Topics in Quantum Chaos and Thermoelectricity

A thesis presented
by

William Edward Bies

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
October 2000

Abstract

Topics in Quantum Chaos and Thermoelectricity

William Edward Bies

Thesis Advisors: Profs. E. J. Heller and H. Ehrenreich

This thesis consists of two parts. Part I is concerned with quantum chaos in two model systems in two dimensions, the stadium billiard and a chaotic double-well potential. We study the localization properties of eigenfunctions in the stadium billiard, and conclude that there is more localization of the local density of states as a function of position in phase space than can be accounted for on the basis of random matrix theory. A part of this can be attributed to scars, but most of the excess localization is found to be due to symmetry effects originating in the parity and time-reversal symmetries. As for the double-well potentials, we study the connection between scarring of eigenfunctions and tunneling through the barrier, as reflected in the level splittings. The level splittings oscillate as a function of energy, with period h in units of action, as expected if scarring plays a role in determining the size of the splitting. Indeed, the size of the splitting is directely correlated with the strength of scarring. Our results are interpreted within the framework of semiclassical theory.

In Part II we study thermoelectric effects in anisotropic materials. As a direct consequence of anisotropy we find that there will be induced electric fields and thermal gradients in order to ensure steady-state current flow through a device made of such materials. It is shown that the transport coefficients must be replaced by effective transport coefficients, which are reduced in magnitude because of the induced fields. The thermoelectric figure of merit can be computed in terms of the effective transport coefficients. We prove an upper bound on the figure of merit and show that, for an isotropic lattice thermal conductivity, it is maximized for the sample orientation in which current flows along the direction of greatest electrical conductivity. These ideas are applied to Bi_2Te_3 and $HgTe/Hg_{1-x}Cd_xTe$ superlattices. Finally, we investigate the experimental reduction in thermal conductivity in the growth direction of GaAs/AlAs superlattices with a realistic model of the interatomic forces and find that it can be explained as a combination of phonon dispersion effects and interface scattering.

Contents

List of Figures

List of Tables

Acknowledgments

I would like, first of all, to thank Profs. E. J. Heller and H. Ehrenreich, without whose support and encouragement this dissertation would not have come to completion. Also, I am happy to thank Drs. Lev Kaplan, Michael Haggerty, R. J. Radtke and Erich Runge, with whom the research presented here was done in collaboration.

The members of the Heller group were always available for discussions of physics and for help with computers; in particular, Alex Barnett, Greg Fiete, Ragnar Fleischman, Areez Mody, Adam Lupu-Sax and Scot Shaw should be mentioned.

Lastly I would like to thank my parents, Richard Martin and Jacqueline Ann Bies, for all they have done for me over the years.

Citations to Previously Published Work

Parts of this thesis have been previously published:

- "Induced electric fields in anisotropic thermoelectric materials," W. E. Bies, R. J. Radtke and H. Ehrenreich, *J. Appl. Phys.* **86**, 5065 (1999).
- "Phonon dispersion effects and the thermal conductivity reduction in GaAs/AlAs superlattices," W. E. Bies, R. J. Radtke and H. Ehrenreich, *J. Appl. Phys.* **88**, 1498 (2000).

Chapter 1

Introduction to Part I: Quantum Chaos

Quantum chaos is the study of quantum-mechanical systems whose classical analogues display deterministic chaos. It is of interest to ask what features, if any, of the classical chaotic dynamics manifest themselves in the quantum mechanics. In fact, the presence or absence of chaos is clearly revealed in the properties of the eigenfunctions and their eigenvalues. The eigenfunctions of classically integrable systems are regular, with a pattern of nodal lines somewhat like a checkerboard, and in phase space they follow the level sets of the classically isolating integrals of motion. In contrast, the eigenfunctions of classically chaotic systems are very irregular, with nodal lines that curve erratically and do not follow any simple form; and further, the nodal lines of successive eigenfunctions bear no simple relation to each other. In systems whose classical phase space is of mixed character, both types of eigenfunction are found. This leads to the supposition that, in a suitable ensemble of systems with Hamiltonians that are different enough to ensure that their eigenfunctions are independent (for this it is sufficient that they differ markedly over a region of size at least a wavelength), the value of a given eigenfunction at a given point, viewed as a function of the parameters describing the Hamiltonian, will be a Gaussian random process. This would indeed be the case if the Hamiltonians were drawn from a random distribution of matrices.

In connection with such randomness there arises the idea of ergodicity. Schnirelman, Zelditch and Colin de Verdiere [103, 32, 126, 127] prove the following result for classically ergodic systems: let ψ_E be an eigenfunction with energy E, $\hat{O}$ be a quantum observable and $O(p,q)$ its symbol in classical phase space, then in the semiclassical limit the quantum expectation $\langle\psi_E|\hat{O}|\psi_E\rangle$ almost always agrees with the classical expectation in the microcanonical ensemble, $\int_{(E,E+\delta E)} O(p,q)dpdq$. Thus, in the semiclassical limit, almost all quantum eigenstates are uniformly distributed over the energy hypersurface (when smoothed over a quantum-mechanically large though classically small region of phase space), and in this sense are ergodic. (The assumption of Gaussian randomness is not necessary; Kaplan and Heller [67] investigate a class of quantum systems which have strongly non-Gaussian eigenstate fluctuations and still satisfy, as they must, the condition of ergodicity in the sense of Schnirelman, Zelditch and Colin de Verdiere.)

Random matrix theory thus appears plausible as a first approximation to quantum-chaotic eigenfunctions. This point of view was originally proposed by Berry

[5]. In addition to ergodicity in the sense of Schnirelman, Zelditch and Colin de Verdiere, random matrix theory also makes statistical predictions about the distribution of level spacings in the semiclassical limit. In an integrable system, the spacings follow a Poisson distribution, while in a system which is classically chaotic there are far fewer small spacings and the distribution follows an analytical form given approximately by the Wigner surmise. This dramatic qualitative difference between the quantum mechanics of integrable systems and that of chaotic systems is confirmed by numerical experiments.

As good a guide to the statistical behavior of eigenfunctions and eigenvalues of classically chaotic systems as it may be, random matrix theory does disregard all information relating to the *dynamics*. To gain a better view of specifically dynamical effects in quantum chaotic systems, it is useful to turn to a semiclassical point of view. The effectiveness of semiclassical techniques in quantum-chaotic systems was first demonstrated by Gutzwiller [45], who derived a formula relating fluctuations in the quantum density of states to the short periodic orbits of the system. This formula has been successfully applied, for instance, to the quantization of the helium atom [125, 38]. Another approach illustrating the effectiveness of semiclassical techniques was taken by Tomsovic and Heller [117, 118]. They found that, when all trajectories are included with appropriate phase factors, the semiclassical propagator reproduces certain quantum-mechanical quantities, such as the autocorrelation function $A(t) = \langle\psi(0)|\psi(t)\rangle$, surprisingly well out to times long compared to the log-time ($\log 1/h$), when classical–quantum correspondence fails and the semiclassical approximation was also naively expected to break down.

An important prediction of semiclassical theory, which is known to be valid at least for short enough times (say, several crossing times), is that some of the eigenfunctions should display a phenomenon called scarring, which is a concentration, above the statistically expected density, of amplitude near short unstable periodic orbits of the classically chaotic system. Scars were originally observed in the stadium billiard by Heller [49]. They have been experimentally observed in various systems, including microwave cavities [109, 112, 110] and tunneling diode junctions [40, 123, 41], and have also been studied numerically in systems such as the hydrogen atom in a strong magnetic field [124, 91]. The reason for the scarring phenomenon can easily be explained in the short-time semiclassical point of view. If T is the period of an unstable periodic orbit with instability exponent λ, a coherent state ϕ launched along this periodic orbit must return to itself at intervals $T, 2T, 3T, \ldots$ and thus the autocorrelation function $A(t) = \langle \phi(0) | \phi(t) \rangle$ will have recurrences at times which are multiples of T, diminished due to spreading by powers of e^{λ} in the transverse direction. The Fourier transform of the autocorrelation function is the local density of states; thus, if we cut off $A(t)$ at some finite time large compared to the time over which it decays exponentially ($O(T/\lambda)$), but still small compared to the log-time scale at which classical–quantum correspondence fails, we will obtain a smoothed-out version of the local density of states which we will call the short-time spectral envelope. Due to the properties of the Fourier transform, the short-time envelope will be a periodic function over a large energy interval, with a sequence of peaks at intervals

$\Delta E = \hbar/T$. Now, whatever the exact form of the local density of states may be, we know that when smoothed out it must give the short-time envelope, and therefore we can say that there is a tendency for eigenstates ψ_E whose energy E lies under a peak of the short-time envelope to have statistically an enhanced overlap $|\langle \psi_E | \phi \rangle|^2$ with the initial Gaussian ϕ, while states whose energy lies under a minimum of the short-time evelope will tend to have smaller intensities on the orbit. In the former case we have scarring (even though the enhancement need not necessarily be visible in a coordinate-space plot of ψ_E, except at focal points), while in the latter we speak of antiscarring, a tendency for amplitude to avoid the neighborhood of the unstable periodic orbit. The prediction of random matrix theory would be, of course, a uniform short-time envelope. This is recovered in the limit of large Lyapunov exponent λ—hence, scarring is not important for long, or otherwise very unstable, periodic orbits. It was also shown that the scar strength, measured as above by the squared amplitude with a coherent state living on that orbit, is a function of Lyapunov exponent λ only and scales as $1/\lambda$ for small λ. The work by Heller may be thought of as a Husimi-space theory. A better understanding of the phenomenon of scarring in coordinate space or in Wigner phase space emerges from subsequent work by Bogomolny [10] and Berry [7, 6], respectively.

Semiclassical theory thus provides an important corrective to the random-matrix picture of quantum chaos, which fails to include scars at all. We have so far discussed only the short-time implications of the semiclassical theory. There are also implications at longer times, as discussed by Kaplan and Heller [68, 65].

For example, the short-time picture relies on linearizing the dynamics around the unstable periodic orbit (in the semiclassical limit, the linearizable region will always be large compared compared to Planck's constant, which is the area in phase space covered by a minimum-uncertainty coherent state). As will be discussed below, the linearized dynamics can be used to derive expressions for the autocorrelation function and the short-time spectral envelope, hence for the strength of scarring. There could, however, be a strong non-linear recurrence due to a homoclinic orbit at some time large compared to the log-time but small compared to the Heisenberg time after which the autocorrelation function becomes quasi-periodic and individual eigenstates can be resolved. Then the intermediate-time spectral envelope would acquire additional oscillations under the short-time peaks. While we still expect scarring to take place, a statistical measure such as the inverse participation ratio would detect the additional fluctuation originating in the non-linear homoclinic recurrence, as compared to the amount of fluctuation in the short-time envelope without a homoclinic recurrence (which in turn is greater than the fluctuation in the flat random-matrix short-time envelope, which is zero).

Another area where the semiclassical theory affords understanding of the long-time effects is in the fluctuations of the autocorrelation function at long times. One might expect that, after the short-time correlations have decayed away, it approaches, say, Gaussian white noise. This is not so; the reason for this is that the homoclinic recurrences are correlated. Kaplan and Heller analyze this phenomonen semiclassically and show that the recurrences come in families that add exactly in phase, thus resulting in an enhanced, relative to the classical expectation, probability to come back to an unstable periodic orbit, which is the *dynamical* origin of the phenomenon of scarring which we have discussed above as a stationary property in the eigenstate domain. Qualitatively, this effect can be understood in terms of a reloading picture: any long-time recurrence at time t will have a tendency to repeat itself at times $t+T, t+2T$, etc., because the short-time dynamics returns to the vicinity of the periodic orbit at intervals of T. Thus, the autocorrelation function at long times can be regarded as a convolution of Gaussian white noise with the short-time autocorrelation function. This enables one to relate the quantum-mechanical mean return probability—a long-time phenomenon—to the short-time autocorrelation function. Amazingly, in quantum mechanics the memory of the short-time behavior persists forever whereas classically the initial conditions are forgotten and one is not any more likely to come back to one's starting point than if one had started anywhere else (on the energy hypersurface). This semiclassical measure of the mean return probability can, and will below, be compared to the exact quantum return probability if the eigenfunctions of a chaotic system are known over a sufficiently large range of energy. Also, it implies that scarring is a correction to complete ergodicity of quantum eigenstates. It is not in conflict with the results of Schnirelman, Zelditch and Colin de Verdiere because the fraction of phase space affected by scarring tends to zero in the semiclassical limit. Only the short periodic orbits are important for scarring and only a region of $O(h)$ in phase space around each periodic orbit sees the scar.

The short-time envelope is a powerful concept for analyzing the semiclassics of chaotic quantum systems. It implies that there is a qualitative difference between the statistical behavior of wavefunctions near short unstable periodic orbits versus that near a generic point in phase space. For a wavepacket on a periodic orbit, the oscillations in the short-time envelope require that there be many more large and small intensities, compared to the chi-squared prediction of random matrix theory. The statistics of wavefunction intensities was investigated by Kaplan [64]. He found that the tail of the wavefunction intensity distribution was indeed dominated by scarring and, after ensemble averaging over systems with classical orbits of differing instability, falls off only as a power law in contrast to the exponential decay in random matrix theory. In the quantum systems to be investigated in Chapters 2 and 3 we shall have the opportunity to verify numerically this departure of quantum chaos from the predicitions of random matrix theory.

The ideas discussed so far, scarring in particular, also have implications for transport in quantum chaotic systems. There, one can consider [47, 50] the mean probability P_{ab} per unit time for going from channel $|a\rangle$ to $|b\rangle$ at long times, and, as a special case, the probability P_{aa} to stay in channel $|a\rangle$. We have already mentioned that the mean long-time quantum return probability P_{aa}, which would seem to depend on details of the long-time behavior, is in fact a function only of the short-time dynamics and is strongly affected by scarring or other forms of short-time localization. Another surprising feature of quantum transport which can be treated semiclassically is discussed by Kaplan [66], who studied the probability to remain in an open chaotic quantum system which is coupled to the environment by a narrow lead. Classically one would expect an exponential decay, if the escape time from the system is large (as will be the case for a sufficiently small channel) compared to the mixing time. Quantum mechanically, random matrix theory can be shown to result in a slower, power-law, fall-off of the probability to remain, if the exit channel is at a generic point in phase space, but if it happens to lie on a short periodic orbit, the short-time dynamics plays a role and scarring leads to an additonal enhancement in the probability to remain at long times. This difference between the actual quantum mechanics and the prediction of random matrix theory highlights the role of antiscarring; due to antiscarring, it is harder for probability to be transported into the region of phase space around the exit channel, and this persists for all time. One can also look at the probability to remain in a state $|b\rangle$ after a long time given that one starts in another channel $|a\rangle$. The result depends on whether they are at generic points in phase space, on the same periodic orbit, or on different periodic orbits. The enhancement or suppression of the transport probability, taking into account all these factors, is purely due to quantum interference and illustrates, once again, how semiclassical techniques allow one to go from short-time dynamics to long-time behavior. These observations are important for mesoscopic devices which may have two or more leads.

Many of the above results have been previously demonstrated for model systems only, for example generalizations of the quantum baker's map. There is a need to test predictions of scar theory for more realistic systems. A first example of a more realistic system is the stadium billiard, by which we mean a region in the plane with a shape given by two semicircles with a rectangular region in between (a 'stadium' shape), and a potential idealized to be zero in the interior and infinite on the boundaries (i.e., we impose Dirichlet boundary conditions). In nearly two-dimensional microwave cavities the Helmholtz equation for the electric field is formally identical to the Schrödinger equation. The modes of this cavity can easily be probed and compared with computed eigenfunctions; scarring is seen, cf. Ref. [112]. Another experimental realization of the stadium billiard is found in micron-scale semiconductor heterostructure devices in which, when very cold, electrons reach the ballistic transport regime and quantum coherence over the scale of the device can be important. For the experimental details, see Marcus *et al.* [86]. Another class of chaotic systems are those with smooth potentials in two or more dimensions, such as arise in the configuration space of many chemical reactions [102, 111] or in quantum dots [71] where the bias voltage on the gates defining the dot provides a smooth potential barrier around the perimeter of the device. In both cases there can be saddle points and this raises the possibility of tunneling between one part of configuration space and another, both of which may be chaotic. Then it is of interest how semiclassical theory relates the chaotic dynamics on either side with the tunneling process which determines the reaction rate, in one case, or the mesocopic conductance, in the other.

We now review Chapters 2 and 3. In Chapter 2 we revisit the stadium billiard and ask what kinds of localization the eigenfunctions display. As discussed above, this is important for quantum transport and thus for any realizations of the stadium in mesoscopic devices. Scarring, of course, may be thought of as a localization phenomenon but there may be other kinds of localization present. A large sample of eigenstates was collected and, for each state, the degree of localization in Husimi space was measured using the inverse participation ratio (IPR). After exclusion of bouncing ball states, random matrix theory predicits an IPR slightly greater than 2, already enhanced relative to the classical value of 1. Yet we find that there is on average even more localization than this. Scars, by themselves, do not suffice to explain the excess localization. In order to disclose the source of the excess localization we study the local density of states as a function of position in phase space. An enhancement in the local density fluctuations is found, as expected, at the positions of the principal short unstable periodic orbits, but, in addition to this, we find the dominant excess fluctuations to arise near certain symmetry-related lines in phase space. The symmetry lines are related to the parity and time-reversal symmetries of the stadium. At certain places in phase space, a Gaussian test state may become purely real when symmetrized with respect to parity, and thus will yield an IPR enhanced from 2 to 3, corresponding to Porter–Thomas statistics. The correlation giving rise to an enhanced IPR persists to the iterates under the Poincare mapping of the symmetry points and thus we find additional symmetry lines, which are curved in general. An analytical form for the IPR near symmetry lines is derived. Also, the analytical predictions of semiclassical theory regarding the mean return probability and the distribution of wavefunction intensities on or off periodic orbits are tested against numerical results.

Chapter 3 investigates the relation in semiclassical theory between chaotic dynamics and tunneling in smooth potentials. A chaotic double-well potential in two dimensions is studied. The level splittings provide a measure of the degree of tunneling through the barrier which is more convenient to compute than, say, the resonance widths in a metastable well (in fact, in one-dimensional WKB theory the resonance widths and the level splittings are related, up to a factor of e^{-S} versus e^{-2S}, where S is the action integral along the optimal tunneling route under the barrier; we believe that such a relationship will continue to hold in two dimensions). A semiclassical theory due to Creagh and Whelan [29] states that the splitting of an eigenfunction is given by its inner product, in a Poincare section transverse to the real continuation of the optimal tunneling trajectory under the barrier, with a Gaussian which is a function of the monodromy matrix for crossing the barrier. Thus, if the real continuation happens to be a short unstable periodic orbit, scarring of eigenfunctions along that orbit should be associated with a non-statistical distribution of splittings. We restrict ourselves to the case where there is a reflection symmetry in y which ensures that the x-axis is a periodic orbit at all energies. The numerical results strikingly confirm that the mean splitting is larger at energies corresponding to scarring, and smaller at intermediate energies, corresponding to antiscarring. Independently, the size of splittings is strongly correlated with scarring as measured by amplitude with a test-state Gaussian living on the horizontal periodic orbit. The short-time envelope is reproduced with some success using the Creagh–Whelan Gaussian together with the formula from Chapter 2 for the autocorrelation function. Finally, the distribution of splittings is shown to deviate from the Porter–Thomas form, with many more large and small splittings relative to the mean, as predicted by scar theory. Thus, we demonstrate that the quantum mechanics of chaotic double-well potentials is consistent with semiclassical theory, as regards both scarring and tunneling across the barrier.

Chapter 2

Localization of Eigenfunctions in the Stadium Billiard

2.1 Abstract

We present a systematic survey of scarring and symmetry effects in the stadium billiard. The localization of individual eigenfunctions in Husimi phase space is studied first, and it is demonstrated that on average there is more localization than can be accounted for on the basis of random-matrix theory, even after removal of bouncing-ball states and visible scars. A major point of the Chapter is that symmetry considerations, including parity and time-reversal symmetries, enter to influence the total amount of localization. The properties of the local density of states spectrum are also investigated, as a function of phase space location. Aside from the bouncing-ball region of phase space, excess localization of the spectrum is found on short

periodic orbits and along certain symmetry-related lines; the origin of all these sources of localization is discussed quantitatively and comparison is made with analytical predictions. Scarring is observed to be present in all the energy ranges considered. In light of these results the excess localization in individual eigenstates is interpreted as being primarily due to symmetry effects; another source of excess localization, scarring by multiple unstable periodic orbits, is smaller by a factor of $\sqrt{\hbar}$.

2.2 Introduction

According to scar theory, the quantum eigenfunctions of a classically chaotic dynamical system do not always look locally like random superpositions of plane waves with fixed energy, as predicted by Berry [8]; instead, many eigenfunctions display a concentration of amplitude around short unstable periodic orbits greater than that expected on the basis of random matrix theory fluctuations. The first examples of scarring in the stadium billiard were presented by Heller [49] (see also unpublished numerical work by McDonald [88]). In Heller's work a semiclassical theory of scarring was given, based on dynamics in the linearizable region around the periodic orbit. Recent developments [68, 65] have extended the theory of scars to the non-linearizable regime, to include the effects of homoclinic recurrences at long times. Scarring is then seen to be a weak localization phenomenon coming from the short-time correlations associated with an unstable periodic orbit; in the energy domain, this means that a wavepacket centered on the orbit may have non-random overlaps with the eigenstates of the system. Quantitative measures of the strength of scarring have been developed and tested numerically [68, 70].

In this Chapter we ask whether the localization caused by scarring on not-too-unstable (i.e., short) periodic orbits, and by atypical regions such as the "bouncing ball" [4, 93, 113, 105, 95] modes is adequate to predict measures of localization in eigenstates of chaotic systems. Quantitative numerical confirmation of scar theory [68, 70] was limited to discrete-time maps, and it would be desirable to study scarring quantitatively in the context of more realistic and experimentally realizable systems, such as the stadium (Bunimovich) billiard. In what follows we present a systematic study of eigenstate localization in the stadium billiard (see Refs. [3, 25, 34] for some other recent analyses of this system). In Section 2.3, we first present the numerical method used to find the eigenstates. Then, in Section 2.4, we study the localization properties of individual eigenstates, followed by an investigation in Section 2.5 of the localization properties of the local density of states. We find evidence of scarring as a ubiquitous phenomenon, in all the energy ranges considered. We examine quantitative predictions of scarring strengths based on the classical structure around the unstable periodic orbits [68]. However, scar theory is not sufficient to explain all the observed localization. Symmetry effects, including the imprints of both time-reversal and parity symmetries, are ultimately found to dominate the excess wavefunction localization in the stadium.

2.3 Method

We study eigenstates of the time-independent Schrödinger equation in an infinite 2-dimensional potential well in the shape of a stadium, taken here to consist of a square of side 2 with semicircular endcaps of radius 1. We use the plane wave method [52] to find states with even parity with respect to reflection about each of the two symmetry axes of the stadium. It is assumed that eigenstates at energy E can be approximated as a superposition of plane waves at that energy; i.e., with $|\mathbf{k}|^2 = E$ (we use $\hbar = 1$ and $m = 1/2$ here and throughout). Therefore we use as a basis plane waves with even-even parity, namely $\cos(kx\cos\theta_j)\cos(ky\sin\theta_j)$, with N values of θ_j chosen uniformly between 0 and $\pi/2$. Then the coefficients of these N plane waves are determined by minimizing the squared value of the wavefunction at $M \geq N$ equally spaced points on the boundary. The wavefunction is set to an arbitrary non-zero value at a point in the interior in order not to have a singular system of equations, and M is kept slightly larger than N so that the system of equations is overdetermined. In practice M must be chosen so that there are several points per wavelength.

Since the eigenvalues are not known *a priori*, we search for them using the 'tension' method: at each energy E we solve the linear system for the coefficients of the plane waves by singular-value decomposition; then, we compute the integral of $|\psi|^2$ along the boundary. If E is an eigenvalue this integral should vanish, to within numerical precision. In practice one finds sharp minima in the tension as a function of k that are identified with the eigenvalues. Since it is not known in advance how many plane waves are needed to give an accurate representation of the wavefunction at a given energy, we repeat the search for minima with increasing numbers of plane waves N and boundary points M until the results of subsequent iterations agree; we find, for instance, that for k around 100, $N = 300$ and $M = 310$ are necessary. This corresponds to eight points per wavelength along the part of the boundary lying in the first quadrant. The number of plane waves needed scales linearly with k.

The computed eigenvalues and eigenfunctions satisfy three diagnostic tests: (1) the total density of states agrees well with the Weyl area rule; the deviation is never more than about 5 states out of 5000, or 0.1% (the bulk of this deviation is due to a small periodic modulation in the level staircase function, which has the right period to be attributable to the bouncing-ball states [113]; after taking out this modulation we find that the total number of possible missing or extra states in the region $100 < k < 200$ is not more than one or two); (2) the histogram of level spacings is consistent with the Wigner surmise for random-matrix statistics; (3) the overlaps between eigenstates are found to be less than 1%. A problem with our procedure is that it produces many shallow tension minima, by which we mean that the contrast between the tension at a local minimum and the background tension is less than a factor of 100. The frequency of such shallow minima grows with increasing energy. Moreover, these poor minima are not improved by increasing the parameters M and N because doing so makes the matrix more ill-conditioned. This suggests that there is an intrinsic shortcoming in the ability of our basis to represent eigenstates; to have a perfectly accurate representation one would presumably need to include evanescent waves of total energy E as well. Although it has been shown by Berry [8] that an evanescent wave may be represented by a singular superposition of plane waves, this representation is numerically ill behaved, giving poor convergence properties.

In spite of this difficulty, essentially all of the questionable states must correspond to true eigenstates, in order for the counting of states to come out right. Also, many of the shallow minima we assume correspond to eigenstates occur in close conjunction with other minima, and yet the numerically computed histogram of level spacings decreases linearly toward zero at small spacings, in good agreement with the theory.

Another difficulty with the present method is that the computation time per eigenstate scales as k^3, making the collection of extensive eigenstate statistics above $k = 300$ almost prohibitively expensive. The plane-wave method described here has been improved upon by Li and Hu [75, 74], who expand the derivative of the tension analytically in k in order to predict the position of the next minimum. Li and Hu's approach improves the speed of the plane-wave method by about a factor of five, but does not change the way the computation time scales with k. Thus, only slightly higher values of k could be reached for a given computation time, and we have not needed to employ this approach here. Another approach to the stadium billiard has been outlined by Saraceno's group [122, 106]. Their method, suitable for very high energies, obtains all the eigenvalues in a narrow range of k simultaneously by solving a generalized eigenvalue problem, rather than searching for them step by step. The energies to be studied in this Chapter, however, are not so high as to necessitate the use of such an algorithm.

Lists of eigenstates of the even-even symmetry class in the three energy ranges of $k = 20$ to 30, $k = 100$ to 150 and $k = 200$ to 225 were generated according to the method outlined above. Some examples are given in Fig. 2.1 of eigenfunctions $\Psi_n(x, y)$ in coordinate space, where the graylevel represents $|\Psi|^2$.

2.3.1 Surface wavefunctions

Classically, one typically uses the boundary of the billiard as a Poincaré surface of section. The variables parameterizing the surface are q, the arclength along the boundary, and its conjugate momentum p, which is the component of momentum parallel to the wall (both positive in the clockwise sense). q is taken modulo the length $L \equiv 4 + 2\pi$ of the perimeter, while p is limited by energy considerations to $-k \leq p \leq k$. Then the classical billiard dynamics is reduced to a nonlinear one-bounce map, $(q', p')^T = M_{\mathrm{Cl}}(q, p)^T$.

A natural way of reducing the quantum problem from two to one dimension is to characterize the Dirichlet eigenfunctions $\Psi_n(x, y)$, which are defined on the interior of the stadium, by their normal derivatives $\phi_n(q)$ on the boundary:

$$\langle q|n\rangle \equiv \phi_n(q) \propto \hat{\mathbf{n}} \cdot \nabla\Psi_n(x, y). \tag{2.1}$$

We will study the properties of these *surface wavefunctions.* The wavefunctions are normalized according to the convention

$$\int\int dxdy|\Psi_n(x, y)|^2 = 1 \tag{2.2}$$

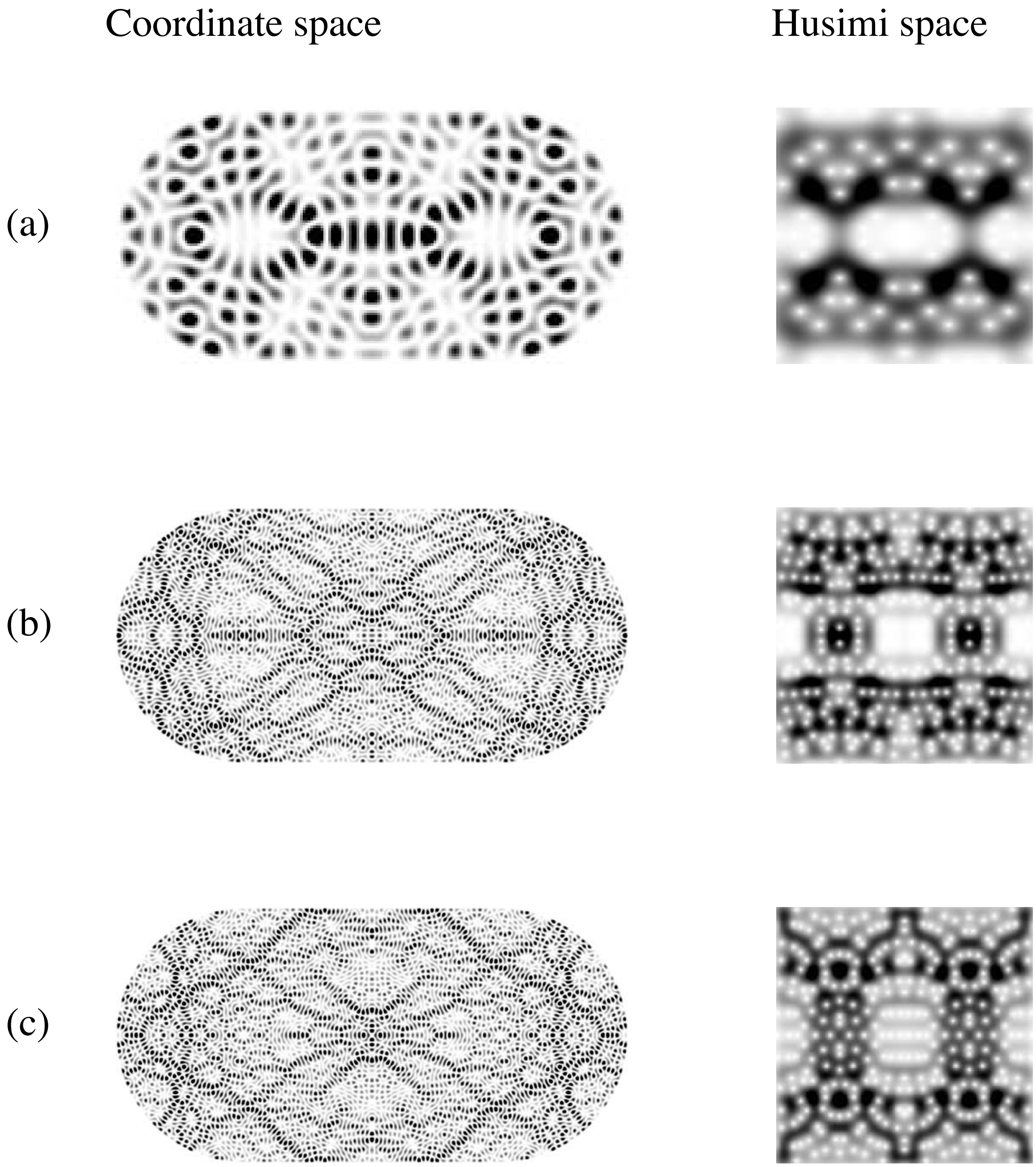


Figure 2.1: Representative eigenfunctions of the stadium billiard in coordinate space (left) with x-axis running from -2 to 2 and y-axis running from -1 to 1, and in Husimi phase space (right) with distance q along the perimeter (horizontal axis) running clockwise from 0 (corresponding to the center of the upper straight segment) to $4+2\pi$ and tangential momentum (vertical axis) running from $-k$ to k, where k is the wavenumber of the eigenfunction. (a) $k = 24.680$, (b) $k = 100.787$ and (c) $k = 105.608$, chosen to illustrate scarring along the bowtie and diamond orbits.

which is equivalent [9] to the following condition on the surface wavefunctions:

$$\frac{1}{2k^2}\int dq(\mathbf{n}(q)\cdot\mathbf{r}(q))|\phi_n(q)|^2 = 1 \tag{2.3}$$

where $\mathbf{n}(q)$ is the unit normal at the position q along the boundary, and $\mathbf{r}(q)$ is the displacement vector from the center of the stadium to the position q.

The Husimi representation of a surface wavefunction $|n\rangle$ is given by its projection $|\langle q_0, p_0|n\rangle|^2$ onto test state Gaussians of the form

$$\langle q|q_0, p_0\rangle = \frac{1}{\sigma^{1/2}\pi^{1/4}}\exp\left[-\frac{(q-q_0)^2}{2\sigma^2} + ip_0(q-q_0)\right]. \tag{2.4}$$

The Gaussian is centered at the point (q_0, p_0) in the boundary phase space. The parameter σ controls the aspect ratio of the Gaussian: its width in position is $\sigma/\sqrt{2}$ while its width in momentum space is $1/(\sigma\sqrt{2})$. To maintain a given aspect ratio of the Gaussian in phase-space, σ must scale as $k^{-1/2}$. Here, we choose $\sigma = [(2+\pi)/k]^{1/2}$, which makes the aspect ratio unity in units where the full phase space on the billiard boundary is taken to be a square. Note that the wavefunctions are real in coordinate space but complex in phase space. Some example Husimi plots are given in Fig. 2.1, where white corresponds to high intensity and black to low intensity.

2.4 IPR statistics for individual eigenstates

We start by investigating the localization properties of individual eigenstates. As a measure of the degree of localization of the eigenstates in phase space, we use the mean squared value of the intensity, also known as the inverse participation ratio (IPR)

$$\mathrm{IPR}_n = \frac{\frac{1}{J}\sum_{j=1}^{J} |\langle q_j, p_j|n\rangle|^4}{\left(\frac{1}{J}\sum_{j=1}^{J} |\langle q_j, p_j|n\rangle|^2\right)^2} \tag{2.5}$$

where q_j and p_j range over the entire phase space as j varies on a sufficiently fine grid, which we take to be 200×200. The range of momenta p_j covered is from $-k$ to $+k$ for each wavefunction; outside this classically allowed region there is almost no wavefunction amplitude. The IPR measures how much fluctuation across phase space there is in the eigenfunction. If $|\langle q_j, p_j|n\rangle|^2$ were completely uniform over phase space, the IPR would equal 1. On the other hand, if all the intensity were concentrated entirely at one point j and was zero elsewhere, the IPR would reach its maximal value J. [Of course, the uncertainty principle prevents the IPR from ever becoming greater than the size of phase space in units of a Planck cell, i.e., $O(kL/2\pi)$.] Random-matrix theory (RMT) predicts a Porter-Thomas distribution of wavefunction intensities, which yields an IPR of 2 (for complex wavefunctions), already larger than the naive classical expectation of 1.

A subtle point is that the wavefunction intensity $|\langle q_j, p_j|n\rangle|^2$, even when averaged over wavefunctions $|n\rangle$, is not uniform over the boundary phase space, being instead a non-trivial function of p_j (see Eq. (2.8) below). Therefore, even in the absence of quantum fluctuations, if the intensities $|\langle q_j, p_j|n\rangle|^2$ for a wavefunction $|n\rangle$ were given simply by the mean value $\sqrt{1-(p/k)^2}$ as in Eq. 2.8, we would obtain an IPR of

$$\left\langle 1-(p/k)^2 \right\rangle / \left\langle \sqrt{1-(p/k)^2} \right\rangle^2 \approx 1.08\,. \tag{2.6}$$

The averages $\langle \cdots \rangle$ here are of course taken over the classically allowed range of p:

$-k < p < k$. If the intensity fluctuations around the smooth value $\sqrt{1-(p/k)^2}$ were instead given by a Porter-Thomas distribution, we would obtain an IPR of $1.08 \times 2 = 2.16$ (since the smooth behavior and the Porter-Thomas fluctuations are independent of one another, the IPR contributions simply multiply).

In Table 2.1 the IPRs of the first 36 eigenfunctions in each energy range are tabulated. The average IPRs for the first 100 states in each energy regime are summarized in Table 2.2. The IPRs cluster around 3, with some much larger. The mean IPR is 3.67, 3.60 and 3.08 in the low, medium and high energy ranges respectively, but with large standard deviations. Visual examination of the Husimi plots for states with the largest IPRs reveals that they correspond to bouncing-ball states. Since this phenomenon is well understood (and is associated with marginally stable periodic orbits), we remove this effect by removing the bouncing-ball states from the average (see column (e) of Table 2.2).

The remaining states avoid the bouncing-ball region, causing their IPR to be increased by an amount that can be estimated. At $k \approx 20$, the bouncing-ball region occupies about 13% of phase space. Assuming the non-bouncing-ball states to vanish identically in that region, the random-matrix theory prediction for the average IPR without the bouncing-ball states becomes $2.16 \times 1.13 = 2.44$. The size of the bouncing-ball region in phase space is energy-dependent (and must of course go to zero in the $k \to \infty$ limit, in order to satisfy the Shnirelman quantum ergodicity condition). As found by Bäcker *et al.* [4] the number of bouncing-ball states scales as $N_{bb}(E) = 0.2E^{3/4}$ and thus the density of bouncing-ball states, which is proportional

k	IPR	k	IPR	k	IPR
20.082	2.03	100.031	2.24	200.015	2.81
20.237	$4.43^\dagger$	100.047	2.88	200.047	3.24
20.439	2.47	100.086	2.73^*	200.054	2.41
20.457	13.36^*	100.099	7.61^*	200.076	2.77
20.728	5.02^*	100.133	2.56	200.099	2.76
20.825	1.47	100.224	$3.51^\dagger$	200.108	17.91^*
21.039	1.88	100.226	2.74	200.113	2.70
21.223	$2.84^\dagger$	100.273	2.02	200.116	2.94
21.285	2.57	100.285	2.05	200.123	3.18
21.322	2.11	100.342	2.02	200.151	2.54
21.671	$5.35^\dagger$	100.363	$3.54^\dagger$	200.190	2.68
21.717	2.61	100.402	2.64	200.195	3.11
21.948	3.12	100.416	2.38	200.212	2.79
22.119	1.71	100.496	2.20	200.255	2.66
22.284	1.69	100.535	7.55^*	200.266	2.45
22.504	3.12	100.552	2.42	200.296	2.21
22.674	3.27	100.613	2.57	200.317	2.86
22.839	2.81	100.628	2.62	200.328	$2.96^\dagger$
23.029	2.41	100.678	1.93	200.334	2.77^*
23.108	$3.39^\dagger$	100.712	2.43	200.351	2.56
23.361	2.44	100.714	2.28	200.370	2.15
23.503	1.82	100.787	2.35	200.386	2.12
23.594	15.79^*	100.830	3.94	200.407	3.12
23.774	2.10	100.846	3.53	200.441	2.30
23.821	3.08	100.854	2.68	200.480	2.37
23.904	2.63	100.933	4.95^*	200.492	2.50
24.061	2.90	100.937	2.88	200.500	2.76
24.277	2.86	100.989	2.02	200.526	$5.67^\dagger$
24.339	5.29	101.017	1.98	200.546	2.27
24.461	2.37	101.027	2.81	200.560	2.94
24.680	2.34	101.082	2.14	200.573	2.91
24.764	$3.60^\dagger$	101.120	2.31	200.588	2.25
24.937	2.84	101.164	2.51	200.604	3.60
25.095	2.62	101.224	2.01	200.626	13.18^*
25.117	2.21	101.237	2.30	200.639	3.10
25.435	2.76	101.290	3.33	200.659	1.97

Table 2.1: Inverse participation ratios of the Husimi plots of the first 36 eigenstates in the low ($k = 20$), medium ($k = 100$) and high ($k = 200$) energy ranges. Bouncing-ball states are labeled with an asterisk, and states visibly scarred on the horizontal-bounce orbit are labeled with a dagger. The random-matrix theory prediction, with bouncing-ball correction, is 2.44, 2.29 and 2.25, respectively, in the three energy ranges.

(a) k_{min}	(b) IPR numerical	(c) bb fraction of phase space	(d) IPR without bb RMT prediction	(e) IPR without bb numerical
20	3.67 (3.53)	13%	2.44	2.71 (0.62)
100	3.60 (3.33)	5.8%	2.29	2.70 (0.65)
200	3.08 (1.98)	4.1%	2.25	2.82 (0.80)

Table 2.2: Averages of the inverse participation ratios of the first 100 states in energy ranges starting at $k = 20$, 100 and 200. Columns: (a) the k value at the beginning of the range; (b) the actual IPR; (c) the estimated fraction of phase space occupied by the bouncing-ball states in this range of k; (d) the predicted IPR when the bouncing-ball region is excluded; (e) the computed average IPR after the bouncing-ball states (e.g., those marked with an asterisk in Table 2.1) are omitted. Standard deviations are shown in parentheses. The large remaining discrepancy between (d) and (e) is mostly due to symmetry effects, as explained in the text.

to the area occupied by them in phase space, scales as $E^{-1/4}$ or $k^{-1/2}$. This scaling law is in agreement with our numerical results. The adjusted RMT predictions are shown in column (d) of Table 2.2.

As seen in Table 2.2, the average IPRs with bouncing-ball states removed are still well above the random-matrix prediction in each energy range. The IPRs were obtained by averaging over 100 states; the statistical uncertainty in the means is thus about $1/\sqrt{100} \approx 0.1$ of the standard deviation, or roughly 0.1, which is comparable to the differences in average IPR between different energy ranges. We see that the IPR does not increase markedly with wavevector k, so the stadium is not an example of weak quantum ergodicity as studied by Kaplan and Heller [69] (this makes sense, since the stadium is strongly chaotic with rapid mixing throughout its phase space with the exception of the bouncing ball region). However, neither is there a marked trend towards the RMT predictions as k increases by a factor of ten.

Thus, we have uncovered systematic evidence of localization of eigenfunctions in the stadium billiard beyond what would be predicted on the basis of random-matrix theory. What is responsible for it? The next highest IPRs, after the bouncing-ball states, are for eigenfunctions visibly scarred along the periodic orbit that bounces horizontally between the centers of the endcaps (these "horizontal-bounce states" are marked with a dagger in Table 2.1). Thus, one might hypothesize that scarring along this and other periodic orbits is partly responsible for the excess localization. In the next section, however, we shall introduce a method that allows the excess localization to be identified with specific classical structures in phase space, and techniques to predict that localization theoretically. We shall find that quantum symmetry effects cause most of the excess localization, while a secondary effect consists of combined scarring contributions from all of the short unstable periodic orbits.

2.5 Spectral localization

2.5.1 Introduction

The IPRs of individual eigenstates establish that there is localization compared with the predictions of random matrix theory. However, it does not permit the localization to be definitively identified with structures in classical phase space because the IPR is an aggregate over all of phase space. But it is also possible to ask a complementary question, namely, how ergodic are the eigenstates at a given point in phase space as one varies the energy? To answer this question, we turn to the *local IPR*—a tool that provides a picture of the regions of phase space that are prone to intensity enhancement or depletion over an ensemble of eigenstates.

Definition

The local IPR ('LIPR') is the mean squared eigenstate intensity at point (q, p), averaged over an ensemble of eigenstates [48, 51]:

$$\mathrm{LIPR(q,p)} = \frac{\frac{1}{\mathrm{N}} \sum_{\mathrm{n=1}}^{\mathrm{N}} \left| \langle q, p | n \rangle \right|^4}{\left(\frac{1}{\mathrm{N}} \sum_{\mathrm{n=1}}^{\mathrm{N}} \left| \langle q, p | n \rangle \right|^2 \right)^2}, \tag{2.7}$$

where N is the number of eigenstates being summed over. Typically the sum is taken over eigenstates in a small energy range around some central value of k. The Gaussian wavepackets $|q, p\rangle$ are adjusted to maintain constant aspect ratio in phase space as k changes (see discussion following Eq. (2.4)); this keeps the classical dynamics fixed as the energy increases. Several LIPR pictures are shown in Fig. 2.2.

Intuitive description

The LIPR at the point (q, p) is a statistical property of the set of 'random' variables $\{\langle q, p | n \rangle : n = 1 \ldots N\}$. The second moment of this quantity [which appears in the denominator of Eq. (2.7)] is a smooth function of q and p, as will be shown in Sec. 2.5.3; it is present as a normalization factor. The lowest moment that is nontrivial is the fourth moment, which appears in the numerator of Eq. (2.7). The LIPR, then, measures the non-uniformity of the local density of states at a (fixed) position in phase space. A large LIPR indicates that a relatively small fraction of eigenstates have large overlaps with the test Gaussian, whereas a small LIPR indicates that the

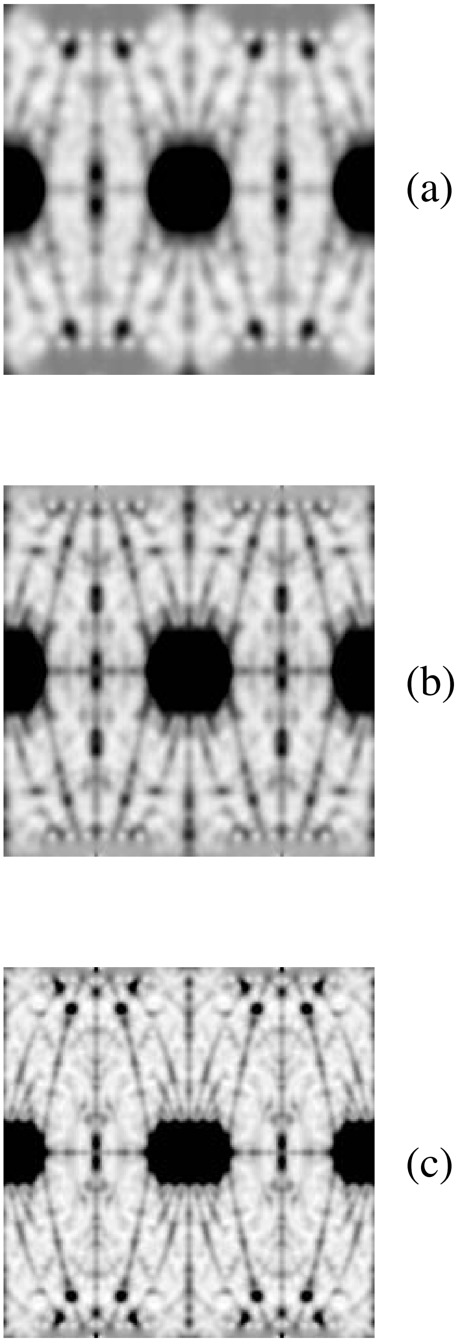


Figure 2.2: LIPR (see Eq. 2.7) in phase space as a function of q running horizontally from 0 to $4 + 2\pi$ and p/k running vertically from -1 to 1. The test-state width σ is chosen so that the test Gaussians have circular cross-sections when the figures are plotted with a square aspect ratio. Darker shading indicates higher values. (a) $k = 50$ to 60, (b) $k = 100$ to 150, and (c) $k = 200$ to 225.

overlaps are distributed in a more ‘egalitarian’ way.

We expect the LIPR to be a sensitive indicator of the presence of scarring, because a wavepacket centered on a periodic orbit has a local density of states that oscillates with energy [49]. Eigenstates in the peak of the oscillation will, on average, have enhanced overlaps with the test state, while eigenstates in the trough will have suppressed overlaps. This nonuniformity will cause the LIPR to have peaks near periodic orbits.

There is a useful interpretation of the LIPR—it is proportional to the long time average of the probability that the wavepacket $|q,p\rangle$, evolved in time, has returned to its original location. A derivation and discussion of this interpretation are presented in Sec. 2.5.7 below.

