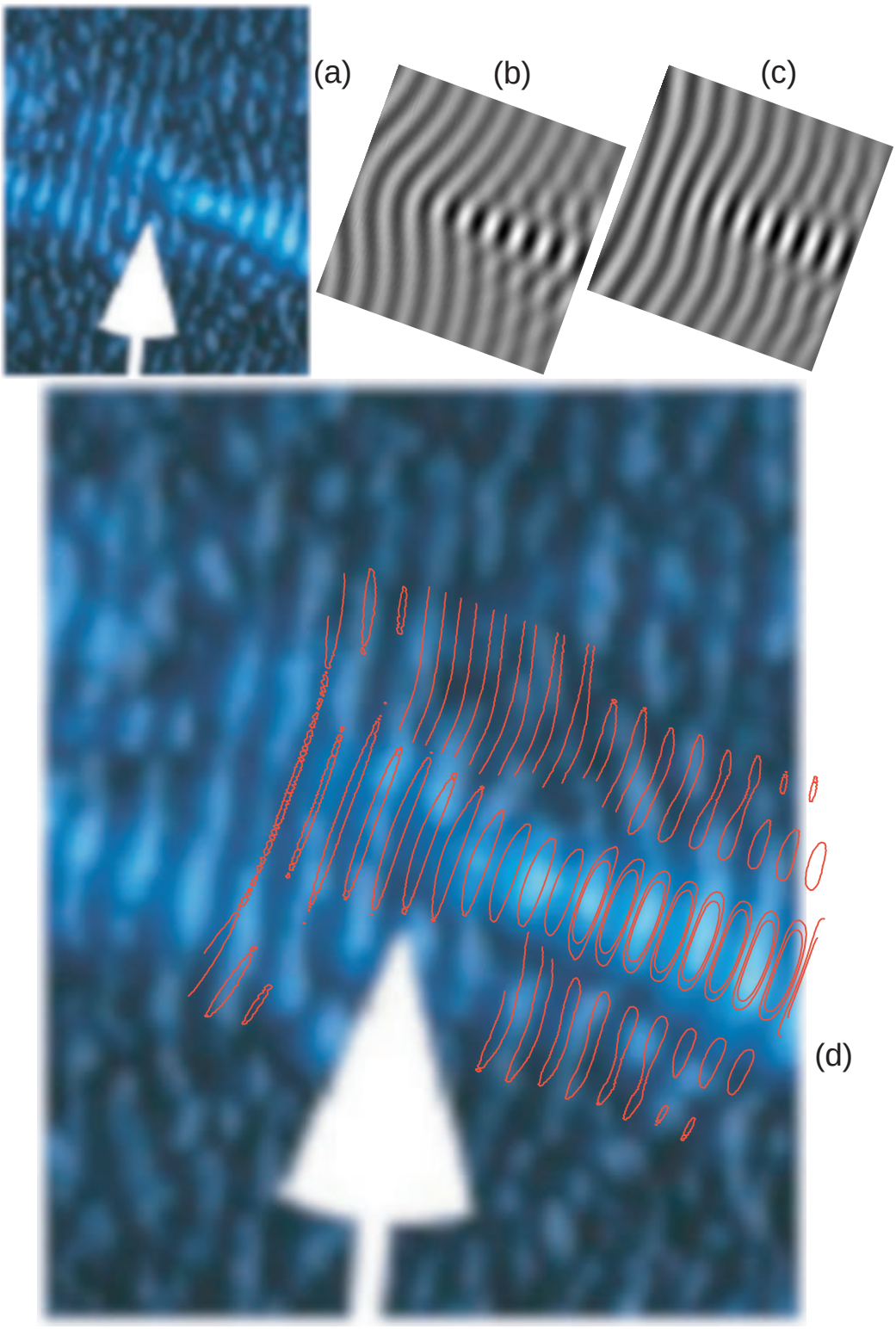

Figure 6.16: (a) A cusp observed in experiments. (b) A cusp produced by Kirchhoff method. (c) Tip scan image of (b). (d) Matching (c) to (a).

Comparing Fig. 6.16(b) and (c), one finds that the wavefront before cusp focus is flattened in tip scan compared to the original wave field. The flat wavefront before focus matches Fig. 6.16(a) very well. Similar flattening effect also exists in Figs. 6.14 and 6.15 in the preceding section, thus seems to be a tip scan effect. The main reason is that, in tip scan, wave travels from Kirchhoff wall to tip and then from tip to a distant QPC behind the Kirchhoff wall. The overall wavefront shape depends on the total distance from the Kirchhoff wall to the tip and then to the QPC, and becomes flattened due to the additional distance from the tip to the QPC.