Since the stadium is strongly chaotic, classical trajectories explore all of phase space (except for a set of measure zero). Therefore the classical return probability for long times is uniform, and the resulting naive classical prediction is that LIPR $= 1$. This is indeed too naive.

Instead, the ‘null hypothesis’ for a chaotic system is that its eigenfunctions are random superpositions of plane waves [8]. Under that maximally random assumption, projections of wavefunctions on test states should follow a χ^2 distribution with one degree of freedom (if the eigenfunctions and test functions are real) or a χ^2 distribution with two degrees of freedom (if they are complex). The corresponding predicted LIPRs are 2 or 3, respectively—already different from the naive classical prediction.

2.5.2 Introduction to pictures

In Fig. 2.2 we present the LIPR computed for even-even eigenstates in the wavenumber ranges $k = 50$ to 60, 100 to 150 and 200 to 225. They display interesting and beautiful localization effects.

The horizontal axis of each picture is q, which runs from 0 to L. $q = 0$ corresponds to the midpoint of one of the straight walls. The vertical axis is p/k, which runs from -1 to 1. The eightfold symmetry of these plots results from the two spatial reflection operations plus time-reversal symmetry. The LIPRs all show a very large amplitude in the bouncing-ball region, explainable by the bouncing-ball and near-bouncing-ball states present in the energy ranges considered. But in addition, there is localization in many other regions of phase space. There are very prominent straight lines of enhanced LIPR running along the lines of symmetry, and filamentary streaks running across other parts of phase space. There are also isolated spots at which the LIPR is unusually large.

Note that, although the details vary somewhat, the main localization features appear in the same positions in each energy range. This lack of $\hbar$-dependence suggests that the streaks are associated with classical phenomena. Our main goal is to give a quantitative explanation for all of these effects. We would also like to understand the relative contributions of each semiclassical effect to the total average IPR enhancement found in Sec. 2.4.

If one varies the aspect ratio of the test state to make it more position-like or momentum-like, a series of bright spots can also be resolved at various points along the streaks, on a scale as small as the smaller axis of the Gaussian test state (see Fig. 2.3).

2.5.3 Mean wavefunction intensity

The mean wavefunction intensity appears in the denominator of Eq. (2.7) in the role of a normalization factor. It can be determined from classical phase space arguments similar to those used in the derivation of Weyl's law. It depends on p for two reasons: (1) we are using normal derivatives, which introduces a contribution of $p_\perp = k\sqrt{1-(p/k)^2}$ multiplying the coefficients of the plane waves (since the coefficients are squared, the resulting factor is $k^2(1-(p/k)^2)$), and (2) there is a geometrical factor coming from the projection from the circle $k_\perp^2 + p^2 = k^2$ onto the boundary of the stadium. Because the plane waves are evenly distributed around the circle, there are more near $p = k$ than near $p = 0$; the resulting geometrical factor is $[1-(p/k)^2]^{-1/2}$. Together, (1) and (2) give

$$\left\langle \left|\langle q, p|n\rangle\right|^2 \right\rangle_n \sim \sqrt{1-(p/k)^2}. \tag{2.8}$$

This smooth factor is divided out of the LIPR in Eq. (2.7), giving the LIPR a flat background.

We checked the mean wavefunction intensity for the computed even-even states in the interval from $k = 100$ to 110. It agrees well with Eq. (2.8), except at $p = 0$ around the centers of the straight segments and the centers of the endcaps, where sharp peaks are seen. The peaks occur because only even-even states have been included in the analysis; since the odd states must vanish at the symmetry points,

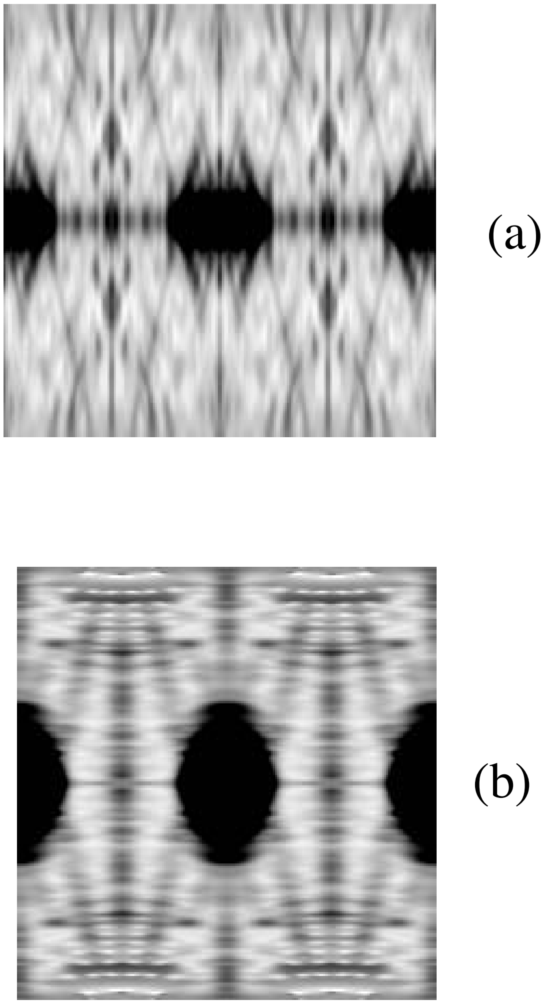


Figure 2.3: LIPR in phase space for eigenfunctions in the range from $k = 100$ to 150, coordinates as in Fig. 2.2(b), for (a) $\sigma^2 \approx 0.01$, position-like test Gaussians, to be compared with $\sigma^2 \approx 0.05$ used in Fig. 2.2(b), and (b) $\sigma^2 \approx 0.25$, momentum-like test Gaussians.

the even-even states compensate by having double the average intensity there. Apart from this, the agreement of the computed mean intensity with Eq. (2.8) indicates that we have included enough eigenfunctions in the energy ranges chosen for Fig. 2.2.

2.5.4 Symmetry lines

What is the explanation for the streaks in the LIPR plots of Fig. 2.2? The most prominent streaks are shown schematically in Fig. 2.4. The streak labeled (1) in Fig. 2.4 corresponds to trajectories emerging from the center of the endcap. The streak labeled (2) corresponds to a family of orbits emerging from the center of the straight segment. Streak (3), which starts at small momentum at the edge of the bouncing-ball region and extends to large momentum at the center of the endcap, corresponds to orbits that have a vertical segment in the endcap region. The box and the bowtie are the two most prominent periodic orbits in this family. The streak labeled (4) starts at small momentum at the center of the straight segment, curves up to larger momentum, and finally comes back down to zero momentum at the center of the endcap. It corresponds to orbits that pass through the center of the stadium. The most prominent periodic orbit in this family is the Z orbit. Finally, there is the horizontal line (5) going across the plot at zero momentum, corresponding to normal incidence on the billiard boundary. All of these families can be understood by considering time reversal and parity symmetry effects in combination with dynamics, as we will now see.

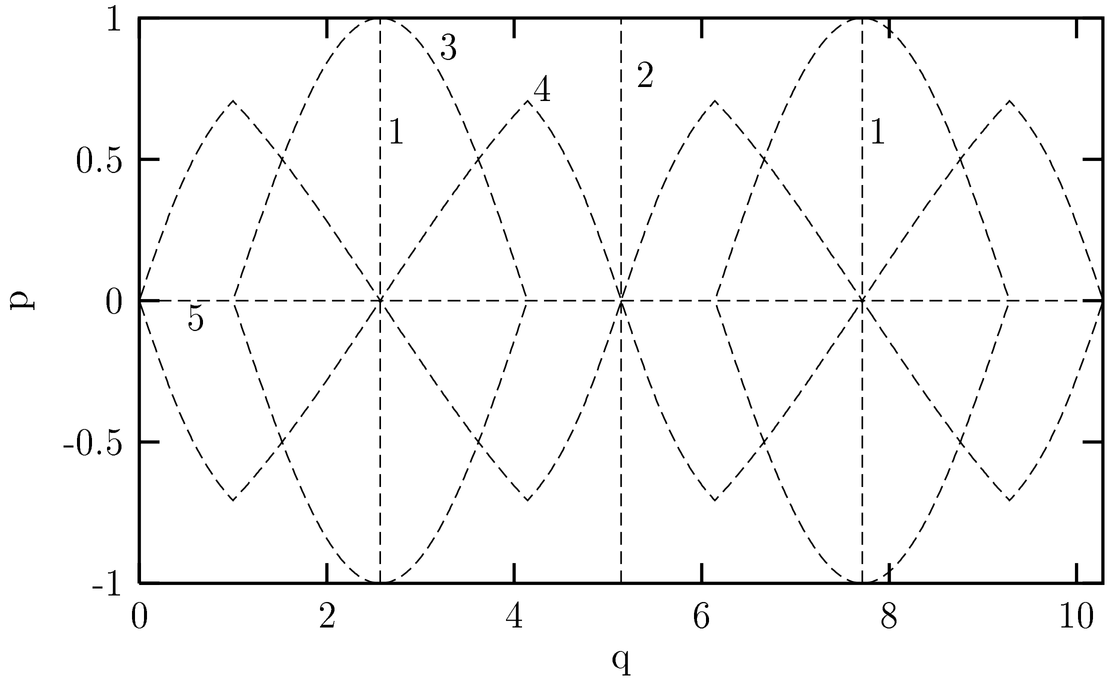


Figure 2.4: Principal symmetry-related structures in phase space of the inverse participation ratio plotted in Fig. 2.2 and the average-return probability plotted in Fig. 2.9. Schematic depiction of the symmetry lines in Fig. 2.2, labeled in the order in which they are described in the text.

Simple explanation

The intuitive reason for the symmetry lines is that the system, having time-reversal symmetry, has real eigenfunctions. However, the Gaussian test states [Eq. (2.4)] are intrinsically complex. Therefore the overlap $\langle q, p|n\rangle$ has both real and imaginary parts which, for most choices of q and p, are not trivially related to one another. Therefore the typical computed LIPR for the system is 2, characteristic of complex eigenfunctions. But in a sense the 'correct' value LIPR $= 3$ (characteristic of real eigenfunctions) is obscured by a non-ideal choice of test states whose properties do not match those of the eigenstates.

However, due to the symmetry of the system, there are certain values of q and p for which the real and imaginary parts of $\langle q, p|n\rangle$ are simply correlated. For example, when $p = 0$ (normal incidence to the wall), the test states and their projections $\langle q, p = 0|n\rangle$ on the eigenstates become pure real and therefore near $p = 0$ the LIPR increases to 3. [This explains the symmetry line labeled (5) in Fig. 2.4.] Similarly, the eigenfunctions are symmetric about the centers of the straight walls, and therefore $\langle q = 0, p|n\rangle$ is pure real and again the LIPR increases to 3. (Similarly for the centers of the endcaps.) This simple argument explains the straight symmetry lines (1), (2) and (5) that appear in the LIPR pictures. A rigorous derivation of their heights and widths will be given in the next section.

It is interesting to note that the same LIPR enhancements would be present if we had used eigenstates of other symmetry classes. For example, for odd-odd eigenstates, the eigenfunctions are antisymmetric about the centers of the straight walls, and therefore $\langle q = 0, p|n\rangle$ is pure imaginary, again giving a χ^2 distribution with one degree of freedom, and the enhancement LIPR = 3.

Derivation

In this section we show how to compute the enhancement of the LIPR that appears near the symmetry lines of the system. We do this in order to verify the conclusions of the preceding simple arguments, and also to deduce important details such as the profiles and widths of the symmetry lines.

The derivation relies on the fact that the eigenfunctions are chosen to be be real (in a coordinate basis), which can be done as a consequence of time-reversal symmetry. We denote time-reversal with T, an antiunitary operator. On the surface of section, $T\,|q, p\rangle = |q, -p\rangle$.

Furthermore, the eigenfunctions are symmetric or antisymmetric with respect to the reflections $x \to -x$ and $y \to -y$, which we will denote with the unitary operators R_x and R_y respectively. On the surface of section, this corresponds to symmetry with respect to reflection about two values $\{q_x, q_y\}$ corresponding to an intersection of the billiard wall with the y and x axes, respectively. Then for example, $R_x\,|q_x + \Delta q, p\rangle = |q_x - \Delta q, -p\rangle$ (where q is always taken modulo the billiard perimeter). For simplicity in the following derivation we will consider the simpler case of a single reflection symmetry, R, about $q = 0$.

It is not correct to model the eigenfunctions as uncorrelated Gaussian random variables, because their reflection symmetry correlates their values at different points. However, we can generate random wavefunctions $|n, \pm\rangle$ with positive ('+') or negative ('$-$') symmetry by taking completely random real wavefunctions $|n\rangle$ (with no symmetry) and projecting them onto the correct symmetry subspace:

$$|n, \pm\rangle = \frac{1}{\sqrt{2}} \left(|n\rangle \pm R\,|n\rangle \right). \tag{2.9}$$

To substitute into Eq. (2.7), we will need to compute $\langle q, p | n, \pm\rangle$. This quantity is complex, so we decompose it into two real random variables $\mu_\pm$ and $\nu_\pm$ as

$$\frac{1}{\sqrt{2}} \left[\langle \phi_{q,p} | n \rangle \pm \langle \phi_{q,p} | R | n \rangle \right] \equiv \mu_\pm + i\nu_\pm. \tag{2.10}$$

If $\mu_\pm$ and $\nu_\pm$ had the same variances and were uncorrelated, the LIPR would uniformly equal 2, like that of any Gaussian random process with two degrees of freedom. This null result would hold even if the variance depended on p and q, because the square of the variance appears both in the numerator and in the denominator of the definition of the LIPR.

However, in reality the symmetries cause the variances of the real and imaginary parts $\mu_\pm$ and $\nu_\pm$ to depend differently on phase space location. The effective number of degrees of freedom varies from two (when $\langle \mu_\pm^2 \rangle = \langle \nu_\pm^2 \rangle$) to one (when $\langle \mu_\pm^2 \rangle \neq 0$, $\langle \nu_\pm^2 \rangle = 0$), and correspondingly the LIPR varies from 2 to 3.

In terms of the three quantities $\langle \mu_\pm^2 \rangle$, $\langle \nu_\pm^2 \rangle$, and $\langle \mu_\pm \nu_\pm \rangle$,

$$\left\langle |\mu_\pm + i\nu_\pm|^2 \right\rangle = \langle \mu_\pm^2 \rangle + \langle \nu_\pm^2 \rangle, \tag{2.11}$$

and $\left\langle |\mu_\pm + i\nu_\pm|^4 \right\rangle$ can be expanded and then contracted pairwise, giving

$$\left\langle |\mu_\pm + i\nu_\pm|^4 \right\rangle = 2 \left(\langle \mu_\pm^2 \rangle + \langle \nu_\pm^2 \rangle \right)^2 + \left(\langle \mu_\pm^2 \rangle - \langle \nu_\pm^2 \rangle \right)^2 + 4 \langle \mu_\pm \nu_\pm \rangle^2 \tag{2.12}$$

The problem is reduced to the computation of the required variances and correlations in Eq. (2.12).

Using $\langle q,p|n\rangle^* = \langle q,-p|n\rangle$ and $R\,|q,p\rangle = |{-q},-p\rangle$, we obtain

$$\mu_\pm = \frac{1}{2\sqrt{2}}\left[\langle q,p|n\rangle + \langle q,-p|n\rangle \pm \langle -q,-p|n\rangle \pm \langle -q,p|n\rangle\right] \tag{2.13}$$

$$\nu_\pm = \frac{1}{2i\sqrt{2}}\left[\langle q,p|n\rangle - \langle q,-p|n\rangle \pm \langle -q,-p|n\rangle \mp \langle -q,p|n\rangle\right]. \tag{2.14}$$

Because $|n\rangle$ represents not a wavefunction but rather the projection of a wavefunction onto the surface of section, the closure relationship does not hold: $\sum_n |n\rangle\langle n| \neq 1$. But as explained above in Sec. 2.5.1, it is still approximately true that

$$\sum_n |n\rangle\langle n|q,p\rangle \approx f(p)\,|q,p\rangle\,, \tag{2.15}$$

where $f(p) = f(-p) \sim \sqrt{1-(p/k)^2}$. It follows that

$$\begin{aligned}\left\langle \mu_\pm^2 \right\rangle &= \frac{1}{8N}\sum_n \Big[\langle q,p|n\rangle + \langle q,-p|n\rangle \pm \langle -q,-p|n\rangle \pm \langle -q,p|n\rangle\Big] \\ &\qquad \times \Big[\langle n|q,p\rangle + \langle n|q,-p\rangle \pm \langle n|-q,-p\rangle \pm \langle n|-q,p\rangle\Big] \\ &= \frac{f(p)}{2N}\Big[\langle q,p|q,p\rangle + \mathrm{Re}\,\langle q,p|q,-p\rangle \\ &\qquad \pm \mathrm{Re}\,\langle q,p|-q,p\rangle \pm \langle q,p|-q,-p\rangle\Big];\end{aligned} \tag{2.16}$$

the last line follows from $\langle q,-p|-q,p\rangle = \langle q,p|-q,-p\rangle$, which is real. The other correlations can be worked out similarly; the results are

$$\begin{aligned}\left\langle \nu_\pm^2 \right\rangle &= \frac{f(p)}{2N}\Big[\langle q,p|q,p\rangle - \mathrm{Re}\,\langle q,p|q,-p\rangle \\ &\qquad \mp \mathrm{Re}\,\langle q,p|-q,p\rangle \pm \langle q,p|-q,-p\rangle\Big]\end{aligned} \tag{2.17}$$

$$\langle \mu_\pm \nu_\pm \rangle = \frac{f(p)}{2N}\Big[\mathrm{Im}\,\langle q,p|q,-p\rangle \pm \mathrm{Im}\,\langle q,p|-q,p\rangle\Big]. \tag{2.18}$$

Even and odd subsets vs. whole set

At this point we must distinguish between two types of LIPR. First, one can evaluate the LIPR by averaging (in the numerator and the denominator) over only the even states (as done in this Chapter) or only the odd states. In that case one obtains

$$\begin{aligned} \mathrm{LIPR}(\mathrm{q},\mathrm{p};\{|\mathrm{n},\pm\rangle\}) &= 2+\frac{\left|\langle q,p|q,-p\rangle \pm \langle q,p|-q,p\rangle\right|^2}{\left(\langle q,p|q,p\rangle \pm \langle q,p|-q,-p\rangle\right)^2} && (2.19)\\ &= 2+\frac{\left|\langle q,p|T|q,p\rangle \pm \langle q,p|RT|q,p\rangle\right|^2}{\left(\langle q,p|q,p\rangle \pm \langle q,p|R|q,p\rangle\right)^2}. && (2.20) \end{aligned}$$

On the other hand, one can evaluate the LIPR by averaging over both the even and the odd states. In that case,

$$\begin{aligned} \mathrm{LIPR}\left(\mathrm{q},\mathrm{p};\{|\mathrm{n},+\rangle\}\cup\{|\mathrm{n},-\rangle\}\right) &= 2+\frac{\left|\langle q,p|q,-p\rangle\right|^2+\left|\langle q,p|-q,p\rangle\right|^2}{\langle q,p|q,p\rangle^2+\langle q,p|-q,-p\rangle^2} && (2.21)\\ &= 2+\frac{\left|\langle q,p|T|q,p\rangle\right|^2+\left|\langle q,p|RT|q,p\rangle\right|^2}{\langle q,p|q,p\rangle^2+\langle q,p|R|q,p\rangle^2}. && (2.22) \end{aligned}$$

The required matrix elements can be worked out from Eq. (2.4):

$$|\langle q,p|q',p'\rangle| = \exp\left[-\frac{(q-q')^2}{4\sigma^2}-\frac{\sigma^2(p-p')^2}{4}\right]. \tag{2.23}$$

The three LIPRs—Eq. (2.5.4) for even or odd symmetry and Eq. (2.5.4) for both symmetry classes put together—all can be shown to give results between 2 and 3. In fact, outside the region near $q=p=0$ the three versions are identical; they all predict bright symmetry lines near $q=0$ (or near any q associated with a parity symmetry) of the form

$$\mathrm{LIPR}(\mathrm{q},|\mathrm{p}|\gg 1/\sigma) = 2+\mathrm{e}^{-2\mathrm{q}^2/\sigma^2} \tag{2.24}$$

and bright symmetry lines near $p=0$ of the form

$$\mathrm{LIPR}(|\mathrm{q}|\gg\sigma,\mathrm{p}) = 2+\mathrm{e}^{-2\mathrm{p}^2\sigma^2}. \tag{2.25}$$

We note that the point $p = q = 0$, near which the three LIPR definitions given above do not agree, will always be on a short periodic orbit (e.g., the horizontal bounce orbit of the stadium billiard), and so the behavior there in any case cannot be determined without taking dynamical scar effects into account (see Section 2.5.5).

Curved symmetry lines

The curved streaks (3) and (4) also correspond to classical structures related to the system's symmetry. Consider first the streak labeled (3) in Fig. 2.4. It corresponds to trajectories with a vertical segment in the endcap region. In the desymmetrized quarter-stadium, these trajectories are incident normally on the $y = 0$ horizontal boundary, so a Gaussian test state with zero tangential momentum placed at the point of normal incidence $(q', p = 0)$ would give rise to an enhanced LIPR of 3, as for the zero-momentum line (5) above. Of course, this point of normal incidence exists only on the boundary of the *quarter*-billiard. At other points along the same orbit, including points that also live on the boundary of the full stadium, a Gaussian with momentum aligned along the trajectory will be close to, but not exactly equal to, a time iterate of this zero-momentum Gaussian living on the lower boundary. The exact time iterate of this purely real Gaussian will in general be a Gaussian centered on the other periodic point (q, p) but with a (possibly complex) width σ somewhat different from that used in our test state. Thus, our test Gaussian at (q, p) will have significant overlap with an iterate of a purely real state at $(q', p = 0)$ and result in an enhanced LIPR somewhere between 2 and 3. Very little wavepacket rotation or stretching occurs during the short vertical trip between the $y = 0$ line and the endcap;

it is for this reason that the streak (3) is so strong. Further iterations of the dynamics will result in additional streaks of enhanced IPR; these however will be much weaker due to the additional stretching which makes a circular wavepacket centered at these distant points less closely related to the evolved version of the $(q', p = 0)$ real wavepacket. All streaks in the LIPR plots which have not been explicitly identified in Fig. 2.4 may be explained as further iterates of symmetry lines we have discussed explicitly. The rather strong streak (4), associated with trajectories going through the center of the stadium, has a completely analogous explanation.

Quantitative contribution of the symmetry lines to the IPR

We can now estimate the contribution of the symmetry lines to the predicted average IPR. There are twelve strong roughly vertical streaks (including the curved streaks (3) and (4)) in the region $p \neq 0$ and one horizontal streak at $p = 0$. We assume temporarily the curved symmetry lines to have a central height near 3 and width of order σ just as for the $x = 0$ symmetry line. Integrating the area under these streaks assuming Eqs. (2.24) and (2.25) gives a predicted IPR from symmetry effects alone of 2.51 (in the energy range from $k = 100$ to 150). Since the bouncing-ball region, which covers about 6% of phase space, is excluded in the above analysis, we should increase the prediction by 6% to 2.66 for the purpose of comparison with the numerical value of 2.70 ± 0.07. In the energy range from $k = 200$ to 225 the corresponding predicted IPR is 2.40, or 2.51 corrected for the bouncing-ball region, whereas the numerical value is 2.82 ± 0.08 (the discrepancy in the higher-energy region remains unexplained). Thus, it seems that inclusion of symmetry and bouncing ball effects is sufficient to yield a quantitative understanding of much of the excess average IPR in Table 2.2. Interestingly, scarring plays little role in the excess average IPR, but a very important role in understanding the phase-space structure.

LIPR using real wavepackets

In light of our claim in Sec. 2.5.4 that complex Gaussian test states living on the boundary of the billiard are not ideal for probing eigenstates in a system with time reversal symmetry, it is natural to ask what test states would be more appropriate. A natural approach would be to symmetrize the test wavepackets with respect to time reversal symmetry, by taking their real or imaginary parts. Equivalently, since the eigenstates are chosen to be pure real, one may simply take the real or imaginary parts of the overlaps $\langle q,p|n\rangle$. We note that one might also consider symmetrizing the test Gaussians with respect to parity symmetry, which is also present in our system. This, however, would have no effect on any observed quantities, since all of the eigenstates already respect parity symmetry. We then define

$$\mathrm{LIPR_{real}(q,p)} = \frac{\frac{1}{\mathrm{N}}\sum_{\mathrm{n=1}}^{\mathrm{N}} \left|\mathrm{Re}\,\langle q,p|n\rangle\right|^4}{\left(\frac{1}{\mathrm{N}}\sum_{\mathrm{n=1}}^{\mathrm{N}} \left|\mathrm{Re}\,\langle q,p|n\rangle\right|^2\right)^2}\,. \tag{2.26}$$

This quantity is plotted in Fig. 2.5, and should be compared with the corresponding pictures in Fig. 2.2 for the LIPR of the original complex wavepackets. Because the new test states are real, the background LIPR rises from 2 to 3, while the LIPR along the straight symmetry lines remains unchanged (at 3).

Naively, one might expect the desymmetrization process described here to

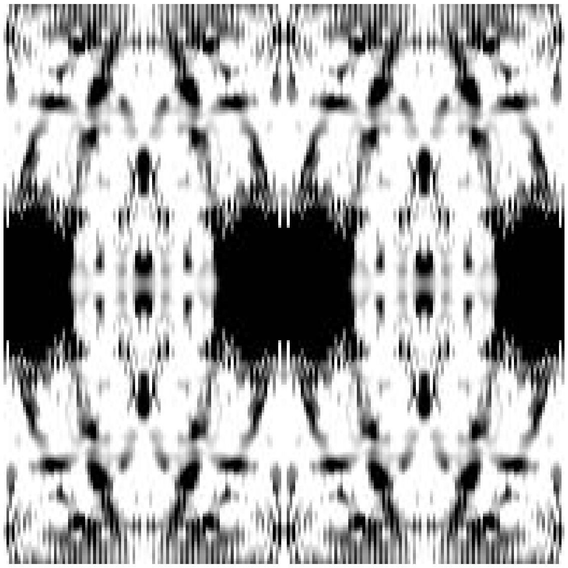


Figure 2.5: LIPR for real test wavepackets (see Eq. (2.26)) in phase space as a function of q running horizontally from 0 to $4+2\pi$ and p/k running vertically from -1 to 1. The axes and parameters are the same as those used in Fig. 2.2.

completely eliminate all symmetry effects, leading to flat $\mathrm{LIPR_{real}(q,p)} = 3$ behavior, with the exception of dynamics-related enhancement in the bouncing-ball region and near isolated unstable periodic orbits. Instead, one finds that although the straight symmetry lines ($p = 0$ and the parity symmetry lines) have indeed disappeared in Fig. 2.5, $\mathrm{LIPR_{real}}$ along the curved symmetry lines (e.g. the streaks identified as (3) and (4) in Fig. 2.4) is enhanced from 3 to ≈ 4.5, making these curved streaks just as visible as in the original plot of Fig. 2.2. This is easy to understand based on our analysis of these streaks in Sec. 2.5.4. We saw there that wavepackets located on the curved streaks are time-evolved versions of real wavepackets. So let r_n be the overlap of eigenstate $|n\rangle$ with a real wavepacket (i.e. either one having $p = 0$ or one that lives on a parity symmetry lines). r_n of course is distributed as a real Gaussian variable. Now the overlap of the same eigenstate $|n\rangle$ with a time-evolved copy of the real wavepacket will be given by $r'_n = r_n \exp i\theta_n$, where in the semiclassical regime θ_n can be taken to be a random phase. The LIPR for r'_n is of course equal to that of r_n (i.e., 3), as the phase has no effect on the intensity $|r'_n|^2 = |r_n|^2$. If, however, we take the real part of the wavepacket living on the curved symmetry line, we must look at the quantity

$$\mathrm{LIPR}_{\mathrm{real}} = \frac{\frac{1}{\mathrm{N}} \sum_{\mathrm{n}=1}^{\mathrm{N}} (\mathrm{Re}\ \mathrm{r}'_{\mathrm{n}})^4}{\left(\frac{1}{\mathrm{N}} \sum_{\mathrm{n}=1}^{\mathrm{N}} (\mathrm{Re}\ \mathrm{r}'_{\mathrm{n}})^2\right)} = \frac{9}{2} \tag{2.27}$$

because $r'_n = r_n \cos\theta_n$, where r_n is Gaussian-distributed and θ_n is a random phase. This is in contrast with the LIPR of 3 obtained on the straight symmetry lines as well as in the background (the LIPR on the straight symmetry lines would also be enhanced to 9/2 if we had let the eigenstates have random phases instead of adopting the convention where they are all real). Thus, there appears to be no natural and simple way to eliminate all symmetry effects on the LIPR (and thus, also on the IPR) in this time-reversal invariant system.

2.5.5 Dynamics

In this section we will consider the LIPR for a system with no time-reversal symmetry, where the wavefunctions are complex instead of pure real, and the arguments of Section 2.5.4 concerning symmetry lines do not apply. In such a system the LIPR plot would be far more uniform, and most of the remaining structure would consist of isolated regions of enhanced LIPR, against a background of LIPR = 2 (corresponding to the statistics of the square of a complex Gaussian variable). The streaks would be all gone, including the vertical parity-line streaks, which also depend on time-reversal symmetry for their existence. Thus in this section we study the effect of dynamics alone on the LIPR. First we give a general theory for computing the LIPR classically anywhere in phase space, and then we show an easier way to compute the LIPR in the neighborhood of a periodic orbit. In the process, we will also see how to treat the case of a periodic orbit that happens to lie on a symmetry line in a time-reversal invariant system. In the following section (Sec. 2.5.6), we will describe an alternative way to treat parity symmetry and dynamical effects in a unified framework, to better understand scar effects in a system with symmetry.

Full dynamics treatment

In this section we show how to compute the enhancement of the LIPR due to the short-time dynamics of the system. Since we are considering the surface-of-section wavefunctions, the classical dynamics that is important is the classical Poincaré map M_{Cl}. However, since the effects we are seeking are quantum mechanical, we need to start from the corresponding *quantum* Poincaré map. For this we use the operator $M(k)$ that enters into the boundary integral method:

$$\langle q|M(k)|\phi\rangle \equiv \oint dq' K_{\mathrm{ex}}(q,q';k)\,\langle q'|\phi\rangle \ , \tag{2.28}$$

where the kernel is [9]

$$K_{\mathrm{ex}}(q,q';k) \equiv -\frac{ik}{2} H_1^{(1)}(k\,|\mathbf{r}(q)-\mathbf{r}(q')|)\ \hat{\mathbf{n}}(q)\cdot[\mathbf{r}(q)-\mathbf{r}(q')], \tag{2.29}$$

$\mathbf{r}(q)$ is a point on the billiard wall at arc length q, and $\hat{\mathbf{n}}(q)$ is the inward-pointing unit normal vector at q. Does $M(k)$ as defined generate the correct quantum dynamics?

In a sense that is a philosophical question since the Poincaré map has no obvious exact quantum counterpart. However $M(k)$ satisfies enough properties expected of a quantum Poincaré map that we shall use it:

1. To within a semiclassical approximation it agrees with the semiclassical Poincaré map [9] (which is easier to define [11]).
2. It is approximately unitary within a subspace whose dimension is given by the area of the Poincaré section in units of Planck's constant, and exponentially small beyond that dimension.
3. If a boundary function $|n\rangle$ corresponds to a wavefunction that satisfies the boundary conditions at wavenumber $k = k_n$, then $|n\rangle$ is mapped onto itself under the action of $M(k)$:

   $$M(k_n)\,|n\rangle = |n\rangle\,. \tag{2.30}$$

   Indeed, this is the boundary integral method criterion for an eigenstate.

The following derivation should be considered suggestive rather than rigorous. For example, $M(k)$ is approximately unitary but it has actual eigenvalues both inside and outside the unit circle. This leads to difficulties analytically continuing it to the region of interest.

Given properties 2 and 3 above, it is clear that one crude way to find surface eigenstates of the system would be to apply $M(k)$ repeatedly to an arbitrary initial state $|\phi\rangle$ and average the results:

$$|n\rangle \quad \propto \quad \lim_{\epsilon\to 0^+} \sum_{j=0}^{\infty} [M(k_n) - \epsilon]^j \,|\phi\rangle \tag{2.31}$$

$$= \quad G(k_n)\,|\phi\rangle\,, \tag{2.32}$$

where

$$G(k) \equiv \lim_{\epsilon\to 0^+} \frac{1}{1 - M(k) + \epsilon} \tag{2.33}$$

is similar to a Green's function. Since the component of $|\phi\rangle$ that projects onto $|n\rangle$ is unchanged by the action of $M(k_n)$, whereas the other components are multiplied by a complex phase, the only component that is not averaged out by the above procedure is the projection on $|n\rangle$.

Whenever k passes through an eigenvalue k_n, one of the eigenvalues $e^{i\alpha_n(k)}$ of $M(k)$ passes along the unit circle through 1 and $G(k)$ has a singularity. The singular part of G can be written schematically as

$$\frac{1}{\pi}\,\mathrm{Re}\,G(k) = \sum_n \frac{|n\rangle\,\delta(k-k_n)\,\langle n|}{|d\alpha_n(k)/dk|}. \tag{2.34}$$

The denominator gives the velocity of rotation of the nth eigenvalue as it passes through 1. It can be shown from semiclassical arguments that the average of these velocities is given by the average length of the trajectory segments corresponding to one Poincaré map:

$$\left\langle \frac{d\alpha_n}{dk} \right\rangle_n = \langle \ell \rangle = \pi \frac{A}{L}. \tag{2.35}$$

The second equality follows from a classical theorem well known in acoustics [90], where A is the billiard's area and L the length of its perimeter. For a chaotic system, Eq. (2.35) holds not only in an average sense but also holds approximately for individual eigenphases. Therefore we take $\langle \ell \rangle$ outside of the summation, yielding the useful expression

$$\frac{A}{L}\,\mathrm{Re}\,G(k) \approx \sum_n |n\rangle\, \delta(k-k_n)\, \langle n|\,. \tag{2.36}$$

Note that the sum is over states of the full system, each with its distinct eigenvalue k_n.

It is a simple warm-up exercise to compute the denominator of the LIPR:

$$\begin{aligned}
\sum_n |\langle q,p|n\rangle|^2 &= \sum_n \langle q,p|n\rangle \langle n|q,p\rangle && (2.37)\\
&= \sum_n \langle q,p|n\rangle \int dk\delta(k-k_n)\, \langle n|q,p\rangle && (2.38)\\
&= \frac{A}{L}\int dk\,\mathrm{Re}\,\langle q,p|G(k)|q,p\rangle && (2.39)\\
&= \frac{A}{L}\,\mathrm{Re}\int dk\Big[\langle q,p|q,p\rangle + \langle q,p|M(k)|q,p\rangle + \Big\langle q,p\Big|M^2(k)\Big|q,p\Big\rangle \\
&\qquad\qquad + \cdots\Big]. && (2.40)
\end{aligned}$$

Now note that $M(k)$ involves phases that vary rapidly with k; therefore only the first term in the brackets survives the integral over k. We are left with

$$\sum_n |\langle q,p|n\rangle|^2 \approx \frac{A\Delta k}{L}\,\langle q,p|q,p\rangle\,, \tag{2.41}$$

where Δk is the range of k over which the averaging is done.

To compute the fourth moment, we proceed as follows:

$$\begin{aligned}
\sum_n |\langle q,p|n\rangle|^4 &= \sum_n \langle q,p|n\rangle \langle n|q,p\rangle \langle q,p|n\rangle \langle n|q,p\rangle && (2.42)\\
&= \sum_{n,m} \langle q,p|n\rangle \int dk\, \delta(k-k_n)\, \langle n|q,p\rangle && (2.43)\\
&\quad \langle q,p|m\rangle \int dk'\, \delta(k'-k_m)\, \langle m|q,p\rangle\, \delta_{nm}. && (2.44)
\end{aligned}$$

Because of the $\delta(k-k_n)$ and $\delta(k'-k_m)$ factors, the Kronecker delta δ_{nm} may be replaced by $C\delta(k-k')$ where C is an undetermined constant with dimensions of momentum which is needed to make its product with the delta function dimensionless. Continuing, then, we have

$$\begin{aligned}
\sum_n |\langle q,p|n\rangle|^4 &= C\left(\frac{A}{L}\right)^2 \int dk\, dk'\, \mathrm{Re}\, \langle q,p|G(k)|q,p\rangle \\
&\qquad\qquad \times \mathrm{Re}\, \langle q,p|G(k')|q,p\rangle\, \delta(k-k') && (2.45)\\
&= C\left(\frac{A}{L}\right)^2 \int dk \Big[\mathrm{Re}\, \langle q,p|G(k)|q,p\rangle\Big]^2 && (2.46)\\
&= C\left(\frac{A}{L}\right)^2 \int dk \sum_{j,j'} \mathrm{Re}\langle q,p|M^j(k)|q,p\rangle\, \mathrm{Re}\langle q,p|M^{j'}(k)|q,p\rangle && (2.47)\\
&\approx C\left(\frac{A}{L}\right)^2 \int dk \sum_j \Big[\mathrm{Re}\langle q,p|M^j(k)|q,p\rangle\Big]^2 && (2.48)\\
&\approx \frac{C}{2}\left(\frac{A}{L}\right)^2 \int dk \sum_j \Big|\langle q,p|M^j(k)|q,p\rangle\Big|^2 . && (2.49)
\end{aligned}$$

where in going to Eq. (2.48) we have used the fact that for $j \neq j'$, the two phases in the integrand are approximately uncorrelated, and the integral averages to zero. Eq. (2.49) (in obtaining which we have used the fact that for $j \neq 0$ the phase of $\langle q,p|M^j(k)|q,p\rangle$ is a rapidly changing function of k, and so may be regarded as random in the integration; the $j = 0$ term is irrelevant in the infinite sum) may be interpreted as expressing the proportionality between the LIPR and the mean quantum return probability, averaged over *all* times. As a result of the reloading effect [68] the mean long-time return probability is proportional to the sum of the *short-time* recurrences, which may be cut off at the mixing time, $j_{\mathrm{mix}} \equiv T_{\mathrm{mix}}/T_B$:

$$\sum_n |\langle q,p|n\rangle|^4 \approx \frac{C'}{2}\left(\frac{A}{L}\right)^2 \int dk \sum_{j=-j_{\mathrm{mix}}}^{j_{\mathrm{mix}}} \Big|\langle q,p|M^j(k)|q,p\rangle\Big|^2 . \qquad (2.50)$$

The constant of proportionality changes in going from Eq. (2.49) to Eq. (2.50), as is reflected in our change of notation from C to C'. The mean quantum return probability is inversely proportional to the number of Planck-sized cells in phase space, or inversely proportional to the Heisenberg time, and thus C' is inversely proportional to the Heisenberg time.

But at times shorter than the mixing time, the quantum map $M^j(k)$ can be approximated semiclassically in terms of the classical Poincaré map $M^j_{\rm Cl}$ (which does not depend on k) times a phase that varies rapidly with k:

$$\Big\langle q,p\Big|M^j(k)\Big|q,p\Big\rangle \approx \Big(q,p\Big|M^j_{\rm Cl}\Big|q,p\Big)^{1/2}\exp(ik\ell_j), \tag{2.51}$$

where $|q,p)$ represents a *classical* Gaussian probability distribution centered at (q,p) (notationally distinguished from a quantum wavefunction by the round brackets), $M^j_{\rm Cl}$ is the classical Poincaré map iterated j times, and ℓ_j is the length of the corresponding classical trajectory. $\Big(q,p\Big|M^j_{\rm Cl}\Big|q,p\Big)$ is the classical overlap of the original distribution with the j-times iterated distribution:

$$\Big(q,p\Big|M^j_{\rm Cl}\Big|q,p\Big) \equiv \int dq'\,dp' P_{M^j_{\rm Cl}(q,p)}(q',p')P_{q,p}(q',p') \tag{2.52}$$

So finally we have

$$\sum_n |\langle q,p|n\rangle|^4 \approx \frac{C'}{2}\left(\frac{A}{L}\right)^2 \Delta k \sum_{j=-j_{\rm mix}}^{j_{\rm mix}} \Big(q,p\Big|M^j_{\rm Cl}\Big|q,p\Big)\,. \tag{2.53}$$

Comparing Eq. (2.41) and Eq. (2.53) we obtain for the LIPR

$$\begin{aligned} \mathrm{LIPR} &= \frac{N\sum_n |\langle q,p|n\rangle|^4}{\Big(\sum_n |\langle q,p|n\rangle|^2\Big)^2} & (2.54)\\ &= \frac{C'N}{2\Delta k}\frac{\sum_{j=-j_{\rm mix}}^{j_{\rm mix}}\Big(q,p\Big|M^j_{\rm Cl}\Big|q,p\Big)}{\Big(q,p\Big|q,p\Big)}, & (2.55)\end{aligned}$$

where the projections in both the numerator and in the denominator are classical. The constant C' can now be fixed by requiring that the LIPR, when evaluated far from any short periodic orbits (where scarring plays no role), should give the RMT value which we call LIPR(RMT). [As discussed above, LIPR(RMT) = 2 in a system with no time-reversal symmetry; in the presence of time-reversal symmetry, LIPR(RMT) reaches the value of 3 on certain symmetry lines, as we have seen in Section 2.5.4. The LIPR contribution obtained from short-time dynamics must always be multiplied by the appropriate factor given by symmetry considerations; LIPR(RMT) is inherently an effect arising from Heisenberg-time behavior and so must be considered separately from short-time contributions to the LIPR.]

Away from short periodic orbits there will be no recurrences, so only the $j = 0$ in the above sum contributes. Thus, $C' = (2\Delta k/N)\,\mathrm{LIPR(RMT)}$ and

$$\mathrm{LIPR(q,p)} = \mathrm{LIPR(RMT)} \frac{\sum_{\mathrm{j=-j_{mix}}}^{\mathrm{j_{mix}}} \left(\mathrm{q,p}\middle|\mathrm{M_{Cl}^{j}}\middle|\mathrm{q,p}\right)}{\left(q,p\middle|q,p\right)}. \tag{2.56}$$

The $j \neq 0$ terms of the sum give an enhancement to the LIPR whenever a point (q,p) is mapped near itself by an iterate of the classical Poincaré map; i.e., near periodic orbits. In the next section we will describe how to evaluate the LIPR in the neighborhood of a periodic orbit.

Periodic orbits, linear theory

The dynamical description of the LIPR developed in the previous section provides a method of computing the LIPR classically, anywhere in phase space, by computing overlaps of classical Gaussian probability distributions with themselves iterated under the classical Poincaré map M_{Cl}. We found that the LIPR is enhanced when the short-time overlap is large. This of course is most dramatically the case in the neighborhood of periodic orbits. In this section we show how to compute the LIPR in the neighborhood of a periodic orbit, using only the properties of that one orbit. We also generalize to the case of stable and unstable manifolds with arbitrary orientation.

Fig. 2.2 confirms the omnipresence of scarring in the eigenstates of the stadium billiard, as we now argue (the identification of peaks in Fig. 2.2 with specific short periodic orbits will be made below in Sec. 2.5.7, 2.5.7 in particular). In the general theory of scarring, a wavepacket launched on a periodic orbit has a local density of states with a linear or short-time envelope (coming from dynamics linearized around the orbit) that consists of an oscillating function of energy; isolated homoclinic recurrences, if present, will give rise to further fluctuation multiplying this envelope [68]. The concentration of the local density of states in some energy regions and avoidance of others leads to an enhanced LIPR in Eq. (2.7). Thus the peaks in the LIPR plot in Fig. 2.2 at the positions of short periodic orbits must be attributed to scarring, even though for the most part the scars are too weak to be visible in coordinate-space plots of individual wavefunctions, such as in Fig. 2.1. The scars are there in the sense of non-χ^2 variation of wavefunction intensity on these periodic orbits, as one scans through the eigenstates of the system. Since there are several periodic orbits with appreciable scarring, any given eigenvalue is likely to lie near a maximum in the linear envelope of at least some of them, and the corresponding eigenstate will more likely than not be scarred on these orbits, and antiscarred on the others. Examination of the spectra corresponding to different short periodic orbits reveals that, first, as one varies the periodic orbit the position of the envelopes in the local density of states varies so as to cause any given energy to lie under the peaks of the envelopes for several different orbits. Secondly, we find that most of the envelopes in the local density of states are filled in an egalitarian way (for a discussion of egalitarian versus totalitarian filling of the spectral envelopes, see Kaplan and Heller [68]), which increases the likelihood of at least weak scarring along that orbit for each eigenstate under that envelope. For typical scarred wavefunctions, the scarring will not be visible in coordinate space, even though the phase space intensity is greatly enhanced at the periodic points, as as one can see in the Husimi plots. Occasionally, the scarring may be strong enough to be also visible in a coordinate-space plot; an illustration of strong multiple scarring can be found in the state at $k = 105.608$, which is simultaneously scarred along the bowtie and diamond orbits; cf. Fig. 2.1(c).

In summary, we find a distribution of wavefunction intensities on a periodic orbit which differs from the χ^2 distribution that would be expected on the basis of random-matrix theory, as illustrated in Fig. 2.6. In Fig. 2.6(a) we show a histogram of wavefunction intensities measured using a Gaussian centered on the bowtie orbit for all states lying between $k = 100$ and $k = 150$. Because the bowtie orbit lies on a symmetry line, the random-matrix-theory prediction is a χ^2 distribution in one degree of freedom (where we have set 6% of the intensities to zero, corresponding to the bouncing-ball states), which fails to account for the large number of high-intensity and low-intensity peaks in Fig. 2.6(a). In the presence of scarring, the χ^2 distribution must be modulated by the linear spectral envelope, which, for a Lyapunov exponent of 2.29 for the bowtie orbit, varies from 0.4 to 2.3 periodically in momentum. The period in energy is given by $2\pi/T$ where $T = 2.6T_B$ is the period of the bowtie orbit. As a result of this modulation, there are many more very small and very large intensities in Fig. 2.6(a) than would be expected from the naive Porter-Thomas (RMT) picture. As plotted on a logarithmic scale, the effect is more pronounced for the larger intensities, but even at low intensities we expect an enhancement by about 20%, which is what is indicated by the data between $\langle I \rangle/100$ and $\langle I \rangle/10$, where the statistics are good; below $\langle I \rangle/100$ the statistics are too poor to distinguish reliably between the scar theory and random matrix theory predictions. In particular, the amount of strong scarring observed is quantitatively consistent with the predicted scarring corrections to random-matrix theory: for instance, the number of wavefunction intensities greater than 10 times the mean is found to be 10 numerically, while the scar-corrected prediction is 13 ± 3 and the uncorrected random-matrix prediction is only 3 ± 3. At a generic point in phase space, not lying on a scar or on a symmetry line, we expect a χ^2 distribution in two degrees of freedom, and Fig. 2.6(b) shows the numerical distribution of wavefunction intensities at a generic point in phase space, in satisfactory agreement with the random-matrix-theory prediction for intensities greater than $\sim \langle I \rangle/10$ (without scarring corrections, since there is no scarring at a generic point in phase space, but with 6% of the intensities set to zero, namely those corresponding to bouncing-ball states). The bouncing-ball states appear of course in the numerical data at intensities $I < \langle I \rangle/10$; the total area under the curve at these small intensities is about 6%). Note the smaller number of very small and very large intensities as compared to the scarring case, Fig. 2.6(a). The 2-degrees-of-freedom distribution also falls off much faster at large intensities than the 1-degree-of-freedom prediction appropriate for a symmetry line.