6.11 Conclusion

In conclusion, we have presented a comprehensive and thorough study of classical and quantum cusps and their tip scan images. The ray patterns and phase space representations of different classical cusps were given and compared. The tip scan imaging mechanism of classical cusps is found to consist of two kinds of reflections, self reflection and exchange reflection. The QPC width and the tip size play important roles in forming the three-layer structure of cusp edges in tip scan images, while the softness of tip gives an interesting stretching effect in the phase space representation of tip scattering. A connection between classical and quantum cusps was then established through the correspondence between classical velocity and wavefront velocity. The Kirchhoff method of generating quantum cusps was examined in detail and the three dimensionless parameters were identified. The transition from a quantum cusp to a classical cusp at the short wavelength limit was demonstrated. Additionally, two cusps aligned in series were shown to interfere differently at different initial phase

differences. Finally, tip scan images of quantum cusps were shown to reasonably reproduce the original cusps even at large tip limit. The typical cusp observed in experiments was matched nicely by simulation based on the Kirchhoff method.

Bibliography

[1] J.L. Agudin. *Phys. Rev.*, **171(5)**:1385, 1968.

[2] A. Ahlswede, P. Weitz, J. Weis, K. von Klitzing, and K. Eberl. *Physica B*, **298**:562, 2001.

[3] V.I. Arnold. *Russ. Math. Surveys*, **30(1)**, 1975.

[4] M. Bajcsy, A.S. Zibrov, and M.D. Lukin. *Nature*, **426**:638, 2003.

[5] G. Bal and L. Ryzhik. Preprint `arXiv:nlin.CD/0201011`.

[6] C. Bardos and M. Fink. *Asymptotic Analysis*, **29**:157, 2002.

[7] C.W.J. Beenakker. *Phys. Rev. B*, **44**:1646–1656, 1991.

[8] M.V. Berry. *Proc. Phys. Soc. Lond.*, **89**:479, 1966.

[9] P. Blomgren, G. Papanicolaou, and H. Zhao. *J. Acoust. Soc. Am.*, **111(1)**:230, 2002.

[10] G.F. Carrier. *J. Fluid Mech.*, **24**:641, 1966.

[11] S. Chan and E.J. Heller. *Phys. Rev. Lett.*, **78**:2570, 1997.

[12] Stella Chan. PhD thesis, Harvard University, Cambridge, MA, 1997.

[13] D. Chauvat, O. Emile, F. Bretenaker, and A. Le Floch. *Phys. Rev. Lett.*, **84(1)**:71, 2000.

[14] M.L. Cowan, I.P. Jones, J.H. Page, and D.A. Weitz. *Phys. Rev. E*, **65**:066605, 2002.

[15] R. Crook, C.G. Smith, M.Y. Simmons, and D.A. Ritchie. *Phys. Rev. B*, **62**:51745178, 2000.

[16] R. Crook and *et al. J. Phys. Cond. Matter*, **12**:L167–L172, 2000.

[17] S. Datta. *Electronic Transport in Mesoscopic Systems*. Cambridge University Press, 1995.

[18] A. Derode. Private communication.

[19] A. Derode, A. Tourin, and M. Fink. *Phys. Rev. E*, **64**:036605, 2001.

[20] A. Derode, A. Tourin, and M. Fink. *Phys. Rev. E*, **64**:036606, 2001.

[21] A. Derode, A. Tourin, and M. Fink. *Ultrasonics*, **40**:275–280, 2002.

[22] M.A. Eriksson and *et al. Appl. Phys. Lett.*, **69**:671673, 1996.

[23] P. Fallahi, A.C. Bleszynski, R.M. Westervelt, J. Huang, J.D. Walls, E.J. Heller, M. Hanson, and A.C. Gossard. *Nano Letters*, **5**:223, 2005.

[24] J.J. Faran. *J. Acoust. Soc. Am.*, **23(4)**:405, 1951.

[25] S.C. Feng and P.A. Lee. *Science*, **251**:633, 1991.

[26] G. Fiete and E.J. Heller. *Rev. Mod. Phys.*, **75**:933, 2003.

[27] M. Fink. *Physics Today*, **50**:34, 1997.