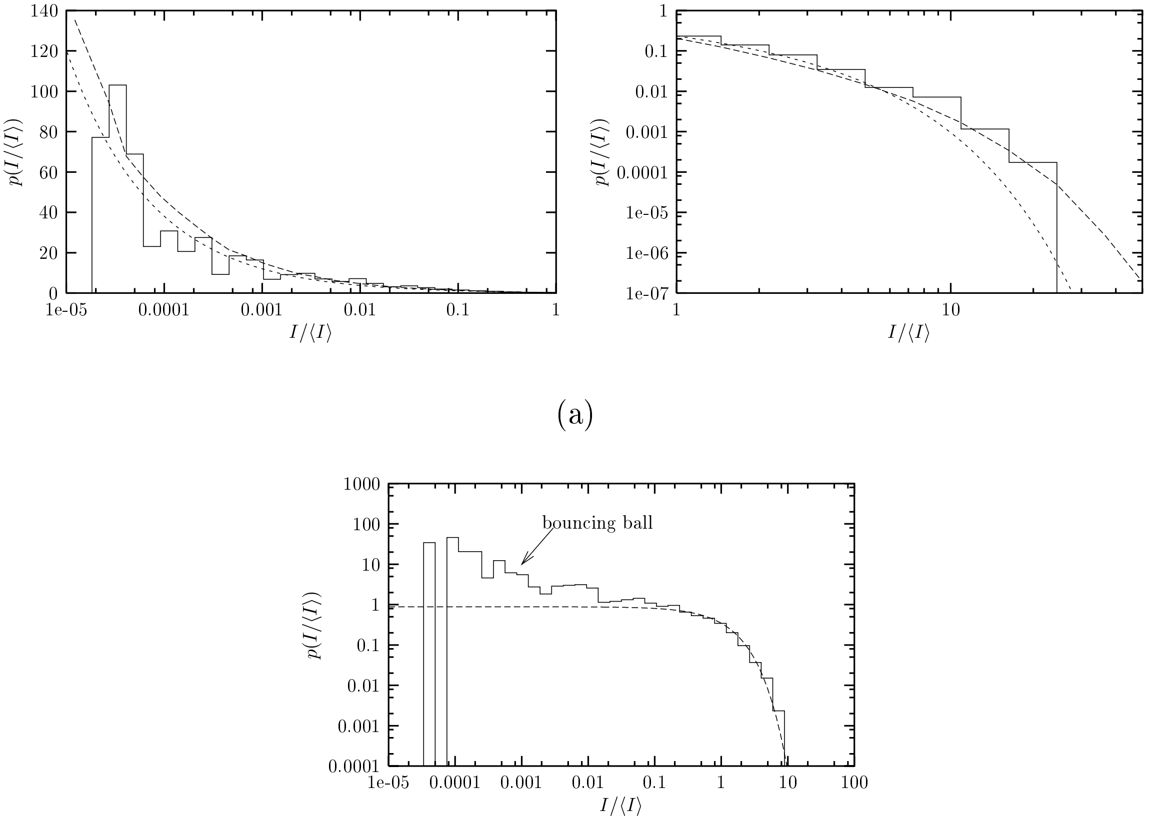
Scar intensity from linear scar theory We now turn to the calculation of the heights of the scarring peaks in Fig. 2.2 based on a linear expansion around periodic orbits. Consider the surface-of-section map in Birkhoff coordinates; we will be interested in the short-time linearized dynamics in the vicinity of a periodic orbit [49]. The stable and unstable invariant manifolds, defined as the locus of points that approach the periodic orbit under infinite iteration of the mapping forwards or backwards in time, respectively, will in general have an arbitrary orientation. In the special case that the invariant manifolds are aligned with the coordinate axes centered on the periodic orbit of interest—the simplest possibility—we have $q(t) = q(0)e^{\lambda t}$ and $p(t) = p(0)e^{-\lambda t}$, where q and p are measured relative to the periodic point. An initial Gaussian corresponding to a classical distribution of probability, $P_0(q,p) = Ce^{-q^2/\sigma^2 - \sigma^2 p^2}$ (where $C = (2/\pi)^{1/2}$ is a normalization constant) will map after n iterations to

$$P_j(q,p) = C \exp\left[-e^{-2\lambda j} q^2/\sigma^2 - e^{2\lambda j}\sigma^2 p^2\right] . \tag{2.57}$$

The overlap with the initial distribution will then be

$$\int dq\; dp\; P_j(q,p) P_0(q,p) = \frac{1}{\cosh \lambda j} . \tag{2.58}$$

We note as an aside that there is no natural way to choose the normalization constant C based on purely classical considerations, since classically the return probability of a distribution is not a meaningful concept. Instead, we are implicitly thinking



(a)

(b)

Figure 2.6: Probability distribution of wavefunction intensities $I_n = |\langle q,p|n\rangle|^2$ for the 1746 energy levels between $k = 100$ and 150. (a) For (q,p) situated on the bowtie orbit. Solid line, numerical distribution of wavefunction intensities; dashed line, scar-corrected prediction of random-matrix theory with one degree of freedom; dotted line, random-matrix prediction, uncorrected for scarring. For purposes of illustration, the low-intensity part is shown in the figure on the left in a semilog plot and the high-intensity tail on the right in a log-log plot. (b) Same as (a) but for (q,p) at a generic point in phase space, and the dashed line is now the random-matrix theory prediction for two degrees of freedom (without scar corrections, but adjusted for the 6% nearly vanishing intensities due to bouncing-ball states).

of Eq. (2.58) as representing the square of a *semi*-classical return amplitude of a Gaussian wavepacket, which is a well-defined physical quantity, and is given by the square root of Eq. (2.58), times some irrelevant phase associated with the action (in units of $\hbar$) of the periodic orbit. This fixes the normalization constant C above.

Due to the reloading effect mentioned in the preceding Section (see also [68]) in which returning homoclinic orbits contribute in phase, the mean return probability for long times at this periodic orbit, and hence the LIPR, which equals a constant times the long-time return probability, will be equal to LIPR(RMT) times

$$S_1 \equiv \sum_{j=-\infty}^{\infty} \frac{1}{\cosh \lambda j} \,. \tag{2.59}$$

Strictly speaking, the sum in Eq. (2.59) should extend only over times short compared to the mixing time, i.e. $|j| < \lambda^{-1} \log(kL)$; however this cutoff is irrelevant in the semiclassical limit $kL \gg 1$ in which we are working.

It is desirable to extend this result to the case when the invariant manifolds are not perpendicular. This is equivalent to considering perpendicular invariant manifolds but with a non-circular Gaussian (as one can go from one case to the other via a canonical transformation). Let $x = (q, p)$ and A be the quadratic form defining the initial Gaussian centered on a periodic orbit,

$$P_0(x) = Ce^{-x^T A x} \,, \tag{2.60}$$

with $\det A = 1$. We define J to be the Jacobian of $M_{\rm Cl}$ at the periodic orbit; then

$$P_j(x) = C \exp\left[-x^T (J^{-j})^T A J^{-j} x\right] \tag{2.61}$$

and the overlap is

$$\int dq\, dp\, P_j(q,p)P_0(q,p) = 2\sqrt{\frac{\det A}{\det[A + (J^{-j})^T A J^{-j}]}} \tag{2.62}$$

(again, we see that using our normalization the overlap is unity at time $j = 0$). Thus, a better prediction of the LIPR peak heights at the periodic orbits—one which takes into account the local stable and unstable directions at the periodic orbit, and not only the stability exponent—is LIPR(RMT) multiplied by

$$S_2 \equiv \sum_{j=-\infty}^{\infty} 2\sqrt{\frac{\det A}{\det[A + (J^{-j})^T A J^{-j}]}} \,. \tag{2.63}$$

Note that S_2 reduces to S_1 when the invariant manifolds are perpendicular, and the Gaussian test state is oriented along these manifolds, as an easy calculation shows. Of course, using a canonical transformation we may easily see that for any orientation of the manifolds there is an infinite one-parameter family of 'optimal' Gaussian test states, all of which display maximum possible scarring in accordance with Eq. (2.59). Both S_1 and S_2 approach unity for long or highly unstable orbits, when the instability exponent λ becomes large.

Comparison of predicted with observed peak heights (table, scatter plot)

We have calculated S_1 and S_2 for the periodic orbits appearing in Fig. 2.7, and compare these theoretical predictions with the actual peak heights in Table 2.3. The predictions of S_1 and S_2 are based on the tangent map in the *desymmetrized* stadium.

All quantities appearing in Table 2.3 are either predicted or actual values of $LIPR(p,q)/2$ at a periodic point (p,q). In other words, we have divided out by the RMT prediction $LIPR = 2$, which is relevant for all of space space away from the

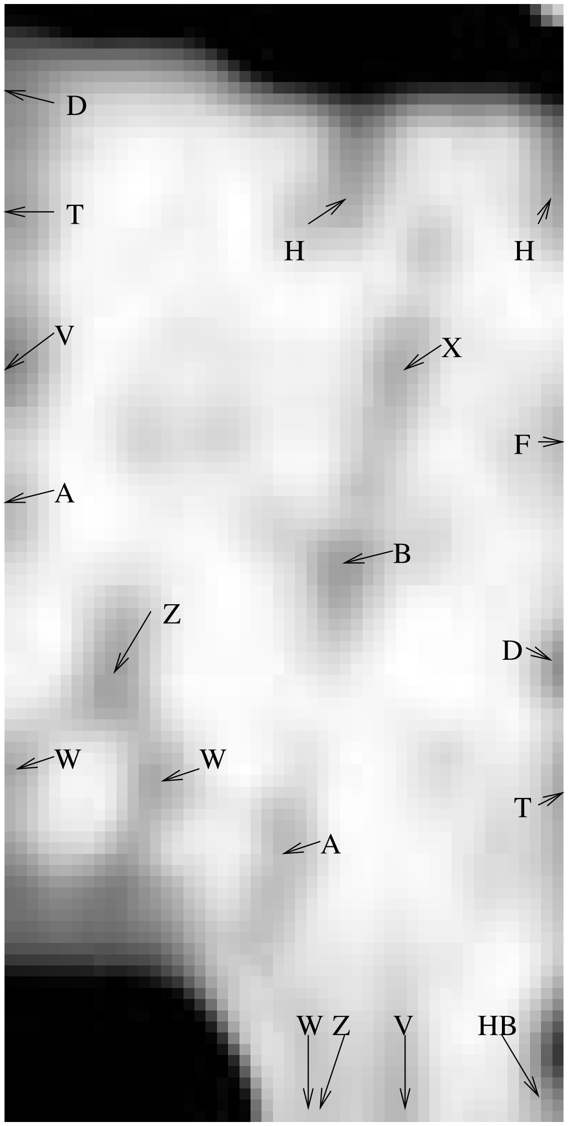


Figure 2.7: Principal dynamics-related structures in phase space of the inverse participation ratio plotted in Fig. 2.2 and the average-return probability plotted in Fig. 2.9. The labeling is according to Table 2.4 notation of scars on the principal short periodic orbits in Fig. 2.9(e), in an irreducible one-eighth region of phase space. The other regions of phase space are obtained by symmetry. The most prominent scarred orbits are W, Z, D, B, X and H.

Orbit	Symmetry	Lyapunov	S_1	S_2	actual	height	classical
		exponent λ			$k =$ 100-150	$k =$ 200-225	
HB	a	1.76	2.72	2.15	1.86	2.01	2.78
V_1	b	2.02	2.40	2.36	1.91	1.93	2.36
V_2	c		2.40	1.95	1.68	1.87	1.97
D_1	a	2.06	2.37	1.97	1.89	1.80	2.03
D_2	b		2.37	2.27	2.15	2.06	2.27
B_1	d	2.29	2.18	2.06	1.86	1.87	2.02
B_2	d		2.18	2.06	1.86	1.87	2.02
Z_1	e	2.38	2.10	2.04	1.69	1.88	2.04
Z_2	e		2.10	2.04	1.69	1.88	2.04
Z_3	c		2.10	1.77	1.53	1.55	1.85
X_1	df	2.58	2.00	1.94	1.91	2.03	2.07
X_2	df		2.00	1.94	1.91	2.03	2.07
A_1	b	2.60	1.98	1.86	1.53	1.55	1.85
A_2	d		1.98	1.86	1.59	1.51	1.90
A_3	d		1.98	1.86	1.59	1.51	1.90
W_1	b	2.68	1.94	1.94	1.71	1.86	1.95
W_2			1.29	1.27	1.68	1.78	1.29
W_3	c		1.94	1.70	1.66	1.81	1.84
W_4			1.29	1.27	1.68	1.78	1.29
H_1	a	3.26	1.74	1.73	2.01	2.37	2.00
H_2	f		1.74	1.74	1.78	2.12	1.99
H_3	f		1.74	1.73	1.78	2.12	2.00
T_1	b	3.83	1.64	1.64	1.74	1.89	1.77
T_2	f		1.64	1.56	1.15	1.26	1.61
T_3	f		1.64	1.56	1.15	1.26	1.61
F_1	a	4.90	1.55	1.53	1.58	1.63	1.95
F_2			1.03	1.04	1.14	1.18	1.02
F_3	d		1.55	1.53	1.50	1.62	1.57
F_4	d		1.55	1.55	1.50	1.62	1.57
F_5			1.03	1.03	1.14	1.18	1.02

Table 2.3: Scar peak heights (values of LIPR/2) on short periodic orbits, as predicted by the linearized semiclassical measures S_1, S_2 and by the exact brute-force classical simulation, defined in the text, to be compared with actual quantum-mechanically computed heights. Periodic orbit labels are defined in Table 2.4. The symmetry family to which each point on each periodic orbit belongs is labeled as follows: a, orbit hits center of the circular endcap; b, hits center of the straight segment; c, incident normal to the boundary; d, possesses a vertical segment in the endcap region; e, passes through the center of the stadium; and f, possesses a horizontal segment. The predictions in columns 3, 4 and 7 include the quantum symmetry correction factor of 3/2 for periodic points with special symmetry, as indicated in column 1.

symmetry lines. However, on the symmetry lines the RMT expectation is 3 rather than 2 (see Section 2.5.4), so the theoretical values S_1 and S_2 have been multiplied by 3/2 for those periodic points which do lie on symmetry lines, in order to compare with the actual LIPR/2 data in the fifth and sixth columns of Table 2.3. In fact, most of the points listed in Table 2.3 do lie on one or another symmetry line; the first column of the table indicates which symmetry class each point lies in (see definition of the symbols in the caption). As expected, S_1 fails to be an accurate measure of the LIPR peak heights because the test states in general are not aligned optimally with the invariant manifolds. With the inclusion of symmetry corrections, the agreement of S_2 is roughly correct, as can be seen in a scatter plot of the predicted versus calculated LIPR/2 peak intensities, Fig. 2.8; as shown in Fig. 2.8(c), the scatter is markedly worse when the symmetry corrections are omitted.

In order to appreciate the significance of these results, we should discuss the factors contributing to the uncertainty in the measured LIPR/2 values listed in columns 6 and 7 of Table 2.3. There are two issues: first, there may be a non-integral number of oscillations of the local density of states in the energy window considered. The size of this effect scales inversely with the number of scar oscillations in the energy window, and is only a few percent for the data presented in Table 2.3. A second, and bigger, source of error arises from the finite number of eigenvalues in the window. For random fluctuations, the uncertainty coming from this effect scales as the square root of the mean level spacing divided by the size of the window. This is why the windows from $k = 100$ to 150 and from $k = 200$ to 225 were taken to be as large as possible.

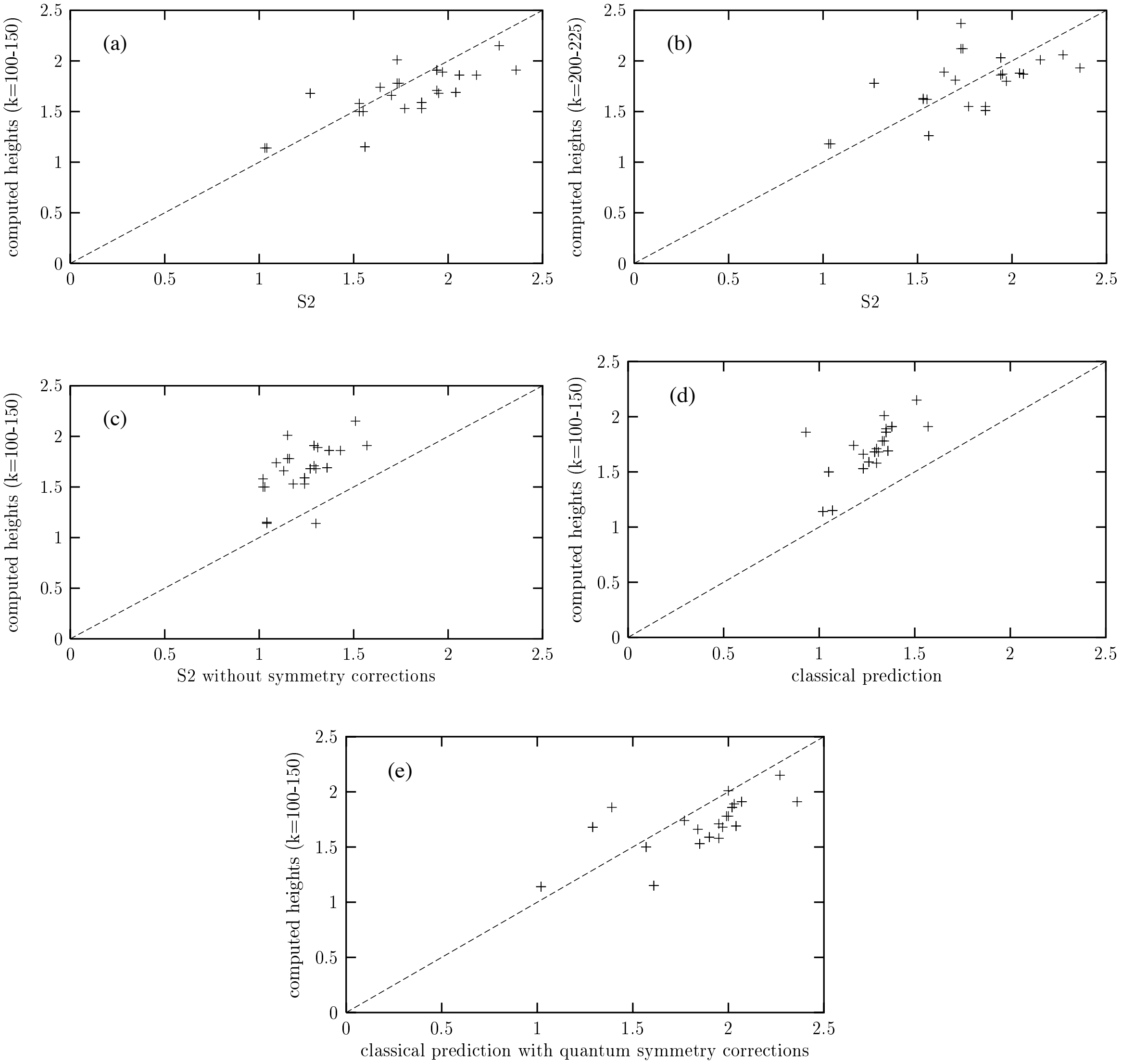


Figure 2.8: Scatter plots of computed peak heights in columns 5 and 6 of Table 2.3 versus the linearized semiclassical predictions S_2 in column 4 and the brute-force semiclassical prediction in column 7. (a) Eigenfunctions ranging from $k = 100$ to 150, symmetry corrections included in S_2; (b) $k = 200$ to 225, symmetry corrections again included in S_2; (c) $k = 100$ to 150, but symmetry corrections *not* included in S_2. Without symmetry corrections the agreement is much worse. (d) $k = 100$ to 150 versus brute-force classical prediction without symmetry correction and (e) with symmetry corrections included.

A monte carlo simulation was done of random wavefunction intensities constrained by the “linear” spectral envelope: it suggests that the statistical uncertainty in the LIPRs should range from 5% to 8% (depending on the stability matrix of the orbit in question) for periodic points on symmetry lines, and should be slightly smaller for those points that are not on any symmetry line. This level of fluctuation is roughly consistent with the dispersion in the scatter plots of Fig. 2.8, and does not alter the conclusion that S_2 together with symmetry factors is a good prediction of the peak heights. Thus, the symmetry lines emerge as essential to a quantitative understanding of the calculated LIPRs due to scarring. A final consideration supporting the validity of our symmetry analysis is the pattern of peak heights for the F orbit, on which periodic points F_1, F_3 and F_4 have a symmetry correction while F_2 and F_5 do not; thus, the predicted peak heights for the points having symmetry corrections are about fifty percent higher than for the others, and this predicted pattern is reproduced in the calculated heights.

Homoclinic recurrences The brute-force classical calculation at times large compared to T_B goes beyond S_1 and S_2 to include the effects of nonlinear homoclinic recurrences, which add to the predicted peak height. For each periodic point, the returning probability for an initial Gaussian distribution centered at that point was integrated up to a cutoff time of $5T_B$, which was chosen to be large compared to T_B but smaller than the mixing time $T_{\mathrm{mix}} \approx 7T_B$, after which the returning probability approaches a constant per unit time independent of position in phase space. In the semiclassical limit, where strong, identifiable recurrences at times between the initial decay of the wavepacket and the Heisenberg time can be neglected, the classical simulation reproduces S_2. The classical simulation reproduces, as it must, one feature of the quantum data: the peak heights it predicts agree for points related by symmetry, namely B_1 and B_2, Z_1 and Z_2, X_1 and X_2, A_1 and A_2, W_2 and W_4, H_2 and H_3, T_2 and T_3 and F_3 and F_4. In predicting the quantum heights, the brute-force classical calculation does no better than S_2 and makes predictions that are systematically high (after including symmetry effects). In order to investigate the energy dependence of the classically predicted peak heights, the classical simulation was run both at $k = 100$ and at $k = 200$, the difference being in the size of the Gaussian as a fraction of the phase space area. For almost all of the periodic orbits the results at the two energies agreed to within a few percent, which is of the same size as the discrepancy between the numerical quantum peak heights in those two energy ranges. This tells us that isolated, identifiable non-linear recurrences beyond the initial decay of the wavepacket away from the periodic orbit are not that important for computing the mean long-time return probability. Any such nonlinear effects would be strongly dependent on the size of the initial wavepacket, i.e. on the wavevector k.

The classical calculation for the average return probability can be performed anywhere in phase space, not only on the short periodic orbits considered in this section; see Sec. 2.5.7 below. There we shall see that we get a picture similar to Fig. 2.2, but with more diffuse peaks and the onset of mixing for $T > T_{\mathrm{mix}}$.

2.5.6 Dynamics plus symmetry

In the previous two sections, we have seen that discrete symmetries have two important and distinct effects on wavefunction localization properties. First, in the presence of time-reversal symmetry, the horizontal line of zero parallel momentum in the surface of section phase space, as well as vertical lines associated with parity symmetry, if any, both have a local IPR of 3, in contrast with the background LIPR of 2 that is seen away from the symmetry lines. The values LIPR $=3$ and LIPR $=2$ correspond to the random matrix theory predictions for real and complex Gaussian random wavefunctions, respectively. Furthermore, short-time iterates of these symmetry lines result in additional (curved) symmetry lines where the LIPR takes values intermediate between 2 and 3. When a periodic point happens to be located on a symmetry line (which in a highly symmetric system such as the stadium happens quite often for the short orbits), the LIPR enhancement associated with scar effects must be combined with the enhancement due to symmetry. In the semiclassical limit where our statistical analysis is relevant, the two effects are independent of one another and simply multiply, the scar effect being associated with short-time dynamics around a periodic orbit, while the symmetry effect appears near the much larger Heisenberg time, where individual eigenlevels are resolved.

Secondly, in the presence of spatial symmetries such as parity, symmetry must be taken into consideration when computing periodic orbit properties (period, classical action, and monodromy matrix) for the purpose of quantifying the scar effect. This holds true even if the system is not time-reversal invariant, so that no symmetry lines are present in the LIPR plot. In the previous section, we have addressed this second issue simply by considering orbits in a desymmetrized version of the billiard. That approach, however, is neither rigorous nor completely general, and in particular is problematic for orbits like the horizontal bounce, which does not exist in the interior of the quarter-stadium. In this section we present an alternative approach to including spatial symmetry in computing scar effects on localization.

In the stadium there are two parity operations, reflection in x and in y, which we represent by the operators R_x and R_y, respectively, and which separate the stadium wavefunctions into four distinct symmetry classes. The sums on the left-hand side in Eqs. (2.41) and (2.53) can be divided into sums over each symmetry class separately; thus, for instance, the left-hand side of Eq. (2.41) becomes

$$\begin{aligned}\sum_n |\langle q,p|n\rangle|^2 &= \sum_{n_{++}} |\langle q,p|n_{++}\rangle|^2 + \sum_{n_{-+}} |\langle q,p|n_{-+}\rangle|^2 + &(2.64)\\ &\quad \sum_{n_{+-}} |\langle q,p|n_{+-}\rangle|^2 + \sum_{n_{--}} |\langle q,p|n_{--}\rangle|^2 &(2.65)\\ &= \sum_{n} |\langle q,p|P_{++}|n\rangle|^2 + \sum_{n} |\langle q,p|P_{-+}|n\rangle|^2 + &(2.66)\\ &\quad \sum_{n} |\langle q,p|P_{+-}|n\rangle|^2 + \sum_{n} |\langle q,p|P_{--}|n\rangle|^2 , &(2.67)\end{aligned}$$

where $P_{\pm\pm}$ is the projector onto the subspace with even or odd symmetry under R_x and R_y, respectively: $P_{\pm\pm} = (1 \pm R_x)(1 \pm R_y)/4$. A similar development of the left-hand side of Eq. (2.53) for the fourth moment is possible. The projection operators may now be considered to act to their left, on the state $|p,q\rangle$. This means simply that $|p,q\rangle$ is to be replaced by $P_{++}|p,q\rangle$, etc., on the right-hand sides of Eqs. (2.41) and (2.53). If we consider the LIPR for the even-even states only we obtain

$$\begin{aligned} \mathrm{LIPR}_{++}(\mathrm{q},\mathrm{p}) &= \frac{N_{++}\sum_n |\langle q,p|P_{++}|n\rangle|^4}{\left(\sum_n |\langle q,p|P_{++}|n\rangle|^2\right)^2} && (2.68)\\ &\approx \mathrm{LIPR(RMT)}\frac{\sum_{j=-j_{\mathrm{mix}}}^{j_{\mathrm{mix}}}\left[\langle q,p|P_{++}M^j P_{++}|q,p\rangle\right]^2}{|\,\langle q,p|P_{++}|q,p\rangle\,|^2} && (2.69)\\ &= \mathrm{LIPR(RMT)}\sum_{j=-j_{\mathrm{mix}}}^{j_{\mathrm{mix}}}\frac{\left[\sum_{i_1,i_2=0,1}\frac{1}{4}\left\langle q,p\middle|R_x^{i_1}R_y^{i_2}M^j\middle|q,p\right\rangle\right]^2}{|\,\langle q,p|P_{++}|q,p\rangle\,|^2} && (2.70)\\ &\approx \mathrm{LIPR(RMT)}\sum_{j=-j_{\mathrm{mix}}}^{j_{\mathrm{mix}}}\frac{\left[\sum_{i_1,i_2=0,1}\frac{1}{4}\left(q,p\middle|R_x^{i_1}R_y^{i_2}M_{\mathrm{Cl}}^j\middle|q,p\right)\right]^2}{(q,p|P_{++}|q,p)} && (2.71) \end{aligned}$$

The constant of proportionality here, $C'_{++} = (2\Delta k/N_{++})\mathrm{LIPR(RMT)}$, is four times larger than C' in Eq. (2.50) because $N_{++} = N/4$; this reflects the fact that the Heisenberg time is four times smaller for the desymmetrized stadium (C'_{++} being inversely proportional to T_{H++}).

Thus we see that in addition to the usual recurrences due to periodic orbits ($M^j(q,p) = (q,p), i_1, i_2 = 0$) there are, for certain initial values (q,p), recurrences when the orbit closes up to a reflection in x or y or both. For example, the V orbit starts at the center of the straight segment, hits the semicircular segment at $j = 1$, and at $j = 2$ returns to where it started up to reflection in x. Another example is the Z orbit, which passes through the center of the stadium; at $j = 3$ it returns to where it started up to a reflection in both x and y.

Similar recurrences up to a symmetry operation would be present in the LIPR summed over all symmetry classes. One last important point to note is that for an orbit located right on a symmetry line (such as the horizontal bounce orbit in the stadium), the corresponding symmetry (R_y in this case) is not included in computing the scar effect. The fact that the wavepacket is located on a symmetry line is of course relevant in setting LIPR(RMT)= 3 instead of 2. The other symmetry, namely R_x, still needs to be included, of course: for the horizontal bounce orbit its effect is to make short-time recurrences in Eq. (2.71) happen for all values of j instead of for even j only.

2.5.7 Derivation as return probability

Finally, we explain the interpretation of the LIPR as the infinite-time average return probability, as in Eq. (2.72) below (where it is assumed that the energy window in the sums over n and n' is large enough to cover the energy scales present in the test state $|a\rangle$). In order to clarify what is happening in the infinite-time limit it is useful to look at finite times, and then let time tend to infinity. Thus we define the average return probability at time T for a wavepacket $|\Psi_a\rangle$ to be [48, 51]

$$\begin{aligned} P_{aa}(T) &= \frac{1}{2T}\int_{-T}^{T} dt \left|\left\langle \Psi_a \middle| e^{-iHt} \middle| \Psi_a \right\rangle\right|^2 \\ &= \frac{1}{2T}\int_{-T}^{T} dt \left|\sum_n \langle \Psi_a | n\rangle \, e^{-iE_n t} \, \langle n | \Psi_a\rangle\right|^2 \\ &= \frac{1}{2T}\sum_{nn'}\int_{-T}^{T} dt \, |\langle \Psi_a | n\rangle|^2 \, |\langle \Psi_a | n'\rangle|^2 \, e^{-i(E_n - E_{n'})t} \\ &= \sum_{nn'} \frac{\sin\left((E_n - E_{n'})T\right)}{(E_n - E_{n'})T} \, |\langle \Psi_a | n\rangle|^2 \, |\langle \Psi_a | n'\rangle|^2 \, . \end{aligned} \tag{2.72}$$

In the limit as T tends to infinity,

$$\frac{\sin\left((E_n - E_{n'})T\right)}{(E_n - E_{n'})T} \to \delta(E_n - E_{n'})$$

and (in the case of a non-degenerate spectrum) we recover Eq. (2.7) up to normalization (see below). The arbitrary state $|\Psi_a\rangle$ living in the interior of the stadium may be replaced by a Gaussian $|q,p\rangle$ living on the boundary if at the same time the eigenstates $|\Psi_n\rangle$ are replaced by their normal derivatives on the boundary, $|n\rangle$. The natural timescale for T is the time T_B for bouncing between the straight segments of the billiard. With $k = 100$ and $m = 1/2$ we have $v = p/m = \hbar k/m = 200$ and so $T_B = 2/v = 0.01$. For $T \ll T_B$ we should have $P_{aa}(T) \sim 1/T$; for intermediate times we expect classical behavior as long as $T < T_{\rm mix}$, where the mixing time $T_{\rm mix}$ is the time it takes for the classical dynamics to spread a given phase-space element throughout the entire phase space (in order of magnitude it is given by T_B times the logarithm of the number of Planck-sized cells in the classical surface-of-section phase space, or $\sim T_B \log(kL)$). The classical behavior will be computed directly below, where based on the time needed for a simulated classical ensemble to begin to spread smoothly throughout phase space we find $T_{\rm mix} \approx 7T_B$ at $k = 100$; that it is so large may be attributed to intermittency due to bouncing-ball orbits. Finally for $T \gg T_H$, that is beyond the Heisenberg time where individual eigenstates are resolved (here, $T_H = 1/\Delta E = 1/(2k\Delta k) = 1/(2 \cdot 100 \cdot 0.05) = 0.1 = 10T_B$ at $k = 100$), we should find that $P_{aa}(T)$ tends to the LIPR plot as shown in Fig. 2.2.

Finite-time return probability

The results for the time-dependent average return probability $P_{(q,p),(q,p)}(T)$ are given in Fig. 2.9, for the energy window at $k = 100$. We are unable to study numerically the limit $T \ll T_B$ because of the finite number of eigenfunctions included in the sum in Eq. (2.72). At $T = T_B$, however, we see large amplitudes coming from the bouncing-ball and whispering-gallery regions of phase space, both of which are short-time classical effects. At $T = 2T_B$ the horizontal-bounce orbit appears. By $T = 3T_B$ several discrete peaks appear, and more of them are present at $T = 4T_B$. These will be shown below to correspond to short periodic orbits of the classical dynamics in the desymmetrized quarter-stadium billiard. As T becomes as large as $5T_B$ and then $10T_B$ the finite-time average return probability indeed approaches the infinite-time LIPR. The non-uniform structure of the infinite-time return probability is a purely quantum-mechanical effect, because according to classical mechanics the average return probability must become uniform for $T \gg T_{\text{mix}}$.

Identification of peaks

Let us work with the data at $T = 4T_B$, where the strong peaks are all present and readily distinguishable from the background [see Fig. 2.9(d)]. We identify these peaks with the short periodic orbits in Table 2.4 and Fig. 2.7. To avoid having to deal explicitly with symmetry effects, we use the properties of the reduced orbits in the fundamental domain—one quarter of the stadium. (This approach has been justified rigorously in Sec. 2.5.6.) Thus the periods of the most important orbits, as tabulated in Table 2.4, range from $2T_B$ to $5T_B$ although their periods in the full stadium may be longer. Although the triangle is a short periodic orbit, it does not appear until $5T_B$ because its desymmetrized version is just as long as the orbit in the full stadium. As T increases beyond $T_H \approx 10T_B$, many of the short periodic orbits become incorporated into the streaks that dominate the IPR plots.

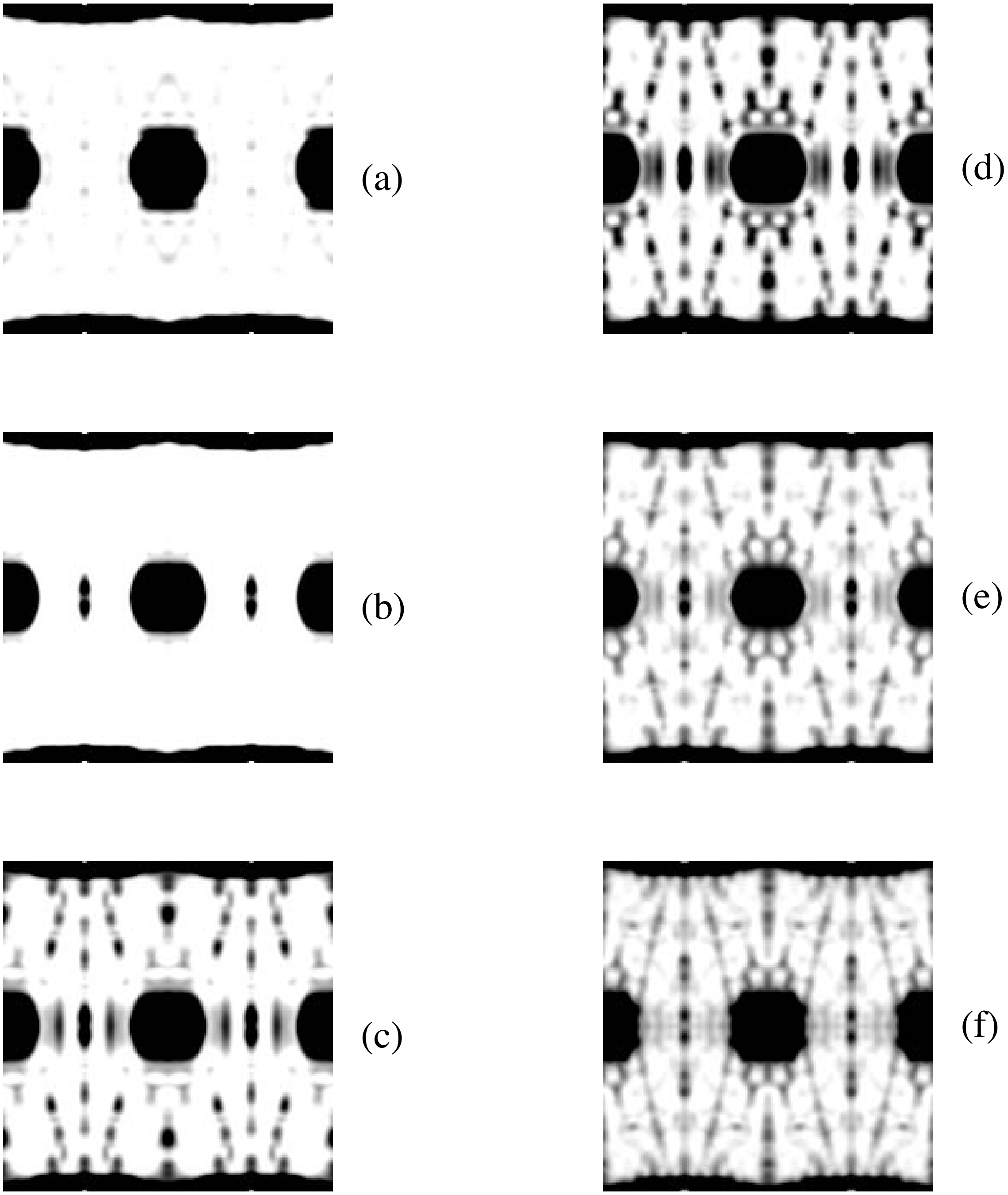


Figure 2.9: Quantum-mechanical average-return probability $P_{aa}(T)$ as defined in Eq. (2.72), near $k = 100$; q and p as before. (a) $T = T_B$, (b) $T = 2T_B$, (c) $T = 3T_B$, (d) $T = 4T_B$, (e) $T = 5T_B$ and (f) $T = 10T_B$. Note the emergence of periodic orbits at $T = 2T_B$ to $T = 3T_B$. For $T = 10T_B$ the figure is already beginning to resemble the LIPR plotted in Fig. 2.2(b).

Label	Description	q	p/k	Length (desymmetrized)	Period/T_B (desymmetrized)
HB	horizontal bounce	2.5708	0	1	2.00
V	V-shaped	0	0.70711	2	2.41
D	diamond	2.5708	0.44721	2	2.24
B	bowtie	1.5236	0.5	2	2.60
Z	Z-shaped	0.5	-0.44721	3	3.24
X	box	1.7854	0.70711	2	2.41
A	accordion	0	0.5491	3	3.30
W	W-shaped	0	0.31623	4	4.16
H	hexagon	2.5708	-0.86603	3	2.50
T	triangle	0	0.83573	3	4.30
F	fish	2.5708	0.6478	5	5.40

Table 2.4: Principal short periodic orbits appearing in Figs. 2.2 and 2.9. The initial coordinate q is the distance around the perimeter of the stadium measured from the center of the upper straight segment; the initial momentum p is positive in the clockwise sense. The lengths and periods are given for the desymmetrized stadium, which is relevant to our calculations.

Classical finite-time return probability

The persistence to long times of short-time classical information is a remarkable property of the quantum mechanics. Classically, we would expect the recurrences to be washed out at times large compared to the mixing time. Indeed, we have computed a classical analogue to the time-dependent average return probability, namely the overlap of an initial Gaussian in classical (Birkhoff) phase space with its iterates under the classical mapping. In order to get results in continuous time, the overlap of the iterate at (real) time t with the initial Gaussian was stored as a delta function at time t and at the end of the calculation the integral over t from $-T$ to $+T$ was taken, as in the quantum calculation. As shown in Fig. 2.10, by $T = 10T_B$ all of the periodic orbits seen in the quantum calculation are present, although the peaks are more diffuse classically, but by $20T_B$ the classical picture is beginning to be uniformly distributed throughout phase space, losing memory of the short-time behavior. Quantum-mechanically, of course, the peaks do not become washed out at long times and persist in the infinite-time limit, Fig. 2.2.

2.6 Conclusion

The advantage of the approach implemented in Section 2.5 is that all of the scars can be observed simultaneously in the plot of inverse participation ratio as a function of phase space location. Although scars are associated with the classical structure of unstable periodic orbits, the persistence to long times of an excess in the average return probability is a purely quantum-mechanical phenomenon. The brightnesses of the peaks in the LIPR plot can be predicted semiclassically with some success (to accuracy of about $10 - 15\%$), the size of the deviations being roughly consistent with expected statistical fluctuations. More precise comparison with the scar theory predictions would require producing a greatly increased set of eigenstates: this could be done by going to higher energies or (with far less computational expense) by using an ensemble of chaotic systems with a given unstable orbit kept fixed [68]. The predictions of brightnesses based on exact classical phase-space evolution at the end of Section 2.5 were no more successful than those based on linearized behavior around the periodic orbit.

Most of the extra localization of individual eigenfunctions in Husimi phase space, over and above Gaussian random fluctuations, may be satisfactorily accounted

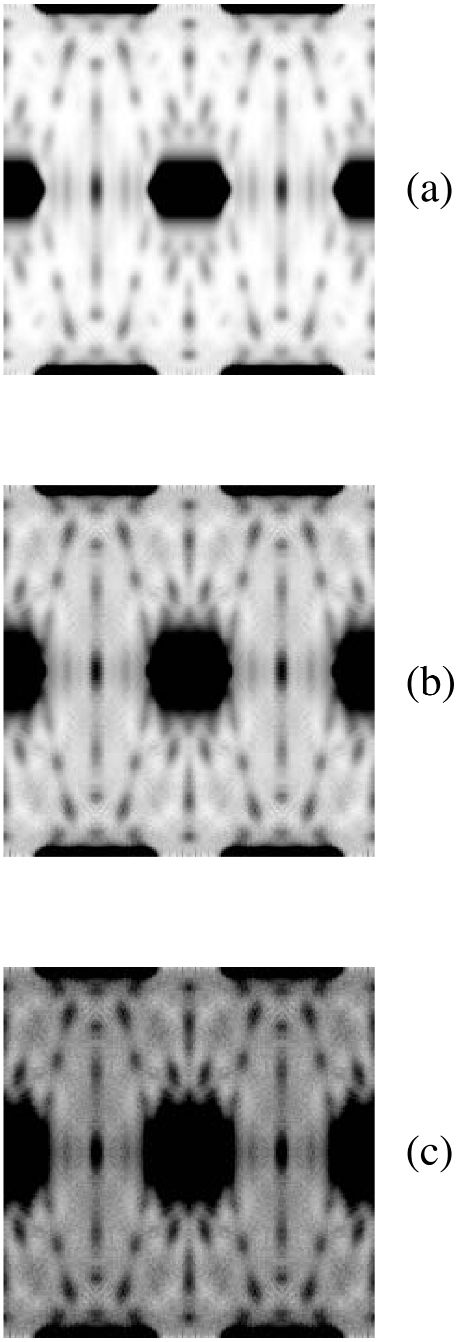


Figure 2.10: Results of a classical simulation of the return probability versus time, q running horizontally from 0 to $4+2\pi$ and p vertically from -1 to 1. Initial Gaussians were chosen with $\sigma^2 = 0.05$, corresponding to a spread in position space of 0.2 out of $4+2\pi$, or 2% of the perimeter. All of the major periodic orbits are present here as short-time recurrences. (a) $T = 5T_B$, (b) $T = 10T_B$, (c) $T = 20T_B$.

for by quantum symmetry effects. Scarring is also seen to be present based on an analysis of the spectra of wavepackets located on short periodic orbits. Symmetry effects on the overall mean IPR scale as $O(\sqrt{\hbar})$ (i.e., $O(\sqrt{1/k})$), while the effect of scarring on the overall mean IPR scales as $O(\hbar)$ (i.e, $O(1/k)$). The reason for this, for the symmetry effects, is that the widths in position and momentum of the test-state Gaussian scale as $\sigma \sim \sqrt{\hbar}$, setting the width of the symmetry lines (Eqs. (2.24), (2.25)), while the length of the symmetry-enhanced lines is $\hbar$-independent. Thus, the total fraction of phase space covered by the symmetry lines scales as $\sqrt{\hbar}$. The scar effect, on the other hand, is significant for those phase-space Gaussian wavepackets that have large intensity on the periodic orbit [94], and the fraction of such Gaussians goes as the area of the Gaussian; i.e., as $\sqrt{\hbar} \times \sqrt{\hbar} = \hbar$. We note that although the size of both these effects on the overall wavefunction IPR is expected to go to zero in the semiclassical limit (because periodic orbits and symmetry lines both affect a smaller and small fraction of phase space surrounding them, in the $\hbar \to 0$ limit), the local IPR (LIPR) either on a symmetry line or on a periodic orbit is $\hbar$-independent. Thus, as measured using the local IPR, deviations from naive RMT predictions persist to arbitrarily high energies and do not decay in the semiclassical limit. As for other possible kinds of eigenfunction localization, not associated with the short periodic orbits of the system or with symmetry lines, their explanation remains an open question.

Chapter 3

Scarring Effects on Tunneling in Chaotic Double-Well Potentials

3.1 Abstract

The connection between scarring and tunneling in chaotic double-well potentials is studied in detail through the distribution of level splittings. The mean level splitting is found to have oscillations as a function of energy, as expected if scarring plays a role in determining the size of the splittings, and the spacing between peaks is observed to be periodic of period $2\pi\hbar$ in action. Moreover, the size of the oscillations is directly correlated with the strength of scarring. These results are interpreted within the theoretical framework of Creagh and Whelan. The semiclassical limit and finite-$\hbar$ effects are discussed, and connections are made with reaction rates and resonance widths in metastable wells.

3.2 Introduction

Many chemical reactions must proceed through a potential barrier before the final dissociation of the products of the reaction can take place. Thus, the reaction rate is governed by tunneling [53, 89, 102, 111]. Radioactive decay of nuclei also involves crossing a potential barrier, dictated by a combination of short-ranged strong binding forces and a longer-ranged Coulomb repulsion [79]. A typical experimental situation is the absorption of a slow neutron; the resulting metastable nucleus decays predominantly by tunneling through a saddle point, and the distribution of resonance widths is well described by random matrix theory [98]. Another important example of quantum tunneling is the conductance of mesoscopic devices [71, 87]. In more than one dimension, when the classical dynamics can be chaotic, quantum chaos plays a role. For instance, in the tunneling diode junction, in which electrons are driven by an applied electric field but must tunnel through potential barriers at either end of the device, the dynamics, in the presence of a magnetic field, is chaotic and scarring of a short periodic orbit is known to dominate the conduction [40, 123, 41]. In this Chapter we study chaotic tunneling in a simple model potential, and establish the connection to scar theory, both qualitatively and quantitatively. In related work, Tomsovic [116] considers the problem of chaos-assisted tunneling in systems with mixed phase space and weak breaking of the reflection symmetries. We prefer to study tunneling by calculating splittings in a double-well potential rather than resonance widths in a metastable well, because the former is an easier computational task, as we will see in the Sec. 3.3. The relation between splittings in a double-well potential and resonance widths in a metastable well will be discussed immediately below, where we will observe that the two quantities are related by known overall normalization factors.

The double-well potential we will work with, in two dimensions, has the very simple form

$$V(x,y) = x^4 - x_0^2 x^2 + ay^2 + \lambda x^2 y^2 + \sum_i b_i e^{-((x-x_i)^2+(y-y_i)^2)/\sigma_i^2} \,. \qquad (3.1)$$

The parameters x_0, a, λ, b_i, x_i, y_i and σ_i will be specified below. For $x_0^2 > 0$, a barrier along the y-axis separates the potential at low energies ($E < 0$) into two wells. The $\lambda x^2 y^2$ term ensures that the potential is not separable, while the Gaussian perturbations, the parameters of which may be changed at will, allow us to generate an ensemble of statistically independent eigenstates near any given energy. If the positions and heights of the Gaussians are chosen suitably, the potential has reflection symmetries in the x and y directions. The symmetry under reflection in y means that there will always be a short periodic orbit on the x-axis, in each of the two wells.

The principal result of this Chapter is that the size of the level splittings in the two-dimensional double well, the classical dynamics of which is chaotic, is directly correlated with the scarring of eigenfunctions along the x-axis, which we believe to be the primary channel for tunneling. The evidence for this is that (1) the distribution of splittings, once the average exponential trend in energy is scaled out, displays oscillations as a function of energy of the sort expected by scar theory. In particular, the action distance between successive peaks is precisely h. However, the dependence of the mean splitting on action is not well reproduced quantitatively by linear scar theory, because non-linear effects appear to be important. (2) The rescaled splittings are correlated with the overlap of the eigenstate with a Gaussian test state lying on the periodic orbit on the x-axis, a measure of scarring. (3) The correlation of the splittings with the overlaps is further supported by the fact that the distribution of overlaps has the same energy-dependent oscillations as the distribution of splittings.

According to one-dimensional WKB theory, the splitting $\Delta E = E_{\rm anti-symm} - E_{\rm symm}$ in a symmetric double well is given in the semiclassical limit by

$$\Delta E = \left(\frac{\hbar\omega}{\pi}\right) e^{-S/\hbar}\,, \tag{3.2}$$

while the resonance width Γ for the state at the same energy in a metastable well is

$$\Gamma = \left(\frac{\hbar\omega}{4\pi}\right) e^{-2S/\hbar}\,. \tag{3.3}$$

Here ω is the frequency of the classical periodic motion at energy $E_{\rm symm}$ and the imaginary action for going under the barrier is

$$S = \int_{-x_{\rm tp}}^{x_{\rm tp}} dx\,\sqrt{V(x) - E}\,, \tag{3.4}$$

$x_{\rm tp} > 0$ being the position of the classical turning point at energy $E_{\rm symm}$ [12]. In order for the semiclassical theory to apply, $E_{\rm symm}$ must be sufficiently far below the barrier that $S/\hbar \gg 1$. Note the factor of two in the exponent of Eq. (3.3).

These one-dimensional formulae may be generalized to the two-dimensional potential well as follows. We expect the tunneling to be dominated by paths that cross the barrier close to the x-axis, which has the smallest action integral S. The splitting ΔE will then be proportional to the exponential factor $e^{-S/\hbar}$.

The correct generalization of the frequency of attempts to cross the barrier, ω, to two or more dimensions, is the frequency with which one returns to a Planck-sized cell in phase space that lies on the horizontal periodic orbit. This horizontal periodic orbit is the real continuation of the least-action path across the barrier. The time for returning to such a cell (or to any other cell in an ergodic well) is the Heisenberg time $T_H = h/\Delta(E)$, where $\Delta(E)$ is the mean level spacing near energy E (i.e., the spacing between doublets in the double-well system). Then the frequency of attempts to cross the barrier is just proportional to the mean level spacing $\Delta(E)$. Thus, we expect on general grounds that the splitting ΔE is given in order of magnitude by $\Delta E \sim \Delta(E) e^{-S(E)/\hbar}$, which gives us the trend as a function of energy. This expression for the splitting will be confirmed by the exact semiclassical theory to be discussed below.

For any given state, we should expect that its splitting will be large or small compared with the mean value at that energy according to whether its amplitude is large or small on the horizontal periodic orbit which leads to optimal tunneling. For simplicity we can study the wave function amplitude near the turning point of the horizontal periodic orbit. The value of the wave function at the turning point in the two-dimensional chaotic system is (ignoring scar-related effects which are the main focus of this Chapter) given approximately by a Gaussian-distributed random variable, as random matrix theory would predict. Thus, $|\psi(x_{\rm tp})|^2$ has, according to random matrix theory, a Porter–Thomas distribution for all energies far enough below the barrier (near zero energy the horizontal periodic orbit becomes stable and the distribution of splittings rolls over to one having many more large and small splittings, corresponding to wave functions that live near or avoid this stable orbit).

However, what is relevant to tunneling in d dimensions is not just the value of the wave function exactly at the turning point but rather its behavior in a whole $\hbar^{d-1}$–sized region surrounding the periodic orbit. We shall see below that the right quantity to consider is the inner product of the wave function with a Gaussian centered on the periodic orbit (at the turning point or at some other location). This has as well a Porter–Thomas distribution, within the random matrix theory approximation. Scar-related effects and finite-$\hbar$ effects on the distribution of splittings will be discussed below.

The viewpoint summarized above is in agreement with the theoretical work of Creagh and Whelan [29]. First, they find that the mean splitting $\langle \Delta E \rangle$ at a given energy is given by the product of an exponential factor $e^{-S/\hbar}$, a factor proportional to the mean spacing between doublets, as described above, and a third factor that carries information about the monodromy matrix of the (imaginary time) tunneling orbit. Then they show that, for chaotic and symmetric double wells, the splitting for a particular eigenvalue, relative to the mean splitting, may be written in the semiclassical limit as a matrix element of the wave function ψ near the real continuation $\mathcal{R}$ of the complex trajectory that passes through the barrier with minimum (imaginary) action. This imaginary-time trajectory and its real-time continuation may be thought of as the optimal route for tunneling. The matrix element $\Delta E \sim \langle \psi | \mathcal{T} | \psi \rangle$ involves integration over a Poincaré surface transverse to the real continuation $\mathcal{R}$. The kernel $\mathcal{T}$ is a semiclassical Green's function which, in the approximation that the dominant contribution to the tunneling matrix element comes from the neighborhood of $\mathcal{R}$, becomes a Gaussian centered on the intersection of $\mathcal{R}$ with the Poincaré section. The width of the Gaussian is of $O(\hbar^{1/2})$ in both directions (e.g. y and p_y) tangent to the surface of section. [The results may of course be easily generalized to dimensions $d > 2$, where the resulting Gaussian has width of $O(\hbar^{1/2})$ in all $2d-2$ directions along the surface of section.]

In the case that the real continuation $\mathcal{R}$ happens to lie on a short periodic orbit, which will always be true when a reflection symmetry across the x-axis is present, this matrix element may be regarded as a measure of scarring on the periodic orbit. The Creagh–Whelan theory predicts, therefore, that strong scarring should be correlated with large splittings, confirming the intuitive expectation that high tunneling rates should occur for those wave functions that have large amplitude along the path with optimal tunneling. Neither in Ref. [29] nor in Ref. [30], however, do the authors demonstrate the connection between scarring and splittings on a state-by-state basis. In the latter work, they confirm their formula for the tunneling matrix element by deriving from it an analytical prediction for the statistical distribution of splittings. This prediction is in agreement with numerical calculations for potentials in which the real continuation of the optimal tunneling orbit is *not* a periodic orbit; when it is, the random-matrix assumption in their derivation breaks down due to scarring on this periodic orbit. Thus, the present work, while confirming the predictions of Creagh and Whelan, goes beyond their results by establishing conclusively the link between scarring and tunneling and by showing, with better statistics, that the distribution of scaled splittings indeed becomes approximately Porter–Thomas, but only after scarring effects have been taken into account.

3.3 Method

The wave functions and splittings were calculated numerically using the discrete variable representation [46, 77, 76]. The matrix elements of the position operators X and Y and of the kinetic energy operators K_x and K_y were first evaluated analytically using standard identities, in a basis of up to the first 300 Gauss-Hermite functions in each dimension. Then, in order to take advantage of the two reflection symmetries in x and y, the two operators X^2 and Y^2 were diagonalized. The reason for using X^2 and Y^2 instead of the usual choice of X and Y is that X^2, K_x, Y^2 and K_y are all block-diagonal, connecting only basis elements within one of the four symmetry classes (even–even, even–odd, odd–even, and odd–odd). Since we are interested only in even–even potentials of the form $V(X,Y) = \sum_i f_i(X^2) g_i(Y^2)$ we can just as well compute V at the eigenvalues of X^2 and Y^2 as at those of X and Y, but using the basis obtained by diagonalizing X^2 and Y^2 ensures that the final Hamiltonian $H = K_x + K_y + V$ will be itself also block diagonal. The four symmetry classes may therefore be analyzed separately. In the basis chosen the potential V is of course diagonal, while, in two or more dimensions, the kinetic energy matrix will be sparse. More precisely, if N is the dimension of one of the blocks in the Hamiltonian matrix, the total number of non-zero matrix entries scales as $N^{1+1/d}$, or $N^{3/2}$ in the two-dimensional system. Since we require large values of N in order to observe semiclassical behavior, a sparse matrix routine is the method of choice for diagonalizing H. The accuracy of the computed eigenvalues was tested for convergence under increase of N, and for the parameters given below we found convergence to $\pm 10^{-12}$ for $N \approx 3500$, corresponding to about 200 Gauss-Hermite functions in the x direction and about 100 in the y direction.

The amount of phase space covered by the region $E < 0$, and thus the number of states under the barrier and the computation time, increases very rapidly with x_0. If all other parameters in the Hamiltonian are kept fixed, the number of states grows as x_0^6, and so the largest value of x_0 we can easily attain is about $x_0 = 6$, for $a = 1$ and $\lambda = 10$. At these parameter values, each well has a depth of $x_0^4/4 = 324$ and about 100 bound states, where $\hbar$ is taken to be unity here and in the following. The typical level spacing near the top of the well is ~ 1, and the splittings range from < 1 near the top of the well to $< 10^{-6}$ near $E = -50$; below this energy many of the splittings become too small ($< 10^{-12}$) to be resolved numerically. Therefore we take $E = -50$ to be the lower cutoff for the energies to be analyzed in the next Sec. 3.4.

3.4 Results

The potential given by Eq. (3.1), apart from the Gaussian perturbation, is mostly integrable for all energies except those near the top of the barrier. When Gaussian perturbations are introduced, the classical mechanics becomes more chaotic, but if these perturbations are too small they would not be seen by the quantum mechanics, which would remain effectively integrable. Thus, in order to render the quantum mechanics at energies corresponding to the bound states chaotic as well, it is necessary to introduce a Gaussian perturbation that is at least as large as the wavelength in question, and whose height is comparable in magnitude to the depth of the potential well. The simplest choice is to place large Gaussian perturbations above the minima of the potential well at $(\pm x_0/\sqrt{2}, 0)$, which will be seen by every bound trajectory as it crosses the center of the well and which thus effectively make the dynamics chaotic at energies down to the lowest considered ($E = -50$). This was checked classically by examining the Poincaré surfaces of section and may also be seen to be true quantum-mechanically in Fig. 3.1, where typical eigenfunctions are shown. Here, we have chosen the parameters of the double-well potential to be $x_0 = 6$, $a = 1$, $\lambda = 10$, and for the central Gaussian perturbation we use a height $b = 150$ (to be compared with a well depth of $V(\pm x_0/\sqrt{2}, 0) = -324$) and a width $\sigma = 0.5$.

We generate an ensemble of 625 systems by placing four further Gaussians (and their reflections in x and y) at $x = \pm 2, \pm 3, \pm 4, \pm 5$ and $y = \pm 1$, with heights $b_i = 20 n_i$, $n_1, \ldots, n_4 = 1, \ldots, 5$, and with equal widths $\sigma_i = 0.5$ as above. A contour plot of the potential for a typical member of the ensemble is given in Fig. 3.2; note the symmetrical distortion of the contours due to the perturbation.

We proceed to analyze statistically the splittings between states in the even–even and even–odd sectors. The results, for the parameters described, are given in Figs. 3.3 and 3.4. As expected, the size of the splittings falls off exponentially with decreasing energy in Fig. 3.3, as the barrier becomes wider and tunneling is suppressed. The trend is approximately linear on a semi-log plot, over six decades as E varies from 0 to -50. In Fig. 3.4 we rescale the splittings as a function of energy by

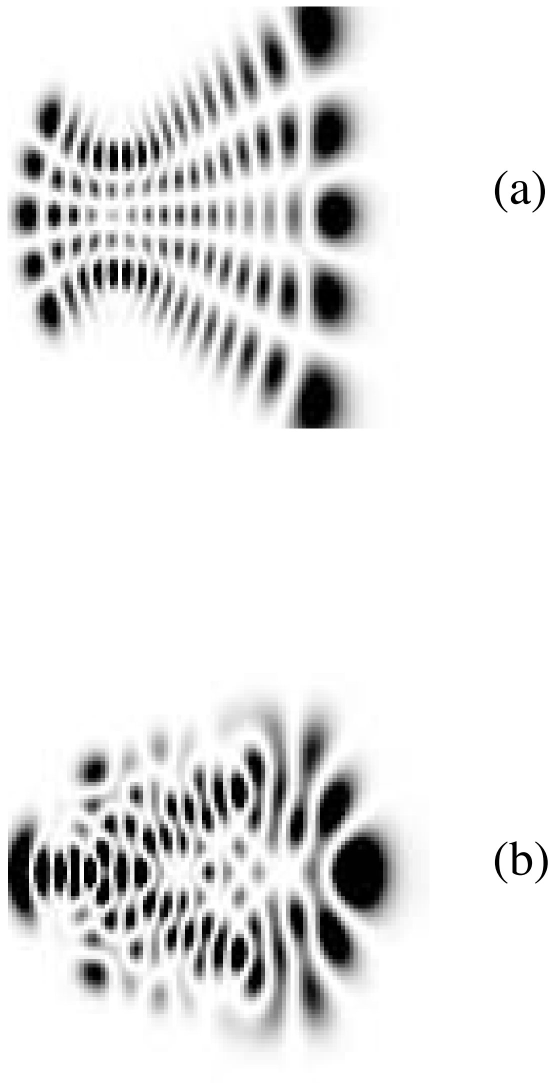


Figure 3.1: Typical eigenfunctions for the double-well potential in the (a) near-integrable case without Gaussian perturbations, $E = -11.055$, and (b) chaotic case with Gaussian perturbations, $E = -12.063$. Only one side of the well is shown in each case, with the x-axis running horizontally from -6 to 0 and the y-axis vertically from -2 to 2. The barrier is located on the right side at $x = 0$.

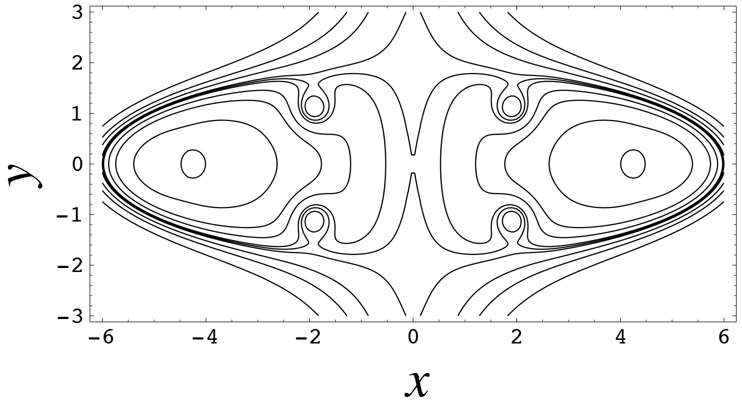


Figure 3.2: Contour plot of the potential for a typical member of the ensemble. The perturbation at $x = \pm 2, y = \pm 1$ leads to a symmetrical distortion of the contours, which would be more rounded in the absence of a perturbation. The contours range from $V = -300$ near the bottom of the potential to $V = +200$ above the top of the barrier.

$s \to s/e^{-S(E)}$. We find very pronounced oscillations in the distribution of splittings as a function of energy. As discussed in Sec. 3.2, we expect theoretically that the rescaled splitting should be proportional to the overlap in a Poincaré section of the eigenfunction with a Gaussian on the horizontal periodic orbit. If the Gaussian may be assumed to have area exactly h in the Poincaré section, such overlaps are described by scar theory [68, 65, 2, 1, 64]. We expect the results to be qualitatively the same even if the Gaussian in the Poincaré section given by the theory of Creagh and Whelan is not a minimum-uncertainty state, as long as it is not too large compared to h (in the latter limit, scar effects must go to zero in accord with the Schnirelman ergodicity theorem [103, 32, 126, 127]). In fact, in the data presented below the area of the Creagh–Whelan Gaussian ranged from $1.5h$ to $4h$. The prediction of scar theory is that, at a given energy, the distribution of splittings should be Porter–Thomas, and that the mean wave function intensity and therefore the mean splitting should oscillate as energy is varied by an amount that depends on the Lyapunov exponent and the monodromy matrix of the unstable periodic orbit. An important confirmation of the scarring picture is obtained when we plot, in Fig. 3.5, the rescaled splittings versus the action (divided by 2π) of the horizontal periodic orbit at the energy eigenvalue. We find that the oscillations are periodic in action with period 2π, which indicates that the scar quantization condition holds. This quantization condition for the action reads $A = 2\pi(n + 1/2 + n_c/4)$ where n is an integer and n_c is the number of conjugate points in one period of the orbit ($n_c = 3$ in the case of the horizontal orbit in our system).

A direct correlation between splittings and scarring is found by plotting, in Fig. 3.6, the rescaled splitting of each eigenvalue versus the overlap of the corresponding eigenfunction with a Gaussian test state lying on the horizontal periodic orbit, a measure of the degree with which this eigenfunction is scarred. The two quantities are correlated, with a slope of 2 on a log-log scale. The correlation coefficient of the logarithms is 0.78; it may be that the degree of correlation would be improved if the Gaussian were chosen to be properly aligned with respect to the monodromy matrix of the optimal tunneling path. The observed correlation nevertheless confirms that there is a direct connection, on a state-by-state basis, between scarring and tunneling, as predicted by the theory of Creagh and Whelan [29]. As a check on our results, we show in Fig. 3.7 that the overlaps display the same energy-dependent oscillations as the splittings do, as they must if the phenomenon of scarring underlies the behavior

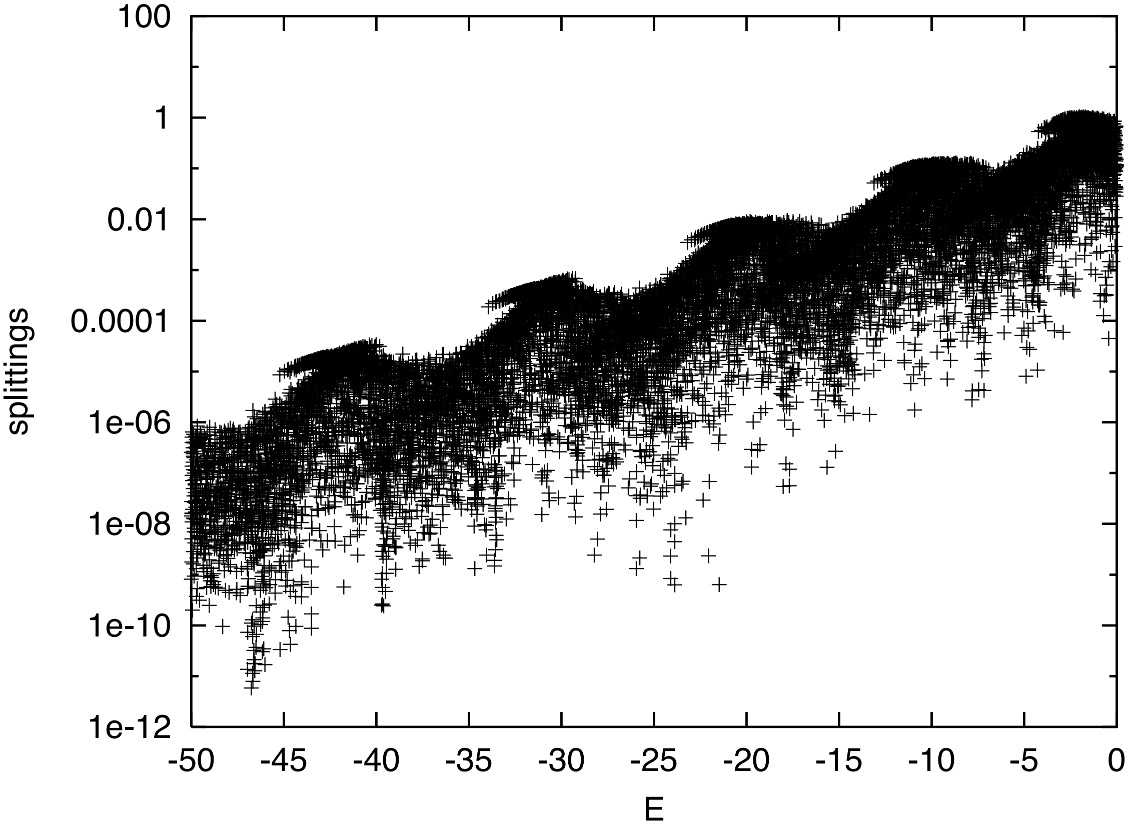


Figure 3.3: Level splitting versus energy E for the 15195 eigenstates between $E = 0$ and $E = -50$ in the ensemble of 625 double-well potentials described in the text.

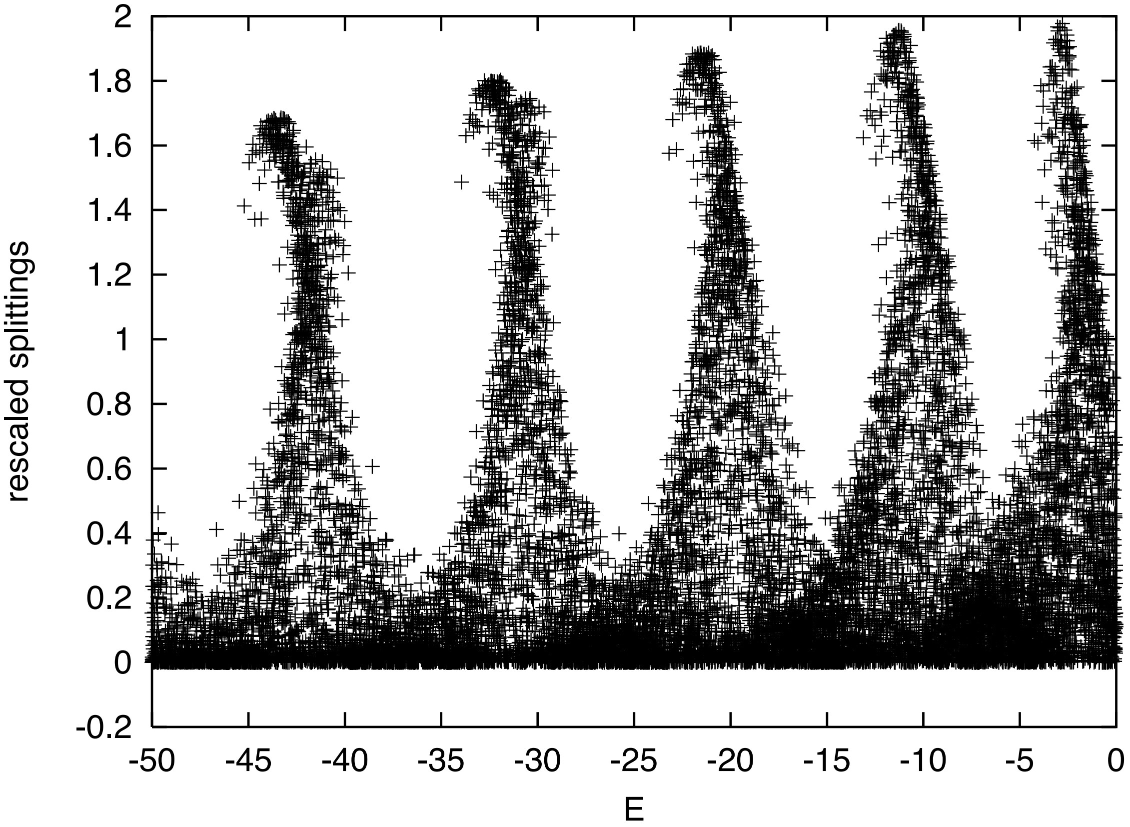


Figure 3.4: Rescaled level splitting versus energy E as in Fig. 3.3.

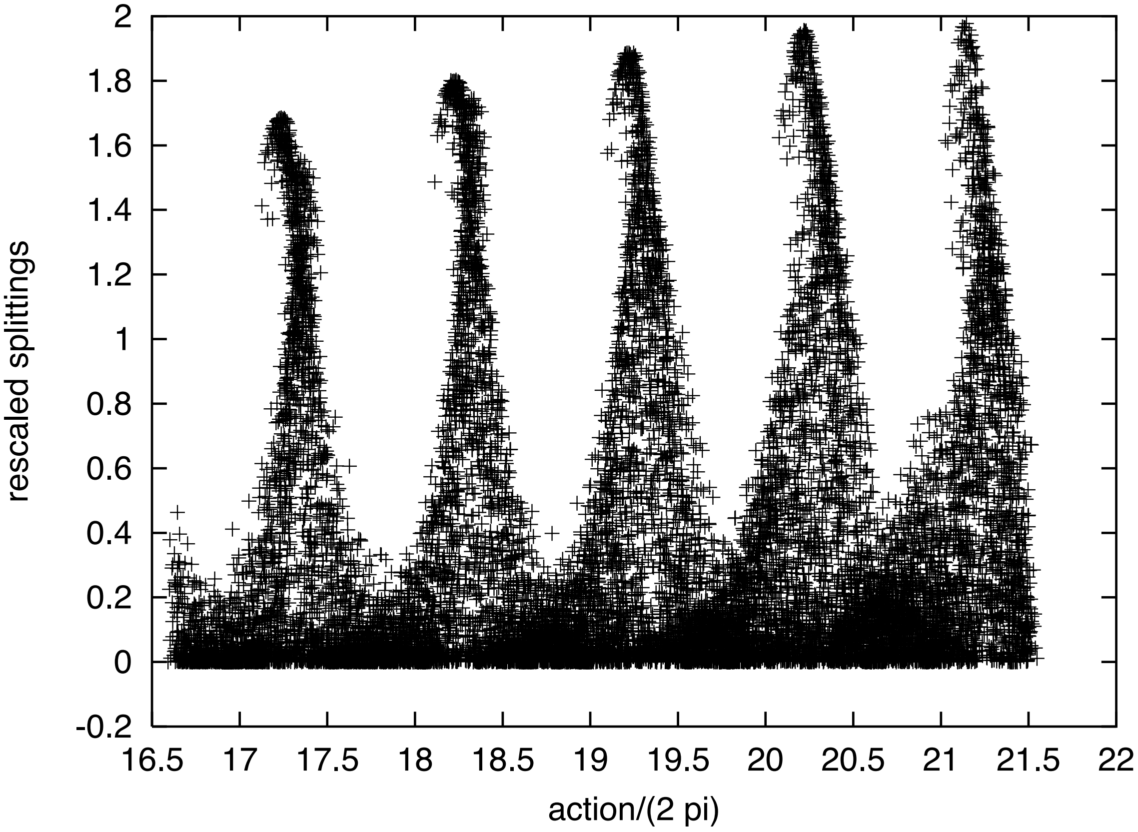


Figure 3.5: Rescaled level splitting versus action/2π with eigenstates as in Fig. 3.3.

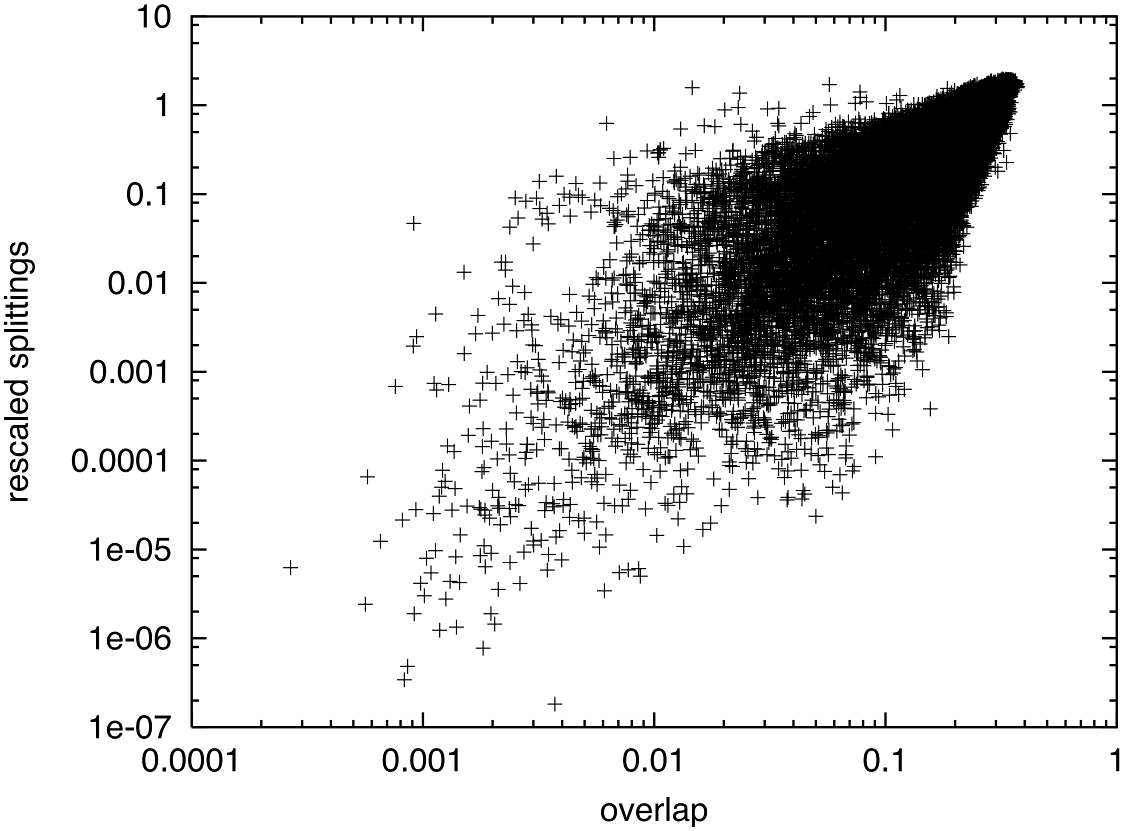


Figure 3.6: Rescaled level splitting versus the overlap of the eigenstate with a Gaussian on the horizontal periodic orbit, with eigenstates as in Fig. 3.3.

of both.

The connection between scarring and tunneling can be tested quantitatively in two ways. First, scar theory in the semiclassical limit predicts that the short-time (smooth) envelope of the oscillations in the mean rescaled splitting versus action is given by the Fourier transform of the autocorrelation function $A(m) = \langle\phi|\phi(m)\rangle$, where ϕ is a Gaussian wave packet (living in the Poincaré section) centered on the horizontal periodic orbit and $\phi(m)$ is its iterate after m bounces. The Gaussian ϕ is chosen to have the same orientation and aspect ratio in the (y, p_y)-plane as the

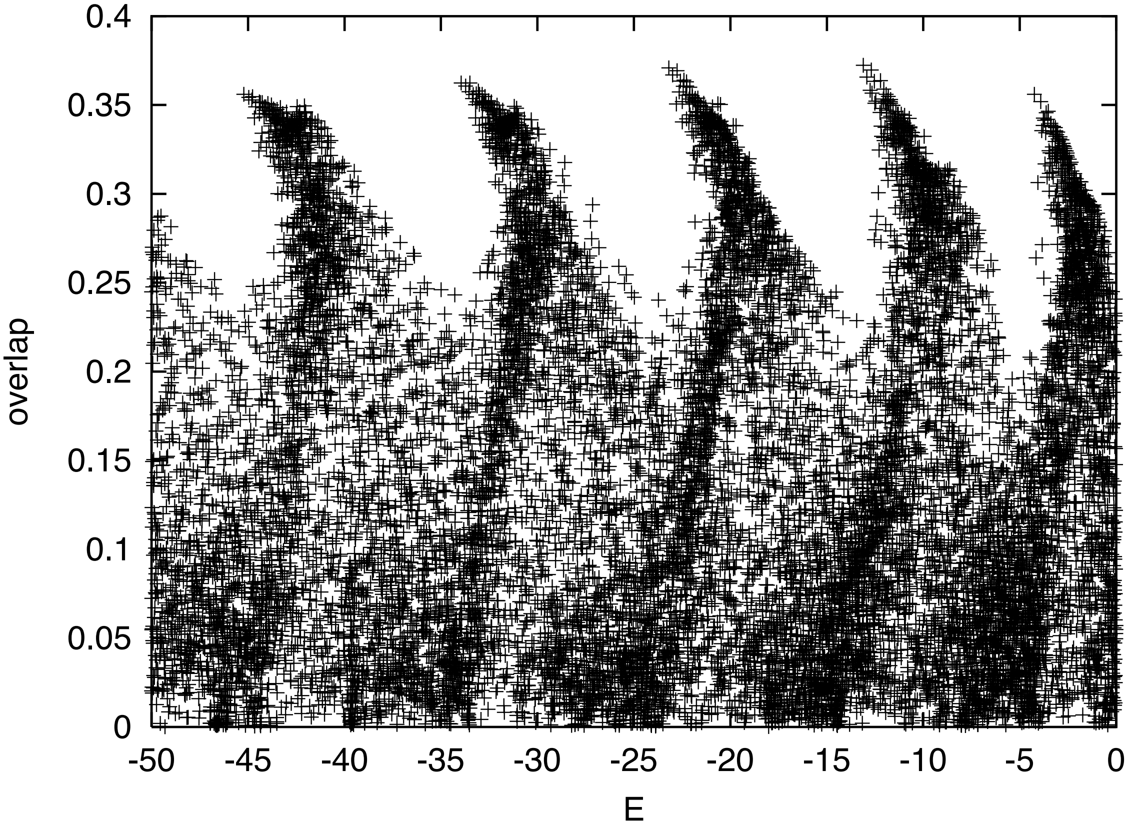


Figure 3.7: Overlap versus energy E, with eigenstates as in Fig. 3.3.

Gaussian called for by the Creagh–Whelan theory, but linearly rescaled so as to have area h as needed for scar theory. Linearizing the dynamics around the horizontal periodic orbit we find, when the Gaussian wave packet is optimally aligned along the stable and usntable manifolds of the orbit,

$$A(m) = \frac{1}{\sqrt{\cosh \lambda m}}\,, \tag{3.5}$$

where λ is the Lyapunov exponent of the periodic orbit. In the special case of orthogonal stable and unstable manifolds, a circular Gaussian will be one example of an optimally aligned wave packet. If the Gaussian is not optimally aligned, this formula may be generalized as follows:

$$A(m) = 2\sqrt{\frac{\det(M)}{\det(M + (J^{-m})^T M J^{-m})}}\,, \tag{3.6}$$

where M describes a Gaussian of the form $(\mathrm{const})\exp(-q^T M q)$ with $q = (y, p_y)^T$ representing the coordinates in the surface of section, and J is the Jacobian of the Poincaré mapping evaluated at the periodic orbit ($\mathrm{Tr}\ J = 2\cosh\lambda$). The matrix M is computed as specified in the Creagh–Whelan theory from the monodromy matrix of the complex orbit that begins at the Poincaré section on the right, goes through the barrier, and ends at the Poincaré section in the left well [29]. Here, the Lyapunov exponent λ and Jacobian J vary with energy over the range $-50 < E < -9$. At higher energies, the trajectory spends less time near the Gaussian at the center of the well, and thus experiences less deflection, leading to greater stability, eventually becoming stable for $E > -9$. The short-time envelope obtained as the Fourier transform of $A(m)$ in either Eq. (3.5) or Eq. (3.6) may be compared with the mean rescaled splitting plotted versus action. As shown in Fig. 3.8, with either form of the autocorrelation function we do find peaks in the predicted envelope of splittings at the right values of action for energies $E < -9$ (for energies $E > -9$ the horizontal periodic orbit becomes stable, so the scar theory does not apply and no prediction about the distribution of splittings can be made), but the heights of the maxima and minima between the peaks are not well reproduced. The contrast predicted by Eq. (3.6) is closer to the numerical data than that predicted by Eq. (3.5). The quantitative failure of semiclassical scar theory is attributable to the fact that, for our parameter values, the linearizable region around the horizontal periodic orbit is not large compared to h. In fact, the size of the linearizable region is only about $0.15h$ for the energies considered. Its size, however, is approximately independent of energy for $-50 < E < 0$, and this may explain the weak dependence of the peak height on energy observed in the numerical data. Nevertheless, we see that not only is scarring associated with larger splittings in the coarse sense of Fig. 3.6 but also the enhancement factor in the distribution of splittings, as a function of energy or of action, does oscillate with energy or action in agreement with the analytical prediction of scar theory; only the precise magnitude of these oscillations remains unexplained within the present linear theory.

A second quantitative test of scar theory in relation to tunneling is to examine the change in the distribution of splittings upon change in the Lyapunov exponent. The horizontal periodic orbit can easily be made more stable by keeping the height of the main Gaussian bump fixed at 150 while increasing its width. An ensemble of eigenstates and associated splittings was computed, just as above, for a larger value

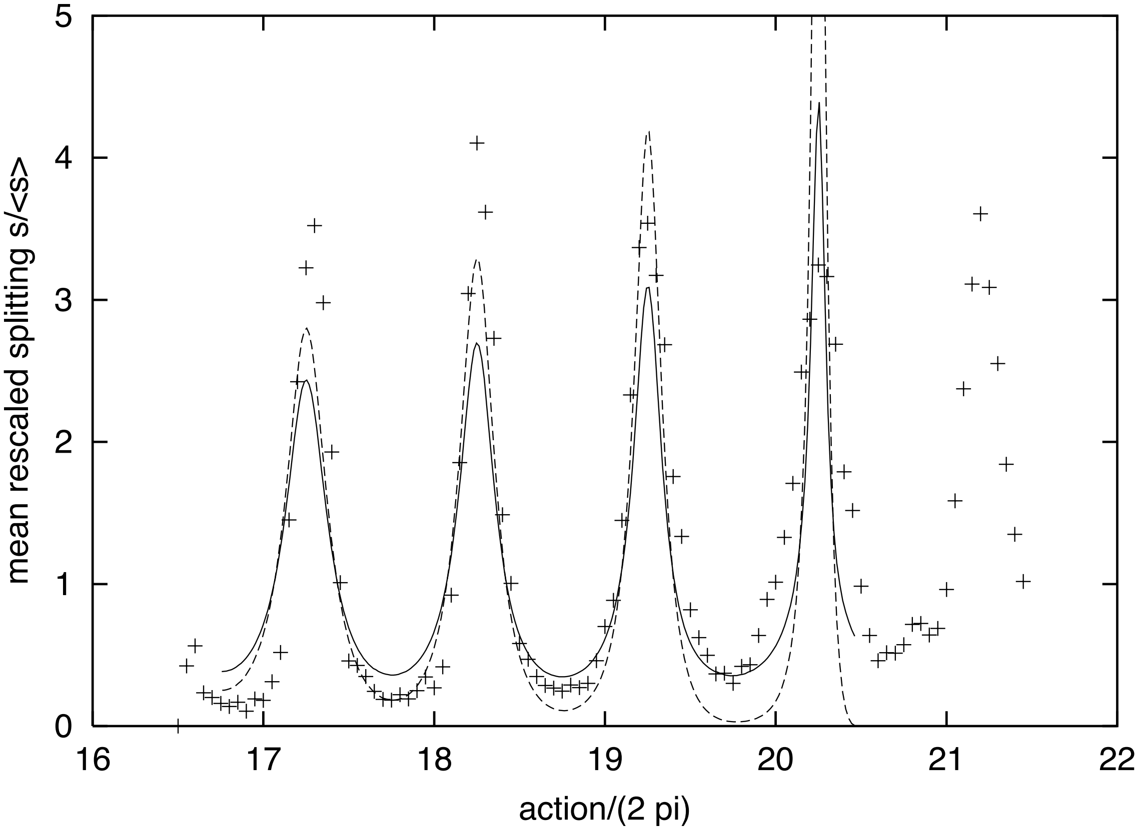


Figure 3.8: Mean rescaled splitting $s/\langle s \rangle$ versus action$/2\pi$, data points; short-time envelope from scar theory using Eq. (3.6), solid line; short-time envelope from scar theory using Eq. (3.5), dashed line.

of the bump width, namely 0.63 instead of 0.50. The expectation from scar theory would be for the distribution of splittings to have many more smaller and larger splittings at the resulting smaller Lyapunov exponent. At $\sigma = 0.5$, the horizontal periodic orbit is stable down to $E = -9.1$; the Lyapunov exponent then increases from zero with decreasing energy to a value of $\lambda = 2.0$ at $E = -50$. For $\sigma = 0.63$ it is stable down to $E = -49.8$ and attains only a value of $\lambda = 0.11$ at $E = -50$. The numerical data, however, show no marked difference between the two computations at $\sigma = 0.5$ and $\sigma = 0.63$; see Fig. 3.10. The lack of any difference between the distributions of splittings despite the difference in stability is an indication that we are not far enough into the semiclassical limit (see discussion below). In both cases the distribution of rescaled splittings (see the histograms in Fig. 3.10) has many more small and large splittings, and consequently fewer splittings around $s/\langle s \rangle = 1$, than a Porter–Thomas distribution would have (except for the sharp cutoff at $s/\langle s \rangle \geq 5$, which will be discussed below). Thus, the prediction of scar theory that there should be many more small and large splittings, relative to the prediction of random matrix theory, is confirmed. Also, the divergence of the probability distribution near zero splitting in the case of scarring on the real continuation of the optimal tunneling path differs markedly from the results, both analytical and numerical, of Creagh and Whelan [30] for the case when the real continuation is not a periodic orbit, which show a probability distribution tending to zero at zero splitting. Our numerical results for the scarring case improve on their statistics and allow us to discern the scar corrections to Porter–Thomas behavior. In particular, we note the excess of very small splittings; these correspond to the phenomenon of antiscarring, as seen in Fig. 3.5 at actions halfway between values of action given by the scar quantization condition for maximal scarring. As studied by Kaplan [66], in an open quantum system coupled to the environment by one channel located on a short unstable periodic orbit, antiscarring causes the probability to remain in the system at times large compared to the Heisenberg time to be substantially enhanced relative to the prediction of random matrix theory. Therefore, we must expect that antiscarring, which we have demonstrated now for the case of level splittings in a smooth chaotic double-well potential, would markedly alter, away from random matrix theory predictions, the distribution of resonance widths in the a chaotic metastable potential, and also the long-time probability to remain in such a well.

In Fig. 3.9 we show the relation between the rescaled splitting $\Delta E/e^{-S}$ and the mean level spacing Δ, which decreases from about 3 at $E = -50$ to about 1 at $E = 0$, aside from some fluctuations. There is no direct correlation between $\Delta E/e^{-S}$ and Δ, thus refuting the intuitive expectation that the tunneling rate should be proportional only to the rate of attempts to cross the barrier given by the classical motion, as discussed above in Section 3.2. In the presence of scarring on the horizontal periodic orbit, tunneling is enhanced by the tendency to remain near the horizontal periodic orbit. At energies for which scarring takes place, the typical wave function intensity measured using a Gaussian at the turning point will be enhanced by a factor of $O(1/\lambda)$ compared with the naive expectation, where λ is the Lyapunov exponent. At energies for which antiscarring, a tendency to avoid the horizontal periodic orbit, takes place, this typical intensity will be strongly suppressed, by an amount that is exponentially small in λ for small λ. The actual distribution of the rescaled splitting $\Delta E/e^{-S}$ versus the level spacing Δ includes the same energy-dependent oscillations seen in Fig. 3.4, as a function of $\Delta(E)$ rather than of E itself. It is evident, then, that chaotic tunneling in two dimensions must be thought of as a quantum-coherent phenomenon, in which the probability of tunneling through the barrier is greater if one comes back in phase when making repeated attempts to cross the barrier, as happens for scarred eigenfunctions. We also note that the horizontal periodic orbit becomes more unstable at lower energies, leading to smaller scar peaks in the mean wave function intensity on the orbit, and thus compensating to some extent for an increase in the mean level spacing at lower energies. This may partly explain the absence a clear trend in the data of Fig. 3.9.

We now discuss how our data are limited by the fact that we must work at finite $\hbar$. First, the sharp cutoff at large splittings in the numerical data relative to the Porter–Thomas distribution in Fig. 3.10 is a finite-$\hbar$ effect. This can be understood as follows. Let the Poincaré surface of section have area N in units of h. Then the expected squared overlap $\langle s\rangle$ of an eigenstate with a Gaussian test state will be $1/N$, because the test state covers an area h in phase space while the eigenstate is, on average, spread evenly over the entire phase space. Now the cutoff arises from the fact that no matter how scarred or otherwise localized the eigenstate is, its overlap with a test state cannot be greater than unity. So $s < 1$ by construction, or $s/\langle s\rangle < N$. Thus the cutoff increases to infinity in the semiclassical limit (i.e., as $\hbar$ tends to zero), even

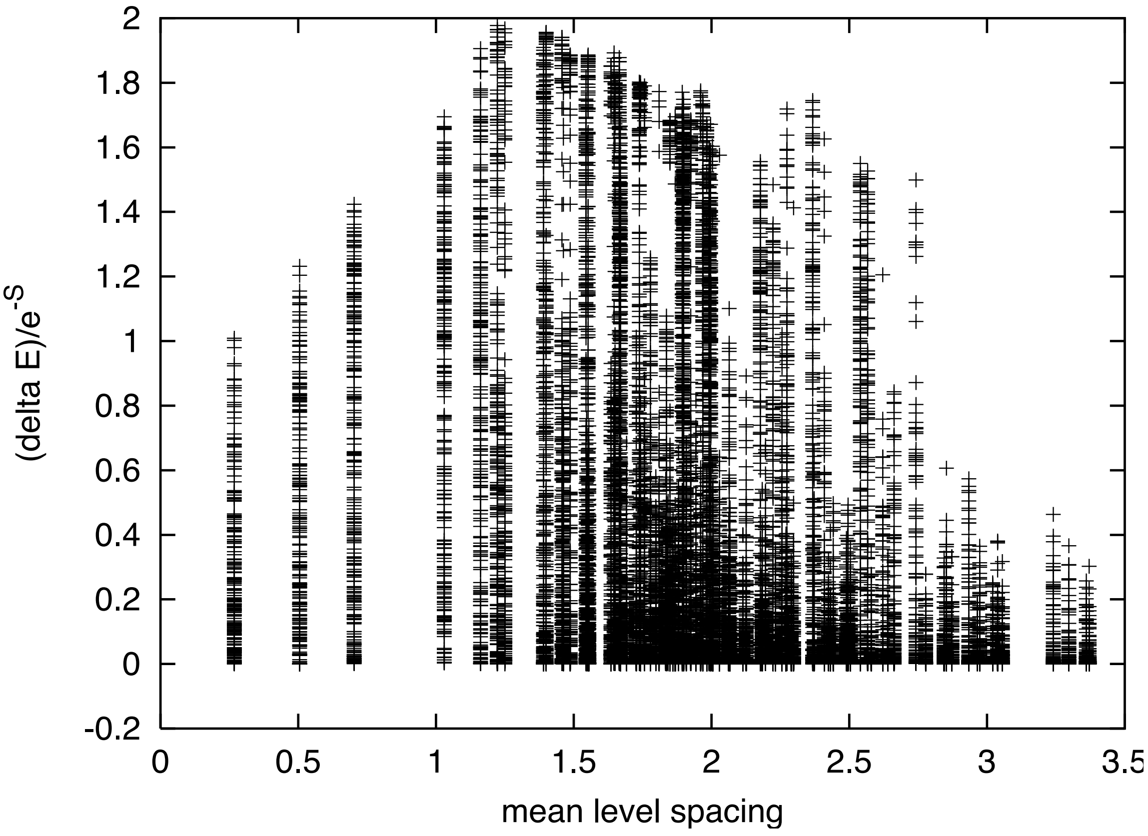


Figure 3.9: Plot of $\Delta E/e^{-S}$ versus mean level spacing Δ.

while $\langle s \rangle$ itself is decreasing. Assuming random matrix theory, the modified form of the Porter–Thomas distribution for finite N can be computed. One takes an ensemble of randomly oriented vectors in N dimensions, normalizes them so they lie on the unit sphere, and takes N times the square of the z-component. This quantity has mean 1 and a sharp cutoff at N. The Porter–Thomas distribution is recovered in the limit $N \to \infty$. For an analytical form for the Porter–Thomas distribution for finite N, see Brody *et al.*, in Ref. [18], especially their Eq. (7.10). In Fig. 3.11 we see that the modified Porter–Thomas distribution for $N = 6$, corresponding roughly to the effective dimension of our Hilbert space, reproduces the cutoff in the numerical data of Fig. 3.10. The scarring corrections (extra splittings at large and small $s/\langle s \rangle$ with fewer splittings around $s/\langle s \rangle = 1$) relative to the modified Porter–Thomas distribution are still present in Fig. 3.11.