[28] M. Fink and J. Rosny. *Nonlinearity*, **15**:R1–R18, 2002.

[29] D.T. Finkbeiner. *Introduction to matrices and linear transformations*. Freeman, San Francisco, 1966.

[30] G. Finkelstein, P.I. Glicofridis, R.C. Ashoori, and M. Shayegan. *Science*, **289**:90, 2000.

[31] M. Froissart, M.L. Goldberger, and K.M. Watson. *Phys. Rev.*, **131(6)**:2820, 1963.

[32] D.J. Griffiths. *Introduction to Quantum Mechanics*. Prentice Hall, 1995.

[33] L. Gurevich, L. Canali, and L.P. Kouwenhoven. *Appl. Phys. Lett.*, **76**:384386, 2000.

[34] A.J. Heinrich, C.P. Lutz, J.A. Gupta, and D.M. Eigler. *Science*, **298**:1381, 2002.

[35] E.J. Heller. *Phys. Rev. Lett.*, **77**:4122, 1996.

[36] E.J. Heller, K.E. Aidala, B.J. Leroy, A.C. Bleszynski, A. Kalben, R.M. Westervelt, K.D. Maranowski, and A.C. Gossard. *Nano Letters*, **5(7)**:1285, 2005.

[37] Jesse Hersch. PhD thesis, Harvard University, Cambridge, MA, 1999.

[38] J. Huang, S. Tomsovic, and E.J. Heller. Unpublished.

[39] J. Huang, J.D. Walls, and E.J. Heller. Unpublished.

[40] T. Ihn, J. Rychen, T. Vancura, K. Ensslin, W. Wegscheider, and M. Bichler. *Physica E*, **13**:671–674, 2002.

[41] T. Jamneala, V. Madhavan, W. Chen, and M.F. Crommie. *Phys. Rev. B*, **61**:9990–9993, 2000.

[42] E. Kerbrat, C. Prada, D. Cassereau, and M. Fink. *IEEE Trans. Ultrason., Ferr., and Freq. Contr.*, **49(8)**:1103–1113, 2002.

[43] A.A. Khan. *Digital Signal Processing Fundamentals*. Da Vinci Engineering Press, 2005.

[44] N.P. Kirk, J.N.L. Connor, P.R. Curtis, and C.A. Hobbs. *J. Phys. A: Math. Gen.*, **33**:4797–4808, 2000.

[45] N.P. Kirk, J.N.L. Connor, and C.A. Hobbs. *Comput. Phys. Commun.*, **132**:142–165, 2000.

[46] L.P. Kouwenhoven, D.G. Austing, and S. Tarucha. *Rep. Prog. Phys.*, **64**:701, 2001.

[47] L.P. Kouwenhoven, C.M. Marcus, P. McEuen, S. Tarucha, R.M. Westervelt, and N. Wingreen. Electron transport in quantum dots. In L.L. Sohn, L.P. Kouwenhoven, and G. Schon, editors, *Mesoscopic Electron Transport NATO ASI series E 345*, pages 105–214, Dordrecht, 1997. Kluwer.

[48] B.J. LeRoy, A.C. Bleszynski, K.E. Aidala, R.M. Westervelt, A. Kalben, E.J. Heller, S.E.J. Shaw, K.D. Maranowski, and A.C. Gossard. *Phys. Rev. Lett.*, **94**:126801, 2005.

[49] B.J. LeRoy, M.A. Topinka, R.M. Westervelt, and K.D. Maranowski. *Appl. Phys. Lett.*, **80**:4431, 2002.

[50] S. Li and E.J. Heller. *Phys. Rev. A*, **67**:032712, 2003.

[51] H.C. Manoharan, C.P. Lutz, and D.M. Eigler. *Nature*, **403**:512–515, 2000.

[52] J.D. Maynard. *Rev. Mod. Phys.*, **73**:April, 2001.

[53] K.L. McCormick and *et al. Phys. Rev. B*, **59**:4654–4657, 1999.

[54] P. Morse. *Vibration and Sound*. McGraw-Hill, NY, 1948.

[55] P.M. Morse and H. Feshbach. *Methods of Theoretical Physics*. McGraw-Hill, NY, 1953.

[56] E.E. Narimanov, A.D. Stone, and G.S. Boebinger. *Phys. Rev. Lett.*, **80**:4024, 1998.

[57] G. Papanicolaou, L. Ryzhik, and K. Solna. Preprint `arXiv:cond-mat/0206095`.