A second test of the effect of finite $\hbar$ is to repeat the calculation at different values of $\hbar$. Since we are near the computational limit already, we consider only the case of larger $\hbar$. This is done by scaling the coordinates $(x, y) \to (x', y') = (cx, cy)$, $0 < c < 1$. Under this transformation the potential becomes

$$V(x', y') = \frac{x'^4}{c^4} - \frac{x_0^2 x'^2}{c^2} + \frac{a y'^2}{c^2} + \frac{\lambda x'^2 y'^2}{c^4} + \sum_i b_i e^{-((x' - cx_i)^2 + (y' - cy_i)^2)/(c\sigma_i)^2} \,, \qquad (3.7)$$

while the kinetic energy remains

$$-\frac{\partial^2}{\partial x'^2} - \frac{\partial^2}{\partial y'^2} \qquad (3.8)$$

since the momenta are not affected by the transformation. The complete transformation of the Hamiltonian may be regarded as the product of three transformations:

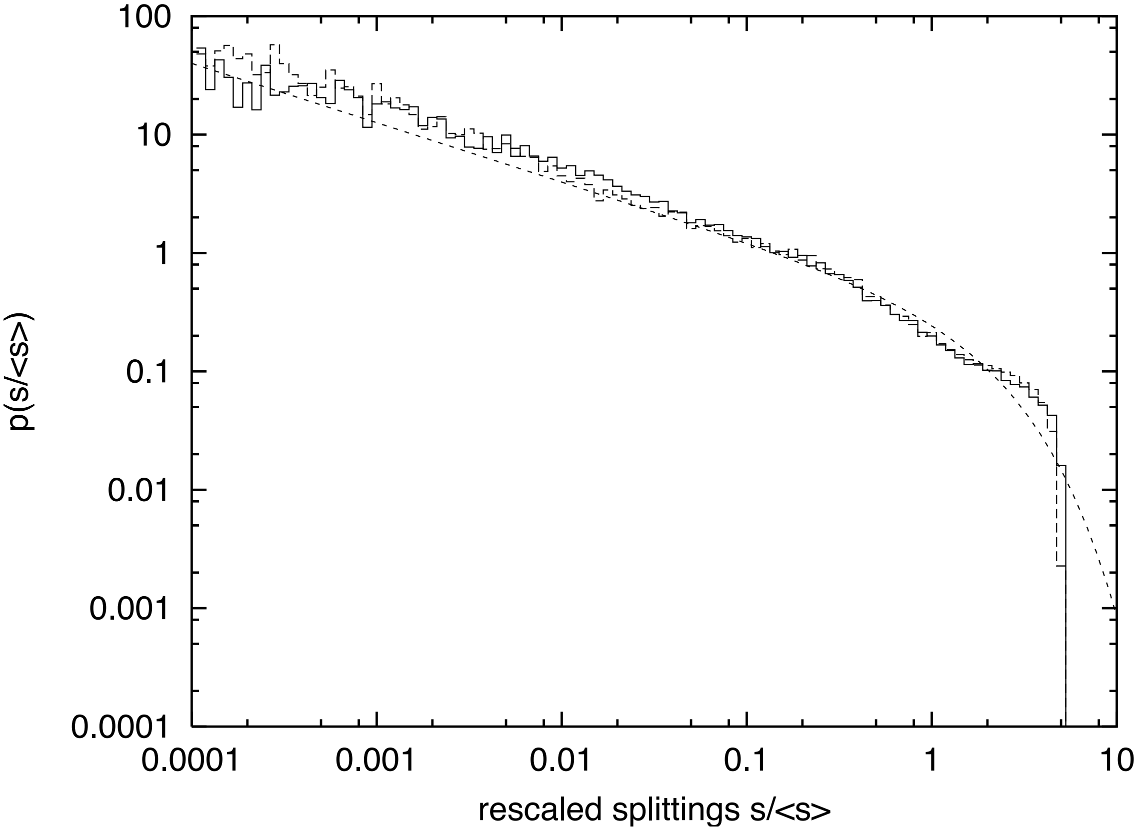


Figure 3.10: Distribution of rescaled splittings for numerical data with $\sigma = 0.5$, solid histogram; numerical data with $\sigma = 0.63$, dashed histogram; Porter–Thomas distribution (without correction for finite $\hbar$, see Fig. 3.11), dashed line.

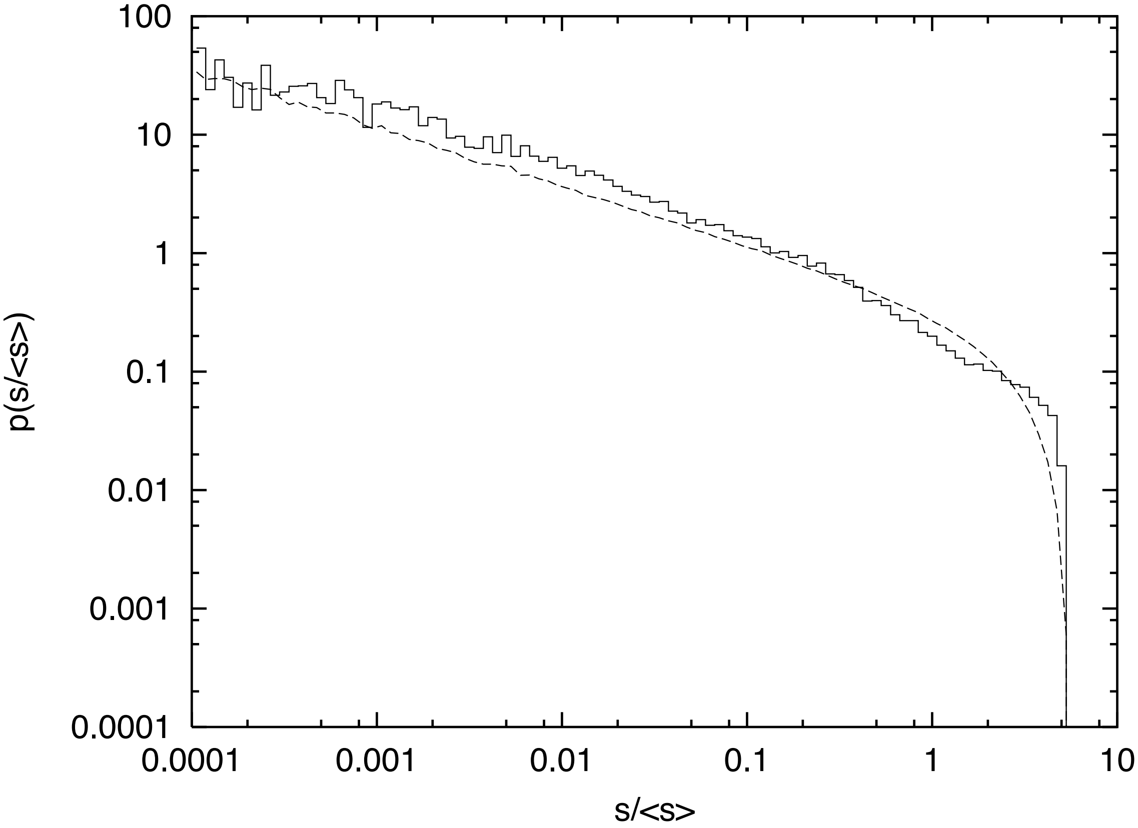


Figure 3.11: Distribution of rescaled splittings for the numerical data with $\sigma = 0.5$, histogram; Porter–Thomas distribution with finite-$\hbar$ correction for $N = 6$, dashed line.

(i) the scaling of coordinates by a factor of $c^{1/2}$ and momenta by a factor of $c^{-1/2}$, which does not change the quantum mechanics, (ii) scaling both coordinates and momenta by a common factor of $c^{1/2}$ while also replacing the Hamiltonian H by cH, which preserves the classical mechanics exactly but is not area-preserving, and thus affects the quantum mechanics by changing the effective value of $\hbar$, and (iii) scaling the Hamiltonian by a factor of $1/c$, which trivially rescales the spectrum back into the original range. The reason we use this transformation is to keep the classical mechanics, all the periodic orbits, their stability properties, etc. unchanged as we change the effective value of $\hbar$, so the results for different values of the effective $\hbar$ (which scales as $1/c$) are directly comparable.

For $c = 0.8$ we find the same oscillations observed peviously in the distribution of rescaled splittings as a function of energy, only now there are four peaks in the range from $E = -50$ to $E = 0$ compared to the five that we saw before, corresponding to a larger effective value of $\hbar$ in the new system. The distribution of splittings is given by the scarring corrections to the modified Porter–Thomas distribution for $N = 4$ now, compared to $N = 6$ above. Thus, the same conclusions continue to hold but with the expected modifications for larger $\hbar$. This indicates that at $c = 1$ we are far enough into the semiclassical regime to see characteristic semiclassical behavior for the locations of the scarring peaks, if not for their precise heights.

3.5 Conclusions

We have demonstrated that scarring on the real continuation of the optimal tunneling path, if it is an unstable periodic orbit, enhances tunneling and thus leads to larger splittings between the symmetric and antisymmetric in x eigenfunctions at energies near the scarring energies (likewise, antiscarring in between the scarring energies leads to smaller splittings). The distribution of splittings displays quantization in action, and the shape of the smooth envelope is roughly consistent with the prediction of scar theory, though the magnitude of the oscillations is not quantitatively predicted by the simple linearized dynamics; a better understanding of the shape of the envelope would require extending scar theory to the non-linear regime. Also, the distribution of splittings is approximately Porter–Thomas with scarring corrections, as we would expect on the basis of scar theory combined with the theory of Creagh and Whelan, discussed in Section 3.2. We do not find, however, the expected dependence on the Lyapunov exponent of the horizontal periodic orbit. This is presumably due to the fact that our calculations do not probe very far into the semiclassical limit, our well being only a few wavelengths across in the transverse (y) direction. Finite-$\hbar$ effects cut off the far tail of the splitting distribution at all energies.

According to Eqs. (3.2) and (3.3), suitably generalized to the chaotic double-well potential in two dimensions as discussed in Section 3.2, the rescaled resonance widths in a single metastable well, the potential of which agreed with the double-well potential we are using for $x < +x_0/\sqrt{2}$, would have the same distribution as the rescaled splittings we have computed. Thus, our results imply a non-statistical distribution of resonance widths in a chaotic metastable well. In view of its importance for chemical physics, this conclusion deserves further investigation.

Finally, we discuss the prospects for many-dimensional systems. If there exists an unstable periodic orbit near the real continuation of the optimal tunneling path in a double well or metastable well, scarring and antiscarring will again play a role. The only question is whether the degree of instability is small enough for scarring to be important; for a Lyapunov exponent λ large compared to unity the short-time envelope approaches the uniform limit of random matrix theory. However, as long as the sum of all instability exponents in directions transverse to the reaction coordinate does not become large, scarring effects are expected to appear, just as in the two-dimensional case discussed in the present chapter.

Chapter 1

Introduction to Part II: Thermal and Thermoelectric Effects in Semiconductor Superlattices

In this introductory chapter we review the work to be presented in the following four related chapters, Chapters 2 to 5 (which include a complete list of references). Thermoelectric effects arise in non-equilibrium transport problems in which either or both an electric field and a temperature gradient are present. Linear phenomenological equations [31] describe the fluxes (electric current $\mathbf{J}^e$ and heat current $\mathbf{J}^Q$) in terms of the generalized forces $\mathcal{E}$ and $-\nabla T/T$:

$$\begin{pmatrix} \mathbf{J}^e \\ \mathbf{J}^Q \end{pmatrix} = \begin{pmatrix} L^{11} & L^{21} \\ L^{12} & L^{22} \end{pmatrix} \begin{pmatrix} \mathcal{E} \\ -\nabla T/T \end{pmatrix}. \tag{1.1}$$

The diagonal terms L^{11} and L^{22} give the usual electrical conductivity and thermal conductivity (at zero electric field). The off-diagonal terms L^{12} and L^{21} describe the properly thermoelectric effects: (1) under open-circuit conditions ($J^e = 0$) a thermal gradient gives rise to a voltage drop, known as the Seebeck effect. It can be understood intuitively in terms of a tendency for hot electrons to spend more time around the cold end of the sample than for cold electrons to spend near the hot end. The coefficient of proportionality S, where $\mathcal{E}^{\mathrm{ind}} = S\nabla T$, is called the Seebeck coefficient. (2) Even in the absence of a thermal gradient, an electrical current gives rise to a heat current $J^Q = \Pi J$, where Π is the Peltier coefficient. The Peltier effect can be observed in a closed circuit consisting of two materials A and B. At the junctions between the two materials a heat proportional to the current J and to the difference in Peltier coefficients, $(\Pi_A - \Pi_B)$, is evolved or absorbed.

Thermoelectricity has important technological applications [42, 81]. The Peltier effect, for instance, serves as the basis for thermoelectric refrigeration. As illustrated in Fig. 1.1, the basic element of a thermoelectric refrigerator is a so-called *pn*-couple. The hot and cold reservoirs are connected by semiconductors with oppositely charged carriers. The sign of the Seebeck coefficient S is the same as that of the charge carriers. Thus, the device geometry can be arranged such that both conductors contribute to the transport of heat from cold to hot when a current is flowing, since, by the Onsager relation, $\Pi = TS$ and the opposite signs of S cancel the opposite signs of J to give a heat current $J^Q = TSJ$ in the same direction for both conductors. In an actual refrigerator the basic *pn*-couple is repeated many times across the surface of the device. By reversing the direction of current flow, of course, a thermoelectric refrigerator can be used to heat rather than to cool. If a simple thermoelectric refrigerator operating between a cold source at temperature T_c and a hot reservoir at temperature T_h is analyzed using Eqs. (1.1) together with the condition of local conservation of energy [81], the coefficient of refrigerator performance $\mathrm{COP} = J_c^Q/P$, where J_c^Q is the rate at which heat is removed from the cold source and P is the input power, can be maximized with respect to variation of the current J^e. The maximum is given by

$$\mathrm{COP} = \frac{T_c\gamma - T_h}{\Delta T(\gamma + 1)} \tag{1.2}$$

where $\Delta T = T_h - T_c$, $\gamma = \sqrt{1 + ZT}$, $T = (T_h + T_c)/2$ and, for an isotropic material, $Z = \sigma S^2/\kappa$. The COP increases as Z increases. For this reason, Z is known as the thermoelectric figure of merit (since the quantity ZT appearing in Eq. (1.2) is dimensionless it is often referred to instead as the figure of merit). The dimensionless quantity ZT is positive and, in principle, unbounded. A value of ZT appreciably greater than 1 would have important technological applications. As ZT tends to infinity the Carnot limit $T_c/\Delta T$ is recovered.

The Seebeck effect enables one to generate electrical power, for example in space probes, given a source of heat energy, such as a radioactive material. The efficiency is defined as $\eta = P/J_h^Q$, P being the power generated and J_h^Q the heat flow from the hot source. A similar analysis to that done for refrigerators gives for the maximum efficiency with respect to variation of the current

$$\eta_{\mathrm{max}} = \frac{\Delta T(\gamma - 1)}{\gamma T_h + T_c}. \tag{1.3}$$

Again, η_{max} increases with ZT and the ideal Carnot efficiency of $\Delta T/T_h$ is recovered in the limit as ZT tends to infinity. Typically the power generation is done in multiple

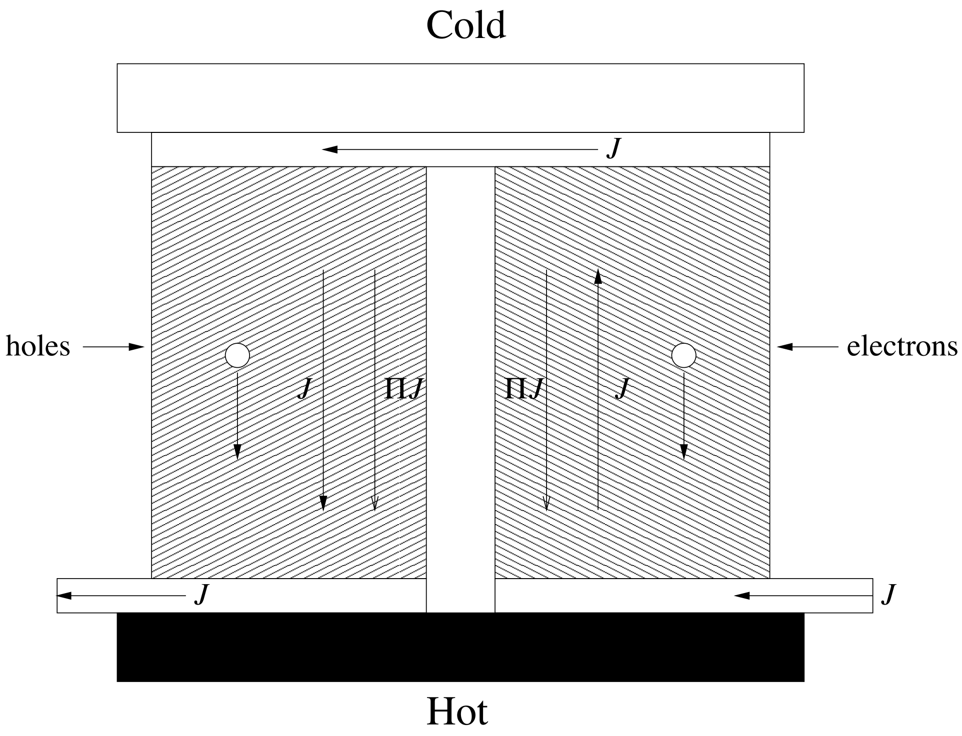


Figure 1.1: A *pn*-couple for a thermoelectric refrigerator. Electric current flow is shown by solid arrows. The component of current in the vertical direction is opposite on the two sides, but the sign of the Seebeck coefficient is also opposite, resulting in a contribution to heat flow from cold to hot in the same direction on both sides. Figure taken from Mahan's review [83].

stages, where at each stage the material is selected to give the optimal performance for the range of temperatures at that stage.

Another application of thermoelectricity is the cooling of infrared detectors to eliminate thermal noise; it is hoped that HgTe/CdTe superlattices could be employed as the thermoelectric material in a refrigerator which could be directly integrated with the detector.

For high ZT one needs σ and S large at the same time. In practice this means semiconductors, because insulators have small σ while metals tend to have small S (with an exception for some transition-metal and rare-earth compounds). Since κ is in the denominator of ZT, one would like to have small κ. However, $\kappa = \kappa_\ell + \kappa_e$ where κ_ℓ is the lattice thermal conductivity and κ_e the electronic thermal conductivity. Now, κ_e is related to σ by the Wiedemann-Franz law, $\kappa_e = LT\sigma$, where for metals the Lorenz number $L = (\pi^2/3)(k_B/e)^2$. Thus, at fixed κ_ℓ, the numerator decreases along with the denominator of ZT. This means that the quantity to minimize is κ_ℓ/σ. For a good thermoelectric, one wants $\kappa_\ell = 1$ W/(m K) or less. Skutterudites offer a strategy for lowering κ_ℓ. These rare-earth compounds have a large unit cell with open spaces into which additional atoms, known as rattlers, can be interpolated. The rattlers absorb thermal energy without propagating it efficiently, leading to low lattice thermal conductivities.

An intensive search for thermoelectric materials with high ZT was carried out in the 1950's and 1960's. Around room temperatures, semiconducting Bi_2Te_3 and its alloys, used for refrigerators, give ZT around 1. At low temperatures (100-

200K) the only known thermoelectric material is Bi-Sb (about 15% Sb). Thermoelectric power generation requires materials that perform well at higher temperatures; there, the best materials are PbTe for n-type elements and TAGS (tellurium-antimony-germanium-silver) for p-type elements (around 700K) and Si-Ge alloys (around 1000K). All of these thermoelectric materials have figures of merit not greatly exceeding 1. For standard temperatures $T_h = 323\text{K}$ and $T_c = 263\text{K}$, $ZT = 1$ translates into a COP of 0.34, whereas conventional cooling technology based on freon compressors has a COP of 1.2 to 1.4. For thermoelectric refrigeration to be competitive with compressor technology $ZT = 3$ to 4 would be required. Suitable materials yielding such ZTs remain to be found. Meanwhile, thermoelectric refrigeration is used in applications in which reliability is more important than efficiency. In the past decade there has been a revival of interest in thermoelectricity based on the belief that novel materials and structures, often involving anisotropy or confined geometries and quantum conduction, could yield improvements in ZT, or perhaps at least a better material at low temperatures, which is needed for many applications. This raises the theoretical question of how thermoelectric materials are to be analyzed in the presence of anisotropy.

Nye [92] discusses the physical situation that arises when one has an electrical conductivity or thermal conductivity which is anisotropic. Let $\mathcal{E}^{\text{ext}}$ be the external applied field. Then $\mathbf{J}^e = \underline{\sigma}\mathcal{E}^{\text{ext}}$ is not parallel to $\mathcal{E}^{\text{ext}}$. An induced field $\mathcal{E}^{\text{ind}}$ arises on the transverse surfaces such that the total field $\mathcal{E} = \mathcal{E}^{\text{ext}} + \mathcal{E}^{\text{ind}}$ yields $\underline{\sigma}\mathcal{E}$ parallel to the axis of the device. The situation is the same in the case of anisotropic heat conduction. The resulting transverse temperature gradient was predicted by Kelvin [115] and observed by Borelius and Lindh [13], Bridgman [16, 17] and Terada and Tsutsui [114]. Another anisotropic effect that arises when the current flow is not uniform is the Bridgman or internal Peltier effect [35, 36]. Following Eq. (1.1) we may write for the local evolution of heat

$$\begin{aligned} \nabla \cdot \mathbf{J}^Q &= T\frac{\partial}{\partial x_\beta}\left(J^e_\alpha S_{\alpha\beta}\right) && (1.4)\\ &= J^e_\alpha T\frac{\partial S_{\alpha\beta}}{\partial T}\frac{\partial T}{\partial x_\beta} + J^e_\alpha T\frac{\partial S_{\alpha\beta}}{\partial x_\beta} + TS_{\alpha\beta}\frac{\partial J^e_\alpha}{\partial x_\beta}. && (1.5) \end{aligned}$$

The first term on the right is the Thomson heat, the second the Peltier heat, which typically arises at the junction between two materials with differing Seebeck coefficients, and the third is the Bridgman heat that may arise in an anisotropic material. It disappears in an isotropic material because $\nabla \cdot \mathbf{J}^e = 0$ in a steady state. While the existence of anisotropic effects in thermoelectric materials has been recognized in the literature, the consequences of anisotropy for thermoelectric devices, especially for the figure of merit, have not been investigated previously.

This Part of the thesis considers thermoelectric performance of devices constructed from anisotropic materials. The principal new results are (1) the demonstration of the existence of sizable induced electric fields and thermal gradients which modify the transport coefficients, as explained in Chapters 2 and 4; (2) a theorem stating that, in the presence of an isotropic lattice thermal conductivity, the device orientation giving maximum ZT will be that in which current flows along the principal direction of maximal electrical conductivity; (3) a physical understanding, in Chapter 3, of the origin of the experimentally observed ten-fold reduction in lattice thermal conductivity along the growth axis of GaAs/AlAs superlattices, proposed as thermoelectric devices with cross-plane current flow.

In Chapter 2, which is based on the general results to be presented in Chapter 4, we investigate the role of anisotropy in thermoelectric materials. As a first consequence of anisotropy, we find that there are induced electric fields and thermal gradients which are needed to ensure that in the steady state conduction of electricity and of heat takes place along the long axis of the device. The induced fields, which are implicit in Nye, can be a substantial fraction of the external fields, and should be easily observable. Effective linear phenomenological equations that replace the microscopic equations are derived. The effective transport coefficients $\sigma_{\mathrm{eff}}, S_{\mathrm{eff}}$ and κ_{eff} appearing in these equations are always reduced in magnitude relative to the microscopic ones because of the induced fields. Another consequence of anisotropy is that the figure of merit ZT depends on the sample orientation, and, remarkably, can be computed from the effective transport coefficients as $ZT = T\sigma_{\mathrm{eff}}S_{\mathrm{eff}}^2/\kappa_{\mathrm{eff}}$. Chapter 2 contains applications of these ideas to some materials of technological interest, $\mathrm{Bi_2Te_3}$ and $\mathrm{Hg_{1-x}Cd_xTe}$. The conduction band of bulk n-type $\mathrm{Bi_2Te_3}$ at room temperature consists of six degenerate valleys related by symmetry operations. The effective-mass tensor in each valley is highly anisotropic, with principal values $0.025m_0$, $0.19m_0$ and $0.25m_0$ (for more detail see Chapter 2). A substantial induced field of 76% of the applied field is predicted. This would result in an effective electrical conductivity as much as 60% smaller than the component of the microscopic electrical conductivity tensor in the transport direction. This effect should readily be seen experimentally.

In the case of HgTe/Hg$_{1-x}$Cd$_x$Te superlattices, both the well and barrier materials have direct Γ-point band gaps. The lowest conduction subband $C1$ has a constant-energy surface which is an ellipsoid of revolution aligned with the growth axis of the SL; thus, the conductivity will be a minimum (σ_{min}) along the growth axis and a maximum (σ_{max}) in the plane of the superlattice. We find that the ratio of induced to external electric field can, for suitable angles between the growth axis of the superlattice and the transport direction, be unbounded as the anisotropy ($\sigma_{\text{max}}/\sigma_{\text{min}}$) increases. The degree of anisotropy is tunable by varying the widths of the two layers. In summary, the effects predicted in Chapter 2 are expected to be important and readily observable. To our knowledge, they do not appear to have been observed thus far.

We now turn to the motivation for Chapter 3. Quantum wells with in-plane current flow comprise an important class of anisotropic thermoelectric materials. Assuming a one-band material with parabolic bands and a perfectly two-dimensional density of states for the electrons, Hicks and Dresselhaus [54] found for Bi$_2$Te$_3$ that ZT is significantly enhanced over the bulk value for thin enough layers, by an amount that is monotonically increasing with decreasing layer thickness. The degree of enhancement depends on the alignment of the plane in which the layers are grown relative to the principal axes of the effective-mass tensor. The enhancement of ZT is due entirely to a two-dimensional density-of-states. Hicks and Dresselhaus [55] find a similar enhancement for quantum wires as long as the width of the wire is narrower than the thermal de Broglie wavelength of the carriers. Reinecke and coworkers [78, 19, 20] improved on these results by including the effects of heat and electronic transport through barriers of non-zero thickness. They find a modest enhancement of ZT by about 20% in a Bi_2Te_3/PbTe superlattice with a period of about 60Å; for larger periods ZT decreases towards its bulk value, while for smaller periods the performance is degraded by tunneling between layers. The more realistic treatment of Reinecke *et al.* fails to reproduce the dramatic enhancements in ZT reported by Hicks and Dresselhaus because the latter work assumes only a single conducting layer. In any multilayer device there must be barrier layers between the conducting layers. Any non-zero thermal conductivity in the barrier layers degrades the performance in bulk. In addition, tunneling lowers the performance for barrier layers that are too thin. In contrast to the disappointing results for superlattices with in-plane current flow, experiment gives a factor-of-ten reduction in the component of lattice thermal conductivity in the growth direction, κ_{zz}, for GaAs/AlAs and Bi_2Te_3/Sb_2Te_3 superlattices [23, 24, 120, 121]. The large reduction in κ_{zz} is not accompanied by a corresponding change in either σ or S. Thus we propose that superlattices operating with cross-plane current flow would make efficient thermoelectric devices, in which the improved performance would be attributable to the reduced lattice thermal conductivity rather than to a higher electrical conductivity. (Mahan and Woods [85, 84] have proposed a method for cooling using cross-plane current flow in multilayer superlattices, but based on thermionic emission rather than any thermoelectric effect.)

To better understand the thermal conductivity reduction in superlattices we undertake first in Chapter 3 a calculation of the superlattice phonon dispersion relations and from these the thermal conductivity. We employ a realistic treatment of the lattice dynamics and phonon spectra. The lattice-dynamical model incorporates both short-range interactions to next nearest neighbors and long-range Coulomb interaction. The short-range forces are treated in Kunc's 11-parameter rigid-ion model [72, 73], which yields the dispersion relation from a phenomenological treatment of force constants to second nearest neighbors determined from experiment. The parameters for bulk GaAs and AlAs are taken from the literature [96, 101]. The Madelung energy is important in III-V compounds which are partly ionic. The Madelung sum and its derivatives are computed exactly using the technique of Ewald transformation, adapted to a superlattice. This approach is computationally intensive and replaces highly simplified and non-physical models, e.g., Hyldgaard and Mahan [59]. The thermal conductivity is obtained from the phonon Boltzmann equation in the relaxation-time approximation. This involves an integration over the superlattice Brillouin zone which is accurate to 1%. The phonon dispersion relation is modified due to the superlattice geometry. Flattening of branches results in lower phonon velocities. This leads to a factor-of-three reduction in κ_{zz}/τ in a 3x3 GaAs/AlAs superlattice. To account for the experimental factor-of-ten reduction we need another factor-of-three reduction in lifetime. Undoubtedly this is due to interface scattering previously considered by Chen [26, 27]. The lifetime for both bulk and superlattice materials can be determined from the experimental thermal conductivities. This was done for 2x2, 3x3 and 6x6 superlattices. We find (1) the phonon-dispersion-related reduction is insensitive to n, while (2) the smaller lifetime inferred for the 2x2 superlattice is consistent with the presence of greater interface scattering. The thermal conductivity reduction is shown to be more sensitive to force-constant differences than to mass differences. Our work leads to the general conclusion that a similar reduction in the contribution of the dispersive part, and perhaps in κ_{zz} as well, may be expected in any superlattice with similar mass or force-constant differences between layers.

In Chapter 4 we return to the anisotropic effects in thermoelectric materials that were discussed first in Chapter 2. A microscopic model, which assumes semiclassical transport theory and the effective-mass approximation, is introduced. The model leads to a factorization of the transport distribution tensor into a tensorial part containing information about the anisotropy and an energy-dependent scalar and hence to a similar factorization of the transport coefficients. A consequence of this factorization is the surprising conclusion that, neglecting lattice thermal conductivity, ZT is a constant independent of sample orientation. Another consequence of the formalism is that, assuming now an isotropic lattice thermal conductivity, ZT is highest in the direction of maximum electrical conductivity. When the lattice thermal conductivity is not isotropic, we expect the maximum of ZT to be in the principal direction in which $\sigma_i/\kappa_{\ell,i}$ is greatest. We do not, however, know of any material that has this property. A third consequence of our formalism is that we can derive an upper bound on ZT that generalizes that obtained by Mahan and Sofo [83] in the isotropic case, see Chapter 4, Appendix A.

The derivations in Chapters 2 and 4 rely on the Onsager relations for Eq. (1.1). Truesdell [119], however, has contended that the Onsager relations might fail to hold for anisotropic crystals of low symmetry, such as triclinic, thus putting our results in question. In response to his objection, we prove in Chapter 5 the Onsager relations for an anisotropic periodic solid. The proof rests on the fundamental principle of microscopic reversibility, which he failed to include in his macroscopic approach. We also extend our formalism to non-rectangular samples cut along the crystalline axes.

Chapter 2

Induced Electric Fields in Anisotropic Thermoelectric Materials

Published as W. E. Bies, R. J. Radtke and H. Ehrenreich, *J. Appl. Phys.* **86**, 5065 (1999).

2.1 Abstract

Transport in an anisotropic material can create potentially large induced transverse fields which reduce the measured electric and thermal conductivities relative to those computed without the induced fields. These affect the thermoelectric figure of merit ZT modestly. The induced electric field in n-type $\mathrm{Bi_2Te_3}$ is predicted to be as much as 76% of the external one and can lower the measured electrical conductivity by up to 60%. In $\mathrm{Hg_{1-x}Cd_xTe}$ superlattices, the anisotropy may be increased by varying the composition and width of the barrier and well to give induced fields much larger than the applied one. These effects should be easily observable. The present work utilizes general results applied to a microscopic model relevant for multi-valleyed materials within the effective-mass and relaxation-time approximations.

2.2 Introduction

Investigations aimed at finding systems with large thermoelectric figures of merit ZT have focused on new materials and to a lesser extent on materials structures or crystallographic anisotropy [81]. As will be demonstrated in Chapter 7, anisotropy can lead to surprising physical effects. These include the formation of possibly large induced transverse electric fields and temperature gradients and the reduction of the measured electric and thermal conductivities from their values in the absence of such fields. Transverse temperature gradients in anisotropic crystal structures were predicted by Lord Kelvin [115] in 1857 and observed as early as 1917 by Borelius and Lindh [13], Bridgman [16, 17], and Terada and Tsutsui [114]. This Chapter considers and predicts the magnitude of the induced electric fields, the effective transport coefficients, and ZT for two materials of current interest: bulk $\mathrm{Bi_2Te_3}$ and $\mathrm{Hg_{1-x}Cd_xTe}$ superlattices (SLs). The SL's anisotropy is tunable through appropriate choice of the composition and width of the well and barrier. It will be seen that the induced fields and reduced transport coefficients should be easily observable.

2.3 Formalism

The calculation proceeds from the Boltzmann equation in the relaxation-time approximation, which yields the transport coefficients

$$\sigma_{ij} = e^2 \int d\varepsilon(-\partial f_0/\partial\varepsilon)\Sigma_{ij}(\varepsilon) \tag{2.1}$$

$$T(\underline{\sigma S})_{ij} = e \int d\varepsilon(-\partial f_0/\partial\varepsilon)\Sigma_{ij}(\varepsilon)(\varepsilon-\mu) \tag{2.2}$$

$$T\kappa_{0,ij} = \int d\varepsilon(-\partial f_0/\partial\varepsilon)\Sigma_{ij}(\varepsilon)(\varepsilon-\mu)^2. \tag{2.3}$$

Here, σ_{ij} and S_{ij} are the components of the electrical conductivity and thermopower tensors, $f_0 = (\exp((\varepsilon-\mu)/k_BT)+1)^{-1}$ is the Fermi-Dirac distribution, μ the chemical potential, T the temperature ($T = 300$K throughout this work), and e the magnitude of the electron charge. $\underline{\kappa}_0$ is related to the total thermal conductivity, $\underline{\kappa}$, by $\underline{\kappa} = \underline{\kappa}_0 - T\underline{\sigma}\underline{S}^2 + \underline{\kappa}_\ell$, where $\underline{\kappa}_\ell$ is the lattice thermal conductivity (throughout this Chapter, underbars will denote tensors).

$$\Sigma_{ij}(\varepsilon) = \int \frac{2d^3\mathbf{k}}{(2\pi)^3} v_i(\mathbf{k})v_j(\mathbf{k})\tau(\mathbf{k})\delta(\varepsilon-\varepsilon(\mathbf{k})) \tag{2.4}$$

are the components of the transport distribution tensor, a generalization of the function discussed by Mahan and Sofo [83]. In Eq. (2.4), $\varepsilon(\mathbf{k})$ is the electronic dispersion relation, $v_i(\mathbf{k}) = \hbar^{-1}\partial\varepsilon(\mathbf{k})/\partial k_i$ the group velocity, and $\tau(\mathbf{k})$ the relaxation time.

Conduction is assumed to take place in a single parabolic band having N degenerate valleys. Intervalley scattering will be neglected. Thus the transport distribution tensor involves just a sum over the N valleys. Assuming the relaxation time to be a function of energy alone, $\tau(\mathbf{k}) = \tau(\varepsilon(\mathbf{k}))$, and independent of crystal orientation,

$$\Sigma_{ij}(\varepsilon) = \sum_{n=1}^{N} \Sigma_{ij}^{(n)}(\varepsilon) = \frac{2^{3/2}\tau(\varepsilon)\varepsilon^{3/2}}{3\pi^2\hbar^3}\left[\sum_{n=1}^{N}\sqrt{\det \underline{M}^{(n)}}\, M_{ij}^{(n)-1}\right], \tag{2.5}$$

where $\underline{M}^{(n)}$ is the effective mass tensor and $\Sigma_{ij}^{(n)}(\varepsilon)$ the transport distribution function of valley n. The separation of the tensor structure [square brackets in Eq. (2.5)] and the energy dependence implies that the Seebeck tensor $(\underline{\sigma S})/\underline{\sigma}$ and the Lorenz tensor $\underline{\kappa}/\underline{\sigma}T$ are necessarily *isotropic.* Furthermore, when the lattice thermal conductivity is neglected, ZT is independent of sample orientation. This is not the case in general, since $\underline{\kappa}_\ell$ cannot be neglected.

When external fields or currents are applied along a non-principal direction in an anisotropic material, transverse fields are induced which can be shown to lower the measured electric and thermal conductivities (see Chapter 7). Under experimental conditions where the heat and electric currents transverse to the external currents vanish, the induced electric field $\mathcal{E}_\perp$ relative to the applied field $\mathcal{E}_\parallel$ is

$$\mathcal{E}_\perp = -\frac{\underline{\sigma}_\perp^{-1}\underline{\sigma}_{od} + \sigma_1 T \underline{S}_\perp \underline{\kappa}_\perp^{-1}\underline{S}_{od}}{1 + \sigma_1 T \underline{S}_{od}^T \underline{\kappa}_\perp^{-1}\underline{S}_{od}}\,\mathcal{E}_\parallel \tag{2.6}$$

and the effective conductivity is

$$\sigma_{\text{eff}} = \frac{\sigma_1}{1 + \sigma_1 T \underline{S}_{od}^T \underline{\kappa}_\perp^{-1}\underline{S}_{od}} \tag{2.7}$$

with

$$\sigma_1 = \sigma_{\|} - \underline{\sigma}_{\mathrm{od}}^T \underline{\sigma}_{\perp}^{-1} \underline{\sigma}_{\mathrm{od}}. \tag{2.8}$$

Here, the conductivity tensor

$$\underline{\sigma} = \begin{bmatrix} \sigma_{\|} & \underline{\sigma}_{\mathrm{od}}^T \\ \underline{\sigma}_{\mathrm{od}} & \underline{\sigma}_{\perp} \end{bmatrix} \tag{2.9}$$

is written in terms of its longitudinal ($\sigma_{\|}$), transverse ($\underline{\sigma}_{\perp}$), and off-diagonal ($\underline{\sigma}_{\mathrm{od}}$) components, and a similar decomposition is used for $\underline{S}$ and $\underline{\kappa}$. Note that the denominator in Eqs. (2.6)-(2.7) does not appear in the effective mass approximation due to the isotropy of $\underline{S}$. Physically, a non-zero $\underline{\sigma}_{\mathrm{od}}$ in the presence of an external longitudinal current leads to a build up of charge on the transverse faces of the sample, creating the induced field $\mathcal{E}_{\perp}$. This field in turn induces a current in the longitudinal direction which opposes the external current $\sigma_{\|}\mathcal{E}_{\|}$, leading to a reduced effective conductivity σ_{eff} such that $0 < \sigma_{\mathrm{eff}} \leq \sigma_{\|}$.

The corresponding general expressions for the effective thermopower S_{eff} and thermal conductivity κ_{eff} are

$$S_{\mathrm{eff}} = S_{\|} - \underline{\kappa}_{\mathrm{od}}^T \underline{\kappa}_{\perp}^{-1} \underline{S}_{\mathrm{od}}, \tag{2.10}$$

$$\kappa_{\mathrm{eff}} = \kappa_{\|} - \underline{\kappa}_{\mathrm{od}}^T \underline{\kappa}_{\perp}^{-1} \underline{\kappa}_{\mathrm{od}}. \tag{2.11}$$

Note that, in the effective mass approximation, $S_{\mathrm{eff}} = S_{\|}$; i.e., the effective thermopower is the same as the longitudinal thermopower since the latter is isotropic. As shown in Chapter 7, the figure of merit ZT is given by

$$ZT = \sigma_{\mathrm{eff}} S_{\mathrm{eff}}^2 T / \kappa_{\mathrm{eff}} \tag{2.12}$$

in analogy with the result applying to isotropic solids. We have stated the unexpectedly simple general results, [Eqs. (2.6)-(2.12)] here without including the lengthy derivation (see Chapter 7).

2.4 Applications to Bi_2Te_3 and $Hg_{1-x}Cd_xTe$

Consider first bulk n-type Bi_2Te_3 at room temperature. The conduction band consists of six degenerate valleys, each described by a highly anisotropic effective-mass tensor [43]. As shown in Fig. 2.1, the constant-energy surfaces at low doping are ellipsoidal. The trigonal-bisectrix plane passes through two of these ellipsoids, which are characterized by effective masses of $0.025m_0$ near the bisectrix, $0.26m_0$ along the binary, and $0.19m_0$ near the trigonal axis [43]. These two ellipsoids are related to each other by inversion through the origin. The remaining four ellipsoids are obtained from the first two by rotations of $\pm 2\pi/3$ about the trigonal axis.

The transport distribution function, Eq. (2.5), is obtained as follows. The relaxation time $\tau(\varepsilon)$ taken to be independent of energy and is determined from the experimental mobility along the bisectrix, 1200 $\text{cm}^2\ \text{V}^{-1}\text{s}^{-1}$ [43]. The $\underline{M}^{(n)}$ are constructed from the experimentally derived inverse mass matrix for a single conduction band ellipsoid [43] combined with the point-group operations of the crystal. ZT is obtained from the resulting transport coefficients by taking the lattice thermal conductivity at 300 K to be 1.5 $\text{W m}^{-1}\ \text{K}^{-1}$ (Ref. [43]) and isotropic. Assuming a carrier density $n = 5.2 \times 10^{18}\ \text{cm}^{-3}$ yields the maximal $ZT = 0.71$ when the longitudinal direction is taken along the bisectrix.

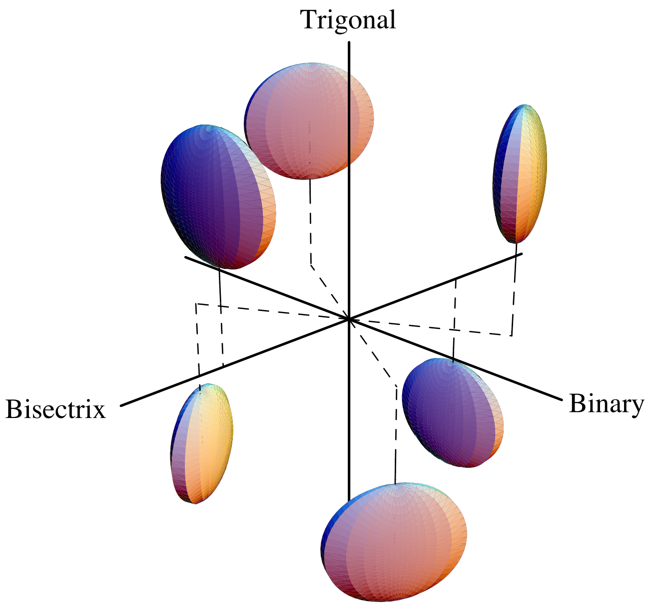


Figure 2.1: Constant-energy surfaces of the conduction band valleys in $\mathrm{Bi_2Te_3}$. The bisectrix, [010], binary, [100], and trigonal, [001], directions are indicated.

Figure 2.2(a) plots the induced electric field relative to the external one [Eq. (2.6)] as θ, the direction of the external currents (the longitudinal direction, defined in inset) is swept through the trigonal-bisectrix plane. The induced field is seen to increase from zero at $\theta = 0$ (trigonal direction) to a maximum of 76% of the external field at $\theta = 0.36\pi$ radians. Increasing θ beyond this point lowers the induced field until it vanishes when $\theta = \pi/2$ (bisectrix direction). Increasing θ further reverses the direction of the induced field. Its magnitude is a mirror image of that in the range $0 < \theta < \pi/2$.

The corresponding effective electrical conductivity σ_{eff} [Eq. (2.7)] is shown in Fig. 2.2(b), along with $\sigma_{\|}$, that computed in the absence of the induced fields. As the longitudinal direction is changed from $\theta = 0$ (trigonal) to $\theta = \pi/2$ (bisectrix),

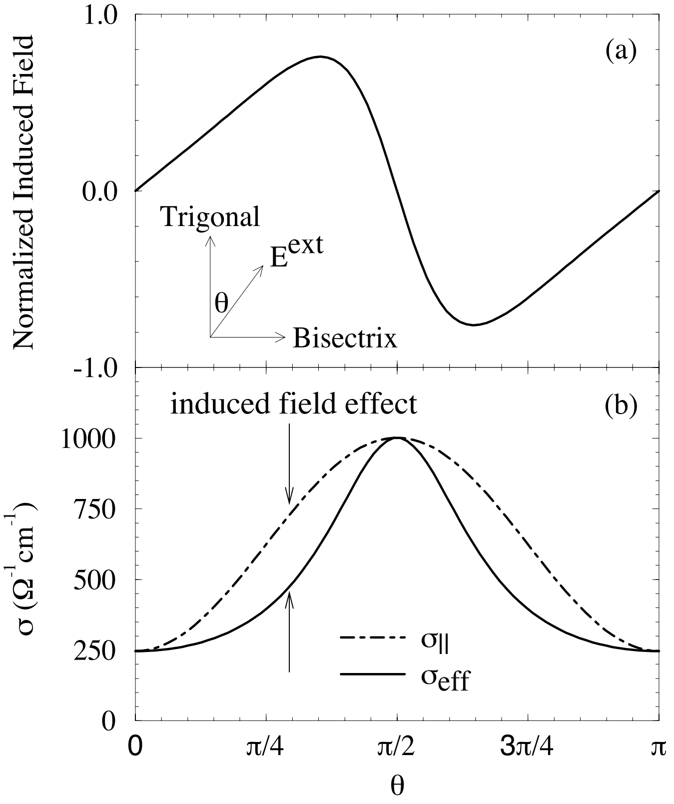


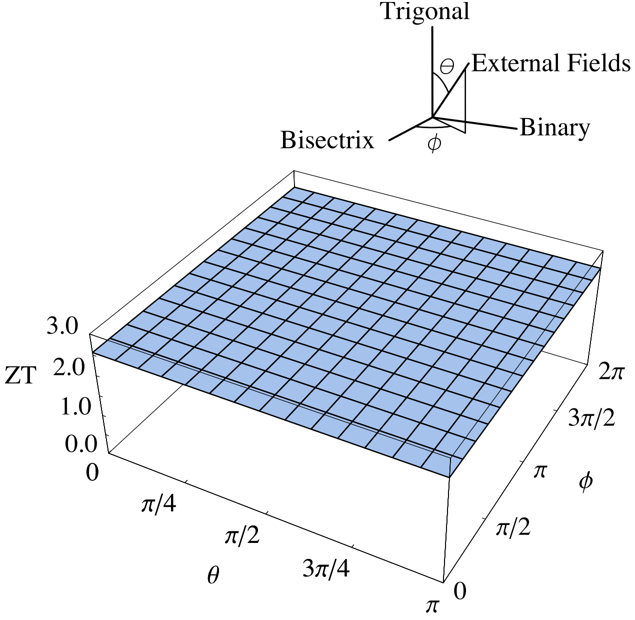
Figure 2.2: Effect of induced transverse field in bulk n-type Bi_2Te_3 at 300 K. (a) Induced / external field $\mathcal{E}_\perp/\mathcal{E}_\|$ and (b) electrical conductivity computed with (σ_{eff}, solid line) and without ($\sigma_\|$, dot-dashed line) the effects of the induced field for the sample geometry shown in the inset.

both conductivities increase by a factor of four due to the intrinsic anisotropy of the material. In general, though, σ_{eff} is smaller than $\sigma_{||}$ by as much as 60%. The influence of the induced field on the conductivity, indicated by the arrows in Fig. 2.2(b), is therefore substantial.

The combination of intrinsic anisotropy and the effects of the induced fields also affect ZT, but only if the lattice thermal conductivity κ_ℓ is non-zero. As shown in Fig. 2.3(a) for hypothetical Bi_2Te_3 having $\kappa_\ell = 0$, $ZT = 2.6$ and is isotropic. As seen in Fig. 2.3(b), when κ_ℓ assumes its bulk value, ZT becomes anisotropic and decreases to a maximum value of 0.71. The maximal ZT applies to samples with external currents along the high-conductivity bisectrix-binary plane ($\theta = \pi/2$). The minimal $ZT = 0.22$ occurs for samples with currents along the low-conductivity trigonal axis ($\theta = 0$). Despite the complicated many-valley band structure, ZT is independent of the azimuthal angle ϕ defined in the inset to Fig. 2.3(a). As seen from Fig. 2.1, when the external field is applied in the bisectrix-binary plane, the valleys present both heavy and light masses along the transport direction. The relationship among these masses is determined by the point-group symmetry of the crystal and is such that the sum in Eq. (2.5), and hence ZT, is independent of ϕ. It is surprising that the induced fields, reduced conductivity, and anisotropic ZT, which should be easily observable in Bi_2Te_3, have not yet been reported.

Now consider $HgTe/Hg_{1-x}Cd_xTe$ SLs, one of many SLs whose well and barrier materials have direct Γ-point band gaps. The constant energy surface of the lowest conduction subband $C1$ consists of a single ellipsoid of revolution aligned with

(a) $\kappa_L = 0$



(b) $\kappa_L = 1.5$ W / m K

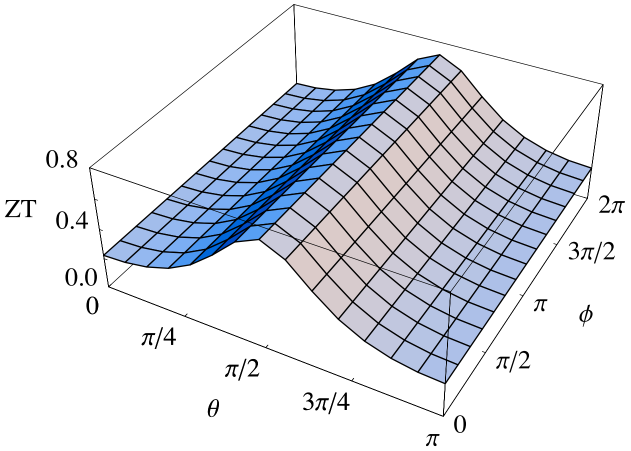


Figure 2.3: ZT of n-type Bi_2Te_3 at 300 K for (a) a hypothetical material with $\kappa_\ell = 0$ and (b) $\kappa_\ell = 1.5$ W m^{-1} K^{-1}, the bulk in-plane value, as a function of the direction in which the external fields are applied (defined in the inset).

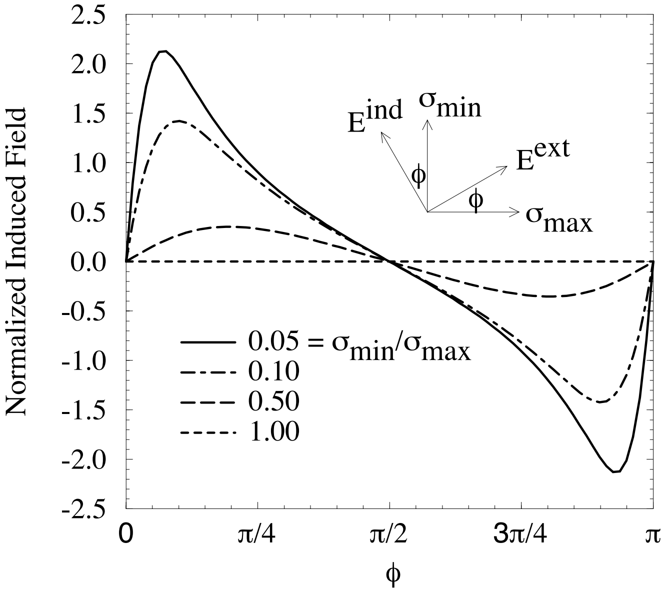


Figure 2.4: Ratio of induced to external electric field $\mathcal{E}_\perp/\mathcal{E}_\parallel$ in a superlattice such as $\mathrm{Hg_{1-x}Cd_xTe}$ as a function of the angle ϕ defined in the inset for several values of the conductivity ratio $\sigma_{\min}/\sigma_{\max}$.

the growth axis of the SL. The conductivity will typically be a minimum along the growth axis ($\sigma_{\min}$) and a maximum within the SL plane ($\sigma_{\max}$). If the external currents are applied at an angle ϕ relative to the SL planes, the induced fields will be non-zero only for sample faces which are not parallel to the SL growth axis, as indicated in the inset to Fig. 2.4. For any material with the symmetry of a SL, use of Eq. (2.6) yields

$$\frac{\mathcal{E}_\perp}{\mathcal{E}_\parallel} = \frac{(\sigma_{\max} - \sigma_{\min}) \sin\phi \cos\phi}{\sigma_{\max} \sin^2\phi + \sigma_{\min} \cos^2\phi}. \tag{2.13}$$

This relation implies that $\mathcal{E}_\perp/\mathcal{E}_\parallel \gg 1$ when $\sigma_{\min}/\sigma_{\max} \ll 3 - 2\sqrt{2} \cong 0.172$. This geometrical effect may be practically useful.

The dependence of $\mathcal{E}_\perp/\mathcal{E}_\parallel$ on ϕ and the conductivity ratio $\sigma_{\min}/\sigma_{\max}$ is shown in Fig. 2.4. For an isotropic system, $\sigma_{\min}/\sigma_{\max} = 1$, and the induced field vanishes for all sample orientations. For an anisotropic system, $\sigma_{\text{min}}/\sigma_{\text{max}} < 1$, and the qualitative features of $\mathcal{E}_{\perp}/\mathcal{E}_{||}$ in Fig. 2.4 are the same as those in Fig. 2.2(a): The induced field is finite for external currents along low-symmetry ($\phi \neq 0,\ \pi/2$) directions and exhibits a maximum near the high-conductivity in-plane direction ($\phi = 0$). From Eq. (2.13), this maximum is given by $\mathcal{E}_{\perp}/\mathcal{E}_{||}\Big|_{\text{max}} = (\sigma_{\text{max}} - \sigma_{\text{min}})/2\sqrt{\sigma_{\text{max}}\sigma_{\text{min}}}$ and occurs at $\phi_{\text{max}} = \sin^{-1}\sqrt{\sigma_{\text{min}}/(\sigma_{\text{max}} + \sigma_{\text{min}})}$. As seen in Fig. 2.4, increasing the anisotropy by decreasing $\sigma_{\text{min}}/\sigma_{\text{max}}$ can lead to induced fields much larger than the applied one.

As specific examples, consider two superlattices: (1) 30-Å HgTe / 20-Å $\text{Hg}_{0.15}\text{Cd}_{0.85}\text{Te}$ and (2) 20-Å HgTe / 40-Å $\text{Hg}_{0.15}\text{Cd}_{0.85}\text{Te}$. Realistic band structures for these materials are obtained through an 8-band $\mathbf{K}\cdot\mathbf{p}$ theory within the envelope function approximation [60, 39] based on the parameters in Tab. 2.1 obtained from Refs. [61, 58], [33], [15], [63] and [104]. Application of this approach to other II-VI and III-V SLs yields accurate band structures which reproduce experimental optical absorption spectra with no adjustable parameters [60, 39]. The computed masses of the $C1$ subband along the growth and in-plane directions required for the transport calculations are given in Tab. 2.2. The room-temperature energy gaps are seen to satisfy $E_g \geq 10k_BT$ in both structures, so hole conduction is negligible. Transport in the next highest conduction subband is also negligible, since it lies more than $20k_BT$ higher in energy than $C1$ in both SLs. The relaxation time and lattice thermal conductivity are estimated from those of an alloy of the same composition as the SL (Tab. 2.2). Thus, interface scattering is neglected.

The thermoelectric performance of these SLs is shown in Tab. 2.2 (see

Table 2.1: Room temperature energy gap E_g, spin-orbit splitting Δ, and heavy-hole effective mass m_{HH} for a HgTe/Hg$_{0.15}$Cd$_{0.85}$Te superlattice. The valence band offset is 0.30 eV from a linear interpolation of the value in Ref. [61, 58], and the momentum matrix element $(2/m_0)| < Z|p_z|S > |^2 = 18.0$ eV (Ref. [33]).

	HgTe	Hg$_{0.15}$Cd$_{0.85}$Te
E_g at 300 K (eV)^a	-0.166	1.191
Δ (eV)^b	1.0	0.9
m_{HH}/m_0^c	0.7	0.7

[a] Ref. [15].

[b] Ref. [63].

[c] Ref. [104].

Refs. [44] and [15]). Since $\sigma_{\min}/\sigma_{\max} = m_{||}/m_{\perp}$ in the effective-mass and constant relaxation-time approximations, SL (1)'s 0.66 mass ratio yields a maximum induced field of only $0.2\mathcal{E}_{||}$. SL (2)'s much smaller 0.39 mass ratio, indicating a larger anisotropy, yields a correspondingly larger induced field of $0.5\mathcal{E}_{||}$. For even larger anisotropies, the effective mass approximation breaks down. Estimates using the model band structure of Esaki and Tsu [37] indicate that induced fields up to $17\mathcal{E}_{||}$ can be obtained in a SL resembling 40-Å HgTe / 50-Å CdTe, for which $\sigma_{\min}/\sigma_{\max} = 8 \times 10^{-4}$. Note that the maximal ZTs shown in Tab. 2.2, although small in magnitude, represent approximately a 20% increase over the equivalent bulk alloy. This enhancement is obtained assuming that κ_ℓ is the same in the bulk alloy and the SL. Recent experiments [23, 120] and theory [59, 28] suggest that κ_ℓ may be considerably reduced in SLs, leading to much larger ZTs than those presented here [99].

Table 2.2: Parameters and thermoelectric performance in HgCd/$\mathrm{Hg_{0.15}Cd_{0.85}Te}$ superlattices (SLs) at 300K. $m_\perp$ and $m_{||}$ are the $C1$ effective masses along and normal to the growth axis, $(\mathcal{E}_\perp/\mathcal{E}_{||})_{\mathrm{max}}$ is the maximal relative magnitude of the induced field, ϕ_{max} is the angle relative to the SL planes at which the external field must be applied to obtain this field, $(ZT)_{\mathrm{max}}$ is the maximal figure of merit for the SL (which occurs for in-plane transport), and $(ZT)_{\mathrm{alloy}}$ is that for an alloy with the average composition of the SL.

	SL (1) 30-Å HgTe / 20-Å $\mathrm{Hg_{0.15}Cd_{0.85}Te}$	SL (2) 20-Å HgTe / 40-Å $\mathrm{Hg_{0.15}Cd_{0.85}Te}$
Band structure:[a]		
E_g (eV)	0.25	0.46
$m_\perp/m_0$	0.038	0.110
$m_{||}/m_0$	0.025	0.043
$m_{||}/m_\perp$	0.66	0.39
Transport parameters:		
τ (s)[b]	5.3×10^{-14}	2.6×10^{-14}
κ_ℓ (mW cm^{-1} K^{-1})[c]	8.2	12
Thermoelectric performance:[d]		
$(\mathcal{E}_\perp/\mathcal{E}_{||})_{\mathrm{max}}$	0.21	0.48
ϕ_{max} (rad)	0.22π	0.18π
$(ZT)_{\mathrm{max}}$	0.14	0.086
$(ZT)_{\mathrm{alloy}}$	0.12	0.061

[a] Based on the parameters in Tab. 2.1.

[b] From the mobility data presented in Ref. [44].

[c] Quadratic fit to the data in Ref. [15].

[d] This work.

Chapter 3

Phonon Dispersion Effects and the Thermal Conductivity Reduction in GaAs/AlAs Superlattices

Published as W. E. Bies, R. J. Radtke and H. Ehrenreich, *J. Appl. Phys.* **88**, 1498 (2000).

3.1 Abstract

The experimentally observed order-of-magnitude reduction in the thermal conductivity along the growth axis of $(\mathrm{GaAs})_n/(\mathrm{AlAs})_n$ (or nxn) superlattices is investigated theoretically for (2x2), (3x3) and (6x6) structures using an accurate model of the lattice dynamics. The modification of the phonon dispersion relation due to the superlattice geometry leads to flattening of the phonon branches and hence to lower phonon velocities. This effect is shown to account for a factor-of-three reduction in the thermal conductivity with respect to bulk GaAs along the growth direction; the remainder is attributable to a reduction in the phonon lifetime. The dispersion-related reduction is relatively insensitive to temperature ($100 < \mathrm{T} < 300$K) and n. The phonon lifetime reduction is largest for the 2x2 structures and consistent with greater interface scattering. The thermal conductivity reduction is shown to be appreciably more sensitive to GaAs/AlAs force constant differences than to those associated with molecular masses.

3.2 Introduction

Superlattice structures have been proposed to be materials with a high thermoelectric figure of merit ZT, for both in-plane [54] and cross-plane [85, 84] current flow. In the latter case, the improvement in thermoelectric performance is attributable to a reduced lattice thermal conductivity [85, 84, 100] rather than to a higher electronic conductivity. Experimentally, a factor-of-ten reduction in the component of lattice thermal conductivity along the growth axis, κ_{zz}, is observed in GaAs/AlAs

[23, 24] and Bi_2Te_3/Sb_2Te_3 [120, 121] superlattices (SLs). Theoretically, the thermal conductivity of Si/Ge SLs has been studied previously by Hyldgaard and Mahan [59] and by Chen [26, 27]. Within the context of a very simplified model of the phonon dynamics, the calculations of Ref. [59] were able to reproduce the factor-of-ten reduction in the thermal conductivity along the growth direction. By contrast, Chen's [26, 27] extensive work focussed on the role of thermal boundary resistance. The present Chapter investigates the reduction associated with SL induced changes of the phonon dispersion based on a realistic, computationally intensive treatment of the phonon spectra and dynamics.

The origin of the observed reduction in thermal conductivity may be explained qualitatively as follows. First, we note that, in the experimental work of Capinski and Maris [23] and Capinski *et al.* [24] on $(GaAs)_n/(AlAs)_n$ SLs for n up to 40, the phonon mean free path inferred from the thermal conductivity, heat capacity and Debye velocity is greater than 370 Å at all termperatures for which measurements exist, always large compared to the size of the SL unit cell, 5.66 Å. Thus, the phonon transport lies in a regime where the SL phonon dispersion relation and lifetime, and not those of the bulk constituents, determine the thermal conductivity. According to the expression for the thermal conductivity derived from the phonon Boltzmann equation in the relaxation-time approximation (see Eq. (3.6)), the thermal conductivity depends on (1) a quantity representing the contribution of the SL phonon dispersion relation, and (2) the lifetime τ, which contains phonon-phonon, interface and defect scattering effects. We shall focus only on item (1), whose effect can be computed within our realistic, albeit complicated, lattice-dynamical model. The effect of the SL geometry is to introduce anticrossings and new gaps in the phonon dispersion relation, when the magnitude of the phonon wavevector along the growth direction equals an integer multiple of π/d, where d is the period of the SL along the growth direction. The consequent flattening of the phonon branches near the Brillouin zone edge leads to a lowering of phonon velocities in the growth direction, and hence a reduction in thermal conductivity.

We describe in Section 3.3 the lattice-dynamical model for the SL, which is a generalization of the 11-parameter rigid-ion model of Kunc [72, 73]. It incorporates short-range interactions to next nearest neighbors and the long-range Coulomb interaction. The construction of the SL dynamical matrix is outlined in Section 3.3 and in Appendix 3.5.

The formalism is applied to a $(\mathrm{GaAs})_3/(\mathrm{AlAs})_3$ SL in Section 3.4. The phonon dispersion relation will be seen to display the flattening expected for a SL. Critical points, especially at the high-symmetry points Γ, X and Z, produce sharp peaks in the density of states in the SL. Miniband formation and anticrossings in the SL phonon dispersion relation lead to a three-fold reduction, relative to bulk, in the contribution of the phonon dispersion relation to the thermal conductivity along the growth direction. The present results contrast with those of Hyldgaard and Mahan [59], who found that, in their simplified model, the order-of-magnitude reduction of κ_{zz}/τ was attributable to effects related to the SL dispersion relation alone. In our more realistic treatment of the SL dynamics, a significant three-fold reduction in the lifetime is also needed to explain the experimental reduction in κ_{zz} by a factor of ten. Finally, the sensitivity of the decrease in κ_{zz} to differences in masses or force constants between the GaAs and AlAs layers is investigated; differences in the force constants are found to play a markedly greater role in the reduction of κ_{zz} than differences in mass.

Section 3.5 is devoted to discussion and conclusions which are broadened by examining the dependence of our results on the period n of the $(\mathrm{GaAs})_n/(\mathrm{AlAs})_n$ SL for $n = 2, 3$ and 6. These results permit some discussion of interface effects on the phonon lifetime and a more detailed comparison of the present work with that of Refs. 9 and 10. The reduction in the contribution of the phonon dispersion relation to the thermal conductivity of the SL relative to bulk is computed, and found to be approximately independent of n for the small values considered here. The lifetime for both the SL and bulk was determined using the experimental thermal conductivities of the SL and of bulk GaAs of Capinski *et al.* [24]. The lifetimes are found to be smaller for the 2x2 SL and larger and roughly equal for the 3x3 and 6x6 SLs, consistent with the presence of greater interface scattering in the 2x2 SLs.

3.3 Formalism

The lattice dynamics can be treated realistically via the 11-parameter rigid-ion model of Kunc [72, 73], which has been successfully applied to zinc-blende-structure compounds. Consider $(\mathrm{GaAs})_3/(\mathrm{AlAs})_3$. In the SL unit cell, consisting of a layer of GaAs above a layer of AlAs in the growth direction, there will be 3 pairs of GaAs followed by 3 pairs of AlAs, or 3 Ga, 3 Al and 6 As atoms in all, which may be indexed by $\kappa = 1, \ldots, 12$. Letting $u_\alpha(\ell\kappa)$ denote the displacement in the direction $\alpha = x, y, z$ of the κ-th atom in the ℓ-th unit cell, plane-wave solutions of the form

$$u_\alpha(\ell\kappa) = M_\kappa^{-1/2} e^{i(\mathbf{k}\cdot\mathbf{x}(\ell\kappa) - \omega_j(\mathbf{k})t)} w_\alpha(\kappa|\mathbf{k}j) \tag{3.1}$$

are assumed, where $\mathbf{x}(\ell\kappa)$ is the equilibrium position of the κ-th atom in the ℓ-th unit cell, and $w_\alpha(\kappa|\mathbf{k}j)$ satisfies the secular equation

$$\omega_j(\mathbf{k})^2 w_\alpha(\kappa|\mathbf{k}j) = \sum_{\beta\kappa'} C_{\alpha\beta}(\kappa\kappa'|\mathbf{k}) w_\beta(\kappa'|\mathbf{k}j). \tag{3.2}$$

The dynamical matrix $\underline{C}$ reflects the interatomic force constants of the crystal. In the present rigid-ion model, the interatomic forces are divided into (1) short-range forces extending to second nearest neighbors, and (2) the long-range Coulomb interaction. Accordingly, the dynamical matrix may be written

$$\underline{C} = \underline{C}_{sr} + \underline{C}_{\mathrm{Coul}}. \tag{3.3}$$

As a result of symmetry of the zinc-blende structure, the short-range forces to second nearest neighbors may be described by ten parameters for each material [73]. For nearest and next nearest neighbor interactions in the SL unit cell, we employ the force constants determined for the constituent bulk materials separately, rotated by the appropriate point-group operation. Bulk GaAs and AlAs parameters are taken from the literature [96, 101]. In the SL, bulk parameters are used within each layer. For the interface atoms and Ga-Al bonds crossing the interface, we employ the average of the bulk parameters following Ren *et al.* [101]. For the Coulomb interaction, the atoms are treated as point charges. The Madelung sum, and its derivatives, are computed using the usual Ewald transformation [14], which has been generalized here for SLs. This is accomplished by separating the sum over the spatial index ℓ into a sum over layers normal to the growth direction, $\ell_{||}$, and a sum along the growth axis, ℓ_z. A two-dimensional Ewald transformation is performed in each layer; these results are then summed over ℓ_z. The resulting expressions for the function $\phi(\mathbf{k}, \mathbf{r})$ and its derivatives (see the Appendix) are similar to those that arise in the usual three-dimensional Ewald procedure; however, the definite integrals differ and must be performed numerically. Detailed expressions for the Coulomb term are given in the Appendix.

The phonon Boltzmann equation in the relaxation-time approximation leads to the following expression for the lattice thermal conductivity:

$$\kappa_{ij} = \int \frac{d^3q}{(2\pi)^3} \sum_\alpha \hbar\omega_{\mathbf{q}}^{(\alpha)} \frac{\partial\omega_{\mathbf{q}}^{(\alpha)}}{\partial q_i} \frac{\partial\omega_{\mathbf{q}}^{(\alpha)}}{\partial q_j} \frac{dn(\omega_{\mathbf{q}}^{(\alpha)})}{dT} \tau_{\text{ph}}(\omega_{\mathbf{q}}^{(\alpha)}, T) \tag{3.4}$$

where $n(\omega_q^{(\alpha)})$ is the distribution function of the phonons, the sum is over branches α, and $\tau_{\text{ph}}(\omega_{\mathbf{q}}^{(\alpha)}, T)$ is the lifetime. Eq. (3.4) can be written in terms of

$$\Sigma_{ij}(\omega) = \int \frac{d^3q}{(2\pi)^3} \sum_\alpha \hbar\omega_{\mathbf{q}}^{(\alpha)} \frac{\partial\omega_{\mathbf{q}}^{(\alpha)}}{\partial q_i} \frac{\partial\omega_{\mathbf{q}}^{(\alpha)}}{\partial q_j} \delta(\omega - \omega_{\mathbf{q}}^{(\alpha)}) \tag{3.5}$$

as

$$\kappa_{ij} = \int d\omega (dn(\omega)/dT) \Sigma_{ij}(\omega) \tau_{\text{ph}}(\omega, T). \tag{3.6}$$

This Chapter will focus on the SL effects on the dispersion relation contained in $\Sigma_{ij}(\omega)$. Relaxation-time effects associated for example with scattering from interfaces, defects, umklapp processes, etc. are not considered explicitly. However, the results will be used to infer some of their properties.

3.4 $(\mathrm{GaAs})_3/(\mathrm{AlAs})_3$ Superlattices

We focus first on the $(\mathrm{GaAs})_3/(\mathrm{AlAs})_3$ SL studied experimentally by Capinski and Maris [23]. Using a picosecond pump-and-probe technique, they observed an order-of-magnitude reduction in the thermal conductivity along the growth direction, κ_{zz}, relative to bulk GaAs. The dispersion relation along the ΓX and ΓZ directions, which was generated numerically according to the method described in Section 3.3, is shown in Fig. 3.1. The significance of the labelled features will be explained in the discussion of Fig. 3.2. Because the SL unit cell contains three unit cells each of bulk GaAs and bulk AlAs, respectively, arranged along the growth axis, the edge of the SL Brillouin zone in the growth direction, Z, is one-sixth as far from the center as it is in the in-plane directions, X and Y. As a result each of the six branches in the bulk material is folded back six times along ΓZ, which is most easily seen for the longitudinal acoustic mode in Fig. 3.1. Bulk GaAs and bulk AlAs optical modes have no frequencies in common, and are therefore localized. This leads to (1) flat SL dispersion in optical modes, and (2) localization of AlAs optical modes to the AlAs layer. The GaAs optical modes are not localized, since they overlap with the acoustic modes as shown in Fig. 3.1 along ΓX. Due to the non-analyticity of $\underline{C}_{\mathrm{Coul}}$ as $q \to 0$ in the SL, $\omega(|q| \to 0)$ differ along ΓX and ΓZ (See Ref. 14).

The density of states $\rho(\omega)$ versus frequency, computed using the tetrahedral integration method as presented in MacDonald *et al.* [80], is given in Fig. 3.2. Note the transverse-acoustic features around $\omega = 80$ to $100\ \mathrm{cm}^{-1}$, the longitudinal-acoustic features around 150 to $200\ \mathrm{cm}^{-1}$, the GaAs optical feature at 220 to $260\ \mathrm{cm}^{-1}$,

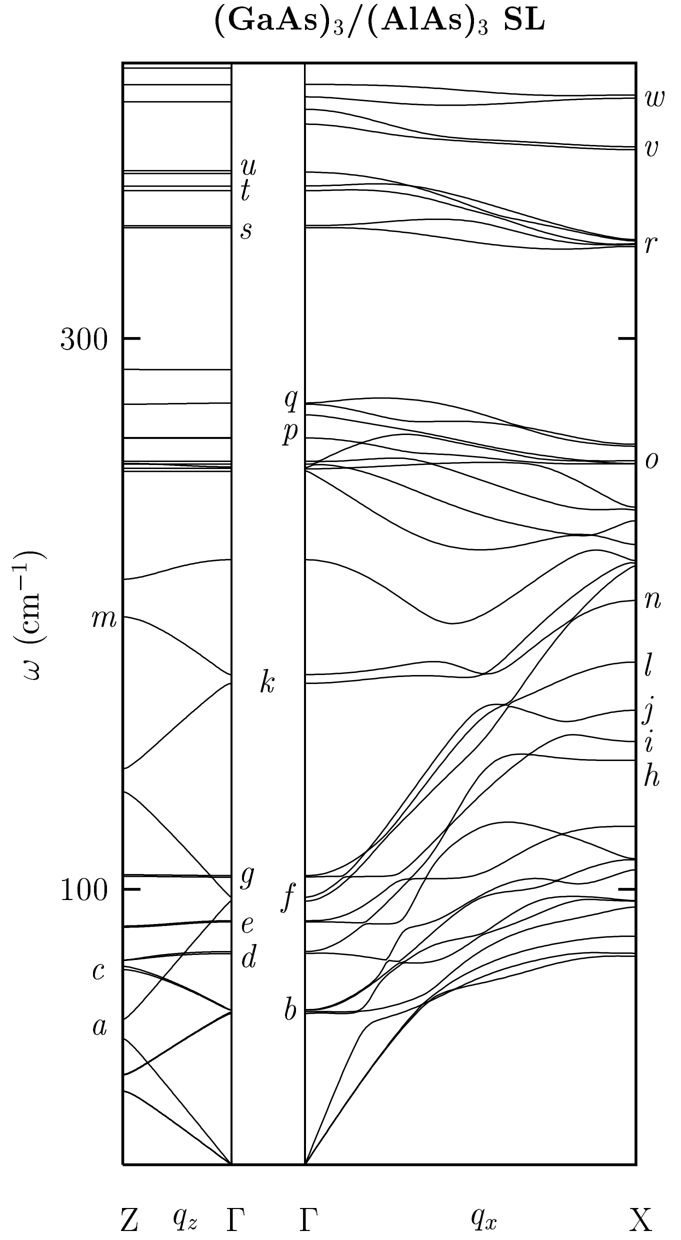


Figure 3.1: (GaAs)$_3$/(AlAs)$_3$ SL dispersion relation along the $\Gamma X = (2\pi/a_0, 0, 0)$ and $\Gamma Z = (0, 0, \pi/3a_0)$ directions; a_0 is the conventional GaAs unit cell size. The labels $a - w$ are defined in Fig. 3.2.

and, separated in frequency at higher frequencies, the AlAs optical feature at 330 to 400 cm^{-1}. Band gaps at the zone edge and anticrossings in Fig. 3.1 yield critical points which, depending on the amount of **q**-space at those frequencies, produce sharp structure in the density of states. As shown in Fig. 3.2, this structure in the density of states can be correlated with features in the dispersion relation, usually at the Γ (for folded-back bands), X and Z points. The structures at Γ, X and Z labelled in Fig. 3.2 are identified with the corresponding features in Fig. 3.1. The strength of each feature depends on the integral of $\delta(\omega - \omega_{\mathbf{q}}^{(\alpha)})$ (cf. Fig. 3.2) at that frequency, which will be large if $|\mathbf{v}|$ is small. Surprisingly, most of the peaks in the density of states appear to be associated with the critical points at Γ, X and Z. No such fine structure in the density of states exists in bulk GaAs or bulk AlAs, as may be seen for instance in Patel *et al.* [96] for GaAs. (The density of states for AlAs is not available in the literature, but our calculations confirmed the absence of fine structure for AlAs as well.)

Fig. 3.3 shows the results for $\Sigma(\omega)$, as defined in Eq. (3.5), for bulk GaAs and the SL. Note that optical modes do not contribute appreciably to $\Sigma_{zz}(\omega)$ in the SL, an effect of localization in the AlAs layers: flat dispersion leads to a vanishing of $\partial\omega_{\mathbf{q}}^{(\alpha)}/\partial q_z$. The fine structure in the density of states is also correlated with that in $\Sigma(\omega)$: peaks in the density of states DOS $\propto \int dS/|\mathbf{v}|$ become dips in $\Sigma \propto \int v^2 dS/|\mathbf{v}|$. (Here, S denotes the surface of constant frequency ω in the Brillouin zone; it consists not only of the surface containing the critical point, but also possibly of other surfaces at the same ω elsewhere in the zone.) Acoustic modes in the SL contribute less than

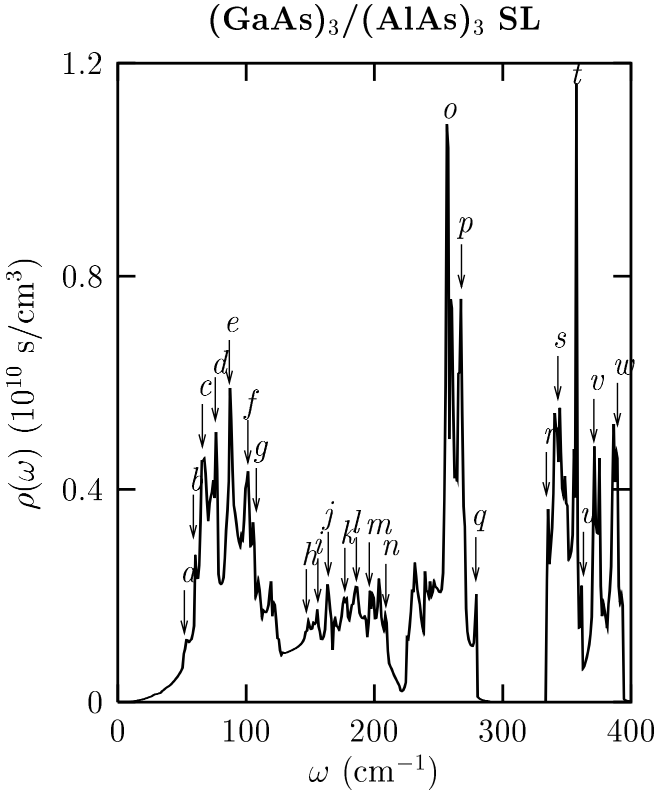


Figure 3.2: Phonon density of states for the $(GaAs)_3/(AlAs)_3$ SL, with labelled critical points identified with features in the dispersion relation in Fig. 3.1.

they would in bulk because of the band-gap and anticrossing-induced reduction in v^2. This leads to a three-fold reduction in κ_{zz}/τ at 300 K, determined here either by integrating Eq. (3.4) directly on a $60 \times 60 \times 20$ grid covering an irreducible wedge of the Brillouin zone, or by integrating Eq. (3.6). Experimentally, Capinski and Maris [23] found a ten-fold reduction factor for $(\mathrm{GaAs})_3/(\mathrm{AlAs})_3$. The full reduction factor is a product of the reduction due to Σ and that due to τ. Assuming τ to be constant at any given temperature, it can be found by requiring equality in Eq. (3.4) or (3.6) to the experimental value for the thermal conductivity in bulk or SL, respectively, if the appropriate dispersion relation is used in computing κ_{zz}/τ. The lifetime in bulk versus lifetime in SL, as well as the layer-width dependence of the results, will be discussed below. Finally, we note that $\Sigma_{zz} < \Sigma_{xx}$ in the SL for the simple physical reason that band flattening along the q_z-direction affects $\partial\omega/\partial q_z$ more than $\partial\omega/\partial q_x$.

The reduction in thermal conductivity due to Σ can be understood by means of an easily visualized picture in $\mathbf{q}$-space. In bulk, the transport quantity $q_x \sum_\alpha (dn(\omega_{\mathbf{q}}^{(\alpha)})/dT)\omega_{\mathbf{q}}^{(\alpha)}(v_{\mathbf{q},z}^{(\alpha)})^2$ represents the contribution of the phonon dispersion relation in Eq. (3.4). In a range Δq_x of the integrand it is weighted by q_x because, the dispersion relation being rotationally symmetric in the (q_x, q_y) plane, we may integrate around the circle of radius q_x to yield a properly weighted function of q_x and q_z alone. The quantity is plotted in Fig. 3.4(a),(b),(c) (bold line) together with the corresponding values for GaAs (light solid line) as a function of q_z for three values of q_x. The SL contribution is reasonably localized in q_z. To gain physical insight, we replace it by the dashed rectangles (of equal area). This approximate localization

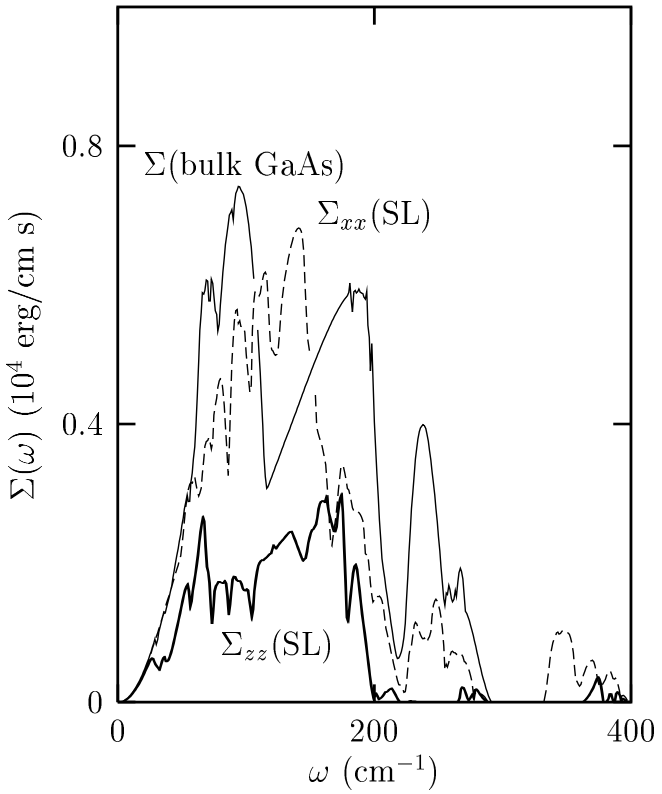


Figure 3.3: The transport quantities $\Sigma_{xx}(\omega)$ and $\Sigma_{zz}(\omega)$, defined by Eq. (3.5) in the text, for bulk GaAs (solid line) and the $(GaAs)_3/(AlAs)_3$ SL (dashed and bold lines).

is related to the reduction at $q_z \approx 0$ and π/d due to miniband formation (that is, flattening of ω vs. q) and has dips from anticrossings, as in Fig. 3.4(a), for instance. The dependence on q_x can then be summarized by the density plot in Fig. 3.4(d), comparing the SL on the left to bulk GaAs on the right. For GaAs the equivalent rectangles extend over the entire range of q_z because there is no localization due to band flattening at the zone edge as in the case of the SL. The shading indicates the weight of each increment Δq_x. We find that points around $q_x = \pi/d$ and at the zone edge ($q_x = 6\pi/d$) contribute most to heat transport. The latter is due to the effect of the weighting by q_x in the annular integration. In addition, this weighting causes the contribution from points near the origin to be very small. A simple numerical estimate leads to a reduction in Σ to 34%, which is to be compared to 36% in the exact calculation.

Finally, we discuss the sensitivity of variation of the mass differences, force constants and effective ionic charge e^* between layers on the thermal conductivity. This manifests itself through variations of the zone edge gap and hence the group velocity. This effect has been studied by interpolating e^*, the force constants K, and cation mass M between their AlAs and their GaAs values in the GaAs layer. Here K refers collectively to the 10 Kunc parameters describing interatomic forces. Explicitly, starting from $(\mathrm{AlAs})_6$ we (i) vary e^{*2} and M linearly in three neighboring AlAs positions to make a $(\mathrm{AlAs})_3/(\mathrm{GaAs})_3$ SL. In the case of the masses we put $M = (1-\alpha)M_{\mathrm{Al}} + \alpha M_{\mathrm{Ga}}$ so that at $\alpha = 0$ and 1 correspond respectively to Al and Ga; (ii) vary e^{*2} and force constants K linearly in α in the same way; and (iii) vary

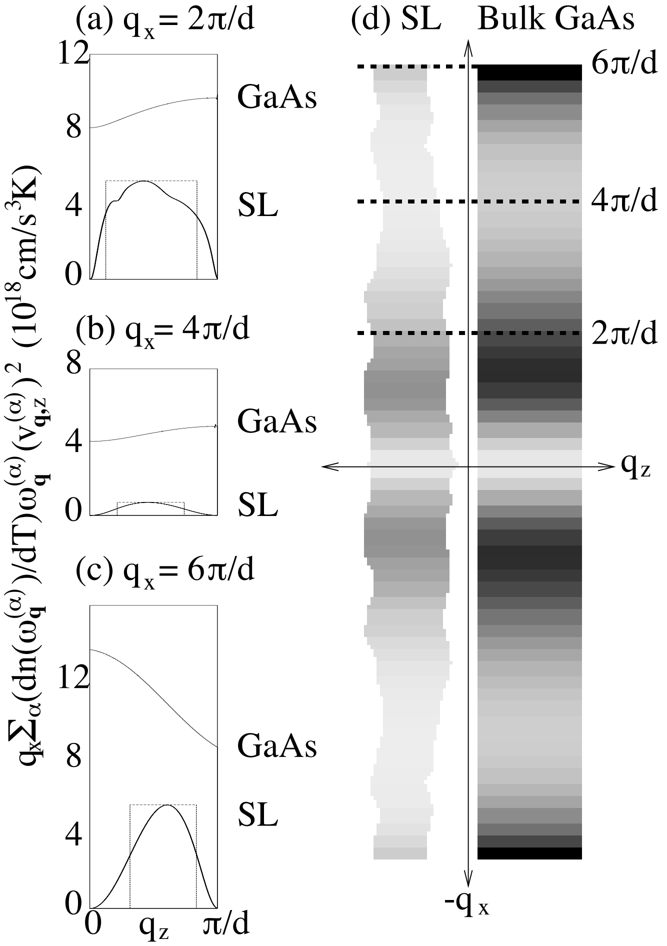


Figure 3.4: The transport quantity $q_x(dn/dT)\sum_\alpha \omega_{\mathbf{q}}^{(\alpha)}(v_{\mathbf{q},z}^{(\alpha)})^2$ for $0 \leq q_z \leq \pi/d$ at fixed q_x, for (a) $q_x = 2\pi/d$, (b) $q_x = 4\pi/d$ and (c) $q_x = 6\pi/d$; solid line: bulk GaAs; bold line: $(\text{GaAs})_3/(\text{AlAs})_3$ SL. (d) Density plot in the (q_x, q_z) plane whose shading indicates the weight of each increment Δq_x along q_x to the value of this transport quantity for the $(\text{GaAs})_3/(\text{AlAs})_3$ SL and bulk GaAs.

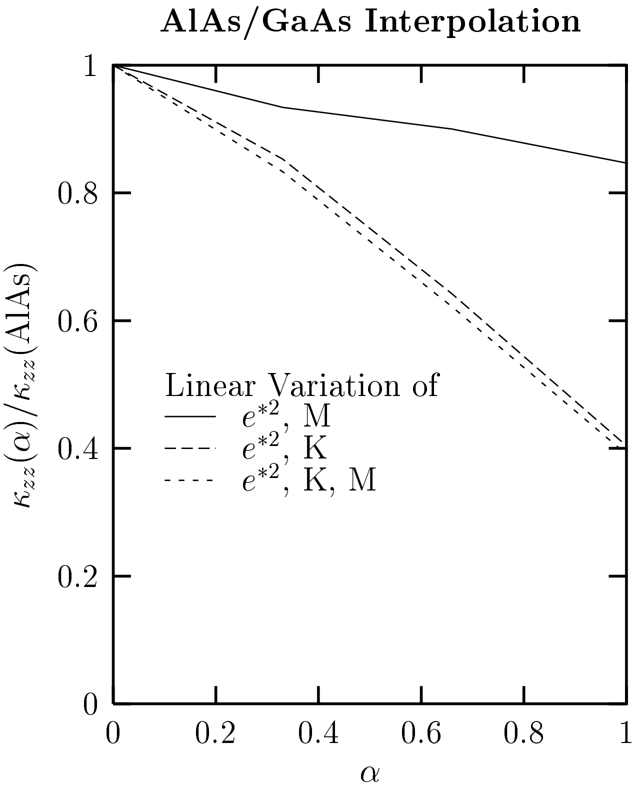


Figure 3.5: Dependence of κ_{zz} as e^{*2} and the mass M (solid line), the spring constants K (long-dashed line) and both M and K (short-dashed line) are interpolated between their AlAs ($\alpha = 0$) and GaAs values ($\alpha = 1$) for the atoms in three contiguous layers in what is initially $(\text{AlAs})_6$ at $\alpha = 0$.

e^{*2}, and both M and K linearly in α. The results are given in Fig. 3.5. The thermal conductivity is seen to be far more sensitive to the variation of force constants than the variation in mass. The variation of the force constants alone produces band flattening which reduces the thermal conductivity by about 60%. The additional flattening due to changing masses leads to only a few percent additional reduction.

3.5 Discussion and Conclusions

Before turning to the final results, we emphasize again that this Chapter is primarily concerned with the effects of superlattice induced changes of the phonon dispersion on the lattice thermal conductivity. Since a SL is a perfect crystal, the ideal structures under discussion here automatically include the coherent effects associated with perfect interfaces. The importance of including such effects was first pointed out by Hyldgaard and Mahan [59] in connection with a simple, highly idealized model for Si/Ge SLs. Extensive work by Chen [26, 27] focussed on diffuse interface scattering of phonons. His model assumes SL layers sufficiently thick that the phonon spectrum in each layer corresponds to that of the bulk. A mixture of spectral and diffuse interface processes is found to be sufficient to explain the observed experimental reduction of the SL thermal conductivity relative to bulk. As already pointed out, the more realistically modeled results of the present Chapter for κ_ℓ/τ lead to a three-fold reduction without any assumptions about the SL phonon scattering mechanisms. Our results must therefore be viewed as complementary to those of Chen [26, 27]. Taken together, they suggest that a combination of phonon spectral changes and imperfect interfaces can account adequately for the observed reduction.

The present calculations permit some statements concerning lifetime effects from the dependence of the SL results on layer width. This dependence was studied numerically for 2x2, 3x3 and 6x6 GaAs/AlAs SLs. The results for

$$\bar{\Sigma}_{zz} \equiv \kappa_{zz}/\tau = \int \frac{d^3q}{(2\pi)^3} \sum_\alpha \hbar\omega_{\mathbf{q}}^{(\alpha)} \frac{\partial\omega_{\mathbf{q}}^{(\alpha)}}{\partial q_i} \frac{\partial\omega_{\mathbf{q}}^{(\alpha)}}{\partial q_j} \frac{dn(\omega_{\mathbf{q}}^{(\alpha)})}{dT} \tag{3.7}$$

are given in Table 3.1. τ is assumed constant. Given an experimental value for

κ_{zz} and the calculated value of $\bar{\Sigma}_{zz}$, the lifetimes listed in Table 3.1 are found as $\tau = \kappa_{zz}/\bar{\Sigma}_{zz}$; only the lifetime itself is given in Table 3.1 and not $\bar{\Sigma}_{zz}$, but the ratios $\bar{\Sigma}_{zz}(\mathrm{SL})/\bar{\Sigma}_{zz}(\mathrm{bulk})$ are listed because we are interested in

$$\frac{\tau_{\mathrm{SL}}}{\tau_{\mathrm{bulk}}} = \frac{\bar{\Sigma}_{zz}(\mathrm{bulk})}{\bar{\Sigma}_{zz}(\mathrm{SL})} \frac{\kappa_{zz}^{\mathrm{SL}}(\mathrm{expt})}{\kappa_{zz}^{\mathrm{bulk}}(\mathrm{expt})}. \tag{3.8}$$

The SL $\bar{\Sigma}_{zz}$ is found to have a value about 40% of the bulk, and to be relatively insensitive to the SL period and temperature. For larger SL periods, there is more folding back, but this is compensated by the smaller size of the gaps at the zone edge, which is found in the computed ΓZ dispersion relations to scale inversely with the SL period. The two effects balance, resulting in an approximately constant reduction factor. The phonon lifetimes for the 2x2, 3x3 and 6x6 SLs are also given in Table 3.1. The bulk lifetime is the same to within 2% for the different SL periods, as it must be. The ratio of SL to bulk lifetimes is significantly smaller for the 2x2 SLs at each temperature while it is larger and roughly the same for the 3x3 and 6x6 SLs. Thus, the reduction in thermal conductivity may be divided into a dispersive part which is insensitive to $n \times n$ and a scattering part which is sensitive to interface scattering. The results for the 2x2 SL are possibly associated with the experimental difficulties associated with achieving sufficiently perfect interfaces for small n.

The present calculations imply: (1) that a similar reduction in the contribution of the SL phonon disperion relation to transport in the growth direction, and perhaps in κ_{zz} itself, may be expected in any SL with similar mass or force-constant differences between layers. (2) If the lifetime is reduced in SLs by increased umklapp scattering, as suggested by Ren and Dow[13], then the present calculations give an

Table 3.1: Reduction in $\bar{\Sigma}_{zz} = \kappa_{zz}/\tau$ relative to bulk GaAs for 2x2, 3x3 and 6x6 SLs at $T = 100, 200$ and 300 K, and SL phonon lifetimes deduced from Eq. (8) for the 2x2, 3x3 and 6x6 SLs at each temperature.

SL	experimental $\kappa_{zz}^{\mathrm{SL}}$ (W/cm K)[a]	theoretical $\frac{\bar{\Sigma}_{zz}(\mathrm{SL})}{\bar{\Sigma}_{zz}(\mathrm{bulk})}$	τ_{bulk} (ps)	τ_{SL} (ps)	$\tau_{\mathrm{SL}}/\tau_{\mathrm{bulk}}$
T=300 K					
$\kappa_{zz}^{\mathrm{bulk}}$=0.45 W/cm K [b]					
2x2	0.040	0.38	36.8	8.7	0.24
3x3	0.068	0.36	37.2	15	0.42
6x6	0.053	0.34	37.2	13	0.35
T=200 K					
$\kappa_{zz}^{\mathrm{bulk}}$=0.64 W/cm K [b]					
2x2	0.055	0.38	55.2	12	0.22
3x3	0.090	0.39	56.1	20	0.36
6x6	0.072	0.35	55.7	18	0.32
T=100 K					
$\kappa_{zz}^{\mathrm{bulk}}$=2.0 W/cm K [b]					
2x2	0.065	0.40	222	18	0.08
3x3	0.110	0.42	227	30	0.14
6x6	0.096	0.37	222	29	0.13

[a] Experimental values for GaAs/AlAs reported by Ref. [24].

[b] Experimental value for GaAs listed in Ref. [23].

upper bound on the SL κ_{zz} and a lower bound on the increase in ZT in a SL relative to bulk.