[58] H.M. Pastawski, P.R. Levstein, and G. Usaj. *Phys. Rev. Lett.*, **75**:4310, 1995.

[59] P.K. Pathria. *Statistical Mechanics (2nd Ed.)*. Butterworth Heinemann, 1996. Appendix E.

[60] T. Pearcy. *Philos. Mag.*, **37**:311, 1946.

[61] W.H. Press, S.A. Teukolsky, W.T. Vetterling, and B.P. Flannery. *Numerical Recipes in C++ (2nd Ed.)*. Cambridge University Press, 2002.

[62] L.S. Rodberg and R.M. Thaler. *Introduction to the quantum theory of scattering*. Academic, NY, 1967.

[63] T.A. Roth. *J. Appl. Phys.*, **45(11)**:4771, 1974.

[64] D. Rugar, R. Budakian, H.J. Mamin, and B.W. Chu. *Nature*, **430**:329–332, 2004.

[65] R. Seydel. *Tools for Computational Finance*. Springer, 2002.

[66] S.E.J. Shaw. PhD thesis, Harvard University, Cambridge, MA, 2002.

[67] S.E.J. Shaw, R. Fleischmann, and E.J. Heller. Preprint `cond-mat/0105354`.

[68] P. Sheng. *Introduction to wave scattering, localization, and mesoscopic phenomena*. Academic, San Diego, 1995.

[69] F.T. Smith. *Phys. Rev.*, **118(1)**:349, 1960.

[70] R. Snieder and J.A. Scales. *Phys. Rev. E*, **58**:5668, 1998.

[71] M. Tanter, J.F. Aubry, J. Gerber, J.L. Thomas, and M. Fink. *J. Acoust. Soc. Am.*, **110(1)**:37–47, 2001.

[72] S. Tarucha, D.G. Austing, T. Honda, R.J. van der Hage, and L.P. Kouwenhoven. *Phys. Rev. Lett.*, **77**:3613, 1996.

[73] R. Thom. *Structural Stability and Morphogenesis*. Benjamin, Reading, MA, 1975.

[74] M.A. Topinka. PhD thesis, Harvard University, Cambridge, MA, 2002.

[75] M.A. Topinka, B.J. LeRoy, S.E.J. Shaw, E.J. Heller, R.M. Westervelt, K.D. Maranowski, and A.C. Gossard. *Science*, **289**:2323–2326, 2000.

[76] M.A. Topinka, B.J. LeRoy, R.M. Westervelt, S.E.J. Shaw, R. Fleischmann, E.J. Heller, K.D. Maranowski, and A.C. Gossard. *Nature*, **410**:183–186, 2001.

[77] M.A. Topinka, R.M. Westervelt, and E.J. Heller. *Physics Today*, **December**:47–52, 2003.

[78] A. Tourin, A. Derode, and M. Fink. *Phys. Rev. Lett.*, **87**:274301, 2001.

[79] A. Tourin, A. Derode, A. Peyre, and M. Fink. *J. Acoust. Soc. Am.*, **108(2)**:503, 2000.

[80] A. Tourin, M. Fink, and A. Derode. *Waves Random Media*, **10**:R31–R60, 2000.

[81] J.Y. Vaishnav, A. Itsara, and E.J. Heller. *Phys. Rev. B*, **73**:115331, 2006.

[82] M.C.W. van Rossum and Th.M. Nieuwenhuizen. *Rev. Mod. Phys.*, **71**, 1999.

[83] J. Vanicek and E.J. Heller. *Phys. Rev. E*, **64**:026215, 2001.

[84] J. Vanicek and E.J. Heller. *Phys. Rev. E*, **67**:016211, 2003.

[85] E.P. Wigner. *Phys. Rev.*, **98(1)**:145, 1955.

[86] L. Wirtz, J.Z. Tang, and J. Burgdorfer. *Phys. Rev. B*, **56**:7589, 1997.

[87] M.T. Woodside and P.L. McEuen. *Science*, **296**:1098–1101, 2002.

[88] A. Yacoby, H.F. Hess, T.A. Fulton, L.N. Pfeiffer, and K.W. West. *Solid State Commun.*, **111**:1, 1999.

[89] M.J. Yoo and *et al. Science*, **276**:579582, 1997.

[90] S. Zhang, B.H. Meier, and R.R. Ernst. *Phys. Rev. Lett.*, **69**:2149, 1992.