The Coulomb Part of the Superlattice Dynamical Matrix

This Appendix presents the detailed formulae for the Coulomb part of the dynamical matrix, derived using the Ewald procedure as described in the text of the paper. Letting κ label the atoms of the SL unit cell, M_κ be the mass of the κ-th atom and $\mathbf{x}(\kappa\kappa')$ be the separation vector from the κ-th atom to the κ'-th atom, we have, for $\kappa \neq \kappa'$,

$$C_{\alpha\beta}^{\mathrm{Coul}}(\kappa\kappa'|\mathbf{k}) = -\frac{e_\kappa e_{\kappa'}}{\sqrt{M_\kappa M_{\kappa'}}} e^{i\mathbf{k}\cdot\mathbf{x}(\kappa'\kappa)} \frac{\partial^2 \phi(\mathbf{k},\mathbf{r})}{\partial r_\alpha \partial r_\beta}\Big|_{\mathbf{r}=\mathbf{x}(\kappa'\kappa)} \tag{.9}$$

where $\phi(\mathbf{k},\mathbf{r})$ is by definition

$$\phi(\mathbf{k},\mathbf{r}) = \sum_\ell \frac{e^{i\mathbf{k}\cdot\mathbf{x}(\ell)}}{|\mathbf{x}(\ell)+\mathbf{r}|}, \tag{.10}$$

$\mathbf{x}(\ell)$ being the position of the ℓ-th unit cell. The result of the Ewald procedure adapted to the superlattice is that $\phi(\mathbf{k},\mathbf{r})$ can be written in the form

$$\begin{aligned} \phi(\mathbf{k},\mathbf{r}) &= R\sum_\ell H(|\mathbf{x}(\ell)+\mathbf{r}|R) e^{i\mathbf{k}\cdot\mathbf{x}(\ell)} + \\ & \frac{2\sqrt{\pi}}{v_{\|}} \sum_{h_{\|},\ell_z} \frac{2e^{-i(\tau(h_{\|})+\mathbf{k}_{\|})\cdot\mathbf{r}_{\|}+ik_z\hat{z}\cdot\mathbf{x}(\ell_z)}}{|\tau(h_{\|})+\mathbf{k}_{\|}|} \end{aligned}$$

$$\times I(\infty, \frac{|\tau(h_{||}) + \mathbf{k}_{||}|}{2R}, \frac{|\tau(h_{||}) + \mathbf{k}_{||}||\mathbf{x}(\ell_z) + z|}{2}) \tag{.11}$$

where R is an arbitrary cutoff (we always take $R = 3/a_0$),

$$H(x) = \frac{2/\sqrt{\pi}}{x} \int_x^\infty e^{-x'^2} dx', \tag{.12}$$

$v_{||}$ is the area of the unit cell in the superlattice plane, ℓ_z labels the layers perpendicular to the growth (z) axis, $h_{||} = (h_x, h_y)$ labels the cells located at reciprocal lattice vectors $\tau(h_{||})$ in each plane, $\mathbf{k} = (\mathbf{k}_{||}, k_z)$, and

$$I(\alpha, \beta, \gamma) = \int_\beta^\alpha e^{-v^2 - \gamma^2/v^2} dv. \tag{.13}$$

For $\kappa = \kappa'$ we have

$$C_{\alpha\beta}^{\mathrm{Coul}}(\kappa\kappa|\mathbf{k}) = \frac{e_\kappa}{M_\kappa} \left[-e_\kappa \sum_{\ell \neq 0} e^{i\mathbf{k}\cdot\mathbf{x}(\ell)} \left(\frac{\partial^2 r^{-1}}{\partial r_\alpha \partial r_\beta} \right)_{\mathbf{r}=\mathbf{x}(\ell)} + \sum_{\ell'\kappa' \neq 0\kappa} e_{\kappa'} \left(\frac{\partial^2 r^{-1}}{\partial r_\alpha \partial r_\beta} \right)_{\mathbf{r}=\mathbf{x}(\ell'\kappa', 0\kappa)} \right]. \tag{.14}$$

The first term on the right is similar to the above expressions in Eqs. (A1)-(A3) but for the $\ell = 0$ in Eq. (A3) term one substitutes $(4/3\sqrt{\pi})\delta_{\alpha\beta}$. The second term is given by

$$\frac{\partial^2}{\partial r_\alpha \partial r_\beta} \left[I_0 + I_1 + I_2 \right]_{\mathbf{r}=0} \tag{.15}$$

with

$$I_0 = R e_\kappa H^0(|\mathbf{r}|R), \tag{.16}$$

$$I_1 = R \sum_{\ell'\kappa' \neq 0\kappa} e_{\kappa'} H(|\mathbf{x}(\ell') + \mathbf{x}(\kappa'\kappa) + \mathbf{r}|R) \tag{.17}$$

and

$$I_2 = \frac{2\sqrt{\pi}}{v_{||}} \sum_{h_{||}, \ell_z, \kappa'} e_{\kappa'} \frac{2}{|\tau(h_{||})|} I(\infty, \frac{|\tau(h_{||})|}{2R}, \frac{|\tau(h_{||})||\mathbf{x}(\ell_z) + \hat{z}\cdot\mathbf{x}(\kappa'\kappa)|}{2}) e^{-i\tau(h_{||})\cdot(\mathbf{x}(\kappa'\kappa)+\mathbf{r})} \tag{.18}$$

where

$$H^0(x) = -\frac{2/\sqrt{\pi}}{x}\int_0^x e^{-x'^2}dx'. \qquad (.19)$$

A similar, but not identical, approach was used in Ref. [101].

Chapter 4

Thermoelectric Properties of Anisotropic Crystals

4.1 Abstract

General effective transport coefficients and the thermoelectric figure of merit ZT for anisotropic systems are derived. Sizable induced transverse fields on surfaces perpendicular to the current flow are shown to reduce the effective transport coefficients. A microscopic electronic model relevant for multi-valleyed materials with parabolic bands is considered in detail. Within the effective-mass and relaxation-time approximations but neglecting the lattice thermal conductivity κ_ℓ, the thermopower and Lorenz number are shown to be independent of the tensorial structure of the transport coefficients and are therefore isotropic. ZT is also isotropic for vanishing lattice thermal conductivity κ_ℓ. A similar result holds in lower dimensions. For non-

vanishing but sufficiently isotropic κ_ℓ, ZT is ordinarily maximal along the direction of highest electrical conductivity σ. Numerical calculations suggest that maximal ZT generally occurs along the principal direction with the largest σ/κ_ℓ. An explicit bound on ZT is derived. Consideration of the Esaki-Tsu model shows that non-parabolic dispersion in superlattices has little effect on the thermopower at the carrier concentrations which maximize ZT. Strong anisotropies develop when the chemical potential exceeds the mini-band width.

4.2 Introduction

In the search to find systems with large thermoelectric figures of merit ZT, the emphasis has been more on new materials than on materials structures or crystallographic anisotropy. Typical of new structures are superlattices, quantum wells [54, 82, 108, 19, 57, 22, 56, 20] and quantum wires [55, 21]. It has been generally assumed that the direction of highest conductivity in an anisotropic material yields optimal thermoelectric properties. The correctness of this assumption is not obvious, since directions may exist along which the lattice thermal conductivity is abnormally low and the thermopower is high enough to result in large ZT even though the electrical conductivity is less than maximum.

This paper develops a more general microscopic transport theory of thermoelectricity in anisotropic systems. The "highest conductivity" assumption is found to be correct for materials having simple band structures typically of the parabolic variety and essentially isotropic lattice thermal conductivities. However, the formalism developed here suggests more generally that the optimal orientation corresponds to the principal direction along which the ratio of the electronic to lattice thermal conductivity σ/κ_ℓ is maximum. There are also several surprising results. These include the formation of possibly large induced transverse electric fields and temperature gradients, the fact that ZT is strictly isotropic in anisotropic systems having parabolic bands if the lattice thermal conductivity is neglected, and that a nearly isotropic thermopower and Lorenz number results under these conditions even if the bands are extremely anisotropic and non-parabolic.

The macroscopic formalism, based on the tensorial form of the usual phenomenological transport equations, is presented in Sec. 4.3. The effects of sample boundaries are included by requiring that the transverse electric and heat currents vanish throughout the volume. Anisotropic and isotropic systems are shown to differ both qualitatively, through the presence of induced transverse fields, and quantitatively, through the magnitude of the transport coefficients.

More detailed statements concerning optimal orientations require use of a microscopic model. Section 4.4 introduces a model commonly used to study transport in semiconductors having multi-valley, anisotropic parabolic band structures. The transport properties follow from the linearized Boltzmann equation in the effective-mass and relaxation-time approximations. Intervalley scattering is neglected. Somewhat surprisingly, the thermopower and Lorenz number turn out to be isotropic. The same is true for ZT when the lattice thermal conductivity κ_ℓ is neglected. Sec. 4.4.2 shows that similar results hold for two- and one-dimensional systems. The realistic case corresponding to finite κ_ℓ is considered in Sec. 4.4.3. If κ_ℓ is sufficiently isotropic, the maximal ZT is shown to occur for samples cut along the direction of highest electrical conductivity. This rather unsurprising result, however, may be modified if the anisotropy of κ_ℓ exceeds that of the electrical conductivity, leading to the conjecture concerning σ/κ_ℓ mentioned above. An explicit expression for the upper bound on ZT is also derived, which is a generalization of that previously obtained [83] to anisotropic systems.

This microscopic description has been applied to several materials of current interest in Chapter 2, in particular, bulk n-type Bi_2Te_3 and HgTe/HgCdTe superlattices (SLs). For sufficiently large anisotropy, the induced fields can be much larger than the external one.

The explicit effects of non-parabolicity and reduced band width in superlattices on the thermopower and Lorenz number are investigated in Sec. 4.5. Both quantities remain nearly isotropic for small doping or carrier concentration. The thermopower is also bounded by its values along the principal axes, supporting the conjecture that optimal ZT is obtained along those directions.

4.3 Electronic Transport Theory

We generalize conventional transport theory to anisotropic media in order to calculate the modified thermoelectric figure of merit. The tensor equations describing the transport of electrons and of heat in the presence of an electric field $\mathcal{E}$

and temperature gradient ∇T are

$$\mathbf{J}^e = \underline{L}^{11}\mathcal{E} + \underline{L}^{21}(-\nabla T/T) \tag{4.1}$$

$$\mathbf{J}^Q = \underline{L}^{12}\mathcal{E} + \underline{L}^{22}(-\nabla T/T). \tag{4.2}$$

Here $\mathbf{J}^e$ is the electric current density, $\mathbf{J}^Q$ is the heat current density and $\underline{L}^{\alpha\beta}$ are the matrices of Onsager coefficients. With the notation $\underline{L}^{11} = \underline{\sigma}$ and $\underline{L}^{21} = T\underline{\sigma}\underline{S}$, where $\underline{\sigma}$ is the electrical conductivity tensor and $\underline{S}$ is the Seebeck tensor, Eq. (4.1) becomes

$$\mathbf{J}^e = \underline{\sigma}(\mathcal{E} - \underline{S}\nabla T) \equiv \underline{\sigma}\mathbf{F}. \tag{4.3}$$

Eq. (4.2) can be put into the form [62, 81]

$$\mathbf{J}^Q = \underline{\Pi}\mathbf{J}^e - \underline{\kappa}\nabla T. \tag{4.4}$$

if we eliminate $\mathcal{E}$ in favor of $\mathbf{J}^e$ and use the notation $\underline{L}^{12} = \underline{\Pi}\,\underline{\sigma}$, where $\underline{\Pi}$ is the Peltier tensor, and define the thermal conductivity tensor to be $\underline{\kappa} = (1/T)\underline{L}^{22} - \underline{\Pi}\,\underline{\sigma}\,\underline{S}$. As yet, we have not assumed the Onsager relations $\underline{L}^{\alpha\beta} = (\underline{L}^{\beta\alpha})^T$, which are equivalent to $\underline{\sigma} = \underline{\sigma}^T$, $\underline{\kappa} = \underline{\kappa}^T$ and $\underline{\Pi} = T\underline{S}^T$. These are not obviously applicable to a purely macroscopically described medium. Truesdell [119] has argued that, in crystalline solids of sufficiently low symmetry, $\underline{\kappa}$ may be non-symmetric. His argument, however, relies only on group theory, and does not take into account the microscopic physics, in particular the fundamental principle of microscopic reversibility, on which the Onsager theory rests. A derivation of the Onsager relations from the linearized Boltzmann equation, in the natural frame for the problem given by the crystalline axes of an anisotropic solid, is given in Chapter 5.

We shall assume that the crystal, whatever its anisotropy, is cut rectangularly [92]. Then Cartesian axes are defined and give the natural coordinate system in which to study the transport.

Transport in an anisotropic medium is modified by the presence of induced fields which arise via Eqs. (4.3) and (4.4) in response to the generalized forces resulting from, for example, the accumulation of charges on the surface of the sample. The induced fields are implicit in the treatment of Nye [92]; as shown there, edge effects will be negligible if we assume that the length of the sample in the direction of transport (longitudinal) is large compared to the size of the transverse directions. Thus $\mathcal{E} = \mathcal{E}^{\rm ext} + \mathcal{E}^{\rm ind}$ and $\nabla T = \nabla T^{\rm ext} + \nabla T^{\rm ind}$ where $\mathcal{E}^{\rm ext}$ is the external aplied electric field, $\mathcal{E}^{\rm ind}$ is the induced electric field, $\nabla T^{\rm ext}$ is the external applied temperature gradient and $\nabla T^{\rm ind}$ is the induced temperature gradient. The fluxes $\mathbf{J}^e$ and $\mathbf{J}^Q$ generated by the total forces will therefore involve an induced part as well, and the net result is that, for a given applied force, the total flux will be smaller than would be the case in the absence of the induced fields. The induced fields vanish in samples cut along a principal axis of the conductivity tensor.

Consider an electric field $\mathcal{E}^{\rm ext} = (\mathcal{E}_{||}, 0, 0)$ and temperature gradient $\nabla T^{\rm ext} = (\nabla_{||} T, 0, 0)$ applied along the longitudinal ($||$) direction. Let $\mathbf{F} = (F_x, F_y, F_z) = (F_{||}, \mathbf{F}_\perp)$ with $F_{||}$ the thermal and electric field along x and $\mathbf{F}_\perp$ the induced transverse fields. We assume experimental conditions in which the transverse ($\perp$) electrical and thermal currents vanish throughout the volume. Then

$$\begin{pmatrix} J^e_{||} \\ \mathbf{0} \end{pmatrix} = \underline{\sigma} \begin{pmatrix} F_{||} \\ \mathbf{F}_\perp \end{pmatrix} = \begin{pmatrix} \sigma_{||} & \underline{\sigma}^T_{od} \\ \underline{\sigma}_{od} & \underline{\sigma}_\perp \end{pmatrix} \begin{pmatrix} F_{||} \\ \mathbf{F}_\perp \end{pmatrix} \qquad (4.5)$$

and

$$\begin{pmatrix} J_{\|}^{Q} \\ \mathbf{0} \end{pmatrix} = \begin{pmatrix} T S_{\|} J_{\|}^{e} - \kappa_{\|} \nabla_{\|} T - \underline{\kappa}_{od}^{T} \nabla_{\perp} T \\ T \underline{S}_{od}^{(1)T} J_{\|}^{e} - \underline{\kappa}_{od} \nabla_{\|} T - \underline{\kappa}_{\perp} \nabla_{\perp} T \end{pmatrix}, \tag{4.6}$$

where $\sigma_{\|}$ is the component of the conductivity along the x-direction, and $\underline{\sigma}_{\perp}$ is a 2×2 tensor for the transverse y, z-directions. The 2×1 off-diagonal term $\underline{\sigma}_{od}$ and its transpose appear since the full conductivity tensor is not block-diagonal. The tensor $\underline{\kappa}$ has a similar decomposition. We have invoked the Onsager relations in Eq. (4.6) by writing $\underline{\Pi} = T\underline{S}^{T}$. The Onsager relations also imply that the conductivity tensor is symmetric, so that $\underline{\sigma}_{\perp}$ in Eq. (4.5) is symmetric as well. The same applies to the tensors $\underline{\kappa}$ and hence $\underline{\kappa}_{\perp}$ in Eq. (4.4). However, the Seebeck tensor

$$\underline{S} = \begin{pmatrix} S_{\|} & \underline{S}_{od}^{(1)} \\ \underline{S}_{od}^{(2)} & \underline{S}_{\perp} \end{pmatrix}, \tag{4.7}$$

need not be symmetric.

To calculate the electrical conductivity note that Eq. (4.5) leads to

$$\mathbf{F}_{\perp} = -\underline{\sigma}_{\perp}{}^{-1} \underline{\sigma}_{od} F_{\|}. \tag{4.8}$$

The off-diagonal elements of $\underline{\sigma}$ induce a current

$$J_{\|}^{\mathrm{e,ind}} = \underline{\sigma}_{od}^{T} \mathbf{F}_{\perp} = -\underline{\sigma}_{od}^{T} \underline{\sigma}_{\perp}{}^{-1} \underline{\sigma}_{od} F_{\|} \tag{4.9}$$

along the x-direction. The total current

$$J_{\|}^{e} = J_{\|}^{\mathrm{e,ext}} + J_{\|}^{\mathrm{e,ind}} = (\sigma_{\|} - \underline{\sigma}_{od}^{T} \underline{\sigma}_{\perp}{}^{-1} \underline{\sigma}_{od}) F_{\|} \equiv \sigma_{1} F_{\|}. \tag{4.10}$$

To calculate the effective thermal conductivity, note that Eq. (4.6) yields the induced temperature gradients

$$\nabla_{\perp} T = \underline{\kappa}_{\perp}^{-1} (T \underline{S}_{od}^{(1)T} J_{\|}^{e} - \underline{\kappa}_{od} \nabla_{\|} T). \tag{4.11}$$

Substituting back into Eq. (4.6) gives

$$J_{\parallel}^{Q,\text{ind}} = -\underline{\kappa}_{od}^{T}\nabla_{\perp}T = -\underline{\kappa}_{od}^{T}\underline{\kappa}_{\perp}^{-1}(T\underline{S}_{od}^{(1)T}J_{\parallel}^{e} - \underline{\kappa}_{od}\nabla_{\parallel}T) \tag{4.12}$$

whence

$$J_{\parallel}^{Q} = T(S_{\parallel} - \underline{\kappa}_{od}^{T}\underline{\kappa}_{\perp}^{-1}\underline{S}_{od}^{(1)T})J_{\parallel}^{e} - (\kappa_{\parallel} - \underline{\kappa}_{od}^{T}\underline{\kappa}_{\perp}^{-1}\underline{\kappa}_{od})\nabla_{\parallel}T \equiv TS_{\text{eff}}J_{\parallel}^{e} - \kappa_{\text{eff}}\nabla_{\parallel}T \quad , \tag{4.13}$$

where S_{eff} and κ_{eff} are "effective" transport coefficients to be discussed further below. Observe that σ_1, Eq. (4.10), and κ_{eff}, Eq. (4.13), have the same form. Since $\underline{\sigma}_{\perp}$ and $\underline{\sigma}_{\perp}^{-1}$ have positive eigenvalues, the quadratic form $\underline{\sigma}_{od}^{T}\underline{\sigma}_{\perp}{}^{-1}\underline{\sigma}_{od}$ is positive definite. Thus, the induced current opposes the external current $J_{\parallel}^{\text{e,ext}} = \sigma_{\parallel}F_{\parallel}$. The induced fields therefore lead to a reduced conductivity: $\sigma_1 \leq \sigma_{\parallel}$. If $\underline{S}$ is diagonal, σ_1 is the effective conductivity; otherwise, the induced temperature gradients give rise to additional terms considered below. In the isotropic case, the induced fields vanish, and $\sigma_1 = \sigma_{\parallel}$. Note that in general $\sigma_1 \geq 0$ because $\sigma_1 F_{\parallel}^2 = J_{\parallel}^{e}F_{\parallel} = \mathbf{F}^T \cdot \mathbf{J}^e = \mathbf{F}^T \cdot \underline{\sigma} \cdot \mathbf{F} \geq 0$. As in the case of σ_1, the induced temperature gradients produce an induced heat current along x which opposes the external heat current $-\kappa_{\parallel}\nabla_{\parallel}T$. The net result is a reduced effective thermal conductivity: $0 \leq \kappa_{\text{eff}} \leq \kappa_{\parallel}$.

Returning to Eq. (4.10), we have

$$F_{\parallel} = \left(\mathcal{E}_{\parallel} - S_{\parallel}\nabla_{\parallel}T - \underline{S}_{od}^{(1)}\nabla_{\perp}T\right) . \tag{4.14}$$

Substituting Eq. (4.11) into Eq. (4.14) leads to

$$J_{\parallel}^{e} \;=\; \frac{\sigma_1}{1 + \sigma_1 T\underline{S}_{od}^{(1)}\underline{\kappa}_{\perp}^{-1}\underline{S}_{od}^{(1)T}}\left(\mathcal{E}_{\parallel} - (S_{\parallel} - \underline{S}_{od}^{(1)}\underline{\kappa}_{\perp}^{-1}\underline{\kappa}_{od})\nabla_{\parallel}T\right) \tag{4.15}$$

$$\equiv \quad \sigma_{\text{eff}}(\mathcal{E}_{||} - S_{\text{eff}}\nabla_{||}T) \tag{4.16}$$

with the help of Eq. (4.10). From the properties of σ_1, $0 \leq \sigma_{\text{eff}} \leq \sigma_{||}$. Therefore, induced fields and temperature gradients always reduce the effective conductivity. Also, S_{eff} in Eq. (4.16) is the same as S_{eff} in Eq. (4.13) because $\underline{\kappa}_{\perp}$ and hence $\underline{\kappa}_{\perp}^{-1}$ are symmetric.

The results of Eqs. (4.13) and (4.16) are of importance because they show that in an anisotropic crystal the parallel component of the electronic current can still be expressed by an expression having the form of Eqs. (3) and (4) but with all transport coefficients replaced by effective quantities σ_{eff}, κ_{eff}, and S_{eff}.

In the steady state, Domenicali's continuity equation for the energy density in an anisotropic medium, as obtained by generalizing Mahan's treatment [81], reads

$$\mathbf{J}^e\underline{\sigma}^{-1}\mathbf{J}^e + \nabla\cdot(\underline{\kappa}\nabla T) = 0. \tag{4.17}$$

It can be shown that this reduces to

$$\sigma_{\text{eff}}^{-1}J_{||}^{e2} + \kappa_{\text{eff}}\nabla_{||}^2 T = 0. \tag{4.18}$$

The thermoelectric figure of merit becomes

$$ZT = T\sigma_{\text{eff}}S_{\text{eff}}^2/\kappa_{\text{eff}}, \tag{4.19}$$

with the help of Eqs. (4.16), (4.13) and (4.18). The form of ZT remains the same, but again the transport coefficients are simply replaced by their effective versions.

4.4 Microscopic Model: Parabolic Bands

4.4.1 Three-Dimensional Structures

According to semiclassical transport theory, the Boltzmann equation in the relaxation-time approximation yields the electronic transport coefficients

$$\begin{aligned} \sigma_{ij} &= e^2 \int d\varepsilon (-\partial f_0/\partial\varepsilon) \Sigma_{ij}(\varepsilon) && (4.20) \\ T(\sigma \cdot S)_{ij} &= e \int d\varepsilon (-\partial f_0/\partial\varepsilon) \Sigma_{ij}(\varepsilon)(\varepsilon - \mu) && (4.21) \\ T\kappa_{0,ij} &= \int d\varepsilon (-\partial f_0/\partial\varepsilon) \Sigma_{ij}(\varepsilon)(\varepsilon - \mu)^2, && (4.22) \end{aligned}$$

where f_0 is the Fermi-Dirac distribution $f_0(\varepsilon) = 1/(\exp((\varepsilon - \mu)/k_B T) + 1)$, μ the chemical potential, and

$$\Sigma_{ij}(\varepsilon) = \int \frac{2d^3\mathbf{k}}{(2\pi)^3} v_i(\mathbf{k}) v_j(\mathbf{k}) \tau(\mathbf{k}) \delta(\varepsilon - \varepsilon(\mathbf{k})) \tag{4.23}$$

are the components of the transport distribution tensor, the generalization of the function discussed by Mahan and Sofo [83]. Here $\varepsilon(\mathbf{k})$ is the electronic dispersion relation, $v_i(\mathbf{k}) = \hbar^{-1}\partial\varepsilon(\mathbf{k})/\partial k_i$ the electronic group velocity, and $\tau(\mathbf{k})$ the relaxation time. Note that $\underline{\kappa}_0$ is the electronic thermal conductivity at zero electrochemical potential gradient inside the sample [83]; the usual electronic thermal conductivity at zero electric current, $\underline{\kappa}_e$, is given in terms of $\underline{\kappa}_0$ by $\underline{\kappa}_e = \underline{\kappa}_0 - T\underline{S}^T \underline{\sigma}\, \underline{S}$. The lattice conductivity, κ_ℓ, is not considered in this section.

The microscopic model to be used here assumes the conduction to be taking place in a single band having N degenerate parabolic valleys centered at $\mathbf{k}^{(n)}, n =$

$1, \ldots, N$ respectively. The dispersion relation for each valley is

$$\varepsilon^{(n)}(\mathbf{k}) = \varepsilon_0 + (\hbar^2/2)\sum_{i,j}(k_i - k_i^{(n)})M_{ij}^{(n)-1}(k_j - k_j^{(n)}) \tag{4.24}$$

with $\underline{M}^{(n)-1}$ the inverse effective-mass tensor. The corresponding group velocity is

$$v_i^{(n)}(\mathbf{k}) = \hbar^{-1}\partial\varepsilon(\mathbf{k})/\partial k_i = \hbar\sum_{j} M_{ij}^{(n)-1}(k_j - k_j^{(n)}). \tag{4.25}$$

Intervalley scattering will be neglected. Thus the transport distribution tensor involves just a sum over the N valleys. Assuming the relaxation time to be a function of energy alone, $\tau(\mathbf{k}) = \tau(\varepsilon(\mathbf{k}))$, and independent of crystal orientation,

$$\begin{aligned}
\Sigma_{ij}(\varepsilon) &= \tau(\varepsilon)\sum_{n=1}^{N}\int \frac{2d^3\mathbf{k}}{(2\pi)^3}\hbar^2 \sum_{i',j'} M_{ii'}^{(n)-1}k_{i'}M_{jj'}^{(n)-1}k_{j'}\delta(\varepsilon - \varepsilon(\mathbf{k}+\mathbf{k}^{(n)})) && (4.26)\\
&= \tau(\varepsilon)\hbar^2\sum_{n=1}^{N}\sum_{i'j'} M_{ii'}^{(n)-1}M_{jj'}^{(n)-1}\int k_{i'}k_{j'}\delta(\varepsilon - \mathbf{k}\underline{X}^{(n)}\mathbf{k})\frac{2d^3\mathbf{k}}{(2\pi)^3} && (4.27)\\
&= \frac{2\tau(\varepsilon)\hbar^2}{(2\pi)^3}\sum_{n=1}^{N}\sum_{i'j'} M_{ii'}^{(n)-1}M_{jj'}^{(n)-1}\left[-\frac{\partial}{\partial X_{i'j'}^{(n)}}\int \Theta(\varepsilon - \mathbf{k}\underline{X}^{(n)}\mathbf{k})d^3\mathbf{k}\right] && (4.28)\\
&= \frac{2^{3/2}\tau(\varepsilon)\varepsilon^{3/2}}{3\pi^2\hbar^3}\sum_{n=1}^{N}\sqrt{\det\underline{M}^{(n)}}M_{ij}^{(n)-1}, && (4.29)
\end{aligned}$$

where $X_{ij}^{(n)}$ are the components of $\underline{X}^{(n)} = (\hbar^2/2)\underline{M}^{(n)-1}$. This transformation relies explicitly on the validity of the effective-mass approximation, the neglect of intervalley scattering, and the relaxation-time approximation used here. The square bracket in Eq. (4.28) is evaluated using the identity

$$\frac{\partial}{\partial X_{ij}^{(n)}}\det\underline{X}^{(n)} = (\det\underline{X}^{(n)})X_{ij}^{(n)-1} \tag{4.30}$$

and the change of variable $\mathbf{k}' = \sqrt{\underline{X}^{(n)}}\mathbf{k}$. Thus,

$$\underline{\Sigma}(\varepsilon) = \underline{A}\mathcal{T}(\varepsilon) \tag{4.31}$$

with

$$\underline{A} = \sum_{n=1}^{N} \left(m_0^{-1/2} \sqrt{\det \underline{M}^{(n)}} \right) \underline{M}^{(n)-1} \tag{4.32}$$

and

$$\mathcal{T}(\varepsilon) = \frac{2^{3/2} m_0^{1/2}}{3\pi^2 \hbar^3} \varepsilon^{3/2} \tau(\varepsilon). \tag{4.33}$$

The constant, dimensionless matrix $\underline{A}$ contains the full tensorial structure associated with the crystal symmetry and separates it from the energy dependence in $\mathcal{T}(\varepsilon)$.

The conductivity then becomes

$$\underline{\sigma} = e^2 \int d\varepsilon (-\partial f_0 / \partial \varepsilon) \mathcal{T}(\varepsilon) \underline{A} \equiv \sigma_0 \underline{A}. \tag{4.34}$$

For silicon and germanium this expression reduces to that given by Smith *et al.* [107].

The Seebeck tensor is therefore necessarily *isotropic* since

$$\underline{(\sigma \cdot S)} = (e/T) \int d\varepsilon (-\partial f_0 / \partial \varepsilon) \mathcal{T}(\varepsilon)(\varepsilon - \mu) \underline{A} \equiv \underline{A} \sigma_0 S_0, \tag{4.35}$$

$$\underline{S} = \underline{\sigma}^{-1} \underline{(\sigma \cdot S)} = S_0 \, \underline{1}. \tag{4.36}$$

Further, using Eqs. (23) and (32), $\underline{\kappa}_0 = \underline{A} \kappa_0$ and hence

$$\underline{\kappa}_e = \underline{\kappa}_0 - T \underline{S}^T \underline{\sigma} \, \underline{S} \equiv \underline{A} \kappa_e \tag{4.37}$$

for $\kappa_e = \kappa_0 - T \sigma_0 S_0^2$.

These results lead to the following surprising conclusion: When the lattice thermal conductivity is neglected, then, within the effective-mass approximation as specified here, the thermoelectric figure of merit ZT is independent of the sample orientation. Note that

$$\underline{\kappa}_e \cdot \underline{\sigma}^{-1} = \underline{A} \kappa_e \cdot \underline{A}^{-1} / \sigma_0 = (\kappa_e / \sigma_0) \underline{1} \equiv L_0 T \underline{1}, \tag{4.38}$$

where the Lorenz number $L_0 = \kappa_e/\sigma_0 T$. Thus, $\kappa_{e,ij} = L_0 T \sigma_{ij}$ and $\kappa_{\text{eff}} = L_0 T \sigma_{\text{eff}}$ as expected. Finally, since $\underline{S}$ is isotropic, $S_{\text{eff}} = S_0$. Combining these effective transport coefficients yields

$$ZT = T\sigma_{\text{eff}} S_{\text{eff}}^2/\kappa_{\text{eff}} = S_0^2/L_0, \tag{4.39}$$

a constant independent of direction. (The isotropy is lost if the lattice thermal conductivity is included.)

4.4.2 Lower-dimensional Structures

Dimensionality enters the transport coefficients through the $\mathbf{k}$-space integrals d^3k, $(2\pi/L_z)d^2k$ and $((2\pi)^2/L_yL_z)dk$ for three, two or one dimensions respectively, where L_y and L_z are the sample sizes in the y- and z-directions. Dimensionality also enters through the confinement energies and effective masses for carriers constrained to move in a lower-dimensional system.

For the two-dimensional case with $\mathcal{E} = (\mathcal{E}_{\|}, \mathcal{E}_{\perp})$ and $\nabla T = (\nabla_{\|}T, \nabla_{\perp}T)$, the induced fields are given by Eqs. (4.8) and (4.11), the transport coefficients by Eqs. (4.13) and (4.16) with $\underline{\sigma}_{\perp}$ replaced by σ_{yy}, $\underline{\sigma}_{od}$ replaced by σ_{yx}; similarly for the other transport coefficients. The analogue of Eq. (4.23) for the components of the transport distribution tensor is obtained within the effective-mass approximation:

$$\Sigma_{ij}(\varepsilon) = \frac{2\tau(\varepsilon)\hbar^2}{4\pi^2 L_z} \sum_{n=1}^{N} \sum_{i'j'} M_{ii'}^{(n)-1} M_{jj'}^{(n)-1} \int d^2\mathbf{k}\, k_{i'} k_{j'} \delta(\varepsilon - \mathbf{k}\underline{X}^{(n)}\mathbf{k}). \tag{4.40}$$

This expression has the same form as Eq. (4.31) with

$$\underline{A} = \sum_{n=1}^{N} \sqrt{\det \underline{M}^{(n)}}\, \underline{M}^{(n)-1} \tag{4.41}$$

and

$$\mathcal{T}(\varepsilon) = \varepsilon\tau(\varepsilon)/\pi\hbar^2 L_z. \tag{4.42}$$

These equations are to be compared with Eqs. (4.32) and (4.33) for the three-dimensional case. Thus, analogously to Eqs. (4.34), (4.36), and (4.37), $\underline{\sigma} = \underline{A}\sigma_0, \underline{S} = S_0\underline{1}$ and $\underline{\kappa}_e = \underline{A}\kappa_e$ where the two-dimensional $\underline{A}$ and $\mathcal{T}(\varepsilon)$ must be used in defining σ_0, S_0 and κ_0. Finally, $\underline{\kappa}_e\underline{\sigma}^{-1} = L_0T\underline{1}$ so that $\kappa_{\text{eff}} = L_0T\sigma_{\text{eff}}$. Also, $S_{\text{eff}} = S_0$ because $\underline{S}$ is isotropic in two dimensions as well. The corresponding figure of merit $ZT = T\sigma_{\text{eff}}S_{\text{eff}}^2/\kappa_{\text{eff}} = S_0^2/L_0$ is again independent of direction, as in three dimensions.

It is seen that the two-dimensional and three-dimensional results are entirely analogous and that the former are obtained from the latter by taking the limit as one of the effective masses tends to infinity. In this limit, the ellipsoidal surface in **k**-space of constant energy becomes increasingly prolate, until it reaches the edge of the Brillouin zone, after which it assumes a cylindrical shape extending from $-\pi/L_z$ to π/L_z upon further increase in the effective-mass parameter. Furthermore ZT in two dimensions can be enhanced due to the factor of $1/L_z$ in the density of states, which, as pointed out in Ref. [54], becomes large for small thicknesses.

In the one-dimensional case there is no transport in the transverse direction. Thus there are no transverse fields and the microscopic and effective transport coefficients are the same. Moreover, all transport coefficients are scalars. The transport distribution function is found to be

$$\Sigma(\varepsilon) = \sum_{n=1}^{N} \frac{2}{\pi L_y L_z}\sqrt{\frac{2\varepsilon}{\hbar^2 m_x^{(n)}}}\tau(\varepsilon). \tag{4.43}$$

Just as in two dimensions, the $1/L_yL_z$ factor leads to an enhancement of the density of states and thus of ZT for thin wires [55].

4.4.3 Implications for ZT

We shall now derive an upper bound for ZT of the Sofo and Mahan form [83] and show that the highest conductivity direction gives optimal values for ZT. An *isotropic* lattice thermal conductivity κ_ℓ, which causes ZT to lose its isotropy, will now be included.

The figure of merit including an isotropic lattice thermal conductivity $\underline{\kappa_\ell} = \kappa_\ell \underline{1}$ may be written in the form of Eq. (4.19) where σ_{eff} and S_{eff} are as in Sec. 4.3 and $\kappa_{\text{eff}} = \kappa_e^* + \kappa_\ell$. Let

$$\underline{\kappa_e} = \begin{pmatrix} \kappa_{||} & \underline{\kappa_{od}}^T \\ \underline{\kappa_{od}} & \underline{\kappa_\perp} \end{pmatrix} \tag{4.44}$$

and (in analogy with Eq. (4.13))

$$\kappa_e^* = \kappa_{||} - \underline{\kappa_{od}}^T (\underline{\kappa_\perp} + \kappa_\ell \underline{1})^{-1} \underline{\kappa_{od}}. \tag{4.45}$$

κ_e^* defines the electronic thermal conductivity in the presence of the non-vanishing κ_ℓ and the sample boundaries. As shown in Appendix A, the upper bound on ZT is given by

$$ZT \leq a_0 \kappa_0 / \kappa_\ell. \tag{4.46}$$

The dimensionless quantity a_0 is defined by Eq. (.54). In the isotropic case [83], $a_0 = 1$; in the present case, a_0 is of order unity. The maximum value is obtained when the parameter ξ, as defined by Eq. (.64), approaches unity. This is the case if and only if $\mathcal{T}(\varepsilon)$ is proportional to a δ-function. In the more physical case $\mathcal{T}(\varepsilon) \propto \varepsilon^r$ with r varying between -0.5 and 2, which encompasses many of the common scattering mechanisms in semiconductors. Numerical computations show that ξ tends to 1 as $\mu/k_BT \to -\infty$ and to zero as $\mu/k_BT \to \infty$. For $\mu/k_BT = 0$, ξ ranges from 0.5 to 0.8. Thus, the upper bound at $\xi = 1$ can be reached at the cost of going to low carrier concentrations, whereas higher carrier concentrations imply smaller ξ.

We now show that, in the effective-mass, relaxation-time, no intervalley scattering, and isotropic-thermal-conductivity approximations, ZT is highest in the direction of maximum electrical conductivity. In the anisotropic case, σ_{eff} and κ_{eff} have an angular dependence. S_{eff} does not because in our microscopic model $\underline{S}$ is isotropic. ZT is therefore also anisotropic. Let now $\underline{P}$ be any symmetric and positive matrix and $\lambda \geq 0$ a positive number; then by the properties of positive matrices, $\underline{P}^{-1} \geq (\underline{P} + \lambda \underline{1})^{-1}$. $\underline{P} = \underline{\kappa}_\perp$ and $\lambda = \kappa_\ell$. Thus

$$\underline{\kappa}_{od}{}^T (\underline{\kappa}_\perp + \kappa_\ell \underline{1})^{-1} \underline{\kappa}_{od} \leq \underline{\kappa}_{od}{}^T \underline{\kappa}_\perp{}^{-1} \underline{\kappa}_{od} \tag{4.47}$$

and consequently for any κ_ℓ

$$\kappa_e^*(\kappa_\ell) \geq \kappa_e^*(\kappa_\ell = 0) = (\kappa_e/\sigma_0)\sigma_{\text{eff}}. \tag{4.48}$$

This shows that the ratio $\kappa_e^*/\sigma_{\text{eff}}$ is minimized when the axis of current flow in steady state lies along one of the principal directions, where, whatever the value of κ_ℓ, the induced-field related terms vanish and so equality obtains in Eq. (4.48). Writing ZT in the form

$$ZT = \frac{S_0^2}{\kappa_e^*/T\sigma_{\text{eff}} + \kappa_\ell/T\sigma_{\text{eff}}}, \tag{4.49}$$

the second term in the denominator is seen to be smallest along the principal direction with largest electrical conductivity. Therefore ZT is maximized for current flow along this direction.

On the other hand, for a sufficiently anisotropic lattice thermal conductivity the favored direction might be determined by κ_ℓ's minimum rather than that of the highest electrical conductivity. For crystals in which two of the principal values of the electrical conductivity are equal, such as Bi_2Te_3 and SLs, numerical results show that as long as the anisotropy of $\underline{\kappa_\ell}$ is smaller than that of $\underline{\sigma}$, the optimum ZT is still to be found in the direction of greatest electrical conductivity. This suggests that, generally speaking, the figure of merit will be maximized in the principal crystal direction in which the ratio $\sigma_i/\kappa_{\ell,i}$ is greatest, where σ_i and $\kappa_{\ell,i}$ are the principal values of the electrical and lattice thermal conductivity tensors obtained from summing over valleys.

4.5 Microscopic Model: Non-Parabolic Bands

Within the effective-mass approximation, the analysis of Sec. 4.4 showed that the thermopower and Lorenz number are isotropic. To see how a non-parabolic band structure affects these conclusions, consider an electronic dispersion relation of the Esaki-Tsu form [37]:

$$\varepsilon(\mathbf{k}) = \hbar^2 k_{\|}^2/2m_{\|} + \Delta(1 - \cos k_z d) \tag{4.50}$$

with wave vector $\mathbf{k} = (k_{\|}\cos\varphi, k_{\|}\sin\varphi, k_z)$. This relation models a superlattice of period d. The in-plane dispersion is parabolic with mass $m_{\|}$, and that along the growth direction z has a tight-binding form with band width $2\Delta = (\varepsilon(0,0,\pi/d) - \varepsilon(0,0,0))$. Since the mass along the growth direction $m_z = \hbar^2/\Delta d^2$, the anisotropy can be increased by reducing Δ.

In the principal frame of the SL, the transport distribution tensor [Eq. (4.23)] is diagonal with components

$$\Sigma_{\rm Growth}(\varepsilon) = \frac{m_{\|} d\tau(\varepsilon)}{2\pi^2\hbar^4} \begin{cases} \Delta^2 \cos^{-1}(1-\varepsilon/\Delta) + (\varepsilon-\Delta)\sqrt{\varepsilon(2\Delta-\varepsilon)}, & \varepsilon < 2\Delta \\ \pi\Delta^2, & \varepsilon \geq 2\Delta \end{cases} \tag{4.51}$$

along the growth direction and

$$\Sigma_{\rm In-plane}(\varepsilon) = \frac{\tau(\varepsilon)}{\pi^2\hbar^2 d} \begin{cases} (\varepsilon-\Delta)\cos^{-1}(1-\varepsilon/\Delta) + \sqrt{\varepsilon(2\Delta-\varepsilon)}, & \varepsilon < 2\Delta \\ \pi(\varepsilon-\Delta), & \varepsilon \geq 2\Delta \end{cases} \tag{4.52}$$

along the planes. The transport coefficients are obtained from Eqs. (4.20)-(4.22) in the principal frame and then transformed into the sample frame. As in Sec. 4.4, the relaxation time $\tau(\mathbf{k})$ is assumed to be a function of energy only. Direct computation with $\tau \propto \varepsilon^r$ indicates that the qualitative features discussed below are independent of the choice of r. Quantitatively, the thermopower is an approximately linear function of r at fixed sample orientation and chemical potential μ and increases by approximately 50% as r goes from 0 to 1.5. In what follows, $r = 0$, $d =$100 Å, $m_{\|}/m_0 = 0.021$, and $\Delta = 57$ meV, corresponding to the C1 subband in a 50-Å $\mathrm{Hg_{0.75}Cd_{0.25}Te}$ / 50-Å $\mathrm{Hg_{0.7}Cd_{0.3}Te}$ SL.

The anisotropy in the resulting effective thermopower $S_{\rm eff}$ [Eq. (4.13)] and Lorenz number $L_{0,\rm eff} = \kappa_{\rm eff}/\sigma_{\rm eff}T$ is shown in Fig. 4.1. For $\mu < 0$, $S_{\rm eff}$ along the growth and in-plane directions differ by $<$ 10%. This near isotropy is expected, since the carriers determining $S_{\rm eff}$ are near the zone center, where the effective-mass approximation is good. The anisotropy increases substantially as μ increases past 2Δ, a region where the bands are appreciably non-parabolic, reaching over 6000 at

$\mu = 0.4$ eV. From Fig. 4.1(b), the anisotropy in $L_{0,\text{eff}}$ is at most 30% and goes to zero in the large-μ limit, approaching the metallic value of $(\pi^2/3)(k_B/e)^2$. As ZT for this band structure is maximal for $\mu \sim 0$, the anisotropy in S_{eff} and $L_{0,\text{eff}}$ in the relevant parameter range are small.

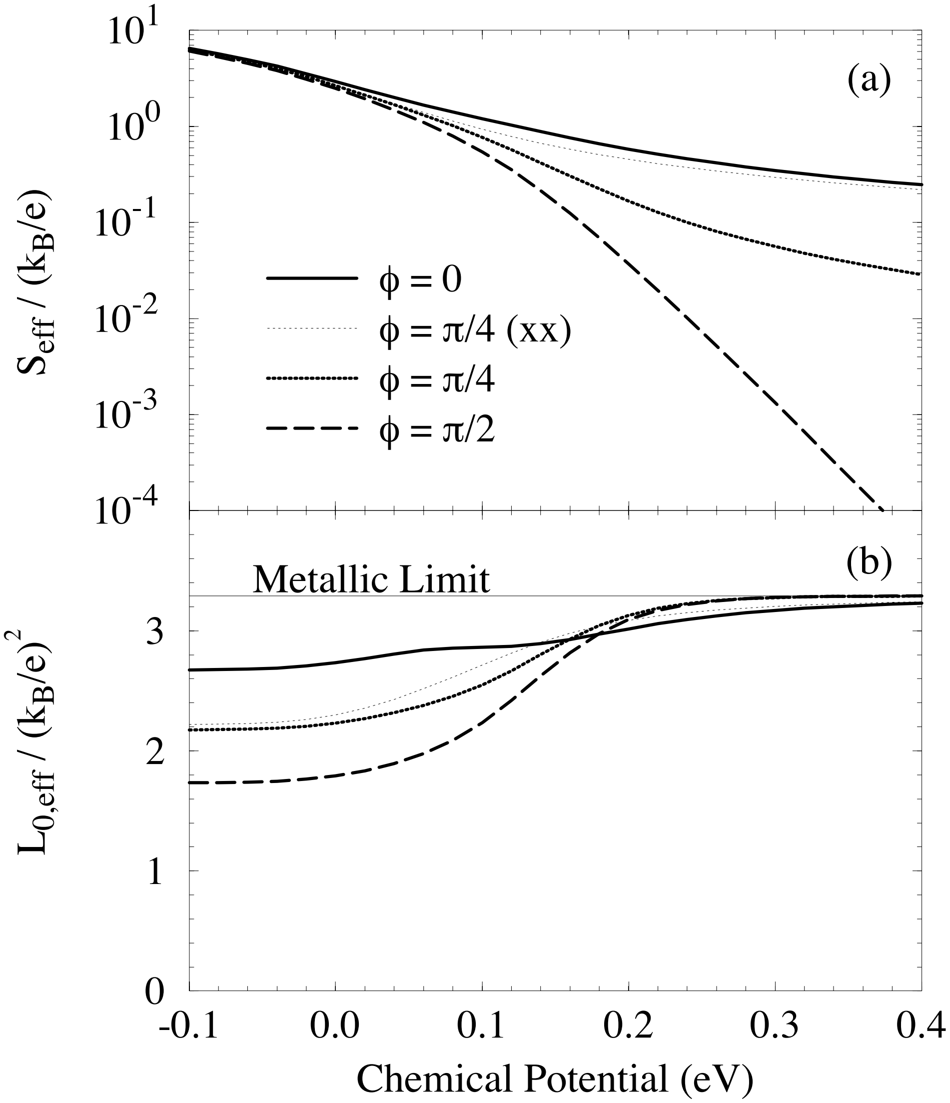


Figure 4.1: (a) Effective thermopower S_{eff} and (b) effective Lorenz number $L_{0,\text{eff}} = \kappa_{\text{eff}}/\sigma_{\text{eff}}T$ as a function of chemical potential for the superlattice described in Sec. 4.5. Shown are samples cut along $\phi = 0$ (heavy solid line), $\pi/4$ (heavy dotted line), and $\pi/2$ (heavy dashed line) where ϕ is the angle between the direction of maximum conductivity $\sigma_{\max}$ and the external field E^{ext}. The results for $\phi = \pi/4$ neglecting the effects of the induced fields are also presented (light dotted line), labeled with (xx) in the legend. The metallic limit of $L_0/(k_B/e)^2 = \pi^2/3$ is indicated in (b).

An Upper Bound on the Thermoelectric Figure of Merit

We present in this Appendix a derivation of an upper bound of the Sofo and Mahan form [83] for an anisotropic material. From $\underline{\sigma} = \underline{A}\sigma_0$ (see Eq. (4.35)) we infer

$$\sigma_{\text{eff}} = \sigma_0 a_0 \tag{.53}$$

with

$$a_0 = A_{xx} - A_{xy}\frac{A_{yx}A_{zz} - A_{yz}A_{zx}}{A_{yy}A_{zz} - A_{yz}A_{zy}} - A_{xz}\frac{A_{zx}A_{yy} - A_{zy}A_{yx}}{A_{zz}A_{yy} - A_{yz}A_{zy}}. \tag{.54}$$

When the κ_ℓ-dependence in Eq. (4.45) is taken into account, we find similarly

$$\kappa_e^* = \kappa_e a(\kappa_\ell) \tag{.55}$$

with $a(\kappa_\ell = 0) = a_0$ and

$$\begin{aligned} a(\kappa_\ell) \quad = A_{xx} \quad & -A_{xy}\frac{A_{yx}(A_{zz}+\kappa_\ell/\kappa_e) - A_{yz}A_{zx}}{(A_{yy}+\kappa_\ell/\kappa_e)(A_{zz}+\kappa_\ell/\kappa_e) - A_{yz}A_{zy}} \\ & -A_{xz}\frac{A_{zx}(A_{yy}+\kappa_\ell/\kappa_e) - A_{zy}A_{yx}}{(A_{yy}+\kappa_\ell/\kappa_e)(A_{zz}+\kappa_\ell/\kappa_e) - A_{yz}A_{zy}}. \end{aligned} \tag{.56}$$

Therefore

$$ZT = \frac{T\sigma_0 S_0^2 a_0}{(\kappa_0 - T\sigma_0 S_0^2)a(\kappa_\ell) + \kappa_\ell}. \tag{.57}$$

Following Mahan and Sofo, introduce dimensionless integrals

$$I_n = \int_{-\mu/k_B T}^{\infty} dx \frac{e^x}{(e^x+1)^2} s(x) x^n, \qquad s(x) = \hbar r_0 \mathcal{T}(\mu + xkT), \tag{.58}$$

where r_0 is the Bohr radius. In terms of these moments,

$$\begin{aligned} \sigma_0 &= \tilde{\sigma}_0 I_0 && (.59) \\ \sigma_0 S_0 &= (k_B/e)\tilde{\sigma}_0 I_1 && (.60) \\ \kappa_0 &= (k_B/e)^2 T \tilde{\sigma}_0 I_2 && (.61) \end{aligned}$$

where $\tilde{\sigma}_0 = e^2/\hbar r_0$ has dimensions of conductivity. Then

$$\begin{aligned} ZT &= \frac{\tilde{\alpha} I_1^2/I_0}{(\tilde{\alpha} I_2 - \tilde{\alpha} I_1^2/I_0)a(\kappa_\ell)/a_0 + 1/a_0} && (.62) \\ &= \frac{\xi}{(1-\xi)a(\kappa_\ell)/a_0 + B} && (.63) \end{aligned}$$

with $\tilde{\alpha} = (k_B/e)^2 T \tilde{\sigma}_0/\kappa_\ell$,

$$\xi = I_1^2/I_0 I_2, \tag{.64}$$

and $B = 1/\tilde{\alpha} I_2 a_0 = \kappa_\ell/\kappa_0 a_0$. By the Cauchy-Schwarz inequality, $0 \le \xi \le 1$. The limit as ξ tends to 1 maximizes the figure of merit by maximizing the numerator and minimizing the denominator in Eq. (.63). From Eq. (.56) it may be seen that for $\kappa_\ell/\kappa_e = \kappa_\ell/(k_B/e)^2 T \tilde{\sigma}_0 I_2 (1-\xi)$ sufficiently large, $a(\kappa_\ell)$ tends to A_{xx}, and certainly this condition is met as $1-\xi$ tends to zero. Thus as $\xi \to 1$ we have that

$$ZT \to \frac{\xi}{(1-\xi)A_{xx}/a_0 + B} \le \frac{1}{B} = \frac{\kappa_0 a_0}{\kappa_\ell}. \tag{.65}$$

Chapter 5

Derivation of the Onsager Relations for Anisotropic Periodic Solids

In this Chapter we derive the Onsager relations for an anisotropic periodic solid. The Boltzmann equation will be applicable on the supposition of translational symmetry, resulting in spatially independent transport coefficients, and Bloch's theorem, which ensures that the stationary states for the electrons are travelling waves. Both hold for an anisotropic periodic solid.

We wish to derive the transport coefficients with reference only to the coordinates in the frame defined by the basis vectors $\mathbf{a}_1$, $\mathbf{a}_2$ and $\mathbf{a}_3$ of the crystalline unit cell in an anisotropic periodic solid. An analysis of physical measurements in this situation will be given below, after the derivation of the Onsager relations. As no symmetry other than translational is assumed, $\mathbf{a}_1$, $\mathbf{a}_2$ and $\mathbf{a}_3$ may be arbitrary as long as they are linearly independent. The dot product of two vectors, say $\mathbf{k}$ and $\mathcal{E}$, appearing in the Boltzmann equation becomes

$$\mathbf{k} \cdot \mathcal{E} = U(k_\xi, k_\eta, k_\zeta) \cdot U(\mathcal{E}_\xi, \mathcal{E}_\eta, \mathcal{E}_\zeta) = \sum_{i,j,\ell} U_{ij} U_{i\ell} k_{\xi j} \mathcal{E}_{\xi \ell} \equiv \mathbf{k}_\xi \cdot \mathcal{E}_\xi \tag{5.1}$$

where we adopt the general notation $k_{\xi 1} = k_\xi$, $k_{\xi 2} = k_\eta$ and $k_{\xi 3} = k_\zeta$ and similarly for $\mathcal{E}$. Here, U is a 3×3 matrix defining the change of coordinates between Cartesian and crystalline frames.

The Onsager relations are a consequence of microscopic reversibility as reflected in the principle of detailed balance. Let $W_{\mathbf{k}_\xi \mathbf{k}'_\xi}$ = transition probability per unit time of scattering from $\mathbf{k}'_\xi \to \mathbf{k}_\xi$. Then, just as for the Cartesian basis we may write the rate of change of $f(\mathbf{k}_\xi)$ as the difference between the number of electrons scattered into the state $\mathbf{k}_\xi$ and the number scattered out:

$$\left(\frac{\partial f}{\partial t}\right)_{\text{coll.}} = \int d^3k'_\xi W_{\mathbf{k}_\xi \mathbf{k}'_\xi} f_{\mathbf{k}'_\xi}(1 - f_{\mathbf{k}_\xi}) - \int d^3k'_\xi W_{\mathbf{k}'_\xi \mathbf{k}_\xi} f_{\mathbf{k}_\xi}(1 - f_{\mathbf{k}'_\xi}). \tag{5.2}$$

As in the Cartesian frame, the physical requirement $(\partial f/\partial t)_{\text{coll.}} = 0$ in equilibrium leads to the principle of detailed balance in the crystal frame:

$$W_{\mathbf{k}_\xi \mathbf{k}'_\xi} f^0_{\mathbf{k}'_\xi}(1 - f^0_{\mathbf{k}_\xi}) = W_{\mathbf{k}'_\xi \mathbf{k}_\xi} f^0_{\mathbf{k}_\xi}(1 - f^0_{\mathbf{k}'_\xi}) \tag{5.3}$$

Note that $f^0_{\mathbf{k}_\xi}$ appearing in the statement of detailed balance for the crystal frame is a different function than $f^0_{\mathbf{k}}$.

As in Cartesian coordinates we may linearize the collision term, keeping terms to lowest order in $f^{(1)}_{\mathbf{k}_\xi}$, where $f_{\mathbf{k}_\xi} = f^0_{\mathbf{k}_\xi} + f^{(1)}_{\mathbf{k}_\xi} = f^0_{\mathbf{k}_\xi} - \varphi_{\mathbf{k}_\xi}(\partial f^0_{\mathbf{k}_\xi}/\partial E_{\mathbf{k}_\xi})$:

$$\left(\frac{\partial f}{\partial t}\right)_{\text{coll.}} = (L\varphi)(\mathbf{k}_\xi) = \frac{1}{k_B T} \int d^3k'_\xi W_{\mathbf{k}'_\xi \mathbf{k}_\xi} f^0_{\mathbf{k}_\xi}(1 - f^0_{\mathbf{k}'_\xi})(\varphi_{\mathbf{k}'_\xi} - \varphi_{\mathbf{k}_\xi}) \tag{5.4}$$

This defines the linearized collision operator L.

In a steady state with no magnetic field the linearized Boltzmann equation in Cartesian coordinates reads:

$$\frac{\partial f^0_{\mathbf{k}}}{\partial E_{\mathbf{k}}}\mathbf{v}_{\mathbf{k}} \cdot (-e\mathcal{E}) + \frac{\partial f^0_{\mathbf{k}}}{\partial E_{\mathbf{k}}}\left(\frac{E_{\mathbf{k}} - \mu}{T}\right)\mathbf{v}_{\mathbf{k}} \cdot \nabla T = (L\varphi)_{\mathbf{k}}. \tag{5.5}$$

Changing now to the crystal coordinates we write $f^0 = f^0(k_\xi, k_\eta, k_\zeta)$, $E = E_{\mathbf{k}_\xi}$ (assumed to be an even function of $\mathbf{k}_\xi$ that does not necessarily correspond to parabolic bands), $\varphi = \varphi_{\mathbf{k}_\xi}$ and the dot product becomes, as above, $\mathbf{v}\cdot\mathcal{E} = \sum_{ij\ell} U_{ij}U_{i\ell}(\hbar/m)k_{\xi j}\mathcal{E}_{\xi\ell}$ and similarly for $\mathbf{v}\cdot\nabla T$. Therefore, in terms of the crystal coordinates, the linearized Boltzmann equation becomes

$$-\frac{e\hbar}{m}\frac{\partial f^0_{\mathbf{k}_\xi}}{\partial E_{\mathbf{k}_\xi}}\sum_{ij\ell} U_{ij}U_{i\ell}k_{\xi j}\mathcal{E}_{\xi\ell} + \frac{\hbar}{m}\frac{\partial f^0}{\partial E_{\mathbf{k}_\xi}}\sum_{ij\ell} U_{ij}U_{i\ell}\left(\frac{E_{\mathbf{k}_\xi} - \mu}{T}\right)k_{\xi j}\nabla_{\xi\ell}T = (L\varphi)_{\mathbf{k}_\xi}. \tag{5.6}$$

As may be seen from the expression for it above, the linearized collision operator has a zero eigenvalue corresponding to the constant eigenfunction $\varphi^{(0)}(\mathbf{k}_\xi) = 1$. If we further assume elastic scattering ($W_{\mathbf{k}_\xi\mathbf{k}'_\xi} = 0$ unless $E_{\mathbf{k}_\xi} = E_{\mathbf{k}'_\xi}$), then L has a second zero eigenvalue corresponding to $\varphi^{(1)}(\mathbf{k}_\xi) = E_{\mathbf{k}_\xi} - \langle E_{\mathbf{k}_\xi}\rangle$ (we subtract the mean energy in order to ensure that $\varphi^{(1)}(\mathbf{k}_\xi)$ is orthogonal to $\varphi^{(0)}(\mathbf{k}_\xi)$). The extension to inelastic scattering is straightforward, but algebraically complicated [97]. The existence of two zero eigenvalues follows from the principle of conservation of electron number and energy; since, in an anisotropic crystal with impurity scattering, there are no other such conservation laws, we do not expect there to be any more zero eigenvalues. Let λ_α denote the eigenvalues of L and $\varphi^{(\alpha)}_{\mathbf{k}_\xi}$ its eigenvectors; L will be shown below to be symmetric. Applying the operator $L^{-1} = \sum_{\alpha\geq 2}(1/\lambda_\alpha)|\varphi^{(\alpha)}_{\mathbf{k}_\xi}\rangle\langle\varphi^{(\alpha)}_{\mathbf{k}_\xi}|$ to both sides, we may solve for $\varphi_{\mathbf{k}_\xi}$ up to the undetermined coefficients of the zero eigenvectors:

$$\begin{aligned}\varphi_{\mathbf{k}_\xi} &= c_0 + c_1(E_{\mathbf{k}_\xi} - \langle E_{\mathbf{k}_\xi}\rangle) - && (5.7)\\ &\frac{e\hbar}{m}\sum_{ij\ell} U_{ij}U_{i\ell}L^{-1}\left(k_{\xi j}\frac{\partial f^0_{\mathbf{k}_\xi}}{\partial E_{\mathbf{k}_\xi}}\right)\mathcal{E}_{\xi\ell} - && (5.8)\\ &\frac{\hbar}{m}\sum_{ij\ell} U_{ij}U_{i\ell}L^{-1}\left(k_{\xi j}\frac{\partial f^0}{\partial E_{\mathbf{k}_\xi}}\left(\frac{E_{\mathbf{k}_\xi}-\mu}{T}\right)\right)\nabla_{\xi\ell}T. && (5.9)\end{aligned}$$

Let

$$\psi_i^{(1)} = \sqrt{n\det U}\frac{e\hbar}{m}\sum_j U_{ij}L^{-1}\left(k_{\xi j}\frac{\partial f^0_{\mathbf{k}_\xi}}{\partial E_{\mathbf{k}_\xi}}\right) \tag{5.10}$$

and

$$\psi_i^{(2)} = \sqrt{n\det U}\frac{\hbar}{m}\sum_j U_{ij}L^{-1}\left(k_{\xi j}\frac{\partial f^0_{\mathbf{k}_\xi}}{\partial E_{\mathbf{k}_\xi}}\left(E_{\mathbf{k}_\xi}-\mu\right)\right) \tag{5.11}$$

be the coefficients of $\mathcal{E}$ respectively $-\nabla T/T$, where n is the density of electrons. The reason for this notation will become clear in a moment.

Now, the current is given by

$$\mathbf{J}^e = ne\langle\mathbf{v}\rangle = \frac{ne\hbar}{m}\int d^3k\mathbf{k}f^{(1)}_{\mathbf{k}}. \tag{5.12}$$

We change to the crystal coordinates, remembering the Jacobian $d^3k = \det U d^3k_\xi$, yielding

$$\begin{aligned}J_i^e &= \frac{ne\hbar\det U}{m}\int d^3k_\xi\sum_j U_{ij}k_{\xi j}f^{(1)}_{\mathbf{k}_\xi} && (5.13)\\ &= -\frac{ne\hbar\det U}{m}\int d^3k_\xi\sum_j U_{ij}k_{\xi j}\varphi_{\mathbf{k}_\xi}\frac{\partial f^0_{\mathbf{k}_\xi}}{\partial E_{\mathbf{k}_\xi}} && (5.14)\\ &= \int d^3k_\xi LL^{-1}\left(\sqrt{n\det U}\frac{e\hbar}{m}\sum_j U_{ij}k_{\xi j}\frac{\partial f^0_{\mathbf{k}_\xi}}{\partial E_{\mathbf{k}_\xi}}\right)\sqrt{n\det U}\varphi_{\mathbf{k}_\xi} && (5.15)\\ &= \left(L\psi_i^{(1)},\psi_j^{(1)}\right)\mathcal{E}_j + \left(L\psi_i^{(1)},\psi_j^{(2)}\right)(-\nabla_j T/T). && (5.16)\end{aligned}$$

The zero eigenfunctions $\varphi^{(0)} = 1$ and $\varphi^{(1)} = E_{\mathbf{k}_\xi} - \langle E_{\mathbf{k}_\xi} \rangle$ appearing in $\varphi_{\mathbf{k}_\xi}$ are even functions of $\mathbf{k}_\xi$ which give zero when integrated against $k_{\xi j}$; hence c_0 and c_1 drop out.

Similarly, the heat current is given by

$$\mathbf{J}^Q = \int d^3k (E_{\mathbf{k}} - \mu) n \mathbf{v} f^{(1)}_{\mathbf{k}} = -\frac{n\hbar}{m} \int d^3k (E_{\mathbf{k}} - \mu) \mathbf{k} \frac{\partial f^0}{\partial E_{\mathbf{k}}} \varphi_{\mathbf{k}}. \tag{5.17}$$

Transforming both sides to the crystal frame yields

$$\begin{aligned} J_i^Q &= -\frac{n\hbar \det U}{m} \int d^3k_\xi (E_{\mathbf{k}_\xi} - \mu) \sum_j U_{ij} k_{\xi j} \frac{\partial f^0}{\partial E_{\mathbf{k}_\xi}} \varphi_{\mathbf{k}_\xi} && (5.18) \\ &= \int d^3k_\xi L L^{-1} \left(\sqrt{n \det U} \frac{\hbar}{m} \sum_j U_{ij} k_{\xi j} \frac{\partial f^0_{\mathbf{k}_\xi}}{\partial E_{\mathbf{k}_\xi}} \left(E_{\mathbf{k}_\xi} - \mu \right) \right) \sqrt{n \det U} \varphi_{\mathbf{k}_\xi} && (5.19) \\ &= \left(L\psi_i^{(2)}, \psi_j^{(1)} \right) \mathcal{E}_j + \left(L\psi_i^{(2)}, \psi_j^{(2)} \right) (-\nabla_j T/T). && (5.20) \end{aligned}$$

Thus the Onsager coefficients in the Cartesian frame are $L_{ij}^{\alpha\beta} \equiv (L\psi_i^{(\alpha)}, \psi_j^{(\beta)})$. Let $V(\mathbf{k}_\xi, \mathbf{k}'_\xi) = W_{\mathbf{k}'_\xi \mathbf{k}_\xi} f^0_{\mathbf{k}_\xi} (1 - f^0_{\mathbf{k}'_\xi})$. From the principle of detailed balance, $V(\mathbf{k}_\xi, \mathbf{k}'_\xi) = V(\mathbf{k}'_\xi, \mathbf{k}_\xi)$. Then the Onsager relations follow from the fact that L is a symmetric operator:

$$\begin{aligned} L_{ij}^{\alpha\beta} = (L\psi_i^{(\alpha)}, \psi_j^{(\beta)}) &= \int d^3k_\xi d^3k'_\xi V(\mathbf{k}_\xi, \mathbf{k}'_\xi) (\psi_i^{(\alpha)}(\mathbf{k}'_\xi) - \psi_i^{(\alpha)}(\mathbf{k}_\xi)) \psi_j^{(\beta)}(\mathbf{k}_\xi) && (5.21) \\ &= \int d^3k_\xi d^3k'_\xi V(\mathbf{k}_\xi, \mathbf{k}'_\xi) \psi_i^{(\alpha)}(\mathbf{k}'_\xi) \psi_j^{(\beta)}(\mathbf{k}_\xi) && (5.22) \\ &\quad - \int d^3k_\xi d^3k'_\xi V(\mathbf{k}_\xi, \mathbf{k}'_\xi) \psi_i^{(\alpha)}(\mathbf{k}_\xi) \psi_j^{(\beta)}(\mathbf{k}_\xi) && (5.23) \\ &= \int d^3k_\xi d^3k'_\xi V(\mathbf{k}'_\xi, \mathbf{k}_\xi) \psi_i^{(\alpha)}(\mathbf{k}_\xi) \psi_j^{(\beta)}(\mathbf{k}'_\xi) && (5.24) \\ &\quad - \int d^3k_\xi d^3k'_\xi V(\mathbf{k}_\xi, \mathbf{k}'_\xi) \psi_i^{(\alpha)}(\mathbf{k}_\xi) \psi_j^{(\beta)}(\mathbf{k}_\xi) && (5.25) \\ &= \int d^3k_\xi d^3k'_\xi V(\mathbf{k}_\xi, \mathbf{k}'_\xi) \psi_i^{(\alpha)}(\mathbf{k}_\xi) (\psi_j^{(\beta)}(\mathbf{k}'_\xi) - \psi_j^{(\beta)}(\mathbf{k}_\xi)) && (5.26) \\ &= (\psi_i^{(\alpha)}, L\psi_j^{(\beta)}) = L_{ji}^{\beta\alpha} && (5.27) \end{aligned}$$

This establishes, among other things, that the electronic thermal conductivity tensor $\underline{\kappa}_e$ is symmetric. We assume that the temperature is high enough for phonon drag to be neglected. Then the phonons relax to equilibrium on a time scale short compared to that on which the electrons relax, so that the coupled electron-phonon Boltzmann equations separate into an electronic part and a phonon part. Treating the latter within the relaxation-time approximation yields for the lattice thermal conductivity

$$(\kappa_\ell)_{ij} = \int \frac{d^3q}{(2\pi)^3} \sum_\alpha \hbar\omega_{\mathbf{q}}^{(\alpha)} \frac{\partial\omega_{\mathbf{q}}^{(\alpha)}}{\partial q_i} \frac{\partial\omega_{\mathbf{q}}^{(\alpha)}}{\partial q_j} \frac{dn(\omega_{\mathbf{q}}^{(\alpha)})}{dT} \tau_{\mathrm{ph}}(\omega_{\mathbf{q}}^{(\alpha)}, T) \qquad (5.28)$$

where the sum is over branches α with dispersion relation $\omega_{\mathbf{q}}^{(\alpha)}$, $\tau_{\mathrm{ph}}(\omega_{\mathbf{q}}^{(\alpha)}, T)$ is the lifetime and $n(\omega_{\mathbf{q}}^{(\alpha)})$ is the Bose-Einstein distribution function. Thus, the lattice thermal conductivity tensor $\underline{\kappa}_\ell$ is symmetric as well; hence, the total thermal conductivity tensor $\underline{\kappa} = \underline{\kappa}_e + \underline{\kappa}_\ell$ is symmetric.

Letting L denote the 6×6 matrix with entries $L_{ij}^{\alpha\beta}$ (including the lattice thermal conductivity where appropriate) we may express the Onsager relations in the Cartestian frame as $L = L^T$. This result suffices for what is needed in the text. We discuss now the physical situation in the crystal frame.

First, the Onsager matrix in the crystal frame, L', may be obtained from the Onsager matrix in the Cartesian frame, L, by a change of coordinate. Let $\mathbf{f} = (\mathcal{E}, -\nabla T/T)$ be the generalized force in the Cartesian frame. Then the generalized force in the crystal frame, $\mathbf{f}'$, is given by $\mathbf{f}' = (U^{-1}\mathcal{E}, U^{-1}(-\nabla T/T))$, or $\mathbf{f}' = \Lambda\mathbf{f}$ with

$$\Lambda = \begin{pmatrix} U^{-1} & 0 \\ 0 & U^{-1} \end{pmatrix}. \qquad (5.29)$$

Accordingly, L goes to $L' = \Lambda L \Lambda^{-1}$, so that $\mathbf{J}' = L'\mathbf{f}'$ expresses the relation between generalized force and flux in the crystal frame. The Onsager relations $L = L^T$ become

$$\begin{aligned}
L' = \Lambda L \Lambda^{-1} &= \Lambda L^T \Lambda^{-1} && (5.30)\\
&= \Lambda\Lambda^T(\Lambda^{-1})^T L^T \Lambda^T (\Lambda^T)^{-1}\Lambda^{-1} && (5.31)\\
&= \Lambda\Lambda^T(\Lambda L \Lambda^{-1})^T(\Lambda^{-1})^T\Lambda^{-1} && (5.32)\\
&= \Lambda\Lambda^T(L')^T(\Lambda^{-1})^T\Lambda^{-1}. && (5.33)
\end{aligned}$$

In the non-orthogonal frame defined by the crystalline axes (i.e., $\Lambda^{-1} \neq \Lambda^T$) the Onsager matrix is no longer symmetric, but nevertheless obeys a definite algebraic relation. Furthermore, this relation is independent of the choice of the Cartesian frame relative to which the matrix U is defined; this may be seen by replacing U with OU in Eq. (5.33), where O is an orthogonal matrix. If Λ is expressed in terms of U the Onsager relations assume the form

$$U(L')^{11}U^{-1} = \left(U(L')^{11}U^{-1}\right)^T \tag{5.34}$$

$$U(L')^{22}U^{-1} = \left(U(L')^{22}U^{-1}\right)^T \tag{5.35}$$

$$U(L')^{12}U^{-1} = \left(U(L')^{21}U^{-1}\right)^T \tag{5.36}$$

which is clearly just the statement that the component tensors making up L' obey the Onsager relations when transformed back to the Cartesian frame.

Now that we have the Onsager matrix in the crystal frame, we discuss the problem of measuring the transport coefficients. Consider a sample of the pure crystal with faces cut parallel to the faces of the unit cell. Suppose it to have length L_1 in the $\mathbf{a}_1$ direction, L_2 in the $\mathbf{a}_2$ direction and L_3 in the $\mathbf{a}_3$ direction. An electric field and temperature gradient are applied along the $\mathbf{a}_1$ direction, $\mathcal{E} = \mathcal{E}_0\mathbf{a}_1/|\mathbf{a}_1|$ and $\nabla T = \nabla_1 T\mathbf{a}_1/|\mathbf{a}_1|$. If we consider open circuit conditions for the two faces parallel to the applied electric field, an induced electric field will result in an effective electrical conductivity for transport in the $\mathbf{a}_1$ direction. We assume $L_1 >> L_2, L_3$ so that edge effects may be neglected in the $\mathbf{a}_1$ direction. As in the rectangular case, the induced field $\mathcal{E}^{\text{ind}}$ will lie in the plane perpendicular to $\mathbf{a}_1$ and be invariant under translations in the $\mathbf{a}_1$ direction. If we chose a Cartesian frame with $\hat{\mathbf{z}}$ along $\mathbf{a}_1$ and $\hat{\mathbf{x}}$ along the projection of $\mathbf{a}_2$ into the plane perpendicular to $\mathbf{a}_1$, then we may write $\mathcal{E}^{\text{ind}} = \mathcal{E}_{\text{ind},x}\hat{\mathbf{x}} + \mathcal{E}_{\text{ind},y}\hat{\mathbf{y}}$ where $\mathcal{E}_{\text{ind},x}$ and $\mathcal{E}_{\text{ind},y}$ are as yet unknown functions of x and y.

Now, as in the rectangular case, in the steady state we must have that the transverse components of the electric current vanish: $J^e_x = J^e_y = 0$. A similar consideration applies to the conduction of heat: in the steady state

$$\nabla T^{\text{ind}} = \nabla_x T^{\text{ind}}\hat{\mathbf{x}} + \nabla_y T^{\text{ind}}\hat{\mathbf{y}} \tag{5.37}$$

and $\mathbf{J}^Q = J^Q_z\hat{\mathbf{z}}$. The column vectors representing $\mathbf{J}^e$, $\mathcal{E}^{\text{ind}}$, $\mathbf{J}^Q$ and ∇T^{ind} in the crystal frame will be

$$U^{-1}\begin{pmatrix} J^e_z \\ 0 \\ 0 \end{pmatrix}, U^{-1}\begin{pmatrix} 0 \\ \mathcal{E}_{\text{ind},x} \\ \mathcal{E}_{\text{ind},y} \end{pmatrix}, U^{-1}\begin{pmatrix} J^Q_z \\ 0 \\ 0 \end{pmatrix}, U^{-1}\begin{pmatrix} 0 \\ \nabla_x T^{\text{ind}} \\ \nabla_y T^{\text{ind}} \end{pmatrix}. \tag{5.38}$$

The phenomenological linear relations between forces and fluxes, expressed using only quantities related to the crystal frame, will thus be

$$U^{-1}\begin{pmatrix} J_z^e \\ 0 \\ 0 \end{pmatrix} = \sigma' U^{-1}\begin{pmatrix} \mathcal{E}_0 \\ \mathcal{E}_{\mathrm{ind},x} \\ \mathcal{E}_{\mathrm{ind},y} \end{pmatrix} - \sigma' S' U^{-1}\begin{pmatrix} \nabla T^{\mathrm{appl}} \\ \nabla_x T^{\mathrm{ind}} \\ \nabla_y T^{\mathrm{ind}} \end{pmatrix} \tag{5.39}$$

$$U^{-1}\begin{pmatrix} J_z^Q \\ 0 \\ 0 \end{pmatrix} = \Pi' U^{-1}\begin{pmatrix} J_z^e \\ 0 \\ 0 \end{pmatrix} - \kappa' U^{-1}\begin{pmatrix} \nabla T^{\mathrm{appl}} \\ \nabla_x T^{\mathrm{ind}} \\ \nabla_y T^{\mathrm{ind}} \end{pmatrix}. \tag{5.40}$$

The left-hand side is independent of position, so the solution for $\mathcal{E}^{\mathrm{ind}}$ and ∇T^{ind} will likewise be constants independent of position. This means for instance that the induced-field component $\mathcal{E}_{\mathrm{ind},x}$ can be measured by the voltage drop between point contacts placed along a line parallel to the $\hat{\mathbf{x}}$ axis in the plane perpendicular to $\mathbf{a}_1$.

Eqs. (5.39) and (5.40) are mathematically identical to

$$\begin{pmatrix} J_z^e \\ 0 \\ 0 \end{pmatrix} = U\sigma' U^{-1}\begin{pmatrix} \mathcal{E}_0 \\ \mathcal{E}_{\mathrm{ind},x} \\ \mathcal{E}_{\mathrm{ind},y} \end{pmatrix} - U\sigma' U^{-1} U S' U^{-1}\begin{pmatrix} \nabla T^{\mathrm{appl}} \\ \nabla_x T^{\mathrm{ind}} \\ \nabla_y T^{\mathrm{ind}} \end{pmatrix} \tag{5.41}$$

$$\begin{pmatrix} J_z^Q \\ 0 \\ 0 \end{pmatrix} = U\Pi' U^{-1}\begin{pmatrix} J_z^e \\ 0 \\ 0 \end{pmatrix} - U\kappa' U^{-1}\begin{pmatrix} \nabla T^{\mathrm{appl}} \\ \nabla_x T^{\mathrm{ind}} \\ \nabla_y T^{\mathrm{ind}} \end{pmatrix}. \tag{5.42}$$

Here, $T\sigma' S' = (L')^{21}$, $\Pi'\sigma' = (L')^{12}$ and $\kappa' = \kappa_0 - \Pi'\sigma' S'$, $\kappa_0 = (L')^{22}$. The Onsager relations in the crystal frame, Eqs. (5.34), (5.35) and (5.36), imply that $U\sigma' U^{-1} = (U\sigma' U^{-1})^T$, $U\kappa_0 U^{-1} = (U\kappa_0 U^{-1})^T$, $U\Pi' U^{-1} U\sigma' U^{-1} = (TU\sigma' U^{-1} U S' U^{-1})^T$ hence $U\Pi' U^{-1} = T(US'U^{-1})^T$ and finally, from these three relations, $U\kappa' U^{-1} = (U\kappa' U^{-1})^T$. These relations will be used below. Let $\sigma = U\sigma' U^{-1}$, $S = US'U^{-1}$ and $\kappa = U\kappa' U^{-1}$.

Then Eqs. (5.41) and (5.42) assume the form

$$\begin{pmatrix} J_z^e \\ 0 \\ 0 \end{pmatrix} = \sigma \begin{pmatrix} \mathcal{E}_0 \\ \mathcal{E}_{\text{ind},x} \\ \mathcal{E}_{\text{ind},y} \end{pmatrix} - \sigma S \begin{pmatrix} \nabla T^{\text{appl}} \\ \nabla_x T^{\text{ind}} \\ \nabla_y T^{\text{ind}} \end{pmatrix} \tag{5.43}$$

$$\begin{pmatrix} J_z^Q \\ 0 \\ 0 \end{pmatrix} = TS^T \begin{pmatrix} J_z^e \\ 0 \\ 0 \end{pmatrix} - \kappa \begin{pmatrix} \nabla T^{\text{appl}} \\ \nabla_x T^{\text{ind}} \\ \nabla_y T^{\text{ind}} \end{pmatrix}, \tag{5.44}$$

which is mathematically the same problem as solved in the text for the rectangular case. Therefore we find

$$J^e = \sigma_{\text{eff}} \left(\mathcal{E}_0 - S_{\text{eff}} \nabla T^{\text{appl}} \right) \tag{5.45}$$

$$J^Q = \Pi_{\text{eff}} J - \kappa_{\text{eff}} \nabla T^{\text{appl}} \tag{5.46}$$

with

$$\sigma_{\text{eff}} = \sigma_1 \left[1 + \sigma_1 T \begin{pmatrix} S_{12} & S_{13} \end{pmatrix} \begin{pmatrix} \kappa_{22} & \kappa_{23} \\ \kappa_{32} & \kappa_{33} \end{pmatrix}^{-1} \begin{pmatrix} S_{21} \\ S_{31} \end{pmatrix} \right]^{-1} \tag{5.47}$$

$$\sigma_1 = \sigma_{11} - \begin{pmatrix} \sigma_{12} & \sigma_{13} \end{pmatrix} \begin{pmatrix} \sigma_{22} & \sigma_{23} \\ \sigma_{32} & \sigma_{33} \end{pmatrix}^{-1} \begin{pmatrix} \sigma_{21} \\ \sigma_{31} \end{pmatrix} \tag{5.48}$$

$$\kappa_{\text{eff}} = \kappa_{11} - \begin{pmatrix} \kappa_{12} & \kappa_{13} \end{pmatrix} \begin{pmatrix} \kappa_{22} & \kappa_{23} \\ \kappa_{32} & \kappa_{33} \end{pmatrix}^{-1} \begin{pmatrix} \kappa_{21} \\ \kappa_{31} \end{pmatrix} \tag{5.49}$$

$$S_{\text{eff}} = S_{11} - \begin{pmatrix} S_{12} & S_{13} \end{pmatrix} \begin{pmatrix} \kappa_{22} & \kappa_{23} \\ \kappa_{32} & \kappa_{33} \end{pmatrix}^{-1} \begin{pmatrix} \kappa_{21} \\ \kappa_{31} \end{pmatrix}, \tag{5.50}$$

with $\sigma = U\sigma' U^{-1}$, $S = US'U^{-1}$ and $\kappa = U\kappa' U^{-1}$ as above, and lastly

$$\Pi_{\text{eff}} = TS_{\text{eff}}. \tag{5.51}$$

Thus, the Onsager relations still hold for the effective transport quantities in the non-rectangular sample geometry.

Bibliography

[1] O. Agam and S. Fishman. *J. Phys. A*, **26**:2113, 1993.

[2] O. Agam and S. Fishman. *Phys. Rev. Lett.*, **73**:806, 1994.

[3] A. Bäcker and R. Schubert. *J. Phys. A*, **32**:4795, 1999.

[4] A. Bäcker, R. Schubert, and P. Stifter. *J. Phys. A*, **30**:6783, 1997.

[5] M. V. Berry. in *Chaotic Behaviour of Deterministic Systems,* ed. by G. Iooss, R. Helleman, and R. Stora. North-Holland, 1983. p. 171.

[6] M. V. Berry. *Proc. Roy. Soc. A*, **243**:219, 1989.

[7] M. V. Berry. *Les Houches Lecture Notes, Summer School on Chaos and Quantum Physics,* ed. M.-J. Giannoni, A. Voros and J. Zinn-Justin. Elsevier Science Publishers B.V., 1991.

[8] M. V. Berry. *J. Phys. A*, **27**:L391, 1994.

[9] P. A. Boasman. *Nonlinearity*, **7**:485, 1994.

[10] E. B. Bogomolny. *Physica D*, **31**:169, 1988.

[11] E. B. Bogomolny. *Nonlinearity*, **5**:805, 1992.

[12] D. Bohm. *Quantum Theory.* Prentice-Hall, New York, 1951. (pp. 264-295).

[13] G. Borelius and A. E. Lindh. *Ann. d. Phys.*, **53**:97, 1917.

[14] M. Born and K. Huang. *Dynamical Theory of Crystal Lattices.* Oxford University Press, Oxford, England, 1954.

[15] J. C. Brice. in *Properties of Mercury Cadmium Telluride,* ed. J. Brice and P. Capper, EMIS Datareview Series No. 3. INSPEC, New York, 1987. (Ch. 5.1).

[16] P. W. Bridgman. *Proc. Nat. Acad. Sci. USA*, **13**:46, 1927.

[17] P. W. Bridgman. *The Thermodynamics of Electrical Phenomena in Metals and a Condensed Collection of Thermodynamic Formulas.* Dover, New York, 1961. (Ch. 4).

[18] T. A. Brody, J. Flores, J. B. French, P. A. Mello, A. Pandey, and S. S. M. Wong. *Rev. Mod. Phys.*, **53**:385, 1981.

[19] D. A. Broido and T. L. Reinecke. *Phys. Rev. B*, **51**:13797, 1995.

[20] D. A. Broido and T. L. Reinecke. *Appl. Phys. Lett.*, **67**:1170, 1995.

[21] D. A. Broido and T. L. Reinecke. *Appl. Phys. Lett.*, **67**:100, 1995.

[22] D. A. Broido and T. L. Reinecke. *Appl. Phys. Lett.*, **70**:2834, 1997. (and unpublished).

[23] W. S. Capinski and H. J. Maris. *Physica B*, **219**:699, 1996.

[24] W. S. Capinski, H. J. Maris, T. Ruf, M. Cardona, K. Ploog, and D. S. Katzer. *Phys. Rev. B*, **59**:8105, 1999.

[25] G. Casati and T. Prosen. *Phys. Rev. E*, **59**:R2516, 1999.

[26] G. Chen. *J. Heat Transfer*, **119**:220, 1996.

[27] G. Chen. *Phys. Rev. B*, **57**:14958, 1998.

[28] G. Chen and M. Neagu. *Appl. Phys. Lett.*, **71**:2761, 1997.

[29] S. C. Creagh and N. D. Whelan. *Ann. Phys.*, **272**:196, 1999.

[30] S. C. Creagh and N. D. Whelan. *Phys. Rev. Lett.*, **84**:4084, 2000.

[31] S. R. de Groot and P. Mazur. *Non-equilibrium Thermodynamics.* North-Holland, Amsterdam, 1963.

[32] Y. Colin de Verdiere. *Commun. Math. Phys.*, **102**:497, 1985.

[33] R. Dornhaus and D. Nimtz. in *Narrow-Gap Semiconductors,* ed. G. Höhler. Springer, New York, 1983.

[34] B. Eckhardt, S. Fishman, J. Keating, O. Agam, J. Main, and K. Muller. *Phys. Rev. E*, **52**:5893, 1995.

[35] P. Ehrenfest and A. J. Rutgers. *Proc. Amst. Acad.*, **32**:698, 1929.

[36] P. Ehrenfest and A. J. Rutgers. *Proc. Amst. Acad.*, **32**:883, 1929.

[37] L. Esaki and R. Tsu. *IBM J. Res. Dev.*, **14**:61, 1970.

[38] G. Ezra, K. Richter, G. Tanner, and D. Wintgen. *J. Phys. B*, **24**:L413, 1991.

[39] M. E. Flatté, C. H. Grein, H. Ehrenreich, R. H. Miles, and H. Cruz. *J. Appl. Phys.*, **78**:4552, 1995.

[40] T. M. Fromhold, L. Eaves, F. W. Sheard, M. L. Leadbeater, T. J. Foster, and P. C. Main. *Phys. Rev. Lett.*, **72**:2608, 1994.

[41] T. M. Fromhold, P. B. Wilkinson, F. W. Sheard, L. Eaves, J. Miao, and G. Edwards. *Phys. Rev. Lett.*, **75**:1142, 1995.

[42] H. J. Goldsmid. *Applications of Thermoelectricity.* Methuen & Co Ltd, London, 1960.

[43] H. J. Goldsmid. *Thermoelectric Refrigeration.* Plenum, New York, 1964.

[44] R. Granger and C. M. Pelletier. *J. Crys. Growth*, **138**:486, 1994.

[45] M. C. Gutzwiller. *J. Math. Phys.*, **12**:343, 1971.

[46] D. O. Harris, G. G. Engerholm, and W. D. Gwinn. *J. Chem. Phys.*, **43**:1515, 1965.

[47] E. J. Heller. *J. Chem. Phys.*, **72**:1337, 1980.

[48] E. J. Heller. *J. Chem. Phys.*, **72**:1337, 1980.

[49] E. J. Heller. *Phys. Rev. Lett.*, **53**:1515, 1984.

[50] E. J. Heller. *Phys. Rev. A*, **35**:1360, 1987.

[51] E. J. Heller. *Phys. Rev. A*, **35**:1360, 1987.

[52] E. J. Heller. *Proc. Les Houches Summer School on 'Chaos and Quantum Physics' (Les Houches Session LII)* ed. M. J. Giannoni, A. Voros and J. Zinn-Justin. Elsevier, North-Holland, Amsterdam, 1991. (p. 602).

[53] R. Hernandez, W. H. Miller, C. Bradley Moore, and W. F. Polik. *J. Chem. Phys.*, **99**:950, 1993.

[54] L. D. Hicks and M. S. Dresselhaus. *Phys. Rev. B*, **47**:12727, 1993.

[55] L. D. Hicks and M. S. Dresselhaus. *Phys. Rev. B*, **47**:16631, 1993.

[56] L. D. Hicks, T. C. Harman, and M. S. Dresselhaus. *Appl. Phys. Lett.*, **63**:3230, 1993.

[57] L. D. Hicks, T. C. Harman, X. Sun, and M. S. Dresselhaus. *Phys. Rev. B*, **53**:R10493, 1996.

[58] P. M. Hui, N. F. Johnson, and H. Ehrenreich. *J. Vac. Sci. Technol. A*, **7**:424, 1989.

[59] P. Hyldgaard and G. D. Mahan. *Phys. Rev. B*, **56**:10754, 1997.

[60] N. F. Johnson, H. Ehrenreich, P. M. Hui, and P. M. Young. *Phys. Rev. B*, **41**:3655, 1990.

[61] N. F. Johnson, P. M. Hui, and H. Ehrenreich. *Phys. Rev. Lett.*, **61**:1993, 1988.

[62] H. Jones. in *Handbuch der Physik,* Vol. 19. Springer-Verlag, New York, 1956. (p. 227).

[63] E. O. Kane. in *Narrow Gap Semiconductors: Physics and Applications,* ed. W. Zawadzki, Lecture Notes in Physics Vol. 133. Springer, New York, 1981. (p. 19).

[64] L. Kaplan. *Phys. Rev. Lett.*, **80**:2582, 1998.

[65] L. Kaplan. *Nonlinearity*, **12**:R1, 1999. (and references therein).

[66] L. Kaplan. *Phys. Rev. E*, **59**:5325, 1999.

[67] L. Kaplan and E. J. Heller. *Physica D*, **121**:1, 1998.

[68] L. Kaplan and E. J. Heller. *Ann. Phys.*, **264**:171, 1998.

[69] L. Kaplan and E. J. Heller. *Physica D*, **121**:1, 1998.

[70] L. Kaplan and E. J. Heller. *Phys. Rev. E*, **59**:6609, 1999.

[71] L. P. Kouwenhoven, C. M. Marcus, P. L. Mceuen, S. Tarucha, R. M. Westervelt, and N. S. Wingreen. in *Mesoscopic electron transport,* ed. L. L. Sohn, L. P. Kouwenhoven and G. Schön. Kluwer, 1997. (pp. 106-214).

[72] K. Kunc. *Ann. Phys. (Paris)*, **8**:319, 1973-1974.

[73] K. Kunc, M. Balkanski, and M. A. Nusimovici. *Phys. Rev. B*, **12**:4346, 1975.

[74] B. Li. *Phys. Rev. E*, **55**:5376, 1997.

[75] B. Li and B. Hu. *J. Phys. A*, **31**:483, 1998.

[76] J. C. Light, I. P. Hamilton, and J. V. Lill. *J. Chem. Phys.*, **82**:1400, 1985.

[77] J. V. Lill, G. A. Parker, and J. C. Light. *Chem. Phys. Lett.*, **89**:483, 1982.

[78] P. J. Lin-Chung and T. L. Reinecke. *Phys. Rev. B*, **51**:13244, 1995.

[79] J. E. Lynn. *The Theory of Neutron Resonance Reactions.* Oxford University Press, Oxford, England, 1968.

[80] A. H. MacDonald, S. H. Vosko, and P. T. Coleridge. *J. Phys. C*, **12**:2991, 1979.

[81] G. D. Mahan. in *Solid State Physics,* Vol. 51, ed. H. Ehrenreich and F. Spaepen. Academic Press, New York, 1998. (pp. 81-157).

[82] G. D. Mahan and Jr. H. B. Lyon. *J. Appl. Phys.*, **76**:1899, 1994.

[83] G. D. Mahan and J. O. Sofo. *Proc. Nat. Acad. Sci. USA*, **93**:7436, 1996.

[84] G. D. Mahan, J. O. Sofo, and M. Bartkowiak. *J. Appl. Phys.*, **83**:4683, 1998.

[85] G. D. Mahan and L. M. Woods. *Phys. Rev. Lett.*, **80**:4016, 1998.

[86] C. M. Marcus, A. J. Rimberg, R. M. Westervelt, P. F. Hopkins, and A. C. Gossard. *Phys. Rev. Lett.*, **69**:506, 1992.

[87] C. M. Marcus, R. M. Westervelt, P. F. Hopkins, and A. C. Gossard. *Chaos*, **3**:643, 1993.

[88] S. W. McDonald. unpublished Ph.D. thesis, University of California at Berkeley.

[89] W. H. Miller, R. Hernandez, C. B. Moore, and W. F. Polik. *J. Chem. Phys.*, **93**:5657, 1990.

[90] F. Mortessagne, O. Legrand, and D. Sornette. *Chaos*, **3**:529, 1993.

[91] K. Muller and D. Wintgen. *J. Phys. B*, **27**:2693, 1994.

[92] J. F. Nye. *Physical Properties of Crystals: Their Representation by Tensors and Matrices.* Oxford University Press, Oxford, England, 1957. (p. 199).

[93] P. W. O'Connor and E. J. Heller. *Phys. Rev. Lett.*, **61**:2288, 1989.

[94] P. W. O'Connor, S. Tomsovic, and E. J. Heller. *Physica D*, **55**:340, 1992.

[95] J. S. Espinoza Ortiz and A. M. Ozorio de Almeida. *J. Phys. A*, **30**:7301, 1997.

[96] C. Patel, T. J. Parker, H. Jamshidi, and W. F. Sherman. *Phys. Stat. Sol. (b)*, **122**:461, 1984.

[97] R. E. Peierls. *Quantum Theory of Solids.* Oxford University Press, 1955.

[98] C. E. Porter and R. G. Thomas. *Phys. Rev.*, **104**:483, 1956.

[99] R. J. Radtke, H. Ehrenreich, and C. H. Grein. *J. Appl. Phys.*, **86**:3195, 1999.

[100] R. J. Radtke, H. Ehrenreich, and C. H. Grein. *J. Appl. Phys.*, **86**:3195, 1999.

[101] S.-F. Ren, H. Chu, and Y.-C. Chang. *Phys. Rev. B*, **37**:8899, 1988.

[102] P. J. Robinson and K. A. Holbrook. *Unimolecular Reactions.* Wiley-Interscience, New York, 1972.

[103] A. I. Schnirelman. *Usp. Mat. Nauk.*, **29**:181, 1974.

[104] J. N. Schulman and Y.-C. Chang. *Phys. Rev. B*, **33**:2594, 1986.

[105] M. Sieber, U. Smilansky, S. C. Creagh, and R. G. Littlejohn. *J. Phys. A*, **26**:6217, 1993.

[106] F. P. Simonotti, E. Vergini, and M. Saraceno. *Phys. Rev. E*, **56**:3859, 1997.

[107] A. C. Smith, J. F. Janak, and R. B. Adler. *Electronic Conduction in Solids.* McGraw-Hill, New York, 1967.

[108] J. O. Sofo and G. D. Mahan. *Appl. Phys. Lett.*, **65**:2690, 1994.

[109] S. Sridhar and E. J. Heller. *Phys. Rev. A*, **46**:R1728, 1992.

[110] J. Stein and H. J. Stöckman. *Phys. Rev. Lett.*, **68**:2867, 1992.

[111] J. I. Steinfeld, J. S. Francisco, and W. L. Hase. *Chemical Kinetics and Dynamics.* Prentice Hall, Englewood Cliffs, N.J., U.S.A., 1989.

[112] H. J. Stöckman and J. Stein. *Phys. Rev. Lett.*, **64**:2215, 1990.

[113] G. Tanner. *J. Phys. A*, **30**:2863, 1997.

[114] T. Terada and T. Tsutsui. *Proc. Imp. Acad. Tok.*, **3**:132, 1927.

[115] W. Thomson. *Trans. Roy. Soc. Edin.*, **21**:153, 1857.

[116] S. Tomsovic. *J. Phys. A: Math. Gen.*, **31**:9469, 1998.

[117] S. Tomsovic and E. J. Heller. *Phys. Rev. Lett.*, **67**:604, 1991.

[118] S. Tomsovic and E. J. Heller. *Phys. Rev. E*, **47**:282, 1993.

[119] C. Truesdell. *Rational Thermodynamics: A Course of Lectures on Selected Topics.* McGraw Hill, New York, 1969.

[120] R. Venkatasubramanian and T. Colpitts. *Mat. Res. Soc. Symp. Proc.*, **478**:73, 1997.

[121] R. Venkatasubramanian, E. Sivola, T. Colpitts, K. Stokes, and B. O'Quinn. *Mat. Res. Soc. Symp. Proc.*, **575**. (to appear).

[122] E. Vergini and M. Saraceno. *Phys. Rev. E*, **52**:2204, 1995.

[123] P. B. Wilkinson, T. M. Fromhold, L. Eaves, F. W. Sheard, N. Miura, and T. Takamasu. *Nature*, **380**:608, 1996.

[124] D. Wintgen and A. Honig. *Phys. Rev. Lett.*, **63**:1467, 1989.

[125] D. Wintgen, K. Richter, and G. Tanner. *Chaos*, **2**:19, 1992.

[126] S. Zelditch. *Duke Math. J.*, **55**:919, 1987.

[127] S. Zelditch and M. Zworski. *Commun. Math. Phys.*, **175**:673, 1996.