[91] N.B. Zhitenev, T.A. Fulton, A. Yacoby, H.F. Hess, L.N. Pfeiffer, and K.W. West. *Nature*, **404**:473–476, 2000.

Appendix A

Asymptotic Expression for 2D Flux

To discuss fringe patterns and make use of Eq. (4.18), a simplified analytical expression of electron flux $\vec{J}(\vec{r},E)$ is desired. In this part let us establish a simple connection between $\vec{J}(\vec{r},E)$ and total reflected wave $\Psi(\vec{r},E)$ at the QPC. Given $\Psi(\vec{r},E)$ of energy E at location $\vec{r}$, the flux through $\vec{r}$ is defined as

$$\vec{J}(\vec{r},E)=\frac{\hbar}{m_e}Im\left(\Psi^*\vec{\nabla}\Psi\right). \tag{A.1}$$

For $k|\Delta\vec{r}| >> 1$, there exist asymptotic forms of the $2D$ Green's function for free propagation from a point scatterer at $\vec{r}_m$ toward the QPC at $\vec{r}_s$: $G_0(\vec{r}_s,\vec{r}_m;E) = -i\frac{m_e}{2\hbar^2}H_0^{(1)}\left(k\left|\vec{r}_s-\vec{r}_m\right|\right) \approx -i\frac{m_e}{2\hbar^2}\sqrt{\frac{2}{\pi k|\vec{r}_s-\vec{r}_m|}}e^{i(k|\vec{r}_s-\vec{r}_m|-\frac{\pi}{4})}$, and its derivative: $\frac{\partial G_0(\vec{r}_s,\vec{r}_m;E)}{\partial|\vec{r}_s-\vec{r}_m|} = i\frac{km_e}{2\hbar^2}H_1^{(1)}\left(k\left|\vec{r}_s-\vec{r}_m\right|\right) \approx i\frac{km_e}{2\hbar^2}\sqrt{\frac{2}{\pi k|\vec{r}_s-\vec{r}_m|}}e^{i(k|\vec{r}_s-\vec{r}_m|-\frac{3\pi}{4})} \approx ikG_0(\vec{r}_s,\vec{r}_m;E)$. In a general s-wave scattering system, $\Psi(\vec{r}_s,E)$ and $\vec{\nabla}_{\vec{r}_s}\Psi(\vec{r}_s,E)$ are

$$\begin{aligned}\Psi(\vec{r}_s,E) &= \sum_{m=1}^{N}\epsilon_m(E)\;G_0(\vec{r}_s,\vec{r}_m;E)\Psi_m(\vec{r}_m,E)\\ &= -i\frac{m_e}{2\hbar^2}\sum_{m=1}^{N}\epsilon_m(E)\;H_0^{(1)}\left(k\left|\vec{r}_s-\vec{r}_m\right|\right)\Psi_m(\vec{r}_m,E)\;,\end{aligned}$$

$$\begin{aligned}\vec{\nabla}_{\vec{r}_s}\Psi(\vec{r}_s,E) &= \sum_{m=1}^{N}\epsilon_m(E)\ \frac{\partial G_0(\vec{r}_s,\vec{r}_m;E)}{\partial\left|\vec{r}_s-\vec{r}_m\right|}\Psi_m(\vec{r}_m,E)\hat{r}_{s,m}\\ &\approx ik\sum_{m=1}^{N}\epsilon_m(E)\ G_0(\vec{r}_s,\vec{r}_m;E)\Psi_m(\vec{r}_m,E)\hat{r}_{s,m}\ ,\end{aligned}$$

where $\hat{r}_{s,m}$ is the unit vector pointing from $\vec{r}_m$ toward $\vec{r}_s$, N is the total number of scatterers (impurities plus a tip), $\epsilon_m(E)$ and $\Psi_m(\vec{r}_m,E)$ are the scattering strength of and the total incident wave on the m^{th} scatterer. Obviously the reflected flux through the QPC at $\vec{r}_s$ includes flux along every orientation originating from the scatterers. Here, of interest is the reflected total flux through $\vec{r}_s$ along the negative Y axis as shown in Fig. 4.1(a),

$$\begin{aligned}J(\vec{r}_s,E) &= \vec{J}(\vec{r}_s,E)*(-\hat{y}) = \frac{\hbar}{m_e}Im\left(\Psi^*\vec{\nabla}\Psi*(-\hat{y})\right)\\ &\approx \frac{\hbar}{m_e}Im\left(\left(\sum_{m=1}^{N}\epsilon_m^*G_{s,m}^*\Psi_m^*\right)\left(ik\sum_{n=1}^{N}\epsilon_nG_{s,n}\Psi_n\left(\hat{r}_{s,n}*(-\hat{y})\right)\right)\right)\\ &= \frac{\hbar k}{m_e}Re\left(\left(\sum_{m=1}^{N}\epsilon_m^*G_{s,m}^*\Psi_m^*\right)\left(\sum_{n=1}^{N}\epsilon_nG_{s,n}\Psi_n\cos\theta_n\right)\right),\end{aligned}$$

where θ_n is the angle that $\hat{r}_{s,n}$ makes with the negative Y axis. For the scattering system studied in this paper, all impurities are located close to each other and have similar θ_n. The SPM tip probes only local region around the impurities and therefore has a similar θ_n as well. For simplicity, let's assume $\theta_n \approx \theta_0$ for all impurities and the SPM tip, where θ_0 is the angle that the center of a group of impurities makes with the negative Y axis as in Fig. 4.1(a). Now the flux can be simplified into

$$J(\vec{r}_s,E) \approx \frac{\hbar k\cos\theta_0}{m_e}Re\left(\left(\sum_{m=1}^{N}\epsilon_m^*G_{s,m}^*\Psi_m^*\right)\left(\sum_{n=1}^{N}\epsilon_nG_{s,n}\Psi_n\right)\right) = \frac{\hbar k\cos\theta_0}{m_e}|\Psi|^2. \quad \text{(A.2)}$$

Eq. (A.2) establishes a proportionality between the reflected flux through $\vec{r}_s$ and the wave density at the same point. The approximation $\theta_n \approx \theta_0$ works well compared to exact numerical simulation in Chapter 4.

Appendix B

2D Scattering Strength

In Eq. (A.2), both the flux and wave density contain scattering strength ϵ_m of various scatterers. It is necessary to obtain an explicit form of the scattering strength for numerical computation in Chapter 4. For the $2D$ scattering strength $\epsilon_m(E)$, there is

$$\epsilon_m(E) = \frac{4i\hbar^2}{m_e}\left(e^{2i\delta_m(E)} - 1\right) = \frac{-8\hbar^2}{m_e}\sin\left(\delta_m(E)\right)e^{i\delta_m(E)} \ , \tag{B.1}$$

where $\delta_m(E)$ is the phase shift in scattering. For a $2D$ circular potential barrier or well with potential energy U ($U < E$) for $r < a$ and 0 for $r > a$, the s-wave phase shift is given by [62]

$$\tan\left(\delta(k)\right) = \frac{\kappa J_0(ka)J_0'(\kappa a) - kJ_0'(ka)J_0(\kappa a)}{\kappa Y_0(ka)J_0'(\kappa a) - kY_0'(ka)J_0(\kappa a)} \ , \tag{B.2}$$

where $J_0(x)$ and $Y_0(x)$ are the zeroth order Bessel and Neumann functions, respectively, $k = \sqrt{\frac{2m_eE}{\hbar^2}}$ and $\kappa = \sqrt{\frac{2m_e(E-U)}{\hbar^2}}$. Let us assume there are two kinds of scatterers in our example. One is an SPM tip with a large radius a_t , which is comparable to the Fermi wavelength but still makes s-wave scattering valid, and the others are identical

impurities with same radius a_1 that is much smaller than the wavelength. To have significant reflection of electron flux, assume further that both create for an electron a potential energy $U \approx E_F$ such that $\kappa \to 0$ and therefore $J_0'(\kappa a) \to 0$. Now the phase shift simplifies to

$$\begin{aligned} \tan\left(\delta(k)\right) = \frac{J_0'(ka)}{Y_0'(ka)} &\approx \frac{\pi k^2 a_1^2}{4} \; for \; impurities; \\ &\approx \cot\left(ka_t + \frac{\pi}{4}\right) \; for \; the \; SPM \; tip, \\ \delta(k) &\approx \frac{\pi k^2 a_1^2}{4} \; for \; impurities; \\ &\approx (n + \frac{1}{4})\pi - ka_t \; for \; the \; SPM \; tip \; , \\ &\quad (n = 0, \pm 1, \pm 2, ...) \; , \end{aligned} \tag{B.3}$$

where approximations at small and large x for $J_0(x)$ and $Y_0(x)$ have been used. Phase shift for a small potential well can be derived similarly and shown to be small as well.